University of Zurich[UZH]

Exploring Multimodal Deep Learning for Contextual Operator Classification in Cartographic Building Generalization

GEO 511 Master's Thesis

Author
Joris Senn
18-705-053

Supervised by
Dr. Zhiyong Zhou
Dr. Cheng Fu
Prof. Dr. Robert Weibel

Faculty representative
Prof. Dr. Robert Weibel

28.01.2025
Department of Geography, University of Zurich

Abstract

Cartographic generalization has proven notoriously challenging to automate, owing to the difficulty of formalizing the implicit knowledge heavily employed throughout the process. Consequently, deep learning has emerged as a promising candidate for a paradigm shift in automated cartographic generalization, as it has the potential to circumvent explicit knowledge formalization by learning from examples. Current studies mainly revolve around ambitious end-to-end generalization approaches, deviating from established cartographic practices by largely dismissing the significance of generalization operators. In addition, they incorporate limited contextual information and predominantly process maps represented as rasters. Therefore, this thesis investigates the feasibility of using deep learning in conjunction with a novel, enriched dataset to predict contextual generalization operators (elimination, aggregation, typification, displacement, and enlargement) that are to be applied to generalize buildings on topographic maps during the transition from 1:25,000 to 1:50,000. To this end, classification models based on vector and raster representations are developed and evaluated. Furthermore, a multimodal model is proposed that exploits both modalities simultaneously to generate predictions. The study also explores the role of contextual map features in the form of surrounding buildings and roads in facilitating operator predictions. The results reveal that deep learning models can effectively predict cartographic generalization operators, particularly those less dependent on contextual information such as enlargement and aggregation. However, performance declines for highly contextual operators such as displacement and typification. The classification of elimination yields the worst evaluation metrics. Moreover, it is shown that the models are particularly adept at predicting operators for buildings located in rural areas. The incorporation of contextual map features is crucial, since their exclusion results in worse performance for certain operators. Across modalities, the multimodal model achieves the best overall classification evaluation metrics. These findings contribute to the advancement of automating cartographic generalization using deep learning, demonstrating its capacity to predict operators and underscoring the importance of contextual features and multimodal approaches. The study lays the groundwork for integrating such models into broader automated generalization workflows driven by deep learning. The code is available at github.com/jorissenn/genops.

Keywords: cartographic generalization, building generalization, contextual generalization, generalization operator, multimodal deep learning, GeoAI, computational cartography

Acknowledgements

I am deeply grateful to the many individuals whose contributions were crucial in shaping the development and outcome of this thesis. Specifically, I would like to express my gratitude to...

- My supervisors *Zhiyong Zhou*, *Cheng Fu*, and *Robert Weibel* for their indispensable guidance and support, both throughout this thesis and during my work on the DeepGeneralization project. Their profound expertise in the domains of deep learning and cartographic generalization and their constructive feedback continuously challenged me to refine my ideas, which proved instrumental in delivering this thesis. Whether it was football games at Letzigrund, endless train journeys to Vienna, or late-night bumper car rides at the Prater, I will always cherish the memories we made over the course of the last year.
- My fellow research assistants for their contributions to the DeepGeneralization project: *Songlin Wang*, *Nicolas Beglinger*, and especially *Jan Winkler* who laid the foundation for the training dataset employed throughout this thesis.
- The team at swisstopo surrounding *Dominik Käuferle* and *Roman Geisthövel* for supplying data and providing valuable feedback during the development of the training dataset.
- *Yves Baumann* for our insightful discussions that greatly enhanced my understanding of the underlying mechanics of graph neural networks, proving vital for the implementation of the vector models.
- My colleagues at Amstein+Walthert for giving me the opportunity to engage in various projects surrounding spatial energy planning and high-alpine solar plants, providing a refreshing change of pace from my studies.
- My fellow students writing their theses alongside me for the countless hours we spent together in Y23-G-10, complemented by enjoyable lunch and coffee breaks that made the journey all the more rewarding.
- All other friends for supporting me through the trials and tribulations of developing this thesis.
- My parents for their continued encouragement and unwavering support throughout my academic endeavors.

Contents

Figures

Tables

Abbreviations

Adam	Adaptive Moment Estimation
AI	Artificial Intelligence
ANN	Artificial Neural Network
AttU-Net	Attention U-Net
AUC	Area Under Curve
BBOX	Bounding Box
BCE	Binary Cross-Entropy
CAM	Class Activation Mapping
CH	Convex Hull
CNN	Convolutional Neural Network
CPU	Central Processing Unit
DCM	Digital Cartographic Model
DL	Deep Learning
DT	Delaunay Triangulation
ERI	Equivalent Rectangular Index
FPR	False Positive Rate
GAE	Graph Autoencoder
GAN	Generative Adversarial Network
GAT	Graph Attention Network
GCNN	Graph Convolutional Neural Network
GeoAI	Geospatial Artificial Intelligence
GNN	Graph Neural Network
GPU	Graphics Processing Unit

GraphSAGE	Graph Sample and Aggregate
HGNN	Heterogeneous Graph Neural Network
HGT	Heterogeneous Graph Transformer
LP	Label Powerset
MBR	Minimum Bounding Rectangle
ML	Machine Learning
MST	Minimum Spanning Tree
NMA	National Mapping Agency
PR	Precision-Recall
RAM	Random-access Memory
ReLU	Rectified Linear Unit
ResU-Net	Residual U-Net
RG	Research Gap
ROC	Receiver Operating Characteristic
ROS	Random Oversampling
RQ	Research Question
RUS	Random Undersampling
SMOTE	Synthetic Minority Oversampling
TLM	Topographic Landscape Model
TPR	True Positive Rate
ViT	Vision Transformer
XAI	Explainable Artificial Intelligence
YOLO	You Only Look Once

1 Introduction

It is evident that cartography is not merely a technical art. It is for the greater part an applied art, an art governed and determined by scientific laws. But how can cartography avoid the rigid rules of mathematical precision? The decisive turning-point lies in the transition from the topographic to the general map. As long as the scale allows the objects in nature to be represented in their true proportion on the map, technical skill alone is necessary.
Where this possibility ends the art of the cartographer begins.
With generalization art enters into the making of maps.
In generalizing lies the difficulty of scientific map-making, for it no longer allows the cartographer to rely merely on objective facts but requires him to interpret them subjectively.

— Eckert (1908, p. 346)

Generalization refers to the process of deriving smaller scale maps from large scale data sources by simplifying the representation of geographic information through the elimination of unnecessary details while maintaining its essential features and relationships to produce maps that are visually clear and functionally effective (Weibel 1995*a*). This process is necessitated by the fact that all maps are abstractions of reality, rendering it impossible to represent every minute detail (Brassel & Weibel 1988). Generalization is used to emphasize pertinent information and to portray the appropriate level of detail based on map purpose. Furthermore, generalization is used to address map features that risk becoming subject to congestion, coalescence, conflict, complication, inconsistency, and imperceptibility due to a reduction in the depicted map surface area induced by a decrease in the map scale (McMaster & Shea 1992, Spiess et al. 2005).

More than a century ago, seminal cartographer Max Eckert identified the conundrum surrounding generalization that still haunts modern cartography well into the digital age: Generalization embodies a semi-structured, subjective problem whose objectives are inherently ambiguous (Armstrong 1991). The generalization process heavily relies on cartographic intuition and expert knowledge, requiring practicing cartographers to strike an optimal balance not only between the level of detail and the available space, but also between scientific rigor and artistic creativity (Mackaness et al. 2014, Slocum et al. 2022). Consequently, map generalization has proven to be notoriously difficult to automate holistically (Harrie et al. 2024).

National mapping agencies (NMAs) commonly publish topographic maps at various scales, as illustrated with the Swiss national map series in Figure 1. To derive the maps, NMAs have to apply generalization to their large scale data sources to obtain adequate representations at the smaller scales. This has historically been a very time-consuming and labor-intensive process, which has established the need for reliable automated solutions. The problem is further exacerbated by the frequent updates of the map necessitated by the acquisition of new data at larger scales that subsequently have to be propagated to smaller scales (Spiess et al. 2005, Li 2007, Duchêne et al. 2014). Automation of the cartographic generalization process has further increased in relevance with the proliferation of digital maps, as users of mobile devices with small screens demand the exploration of data at different scales by zooming in and out (Sester 2020). Additionally, generalization is instrumental for integrating and harmonizing the increasingly diverse data gathered from multiple sources at different scales, such as volunteered geographical information (Sester et al. 2014, Slocum et al. 2022, Touya et al. 2023).

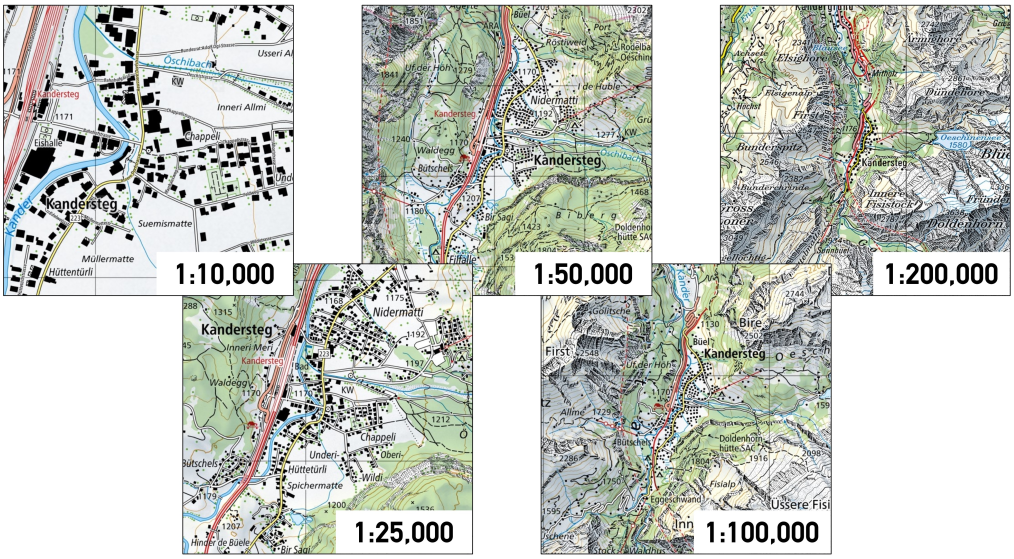


Figure 1 National maps of Switzerland at various scales (© swisstopo).[1]

Coincidentally, Eckert (1908) also recognized the importance of map logic, the principles that dictate cartographic perception and therefore form the basis for the map-making process, for addressing the dilemma between subjectivity and objectivity. Since these concepts are subconsciously applied by expert cartographers during map generalization, the resulting generalized maps can be considered collections of implicit cartographic knowledge employed throughout the generalization process (Kavouras & Kokla 2007, Varanka & Usery 2018). This observation has led to the emergence of a new family of automated generalization approaches based on deep learning (DL) that work under the assumption that highly complex statistical models can exploit existing generalized maps to discover and learn the intricacies that guide the generalization process, emulating the mind and decisions of a trained human cartographer (Touya et al. 2019, Sester 2020).

[1] map.geo.admin.ch

Buildings constitute prominent features within topographic maps. They display a variety of shapes and sizes and often appear in dense clusters, increasingly competing for map space as the scale decreases. Buildings serve as common cartographic landmarks that are paramount for orientational and navigational purposes. Consequently, the preservation of the legibility, the distinct appearance of individual structures, as well as the overall arrangement within a group of buildings when transitioning across scales is of utmost importance. This establishes the generalization of buildings as particularly intriguing (Regnauld 2001, Spiess et al. 2005, Regnauld & McMaster 2007, Yan et al. 2020).

Against the backdrop of the emerging DL paradigm and the significance of buildings in topographic maps, the present thesis seeks to explore DL for the cartographic generalization of buildings. The objectives of the research involve the investigation of the extent to which DL models can be used to predict the generalization operators that should be applied to a given building in order to obtain an adequately generalized representation. Thereby, the proposed approach represents a preparatory step in the map generalization process, facilitating the implementation of a myriad of downstream tasks. Particular emphasis is placed on the investigation of the importance of suitable contextual map features such as surrounding buildings and roads that are provided to the models as additional information for the prediction of the generalization operators. To conduct the operator classification task, models based on vector and raster representations of buildings are implemented. In a second step, the models trained on the individual modalities are integrated in a novel multimodal model capable of generating predictions by leveraging raster and vector representations simultaneously.

The remainder of the thesis is organized as follows. Chapter 2 provides an overview of map generalization by introducing important terms and concepts, reviewing conventional techniques, and outlining open challenges that such approaches are commonly faced with. Chapter 3 discusses the potential of DL for the cartographic generalization of buildings and presents a state-of-the-art literature review of the burgeoning field based on which research gaps and the appropriate research questions to address them are identified. Chapter 4 describes the dataset used to train the DL models and outlines the procedure implemented for its derivation. The methodology developed to answer the research questions involving workflow conceptualization, sampling strategies, data preprocessing, model structure, and the training and evaluation of neural network architectures is introduced in Chapter 5. Chapter 6 presents the results obtained after performing the operator classification task. The results are picked up and critically discussed in Chapter 7 to answer the previously outlined research questions while highlighting limitations and suggesting directions for future research. Finally, Chapter 8 provides concluding remarks and a summary of the work.

This thesis was developed within the scope of the Swiss National Science Foundation project *DeepGeneralization* carried out at the Department of Geography at the University of Zurich. Project partners include the Swiss Federal Office of Topography swisstopo, the French National Institute of Geographic and Forest Information IGN, and the Institute of Cartography and Geoinformatics IKG at the Leibniz University in Hanover, Germany.

2 Conventional map generalization

2.1 Generalization operators

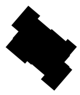
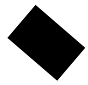
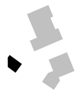
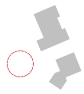
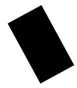
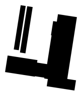
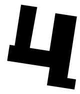
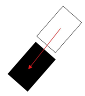
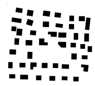
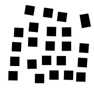
Before the advent of digital cartography, cartographers would painstakingly perform map generalization by hand, a meticulous and labor-intensive process that required considerable skill and patience (Cebrykow 2017). Established approaches for automating the cartographic generalization process are concerned with the development of algorithms that are capable of emulating the logic employed by expert cartographers to produce generalized maps that are on par with manually derived products (Zhang et al. 2024). In order to formalize cartographic knowledge for the automatic generalization process and to better grasp its complexity, researchers focused on developing conceptual models in the early days of the discipline (Sarjakoski 2007, Regnauld & McMaster 2007, Slocum et al. 2022). Brassel & Weibel (1988) propose a model that decomposes map generalization into structure recognition, process recognition, process modeling, process execution, and data display. McMaster & Shea (1992) develop a theoretical framework that delineates the *why*, *when*, and *how* of the generalization process.

The conceptual models stipulate that a profound understanding of the process is required for its successful formalization and subsequent automation. According to Armstrong (1991) and Müller (1991), the knowledge necessary to conduct generalization can be classified according to geometrical, structural, and procedural knowledge. Geometrical knowledge pertains to the location and distribution of map features. Structural knowledge relates to the structure of map features with respect to their meaning, shape, and topological relations. Procedural knowledge represents the knowledge required to control the flow of operations and orchestrate the generalization procedure (Weibel et al. 1995, Mackaness & Edwards 2002, Harrie & Weibel 2007). Generalization can be considered as a process consisting of two stages (Grünreich 1985, Müller et al. 1995): *Model* generalization involves the modification of the representations of geographic information in the database and is outside of the scope of this thesis, while *cartographic* generalization (henceforth occasionally referred to as simply *generalization*) describes the manipulation of the graphical representation of features on the map (Weibel & Dutton 1999, Regnauld & McMaster 2007, Sarjakoski 2007, Roth et al. 2011).

Most conceptual models assume that a given cartographic generalization task can be decomposed into the application of a sequence of logical operations called *generalization operators* (Regnauld & McMaster 2007). A generalization operator embodies a generic descriptor specifying the spatial transformations that are to be accomplished on a set of map features during the generalization process (Weibel & Dutton 1999, Sarjakoski 2007, Regnauld & McMaster 2007, Stanislawski et al. 2014). They represent abstract expressions of how expert cartographers envision cartographic design decisions during manual generalization and are used as means of resolving cartographic conflicts that arise due to reductions in map scale (Förster et al. 2007, Roth et al. 2011). There is no general consensus among researchers on operator taxonomy (Rieger & Coulson 1993, Li 2007). In the context of the present thesis concerned with the generalization of polygonal buildings, the following operator classification is adopted. The transformations induced by the identified relevant operators are illustrated in Table 1.

- **Simplification** denotes the reduction or displacement of vertices in a building boundary, resulting in a simplified building outline while preserving the characteristic shape of the original (Regnauld & McMaster 2007, Stanislawski et al. 2014, Zhou et al. 2023). The act of shifting vertex positions is occasionally summarized under a separate operator termed *smoothing* (McMaster & Shea 1992, Roth et al. 2011).
- **Elimination** represents the omission of insignificant buildings in highly congested areas without replacement (Förster et al. 2007, Stanislawski et al. 2014). The term *selection* is frequently employed to denote the complementary action to elimination (Regnauld & McMaster 2007).
- **Enlargement** pertains to the amplification of the area of a building while preserving its shape and proportions, guaranteeing that each building adheres to the minimum dimensions imposed by map specifications. Due to their small size in relation to other map objects, enlargement is commonly applied to buildings (Regnauld & McMaster 2007).
- **Aggregation** replaces a dense group of adjacent buildings with an artificial building enveloping the original group while maintaining the shape of its outermost geometries (Förster et al. 2007). This operator is occasionally referred to as *amalgamation* or *merge*, whereby aggregation is instead associated with a replacement feature with increased dimensionality (Regnauld & McMaster 2007, Roth et al. 2011, Stanislawski et al. 2014).
- **Displacement** constitutes a shift in the location of a building to avoid coalescence with other map objects while maintaining topological relations among the buildings (Förster et al. 2007, Regnauld & McMaster 2007, Roth et al. 2011, Stanislawski et al. 2014). Displacement of buildings is especially common along traffic and hydrographic features, which have a higher precedence in the generalization order and are superimposed on built-up areas (Kilpelainen 1994, Spiess et al. 2005).
- **Typification** refers to the controlled reduction of building density by replacing the buildings with a smaller subset of representative building polygons, preserving the characteristic distribution patterns of the original building group (Lee 1996, Regnauld & McMaster 2007, Förster et al. 2007, Roth et al. 2011, Stanislawski et al. 2014).

Table 1 Generalization operators applied to buildings (map data © swisstopo).

Operator	Original building	Generalized building
Simplification		
Elimination		
Enlargement		
Aggregation		
Displacement		
Typification		

The choice of the appropriate operator is facilitated by data enrichment, explicitly revealing the implicit relationships and structures of the map (Mackaness et al. 2014). Generalization operators can be roughly classified along a spectrum, ranging from completely independent to highly contextual (Barrault et al. 2001), whereby the degree of context-dependence is often influenced by the map scale. Simplification is an operator that is usually applied indiscriminately to buildings on a map and therefore tends to be considered context-independent. Displacement and typification are referred to as highly contextual operators, as their application is exceedingly contingent on surrounding map features such as other buildings or the road network. Although elimination, enlargement, and aggregation display some elements of independent operators in transitions from large to medium scales, they are selectively applied to buildings when generalizing from medium to small scales, where they are commonly perceived as contextual operators (Basaraner & Selcuk 2008, Wang et al. 2017, Fu, Zhou, Feng & Weibel 2024).

Map features are increasingly competing for space at smaller map scales, as a decrease in scale induces an associated quadratic reduction in the represented map area (Spiess et al. 2005, Sarjakoski 2007). Therefore, the distribution of applied operators is scale-dependent (Cecconi et al. 2002, Roth et al. 2011): Independent operators are dominant in transitions to large scales, whereas buildings are increasingly subject to contextual operators at smaller scales. The relationship between map scale and generalization operator prevalence is illustrated in Figure 2. At very small scales beyond 1:100,000, buildings are represented by continuous built-up areas, rather than individually (Spiess et al. 2005, Regnauld & McMaster 2007, Harrie & Weibel 2007).

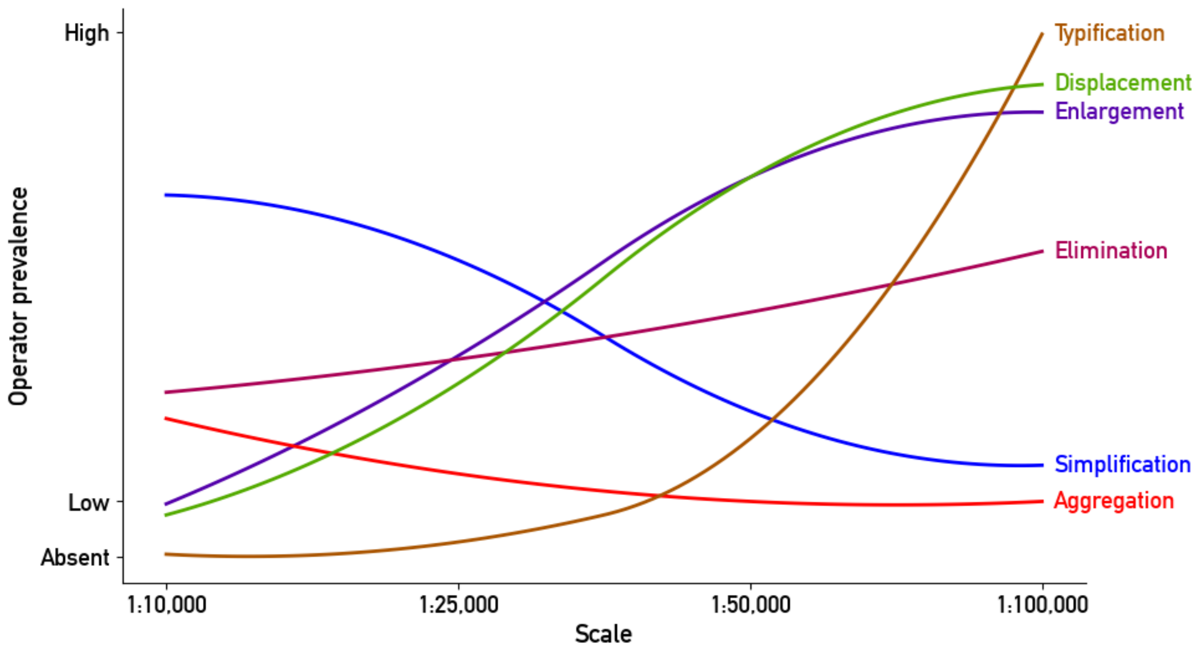


Figure 2 Relationship between scale and generalization operator prevalence (after Cecconi et al. 2002).

2.2 Approaches and techniques

Considerable research efforts have been dedicated to automating map generalization since the 1960s (Li 2007). Initial studies mainly focused on formalizing different parts of the process. Töpfer & Pillewizer (1966) propose the radical law, establishing a relationship between map scale and the appropriate number of cartographic objects that should be selected for display. Li & Openshaw (1993) introduce the natural principle, stating that details of map objects whose spatial variations at a given scale exceed the limits of what human eyes can perceive may be disregarded. The first cartographic generalization approaches were concerned with implementing generalization operators for individual features, such as the Douglas-Peucker algorithm to simplify line features (Douglas & Peucker 1973). Evaluation of the efficiency of generalization algorithms was another important research field in the early days of the discipline (Buttenfield & McMaster 1991). A comprehensive review concerning the generalization of individual map features is given by Li (2007). However, researchers soon identified the need to treat the cartographic generalization process in a more holistic manner by simultaneously considering multiple map features to resolve conflicts (Müller 1991, Sarjakoski 2007).

This observation gave rise to knowledge-based approaches (Müller 1990, 1991, Buttenfield & McMaster 1991, Armstrong 1991). Critical to their success are knowledge bases containing rules that represent the heuristic understanding of the field by an expert to make inferences and decisions (Hayes-Roth et al. 1983, Parsaye & Chignell 1988, Sarjakoski 2007). The rules are combined with structural knowledge to trigger the appropriate generalization algorithms (Harrie & Weibel 2007). To develop knowledge bases, considerable research efforts were dedicated to the task of knowledge acquisition (Buttenfield & McMaster 1991), exploiting a variety of sources of cartographic knowledge, such as cartographers through interviews, existing maps, text documents, and process tracing in interactive systems (Weibel 1995*b*, Weibel et al. 1995). To facilitate rule formalization, Weibel (1991) propose the use of semi-automated, interactive amplified intelligence techniques that involve the incorporation of various tools concerned with the enhancement of the knowledge of the cartographer who contributes the procedural knowledge (McMaster & Mark 1991, Mackaness 1995, Sarjakoski 2007, Harrie & Weibel 2007). Systems based on amplified intelligence can propose appropriate actions to the human expert who remains in charge of guiding the generalization procedure (Harrie & Weibel 2007).

The observation that successful generalization necessitates controlling for interactions between map features led to the emergence of constraint-based approaches (Weibel & Dutton 1998, Jones 2014): Map specifications can be formalized in the form of various constraints that have to be satisfied to the best possible extent during an iterative generalization process (Beard 1991). Constraint-based approaches allow for the evaluation of different strategies that support backtracking should the result not be satisfactory (Harrie & Weibel 2007, Mackaness 1995). The developments up until the turn of the millennium culminated in the emergence of agent-based approaches that involve modeling map features as a set of communicating and interacting cartographic agents whose target is to achieve an optimal generalization state with respect to themselves and to other map features. Agents have the capacity to evaluate their current generalization state using cartographic constraints at the micro, meso, and macro levels, based on which generalization algorithms are applied to resolve possible conflicts (Lamy et al. 1999, Barrault et al. 2001, Ruas & Duchêne 2007, Sarjakoski 2007, Duchêne et al. 2018). The generalization engine of the agent-based approach is depicted in Figure 3. Although recent years have seen a consolidation or perhaps even a decline in research activity surrounding automated cartographic generalization (Sester 2020), other constraint-based approaches based on optimization techniques such as genetic algorithms (Wilson et al. 2003), least-squares adjustment (Sester 2005), and simulated annealing (Ware et al. 2003) have emerged that seek to challenge the dominant agent-based paradigm.

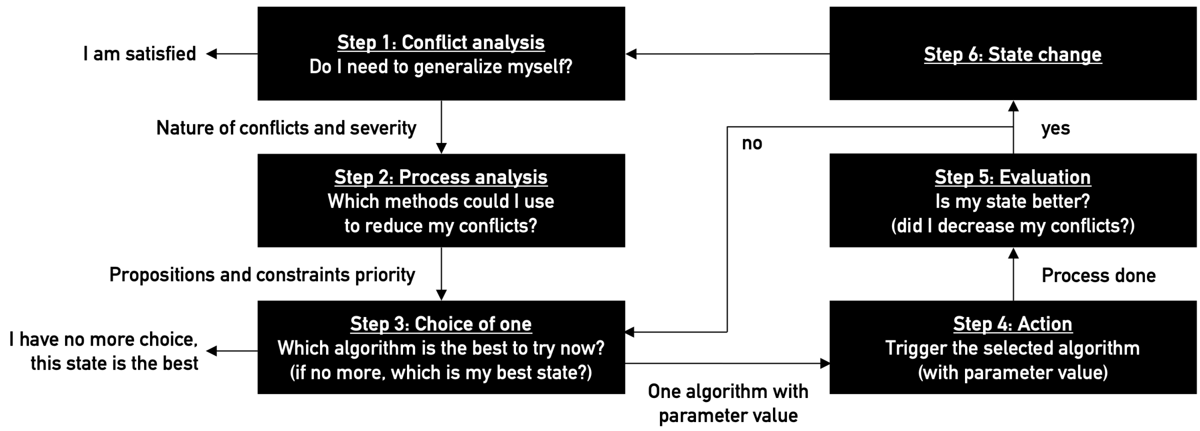


Figure 3 Generalization engine of the agent-based approach (Ruas & Duchêne 2007).

2.3 Limitations and challenges

The generalization engine associated with agent-based approaches illustrated in Figure 3 demonstrates the necessity ubiquitous in conventional approaches of explicitly defining ambiguous terms such as conflict, constraint, algorithm, and parameters to ensure adequate generalization results. In practice, this task is performed by expert cartographers who can rely on their intuition and expertise to determine the most sensible course of action rather than following a strict set of rules, leading to inconsistent results (Lee 1996, Cebrykow 2017, Fu, Zhou, Feng & Weibel 2024). Therefore, the map generalization process is significantly more iterative and non-linear compared to what is implied by rule-based or constraint-based systems (Mackaness 1995, Ruas & Duchêne 2007, Kang et al. 2024). According to Parsaye & Chignell (1988), Nyerges (1991) and Rieger & Coulson (1993), expert cartographers do not consciously understand the framework of their knowledge, and although they find the steps in their reasoning process straightforward, they may fail to articulate a comprehensive and detailed explanation necessary for practical implementation (Sarjakoski 2007). Therefore, map specifications commonly contain only superficial knowledge (Müller et al. 1995). Due to an incomplete understanding of the process and a lack of suitable frameworks for knowledge representation, even the most sophisticated conventional approaches are unable to holistically encapsulate the subjective and intuitive nature associated with cartographic generalization (Li & Su 1995, Su 1996, Xiao et al. 2024).

The difficulty of explicitly formalizing the implicit knowledge employed by expert cartographers during the generalization process has been coined the *knowledge acquisition bottleneck* by Weibel et al. (1995). The scarcity of available cartographic expert knowledge that can be harnessed in the form of rules or constraints has significantly hindered the development of automated generalization approaches (Stanislawski et al. 2014, Zhou et al. 2023). As none of the conventional approaches have been able to effectively resolve the knowledge acquisition bottleneck, automated solutions only tackle subtasks of the generalization process, whereas there are currently no fully automated end-to-end generalization solutions that can rival the quality of manually generalized maps (Duchêne et al. 2014, Harrie et al. 2024, Zhang et al. 2024).

According to Touya et al. (2023, p. 345), map generalization can be considered a "way of modeling the interactions between entities rather than the entity itself". Therefore, the generalization of a given map feature must be performed by taking into account a wider context of mutual dependencies (Spiess et al. 2005). While an abundance of satisfactory solutions exist for generalizing individual features, the orchestration of generalization algorithms, referring to the optimal choice and sequencing of contextual operators for a given situation, remains an unsolved issue (Stanislawski et al. 2014, Sester 2020, Courtial et al. 2021*b*). This problem is especially pronounced at smaller scales, where limited space is available and there are many possibilities for interactions and the associated propagation of conflicts (Ware et al. 2003, Lee et al. 2017, Duchêne et al. 2018, Feng et al. 2019). Therefore, the state-of-the-art for cartographic generalization employed by NMAs continues to consist of semi-automated approaches that involve considerable manual intervention by expert cartographers (Bićanić & Solarić 2017).

3

Deep learning for building generalization

3.1 Rationale

Following the identification of the knowledge acquisition bottleneck, Weibel et al. (1995) propose the use of data-driven strategies in the form of machine learning techniques to address the scarcity of explicitly formalized cartographic knowledge that can be exploited for map generalization. Machine learning (ML) constitutes a specialized area within the broader discipline of artificial intelligence (AI) and refers to models that automatically improve their performance by detecting patterns and structures within large amounts of training data for which the desired output is known (Mitchell 1997, Alpaydin 2020). Artificial neural networks (ANNs) serve as a pivotal technology within the domain of ML. The architecture and learning process of ANNs are illustrated in Figure 4. ANNs operate by processing input data through layers of interconnected neurons, each of which applies specific mathematical operations to the data. The connections between these neurons have weights that are adjusted during training to minimize the difference between the network output and the desired outcome. This process enables the network to learn sophisticated patterns and make predictions based on input data, simulating a simplified version of the way human brains operate (Schmidhuber 2015).

In the context of cartographic generalization, ML techniques can be applied to existing generalized maps to learn the implicitly encoded knowledge infused by expert cartographers during map generalization, essentially reverse engineering the generalization process. As they learn from examples, ML-driven approaches are capable of gradually imitating the behavior of human cartographers (Sester 2020). Therefore, they display the potential to overcome the knowledge acquisition bottleneck by avoiding the explicit formulation of cartographic knowledge altogether (Weibel et al. 1995). ML models are able to develop internal schemata of cartographic knowledge, discovering unique methods to solve problems that differ from human cognition (Zhang et al. 2024). In light of these observations, the application of AI techniques for automated generalization has its roots in the early days of the discipline (Weibel 1991).

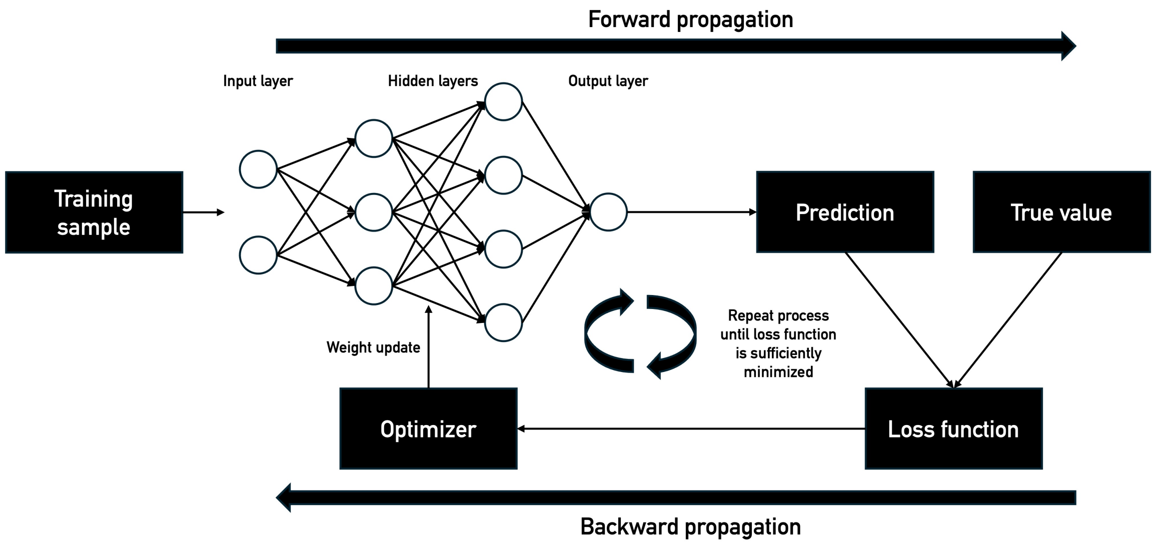


Figure 4 Architecture and learning process of ANNs.

Initial approaches based on ML were mainly concerned with facilitating the decisions made by cartographers through the enrichment of data with implicit structures and relations, the acquisition of procedural knowledge to orchestrate and parameterize algorithms, and the evaluation of generalized maps (Touya et al. 2019, Fu, Zhou, Feng & Weibel 2024). Sester (2000) was among the first to propose a framework that applies ML to maps to explicitly reveal the implicit information contained within that can subsequently be harnessed to improve existing generalization algorithms. Similarly, Lagrange et al. (2000) leverage ML to determine the parameters of cartographic generalization algorithms based on a collection of measures describing the feature undergoing transformation. Steiniger et al. (2008, 2010) propose approaches based on discriminant analysis and decision trees to classify buildings across various geographic settings to aid in the selection of the appropriate generalization operators based on a combination of expert and learned rules. Lee et al. (2017) use different ML models to determine whether a given building is to be eliminated, retained, or aggregated. Cheng et al. (2013, 2015) propose ANNs that are capable of simplifying and aggregating individual building outlines. Yang et al. (2022) develop an ANN that identifies the ideal simplified representation of a building group among four conventional generalization algorithms.

However, it turns out that conventional ML-based approaches are limited in their ability to directly produce generalized output maps, since ML models struggle to process raw map data and typically require extensive feature engineering to achieve satisfactory performance (LeCun et al. 2015). Therefore, the application of ML in automated cartographic generalization has mainly been confined to the steps preceding the actual map generalization (Fu, Zhou, Feng & Weibel 2024). The limitations that conventional ML-based approaches are faced with led to the emergence of deep learning (DL) as a new sub-field of ML and AI. The domain of DL is concerned with the study of deep ANNs that are capable of representation learning, referring to models that automatically identify the features necessary for detection or classification directly from the raw data (LeCun et al. 2015). The influx of large amounts of geospatial data coupled with the widespread availability of high-performance computing has led to the dispersion of DL techniques in the domain of geography, coining the term *Geospatial Artificial Intelligence*, commonly abbreviated as GeoAI (Janowicz et al. 2020, Li 2020, Gao et al. 2023). Methods from the field of GeoAI are already used in cartographic practice to support the production of topographic maps (Usery et al. 2022).

Given large amounts of high-quality data, DL models have proven to excel at tackling problems that are difficult to formally describe and were previously relegated to being solvable only by humans, such as tasks from natural language processing and image recognition (Goodfellow et al. 2016). Since cartographic generalization is a prototypical example of such a problem, DL has emerged as an intriguing candidate for the next paradigm shift to automate the map generalization process, providing a new way of acquiring knowledge (Touya et al. 2019, Yan et al. 2020). The introduction of DL into data-driven cartographic generalization approaches opens up the possibility of end-to-end solutions, as the models are powerful enough to produce generalized maps. The novel DL paradigm is further supported by the abundance of existing generalized map series that can be used to train models (Feng et al. 2019). According to Courtial et al. (2024), DL has the potential to contribute to the automation of cartographic generalization through data enrichment, as a generalization operator, and by holistically generalizing maps.

3.2 Proposed solutions

The emergence of the DL paradigm has resulted in a plethora of studies that apply GeoAI models to an array of problems in cartography, such as pattern recognition and preliminary data enrichment (e.g., Touya & Lokhat 2020, Li et al. 2024), map style transfer (e.g., Kang et al. 2019, Christophe et al. 2022), map labeling (e.g., Li et al. 2020, Oucheikh & Harrie 2024), and cartographic generalization, the latter of which has been identified as the cartographic design task most commonly supported by GeoAI (Kang et al. 2024, Harrie et al. 2024). While the present thesis is concerned with the generalization of polygonal buildings, DL has also been applied to generalize different linear map features, such as roads (e.g., Courtial et al. 2020, 2023, Zheng et al. 2021, Beglinger 2023), rivers (Du, Wu, Yin, Liu & Gong 2022), coastlines (e.g., Du, Wu, Xing, Gong & Yu 2022, Du, Wu, Zhu, Liu & Wang 2022, Jiang et al. 2023), and contour lines (Yu & Chen 2022).

The overarching idea of using DL for the generalization of buildings on maps is to use existing map series to train models that are supposed to emulate the decisions taken by an expert cartographer who is carrying out the generalization process (Sester 2020). Figure 5 illustrates a typical building generalization workflow. Most DL-based end-to-end cartographic building generalization approaches are implemented by adopting a supervised learning technique since the input and output domains tend to be fairly similar: DL models are presented with paired training samples, whereby one sample represents a version of a building (or group of buildings) prior to generalization and the other depicts the corresponding generalized situation.

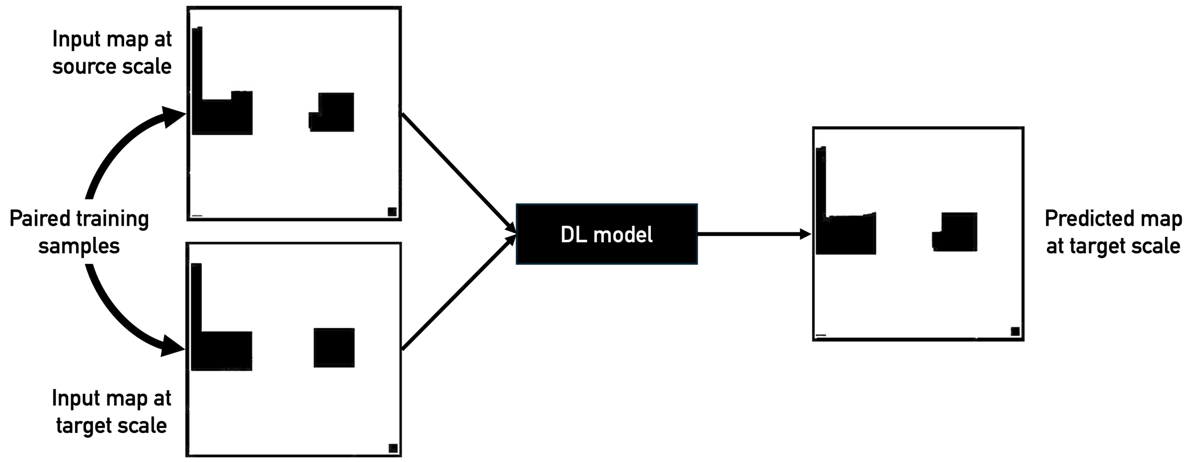
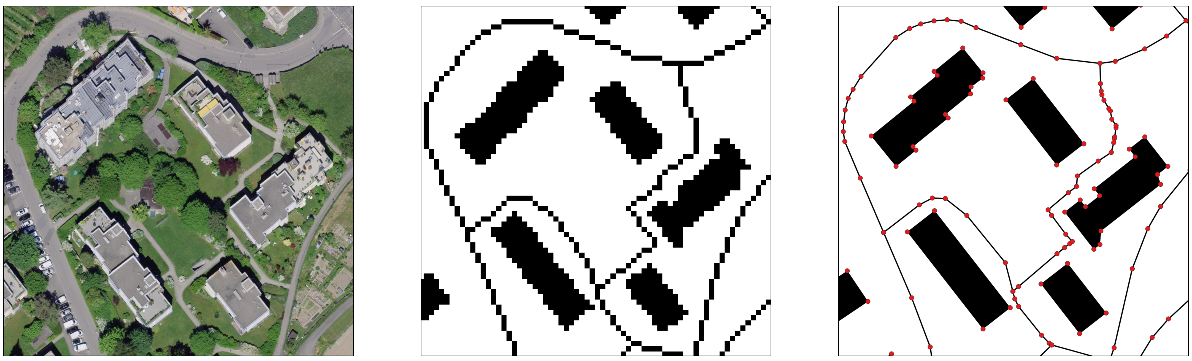


Figure 5 Supervised building generalization workflow using DL (after Fu, Zhou, Feng & Weibel 2024).

Unsupervised training for DL-driven cartographic generalization tasks is less common. It consists of supplying unpaired samples to the model that is subsequently tasked with figuring out the properties of the domains to perform the generalization (Courtial et al. 2021*b*). Based on large abundances of these training samples, the model is expected to adjust its parameters such that it learns to recreate the generalized from the non-generalized map. Throughout the training process, the performance of the model is evaluated by comparing the map generalized by the model with the true map at the target scale. After successful training, the model is supposed to be able to apply the knowledge obtained from the training data to non-generalized maps not processed during training to create a generalized output (Fu, Zhou, Feng & Weibel 2024, Kang et al. 2024, Schmidhuber 2015). Such techniques have recently been explored for developing a variety of end-to-end building generalization solutions that implement a myriad of generalization operations. As illustrated in Figure 6, cartography primarily utilizes two data formats, vector and raster, for the storage and representation of geospatial information (Peter & Weibel 1999). Consequently, existing DL-based building generalization approaches can also be classified according to the two modalities based on the underlying data structure used to derive the encodings of the cartographic data for the DL models (Yan & Yang 2022).



(a) Aerial image **(b)** Raster representation **(c)** Vector representation

Figure 6 Vector and raster representations of buildings and roads in maps (map data © swisstopo).

3.2.1 Raster-based approaches

Raster data consists of a grid of pixels, where each pixel is associated with a value representing the characteristics of the area it covers, as exemplified in Figure 6b (Kang et al. 2024). Existing building generalization approaches based on raster data mainly rely on binary rasters, where each pixel in the image denotes the presence or absence of a building part at the respective location. The fact that maps that are represented as rasters are structured identically to conventional images enables the application of established DL techniques and models from the domain of computer vision (Touya et al. 2019). Computer vision encompasses the principles and development of artificial systems designed to interpret and understand the content of images by extracting information from them (Danuser 2011). Techniques from the domain of computer vision have the potential to contribute to automated map generalization in the form of image classification and segmentation tasks (Touya et al. 2019). Whereas image classification techniques are concerned with assigning a single label to a given input raster, image segmentation allows for pixel-wise predictions (Long et al. 2015). Therefore, segmentation models are capable of directly outputting a generalized map through image-to-image translation (Sester et al. 2018).

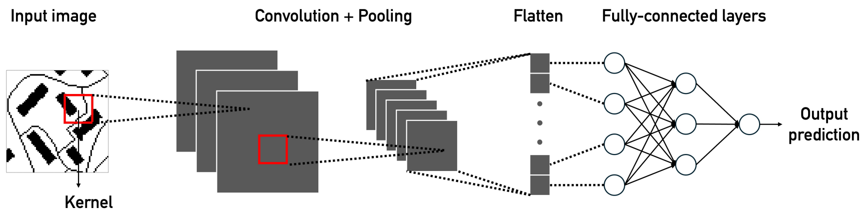
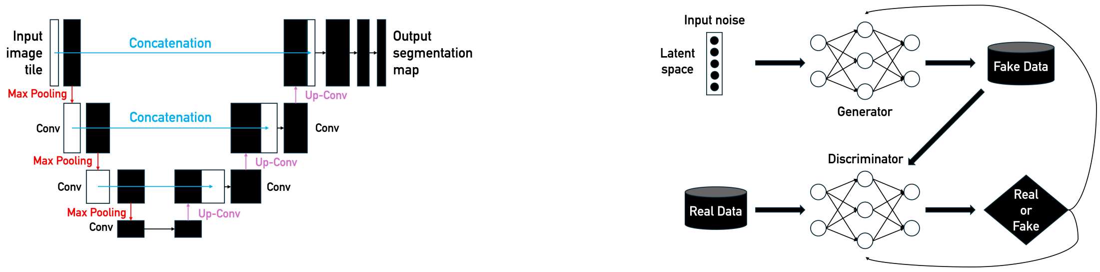


Figure 7 Architecture of a CNN (map data © swisstopo).

Convolutional neural networks (CNN, LeCun & Bengio 1998) are specialized DL architectures whose applications are ubiquitous in the domain of computer vision (LeCun et al. 2015), having demonstrated exceptional performance for a variety of classification tasks, such as image analysis and pattern recognition (Krizhevsky et al. 2017). The architecture of a CNN is displayed in Figure 7. CNNs extend conventional neural networks by incorporating various convolutional and pooling layers. Convolutional layers detect correlated local groups in the image such as edges, whereas pooling layers merge semantically similar features, thereby reducing the dimensionality of the data. A sequence of multiple convolution and pooling operations allows for effective feature extraction that can then be processed and classified by a series of fully-connected layers (LeCun et al. 2015). U-Nets (Ronneberger et al. 2015) are a specialized version of CNN designed for image segmentation; their architectures are illustrated in Figure 8a. U-Net features a symmetric architecture that first contracts the input image using a sequence of convolution and pooling operations, after which the input is expanded again using a series of up-convolution layers, reconstructing the segmentation map. Consequently, U-Nets can output a label for each individual pixel in the raster, rather than a single label for the whole raster. They additionally involve skip connections that connect the layers in the encoder path with the corresponding layers in the decoder path, helping to propagate information throughout the network (Ronneberger et al. 2015). Residual U-Nets (ResU-Net, Zhang et al. 2018) extend traditional U-Nets by including residual units (He et al. 2015) that further facilitate information propagation. Attention U-Nets (AttU-Net, Oktay et al. 2018) incorporate a self-attention mechanism (Vaswani et al. 2017) that emulates cognitive attention, directing the focus of the neural network toward crucial patterns hidden within the data (Fu, Zhou, Feng & Weibel 2024).



(a) U-Net **(b)** GAN

Figure 8 Architectures of DL models applied in raster-based approaches.

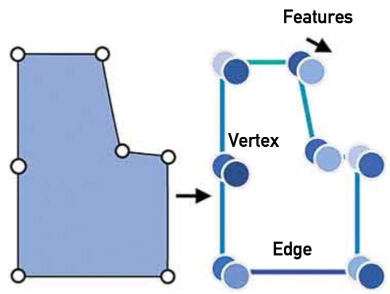
Generative adversarial networks (GAN, Goodfellow et al. 2014) are yet another class of generative models that are commonly used for raster-based building generalization approaches. The GAN architecture is illustrated in Figure 8b. A GAN is composed of two neural networks, a generator and a discriminator, that are trained concurrently in an adversarial manner. The generator (e.g., an image segmentation model such as a U-Net) is tasked with generating fake data that look as real as possible. The discriminator (e.g., an image classification model such as a CNN) is a classifier that learns to classify the generator output into real and fake data. The training process involves the optimization of both models, by which the generator produces progressively more realistic data, while the discriminator improves in distinguishing the fake data from the real data. Training continues until an equilibrium is reached that occurs once the fake data produced by the generator become so convincing that the discriminator cannot distinguish it from the real data (Goodfellow et al. 2014). According to Courtial et al. (2021*b*), GANs can be trained in a supervised (e.g., pix2pix, Isola et al. 2016) or unsupervised manner (e.g., CycleGAN, Zhu et al. 2017).

Sester et al. (2018) were among the first to propose the application of U-Nets for supervised learning of the elimination, simplification, and aggregation of buildings in a single model by providing it with rasterized map tiles at large scales. Their approach is further refined by Feng et al. (2019), extending the models applied to ResU-Nets and GANs. The models outperform various baseline methods, whereby the ResU-Net achieves the best results, improving the preservation of straight building outlines and corners. Kang et al. (2020) apply specialized GAN architectures that can encode cartographic knowledge in the form of various geometric transformations in the DL process (Fu et al. 2019), outperforming conventional models. Courtial et al. (2021*b*) adopt GANs to generalize various map features in urban areas, such as buildings, roads, and rivers simultaneously. Compared to previous studies that focused on map scales larger than 1:25,000, their approach targets medium scales up to 1:50,000, where typification is the dominant operator. The building maps generated by the GANs satisfy various evaluation constraints relating to building structure, orientation, and relative density. Courtial et al. (2022*b*) extend the established approaches by proposing a layered data representation model based on a GAN that allows for the incorporation of contextual information into the DL process in the form of additional map features such as the surrounding road network or semantic information, improving the quality of generalized building maps. Courtial et al. (2024) speculate that a unique model is not sufficient to implement holistic map generalization. Similarly to traditional map generalization approaches, they propose a decomposition of the process into smaller tasks that can be individually solved by specialized, fine-tuned DL models. Their devised workflow further boosts the quality of building generalization results compared to their previous approach. Several of the aforementioned studies report that buildings generalized by DL models suffer from unrealistic shapes in the form of fuzzy and deformed boundaries (Sester et al. 2018, Feng et al. 2019, Courtial et al. 2021*b*, 2024). Fu, Zhou, Feng & Weibel (2024) address this problem by proposing multi-channel ResU-Nets and AttU-Nets that store the building to be generalized and its context buildings in separate channels. In conjunction with an abundance of training samples, the buildings generalized by the models exhibit straight walls while maintaining their characteristic rectangularity and parallelism. Zhou et al. (2024) explore this data model in combination with a spatially aware GAN to determine that DL models need to accurately understand geometrical characteristics and spatial relationships to effectively learn how to generalize buildings. Fu, Zhou, Xin & Weibel (2024) investigate how different pixels on the source map contribute to the prediction of the generalized map. They discover that the models primarily focus on building boundaries and the space between buildings, indicating that the aforementioned DL models account for the spatial building layout in a manner that resembles the cartographic knowledge employed by humans. Recently, approaches based on diffusion models (Feng 2023) and Swin transformers (Winkler 2023) have been proposed, demonstrating the potential of novel DL models for cartographic building generalization.

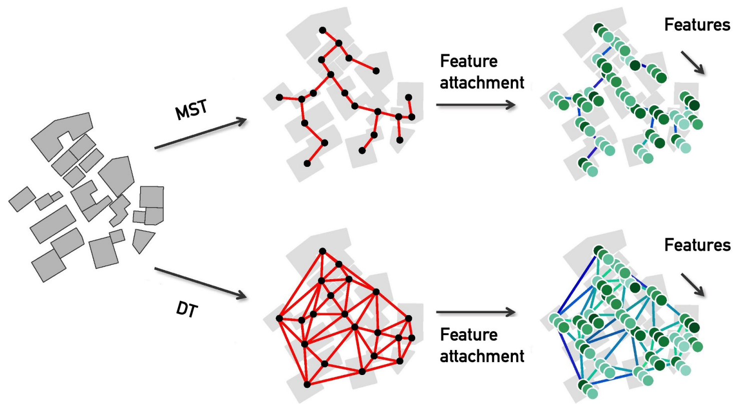
Although the raster representation is the most straightforward way of encoding a map for DL purposes, it is also associated with various issues. The process of converting vector to raster data leads to information loss, as it introduces fuzziness and uncertainty into the generalization procedure (Liao et al. 2012, Knura 2024). Consequently, rasters can only represent implicit spatial relationships, do not allow for the distinction between overlapping features, and restrict the analysis to a fixed image size with a limited number of pixels (Harrie et al. 2024, Touya et al. 2019). Furthermore, important contextual and semantic information cannot be captured using rasters (Courtial et al. 2024) and it remains a challenge to encode prior cartographic knowledge in the DL process (Kang et al. 2020). The application of a constraint-based evaluation to assess the quality of the rasterized output map also proves to be difficult (Stoter et al. 2014, Courtial et al. 2022*a*). The choice of the appropriate pixel-based loss function used during training was identified as a constraining factor in a large number of existing approaches (Knura 2021). Although some of these limitations can be addressed through the use of layered representations (Courtial et al. 2022*b*) or multi-channel models (Fu, Zhou, Feng & Weibel 2024, Winkler 2023), there has recently been an emergence of approaches that directly seek to exploit vector data as input for DL-based cartographic building generalization to overcome these shortcomings.

3.2.2 Vector-based approaches

Vector data represents each geographical feature as a unique object consisting of points, polylines, or polygons, as illustrated in Figure 6c. These objects include both spatial geometry and descriptive attributes and are subject to manipulation through various spatial operations (Kang et al. 2024). Approaches that leverage vector data are more flexible, as they have the ability to process inputs of arbitrary size, which means that they are not confined to fixed geographical extents (Zhou et al. 2023, Harrie et al. 2024). Compared to raster data, vector data contains more information that can be exploited by DL models during the generalization process, such as explicit topological relationships and attribute richness (Touya et al. 2019). However, this also implies that vector-based approaches are inherently more complex, since they require modeling of relationships such as topology and connectivity (Regnauld & McMaster 2007). Additionally, the data that form the foundation of maps are usually provided in the vector format (Knura 2024). Therefore, employing vector data as input to DL models can eliminate the need for the rasterization process and its associated issues, while still retaining elements of the traditional map generalization workflow for subsequent modifications (Knura 2021).



(a) Individual buildings (Yan et al. 2021)



(b) Groups of buildings (Yan et al. 2019)

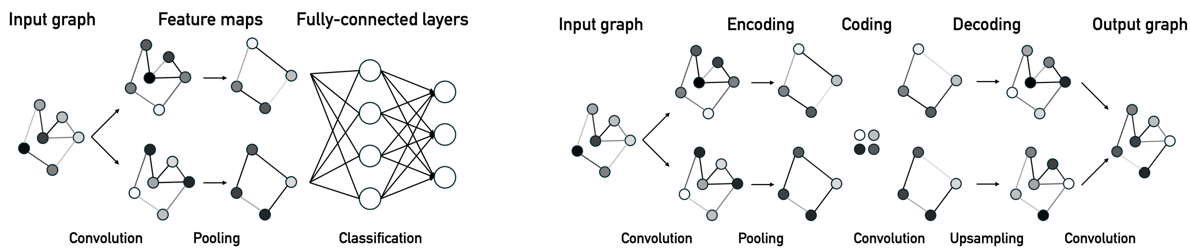
Figure 9 Graph construction techniques applied to vector-based buildings.

However, DL models are characterized by a requirement for regularly structured and normalized input that spatial vector data do not fulfill due to their unstructured nature and non-stationary neighborhood structures (Yan et al. 2019, Knura 2024). Consequently, encoding vector data for DL has proven much harder compared to raster data. Although there have been initial proposals for general-purpose representation learning approaches for polygonal geometries (van 't Veer et al. 2018, Mai et al. 2023), there are currently no widely available DL models that directly process vector data. Therefore, vector-based approaches have received less attention than their raster-based counterparts. Existing solutions leverage undirected graph representations that have shown a significant potential to encode vector data for DL (Mai et al. 2022). Figure 9 shows how individual buildings and building groups can be represented by a graph. To capture spatial relationships among adjacent buildings, neighborhood graphs such as minimum spanning tree (MST, Borůvka 1926) and Delaunay triangulation (DT, Delone 1934) are commonly applied

(Touya et al. 2023). The graph structure conserves the characteristics of individual features, as well as the relationships among them (Yan et al. 2020) and further allows for the attachment of descriptors to individual nodes and edges in the form of a feature matrix that can contribute important information for the generalization process (Knura 2024).

Neural networks that are particularly adept at processing graph-based data commonly employed in existing DL-based building generalization approaches are called graph neural networks (GNN, Scarselli et al. 2009). The core principle of GNNs revolves around the notion of message passing, which involves the aggregation of information from the neighbors of a node, thus updating its representation through multiple layers. As a result of this iterative process, each node develops a dense, lower-dimensional vector called an embedding that captures its intrinsic features in addition to the structural context provided by its connections in the graph. These embeddings allow GNNs to learn complex patterns across the graph, making them highly versatile for node, edge, and graph-level prediction tasks (Sanchez-Lengeling et al. 2021, Bronstein et al. 2021).

Specialized versions of GNNs further extend their applicability and performance. Graph convolutional neural networks (GCNN, Kipf & Welling 2016*a*) represent a generalization of the convolution operation from grid-like structures (as in CNNs) to arbitrary graphs, enabling efficient neighborhood aggregation and feature learning (Daigavane et al. 2021). The structure of GCNNs is illustrated in Figure 10a. Graph autoencoders (GAE, Kipf & Welling 2016*b*) try to learn representations of graph data in an unsupervised manner by encoding the graph into a latent space (often using a GCNN) to generate embeddings for each node. Subsequently, a decoder attempts to reconstruct the graph properties from these embeddings. If the properties are accurately reconstructed, the embeddings generated by the encoder must accurately capture the structure of the graph (Zhang, Chen, Wang, Li, Bai & Hancock 2019). The GAE architecture is illustrated in Figure 10b. Graph attention networks (GAT, Veličković et al. 2017) leverage an attention mechanism to dynamically weigh the importance of neighboring contributions, which enhances the ability of the model to focus on relevant parts of the graph structure. Finally, GraphSAGE (Graph Sample and Aggregate, Hamilton et al. 2017) is an architecture for inductive representation learning on large graphs designed to generate embeddings for previously unseen data using a learned function that aggregates information from the local neighborhood of a node (Daigavane et al. 2021).



(a) GCNN **(b)** GAE

Figure 10 Architectures of DL models applied in vector-based approaches.

The application of vector-based methods is mainly confined to pattern recognition and data enrichment (Zhou et al. 2023). These tasks are considered important precursor steps for map generalization, as the explicit unveiling of the implicit patterns contained within a map can facilitate the choice of the appropriate generalization operators (Harrie et al. 2024). GCNNs were recognized as an effective model for recognizing patterns among building groups in a plethora of studies. Yan et al. (2019) develop a GCNN that manages to distinguish regular from irregular building groups that are represented as graphs using MST and DT. Bei et al. (2019) propose a model that performs both building group division and classification of building groups according to various pattern types without ancillary information such as road and river networks. Zhao et al. (2020) model buildings belonging to street blocks based on a constrained DT to detect building group patterns of varying shapes. Yan et al. (2020) represent building groups as graphs using a tessellation based on Voronoi polygons and various cognitive variables to recognize building groups based on their spatial configurations. Li et al. (2024) propose a method not based on GNNs to classify building groups according to their regularity by modeling them as point clouds and subsequently applying a specialized DL architecture.

In addition to processing groups of buildings, GCNNs are also frequently leveraged to recognize the shapes of individual buildings. Yan et al. (2021) use a GAE to distinguish buildings based on their shape. Liu et al. (2021) propose a method that avoids graph construction and feature extraction by applying a deep point convolutional network directly to the vector data to classify buildings according to their shape. Hu et al. (2022) expand on their work by developing a relation network that predicts building shape types based on few labeled samples. Yan & Yang (2022) propose an encoder-decoder framework for encoding buildings represented as graphs, sequences, and rasters to retrieve building shapes, whereby the graph-based model performed best. Knura (2024) develops various encoding schemes in combination with recurrent neural networks, CNNs, and GCNNs to show how building shape recognition approaches can be used for cartographic generalization by replacing buildings with simplified versions according to a template matching approach (Rainsford & Mackaness 2002, Yan et al. 2017).

Recently, researchers have started exploring vector-based approaches for potential end-to-end solutions. Xiao et al. (2024) propose a GCNN for the generalization of point clusters, which they apply for the selection of buildings abstracted as points at very small scales. Zhou et al. (2023) formulate building simplification in terms of moving and removing polygon vertices. They propose a multi-task learning method based on GraphSAGE to simplify building outlines, outperforming GCNN and GAT. Feng et al. (2023) model polygons as a data sequence and subsequently apply a transformer model for building simplification. Yan & Yang (2024) propose a GAE for the simplification of building outlines allowing for flexible configuration of constraints by adjusting the loss functions, obtaining better results compared to established approaches. Additionally, there are several DL-based approaches conceptualized for the simplification of linear vector features, such as roads (Beglinger 2023), coastlines (Du, Wu, Zhu, Liu & Wang 2022, Jiang et al. 2023), rivers (Du, Wu, Yin, Liu & Gong 2022), and contour lines (Yu & Chen 2022), which could potentially be extended to the simplification of building outlines.

3.3 Research gaps

A summary of DL-driven approaches for the generalization of buildings is provided in Table 2.

Table 2 End-to-end building generalization approaches based on DL with underlying modality, DL architectures, learning techniques, modeled operators, generalization strategies used to derive the smaller scale representations used for training, and scale transitions.

Publication	Modality	Architecture	Learning technique	Operators	Strategy	Scale
Sester et al. 2018	raster	CNN (U-Net)	supervised	aggregation simplification elimination	CHANGE	1:5k to 1:10k 1:5k to 1:25k 1:5k to 1:50k
Feng et al. 2019	raster	CNN (U-Net, ResU-Net) GAN (ImageGAN, PatchGAN)	supervised	aggregation simplification elimination	CHANGE	1:5k to 1:10k 1:5k to 1:25k 1:5k to 1:50k
Kang et al. 2020	raster	GAN (CycleGAN, GcGAN)	unsupervised	simplification elimination	ArcGIS toolbox	1:5k
Courtial et al. 2021*b*	raster	GAN (pix2pix, CycleGAN)	supervised & unsupervised	typification	CartAGen	1:25k to 1:50k
Courtial et al. 2022*b*	raster	GAN (pix2pix)	supervised	typification	CartAGen	1:25k to 1:50k
Courtial et al. 2024	raster	GAN (pix2pix)	supervised	enlargement displacement typification elimination	CartAGen	1:25k to 1:50k
Winkler 2023	raster	CNN (U-Net) Swin Transformer	supervised	holistic	swisstopo CHANGE	1:10k to 1:25k
Fu, Zhou, Feng & Weibel 2024	raster	CNN (ResU-Net, AttU-Net)	supervised	aggregation simplification elimination	CHANGE	1:5k to 1:10k 1:5k to 1:15k
Zhou et al. 2024	raster	GAN (pix2pix)	supervised	aggregation simplification	CHANGE swisstopo	1:5k to 1:10k 1:10k to 1:25k
Zhou et al. 2023	vector	GNN (GCNN, GAT, GraphSAGE)	supervised	simplification	CHANGE	1:5k to 1:10k
Yan & Yang 2024	vector	GAE	unsupervised	simplification	ArcGIS toolbox	1:10k to 1:25k

Based on the existing literature, the following research gaps (RGs) are identified.

- **RG0**: *Purpose-built training dataset with annotated generalization operators*.

As can be seen in Table 2, existing approaches leveraging DL for building generalization almost exclusively utilize training datasets where the smaller scale representations were derived from the source data through established map generalization software such as CHANGE (Powitz 1993) or CartAGen (Renard et al. 2011). Exploiting such data for DL model training can have the effect that the model simply imitates existing, in many regards unsatisfactory algorithmic solutions (Touya et al. 2019). Furthermore, these software packages typically only implement a subset of the generalization operators commonly applied to buildings (Regnauld & McMaster 2007) and do not reflect expert knowledge exercised by human cartographers, implying that models trained on such data may not have access to sufficient information to perform the generalization task (Courtial et al. 2021*b*). The few existing datasets are increasingly reused across multiple studies such as the Stuttgart dataset used for end-to-end generalization by Sester et al. (2018), Feng et al. (2019), Zhou et al. (2023, 2024), Fu, Zhou, Feng & Weibel (2024) or the Shanghai dataset exploited for shape cognition by Yan et al. (2021), Liu et al. (2021), Hu et al. (2022), Yan & Yang (2022), Knura (2024). This suggests a possible lack of diversity and is presumably due to the effort required to derive datasets suitable for training DL models. Since the training datasets leveraged by existing approaches simply reflect the original and generalized maps, they do not explicitly contain information regarding the operators that were applied to individual buildings. Therefore, these datasets may be imbalanced with respect to generalization operators, which can be a serious problem for DL models. This establishes the need for a large, diverse, purpose-built training dataset where building geometries generalized under the supervision of expert cartographers are annotated with the operators applied during the generalization process (Fu et al. 2023, Senn et al. 2024).

- **RG1**: *Prediction of generalization operators*.

Due to the lack of purpose-built training datasets with annotated operators identified in RG0, most of the established approaches neglect individual generalization operators in favor of ambitious end-to-end solutions that directly output a generalized map. Consequently, existing solutions based on DL deviate from established cartographic practice, as they are unable to decompose the generalization process into the application of individual operators (Courtial et al. 2021*b*). Although many of the approaches outlined in Table 2 have successfully leveraged DL to subject buildings to a handful of generalization operators, the question of whether these operators should be applied in the first place has received little attention. For example, the application of a model capable of typifying buildings (Courtial et al. 2021*b*, 2022*b*) may not be sensible for map sections with sparse building distributions, where other operators are more effective. To the best of the author's knowledge, there are currently no DL-driven approaches that seek to predict the generalization operators that should be applied to a given building. Comparable approaches only classify individual generalization operators and either rely on ML techniques requiring extensive feature engineering (Steiniger et al. 2008, 2010, Lee et al. 2017) or focus on map features other than buildings, such as roads (Zheng et al. 2021, Courtial et al. 2021*a*). DL has the potential to make important contributions to the complex question regarding the generalization operators that should be applied to a building in a given situation.

- **RG2**: *Incorporation of contextual information and operators*.

Understanding spatial context is crucial for addressing problems of geographical nature (Goodchild 2018). Therefore, it is necessary to consider surrounding features such as other buildings and the road network in order to apply contextual operators for building generalization (Barrault et al. 2001). Apart from the studies conducted by Courtial et al. (2021*b*, 2022*b*, 2024), the publications outlined in Table 2 only incorporate limited cartographic context in the form of buildings in the immediate vicinity of the features subject to generalization, commonly neglecting the effect of the surrounding road network on the generalization outcome (Courtial et al. 2024, Harrie et al. 2024). Consequently, existing studies focus on scale transitions from large to medium scales (larger than 1:25,000). At these scales, independent operators such as the simplification of individual buildings are dominant, for which there are a plethora of satisfying conventional solutions (Touya et al. 2019). Therefore, not much attention has been paid to the application of DL for modeling the more complex contextual generalization operators such as displacement and typification at the building group level (Fu et al. 2023), which have been eluding researchers concerned with automating the cartographic generalization process for years (Ruas 2001, Regnauld & McMaster 2007).

- **RG3**: *Vector-based and multimodal approaches*.

As evident in Table 2, most existing DL-based end-to-end approaches are focused on exploiting raster data for the cartographic generalization of buildings, since the representation of maps as rasters is intuitive and straightforward (Touya et al. 2019). Vector-based approaches, on the other hand, are underrepresented, since they require considerable feature processing and engineering to obtain encodings that can be effectively processed by DL models, currently confining them to the simplification of individual buildings (Zhou et al. 2023, Yan & Yang 2024). Additional research is needed to investigate optimal vector geometry encodings for DL-driven map generalization (Harrie et al. 2024). Moreover, all established studies are currently either exclusively based on raster or vector data. Monmonier (1986) points out the importance of leveraging hybrid structures that incorporate vector and raster data for conventional cartographic generalization. Although some approaches demonstrate the benefits of integrating raster and vector data for the DL-based simplification of coastlines (Du, Wu, Zhu, Liu & Wang 2022, Jiang et al. 2023), the application of such multimodal techniques to the generalization of buildings in the DL context is unexplored.

3.4 Research questions

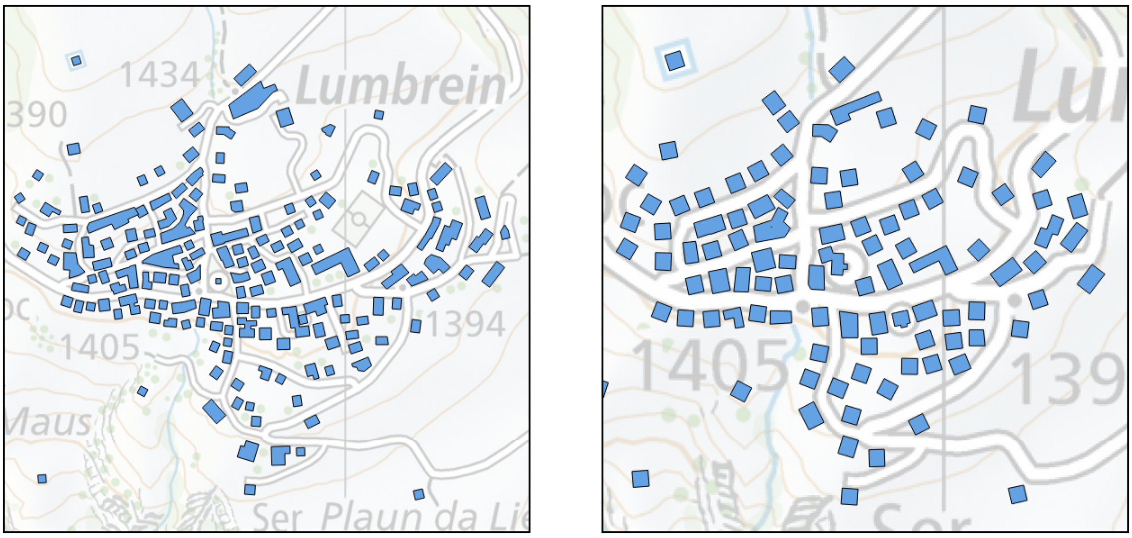
According to Harrie et al. (2024), none of the DL-based building generalization approaches proposed to date can rival the quality of maps generalized semi-automatically through human intervention. Therefore, recent research has shifted towards decomposing the generalization process into a sequence of DL models, each of which is optimized for a certain task, instead of trying to create a single model to carry out ambitious end-to-end generalization (Courtial et al. 2024). In the context of such a workflow, an intermediate model that recommends the generalization operators that should be applied to a given map feature is of special interest. Once the appropriate operators have been identified, specialized models may be incorporated to execute them. Therefore, the explicit identification of generalization operators represents an important stepping stone for many downstream tasks, such as the production of a fully generalized map based on DL (Lee et al. 2017). Understanding the appropriate set of generalization operators for a given situation facilitates the practical implementation of the generalization process and has the potential to make important contributions to knowledge acquisition and to the explainability of DL-driven generalization approaches (Rudin 2019, Fu et al. 2023). In light of the identified RGs, the thesis seeks to investigate the DL-based prediction of the generalization operators that should be applied to individual buildings based on their surrounding cartographic context.

In order to address the absence of purpose-built training datasets with annotated generalization operators identified in RG0, one of the objectives of the *DeepGeneralization* project constitutes the development of a database where individual buildings are matched across adjacent scales and additionally annotated with information regarding the presence or absence of the most important operators applied during the generalization procedure. This thesis aspires to leverage the dataset developed for the hitherto neglected scale transition from 1:25,000 to 1:50,000 to train various DL models to predict whether any of the following contextual generalization operators should be applied to a given building: *elimination*, *aggregation*, *typification*, *displacement*, and *enlargement*. The dataset is described in more detail in Chapter 4. To support the operator classification, the models are additionally provided with the cartographic context features associated with the building for which the prediction is to be conducted, comprising the road network and other surrounding buildings. To encode the building and its contextual map features, three types of DL models are proposed: a raster-based, a vector-based, and a multimodal model that accepts vector and raster features simultaneously. The multimodal approach has the potential to harness the advantages of both data models: the intuitive representation of maps as raster images and the flexibility and explicit spatial relationships contained in vector data. With respect to the outlined operator classification task, the thesis seeks to explore the following research questions (RQs), each designed to address the RG with the corresponding number identified in Section 3.3.

- **RQ1**: *To what extent can DL models be used to predict the generalization operators that should be applied to a given building?*
- **RQ2**: *To what degree can the inclusion of cartographic context enable more informed generalization operator predictions?*
- **RQ3**: *To what extent can a multimodal model integrating vector and raster representations outperform unimodal models based on the individual modalities?*

4
Data

The utility of GeoAI models is directly proportional to the quality of the datasets on which they are trained (Janowicz et al. 2020, Kang et al. 2024). In light of the identified paradigm shift towards DL and the associated absence of purpose-built training datasets for DL-driven cartographic generalization identified in Section 3.3, the *DeepGeneralization* project set out to develop a comprehensive and balanced training dataset in collaboration with swisstopo, the Swiss Federal Office of Topography (Fu et al. 2023, Senn et al. 2024, Fu et al. 2025). A seamless building database supplied by swisstopo containing vector geometries of all buildings across Switzerland generalized to 1:25,000 and 1:50,000 constitutes the foundation of the dataset exploited in the present thesis. Figure 11 illustrates the buildings at the two scales for a small map section. The scale transition from 1:25,000 to 1:50,000 was chosen since contextual operators are predominantly applied when generalizing between medium to small scales (Regnauld & McMaster 2007, Roth et al. 2011). Given that the dataset stems from an authoritative source, it can reasonably be assumed to possess high data quality (Kang et al. 2024).



(a) Source scale 1:25,000

(b) Target scale 1:50,000

Figure 11 Buildings generalized to the investigated source and target scales (map data © swisstopo).

Swisstopo maintains the topographic landscape model (TLM) containing ungeneralized representations of the map features on the topographic maps. From the TLM, a digital cartographic model (DCM, Grünreich 1985) is derived for every map scale using a combination of automatic and manual generalization techniques. The building geometries investigated in the present thesis originate from the DCM25 and DCM50, respectively. DCM50 was provided by swisstopo, as it is not publicly available. The generalization is conducted automatically based on a bespoke system that is configured to reach specific generalization objectives. Afterwards, the results of the automatic generalization process are validated by expert cartographers. Since individual buildings are retained in dense urban areas, the generalization of DCM50 requires vigorous human intervention. The system achieves satisfactory generalization rates for roughly 80% of map features, necessitating manual edits for the remaining 20% (Duchêne et al. 2014). This establishes the dataset as particularly intriguing for DL applications, since the models can capture the expert knowledge employed during the process (Touya et al. 2019).

When generalizing buildings to 1:50,000, Swiss map specifications stipulate that buildings comply with minimum dimensions imposed by the limits of human perception while preserving as much of the settlement structure as possible. Due to their importance for navigation, traffic-related features such as roads and railways are significantly enlarged, forcing the generalization of the buildings to be carried out according to the new circumstances. Relative positions between buildings are considered to be more important than absolute positional precision (Spiess et al. 2005). Figure 12 illustrates that the generalization operators identified in Section 2.1 are used to address cases that are at risk of violating the map specifications.

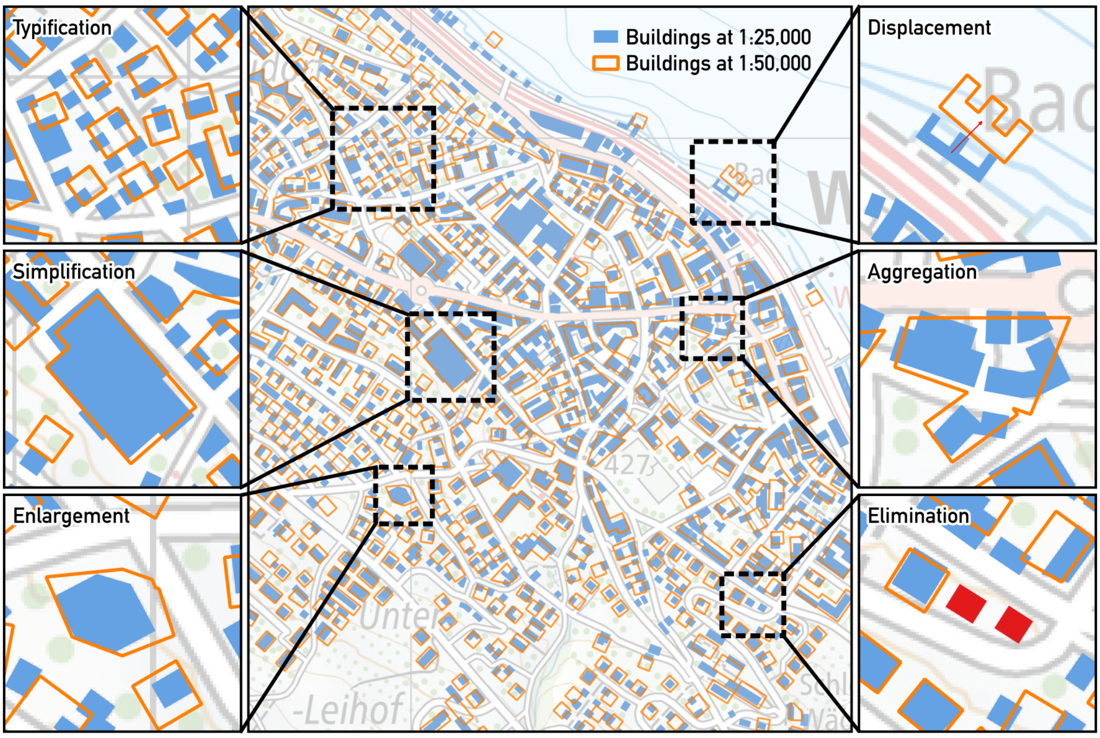


Figure 12 Operators applied during building generalization (map data © swisstopo).

To construct the training dataset, the semi-automated generalization operator annotation framework illustrated in Figure 13 was devised (Senn et al. 2024).

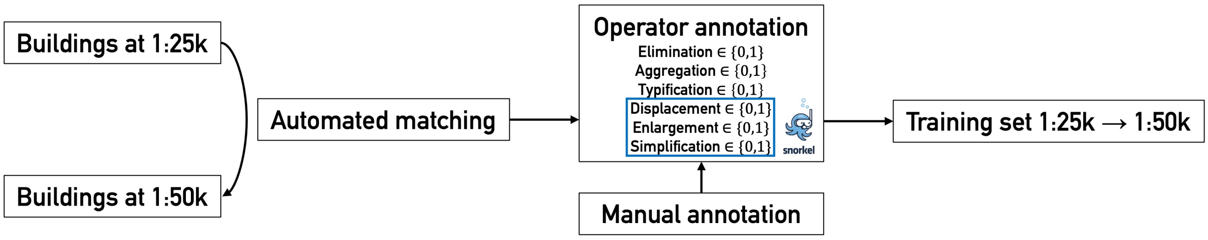


Figure 13 Workflow adopted for deriving the training dataset.

The process of contextual cartographic building generalization is commonly considered to consist of two distinct steps: the recognition of building groups and the subsequent application of generalization operators (Li et al. 2004, Yan et al. 2008, Cetinkaya et al. 2015). Therefore, the first step consisted of partitioning the two building datasets into street blocks according to the neighborhood model from urban morphology (Patricios 2001) by constructing polygons delineated by the road network (Li et al. 2004). The concept of grouping buildings based on street blocks is frequently applied as part of the data enrichment step in conventional generalization techniques, such as agent-based approaches (Barrault et al. 2001, Ruas & Duchêne 2007). The availability of such context-based geographic areas and the associated contextual information facilitates the choice of the appropriate generalization operators that should be applied to conserve the relationships and characteristics of the original street block (Ruas 1999, Stanislawski et al. 2014, Deng et al. 2018). Since roads partition and structure the map space, street blocks act as topological constraints for the generalization process (Bader et al. 2005, Cetinkaya et al. 2015, Courtial et al. 2021*b*). Therefore, it is reasonable to assume that buildings within a street block are generalized independently of buildings in other street blocks (Barrault et al. 2001, Basaraner & Selcuk 2008, Zhang et al. 2014). As roads have a higher precedence in the generalization process compared to buildings, partitioning was carried out based on the Swiss road network generalized to the target scale of 1:50,000, since buildings are expected to conform to the roads after the generalization process (Kilpelainen 1994, Li et al. 2004, Bader et al. 2005, Spiess et al. 2005). Figure 14a illustrates the partitioning of space into street blocks.

In a second step, the buildings within the same street block were matched across the two scales. In order to facilitate efficient propagation of incremental updates throughout the scales, swisstopo maintains feature links between the datasets that can be exploited to determine many-to-one building matches (Duchêne et al. 2014). However, the significant presence of the typification operator in the transition from 1:25,000 to 1:50,000 (Regnauld & McMaster 2007, Harrie & Weibel 2007) additionally established the need to identify many-to-many relationships that were implemented according to the technique proposed by Zhang et al. (2014). Their approach assigns each building at the larger scale its most likely counterpart(s) at the smaller scale based on a relaxation labeling procedure that accounts for contextual information such as relative position, orientation, size, and shape between neighboring buildings.

(a) Street block partitioning **(b)** Building matching and operator annotation

Figure 14 Illustration of the operator annotation workflow (map data © swisstopo).

Based on the matched buildings, the presence or absence of the relevant generalization operators depicted in Figure 12 was identified for every building at 1:25,000. The presence of *elimination*, *aggregation*, and *typification* was determined entirely by the building matching process.

- **Elimination**: The building at the larger scale cannot be matched to any building at the smaller scale.
- **Aggregation**: For a given building at the larger scale, there are other buildings at the larger scale that are matched to the same building at the smaller scale.
- **Typification**: If a building at the larger scale is matched to multiple buildings at the smaller scale, these buildings at the smaller scale are considered the result of a typification. Any additional buildings at the larger scale matched to those deemed typified are, by extension, also subject to typification.

All buildings that were not classified as eliminated were subsequently annotated with the operators *displacement*, *enlargement*, and *simplification* using Snorkel[2] (Ratner et al. 2020), a framework designed to address the bottleneck caused by the labor-intensive task of hand-labeling large datasets for DL models. Instead of relying on manual annotation, the Snorkel framework is based on weak supervision, allowing for the formulation of various labeling functions that express heuristics, patterns, or use external knowledge bases to programmatically assign labels to data points without requiring ground-truth labels. Individually, these labeling functions are noisy and may output conflicting labels for a given observation. To capture the signal of all labeling functions, Snorkel provides a generative model that estimates their accuracies and dependencies based on their outputs over the unlabeled data to produce probabilistic labels. The generative model assigns higher weights to labeling functions with higher estimated accuracies and accounts for pairwise correlations to avoid overcounting (Ratner et al. 2020).

[2]snorkel.org

The workflow adopted to annotate the presence or absence of displacement, enlargement, and simplification using the Snorkel framework is illustrated in Figure 15. Initially, a subset of 1,000 matched buildings was manually annotated with the generalization operators that were applied to the buildings at 1:25,000 to derive the buildings at 1:50,000. Based on this hand-labeled dataset, multiple geometric metrics were identified for each operator, which can contribute to the identification of its presence or absence. For example, the centroid distance between the source and target geometries tends to be smaller for non-displaced buildings as opposed to buildings that were annotated as displaced. By individually visualizing the distributions of the metrics for displaced and non-displaced buildings, a threshold that separates the distributions as optimally as possible was determined through visual inspection. The identified metrics and thresholds were subsequently translated into a set of labeling functions for each operator.

The formulated labeling functions were used to train separate generative Snorkel label models for every operator to produce binary annotations denoting the presence or absence of displacement, enlargement, and simplification for each sample in the training database. To monitor the effectiveness of the different labeling functions, the hand-labeled samples were reintroduced in order to perform an ablation study, choosing the combinations that yielded the highest evaluation metrics on the manually annotated samples. The resulting labeling functions are summarized in Table 3. Finally, the performance of the devised approach was validated using a separate set of hand-labeled samples not incorporated during the calibration of the procedure, indicating that the application of the proposed annotation workflow to previously unseen samples results in satisfactory performance (Senn et al. 2024). Figure 14b displays the results of the building matching and operator annotation workflows within a street block.

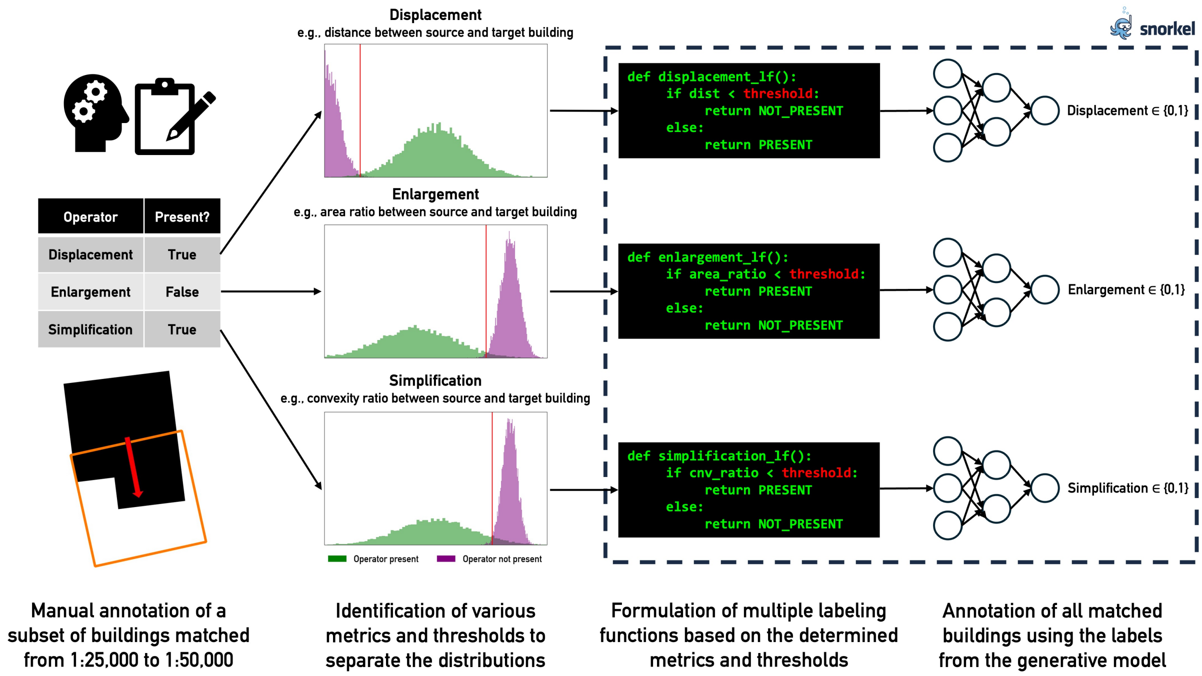


Figure 15 Annotation of displacement, enlargement, and simplification using Snorkel.

Table 3 Labeling functions formulated for the Snorkel operator annotation framework.

Generalization operator	Labeling function
Displacement	Centroid distance Convex hull centroid distance Spatial intersection
Enlargement	Area ratio Area difference Perimeter length ratio
Simplification	Number of vertices Number of notches (Brinkhoff et al. 1995) Convexity (Basaraner & Cetinkaya 2017) Equivalent rectangular index (Basaraner & Cetinkaya 2017)

To derive the training dataset employed throughout this thesis, the semi-automated operator annotation workflow was applied to the building database. After partitioning the dataset into street blocks, all street blocks containing more than 75 buildings at 1:25,000 were discarded, leaving 121,422 street blocks with 1,750,468 buildings. This corresponds to roughly 85% of the buildings in Switzerland. For these buildings, the presence or absence of the identified generalization operators was annotated. The substantial size and large coverage extent of the dataset imply that it encompasses a wide variety of situations, including numerous building types and styles across both rural and urban settings. This makes it especially compelling for training DL models, since it has the potential to improve performance and facilitate generalization to unseen data (Sug 2018). Figure 16 illustrates the distribution of the annotated operators, confirming the notion that contextual operators are dominant at the present medium to small scale transition compared to operators that tend to be context-agnostic, such as simplification (Regnauld & McMaster 2007, Roth et al. 2011). Despite the apparent imbalance, the dataset contains at least 250,000 occurrences of every operator due to the large abundance of data.

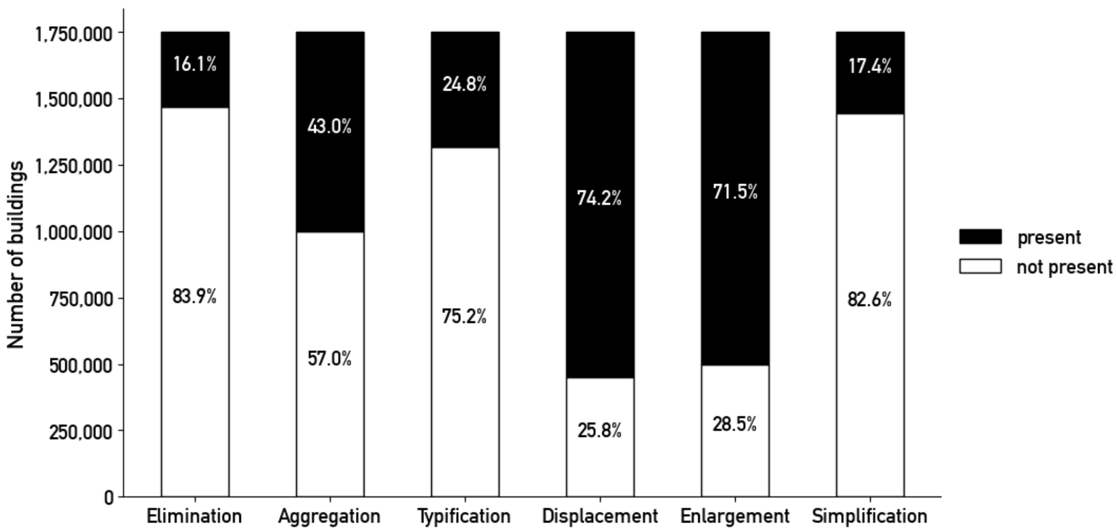
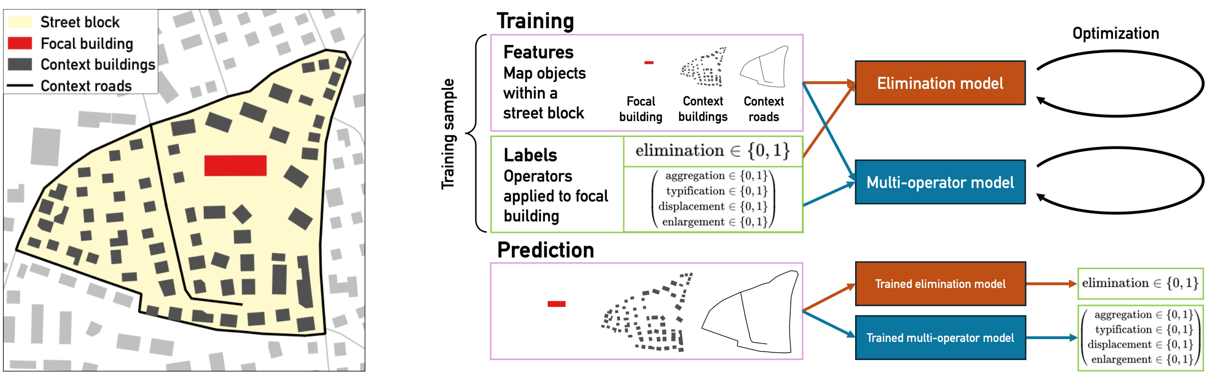


Figure 16 Generalization operator distribution within the annotated training database.

5
Methodology

5.1 Conceptualization

The main goal of the present thesis is to investigate to what extent it is possible to predict the generalization operators that should be applied to a given building during the scale transition from 1:25,000 to 1:50,000 based on its cartographic context. The cartographic context is represented by the surrounding buildings generalized to 1:25,000 and the roads generalized to 1:50,000, as argued in Chapter 4. Borrowing from the terminology introduced by Fu, Zhou, Feng & Weibel (2024) and Fu, Zhou, Xin & Weibel (2024), the building for which the operator prediction should be conducted is referred to as the *focal* building, whereas the surrounding buildings and roads are called the *context* buildings and roads, respectively. According to the justification outlined in Chapter 4, the cartographic context included to facilitate the prediction for a given focal building is confined to its street block, serving as the analysis unit. Consequently, the context buildings represent all the remaining buildings within a given focal building's street block. The context roads are defined as the roads that enclose the street block of the focal building, in addition to any roads that lie within the boundaries of the street block itself. Figure 17a provides an illustration of the terminology employed based on an arbitrary focal building.



(a) Terminology

(b) Overview

Figure 17 Conceptualization of the operator classification approach (map data © swisstopo).

The task of predicting the generalization operators that should be applied to a focal building given its context buildings and roads can be formulated as a supervised, binary multi-label classification problem, since for each building multiple generalization operators may be present simultaneously (Tsoumakas & Katakis 2007). As simplification is not considered a contextual operator, its application is assumed to be independent of the surrounding map features (Barrault et al. 2001, Basaraner & Selcuk 2008). Therefore, only the contextual operators elimination, aggregation, typification, displacement, and enlargement are investigated further. The various DL models proposed in the following are trained using the focal building and its context buildings and roads as features and the operators applied to the focal building as labels. The map features within a street block and the labels associated with the focal building collectively constitute a sample for training the DL models. Provided with a sufficiently large number of training samples, the models are assumed to be able to distinguish the different map features and leverage the context features to facilitate the classification of the generalization operators to be applied to the focal building (Fu, Zhou, Feng & Weibel 2024). The operators to which the context buildings are subjected are not accessible to the model during the training process. Finally, trained models are expected to predict the presence or absence of the aforementioned generalization operators given a focal building and its context buildings and roads.

The presence of elimination and the remaining contextual operators are mutually exclusive, e.g., a building that is eliminated cannot be enlarged, aggregated, typified, nor displaced. Therefore, the operator classification approach is formalized using two distinct models that can be applied consecutively to predict the operators that should be applied to a building. The first model, referred to as the *elimination model*, is tasked with solving a separate single-label classification problem to classify whether a given focal building should be eliminated or retained. The *multi-operator model* on the other hand refers to the multi-label classification model that is concerned with predicting the presence or absence of the remaining contextual operators (aggregation, typification, displacement, enlargement) for the retained buildings simultaneously. The conceptualization of the proposed operator classification approach is illustrated in Figure 17b.

To implement the conceptualized operator classification approach, the workflow outlined in the following is devised. The proposed framework is illustrated in Figure 18. In a first step, data balancing is performed, as the annotated training dataset is largely imbalanced with respect to the generalization operators applied for the scale transition between 1:25,000 and 1:50,000. Section 5.2 outlines the procedure for the generation of a balanced *elimination dataset* (with respect to elimination) and a balanced *multi-operator dataset* (with respect to aggregation, typification, displacement, and enlargement) that are used to train the elimination and multi-operator models, respectively (Section 5.3). As described in Chapter 3, supplying DL models with raw vector data is not trivial. To encode the map features displayed in Figure 17, both a raster-based approach and a vector-based approach are devised. The raster-based approach involves the formalization of the proposed approach as an image classification problem by representing the street block features as a stacked tensor containing the rasterized features, which is subsequently used to train models from the domain of computer vision. For the vector-based approach, the street block features and their characteristic properties are transformed into a heterogeneous graph. Heterogeneous GNNs are subsequently used to perform node-level prediction of the generalization operators applied to the focal building. Once the raster-based and vector-based models have been trained and evaluated, multimodal models are additionally proposed that unify the models capable of processing the individual modalities in a single model able to learn from the raster and graph representations simultaneously. Finally, the performance of all models with respect to the operator classification task is analyzed by assessing various evaluation metrics, both in a holistic and stratified manner.

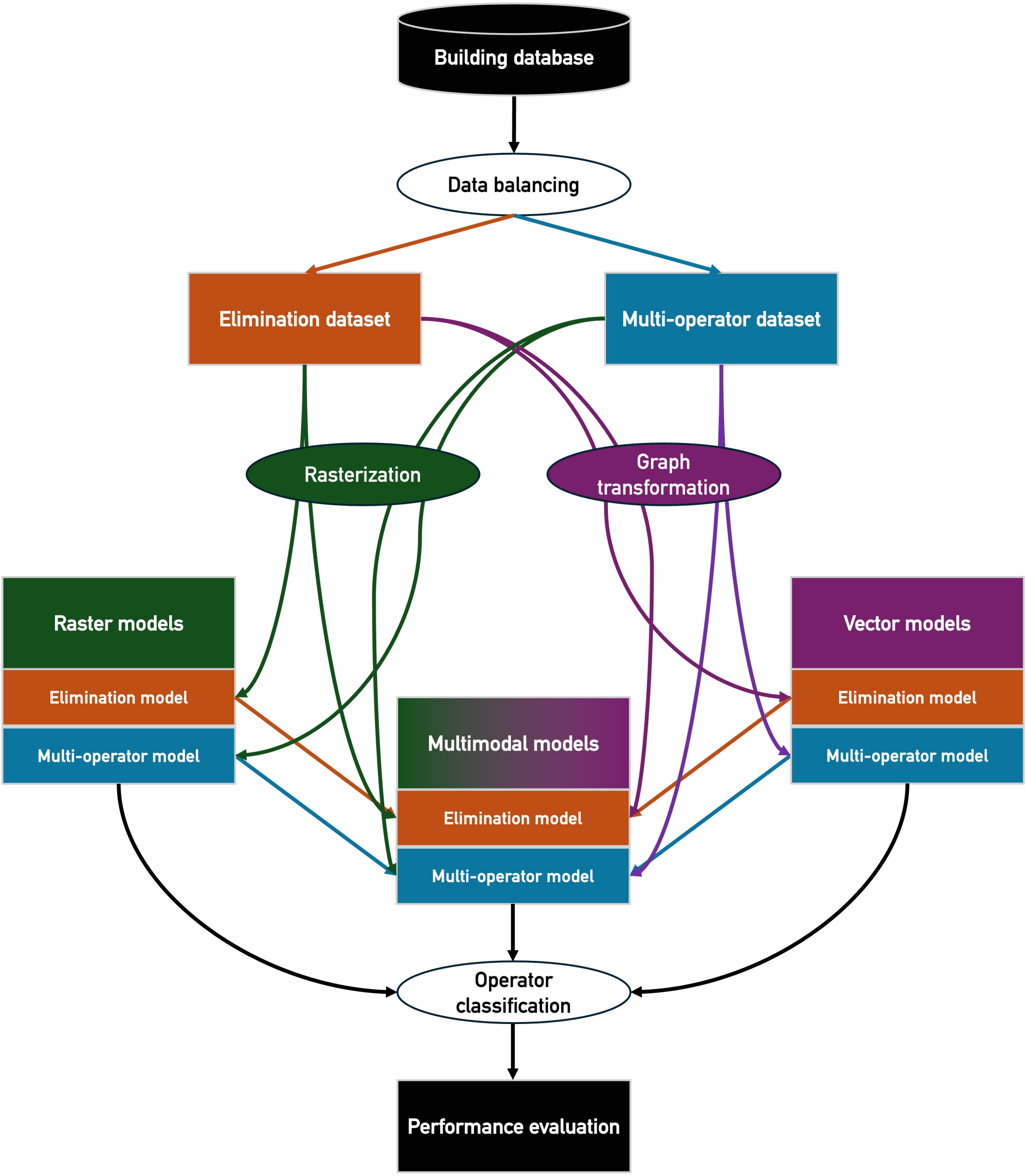


Figure 18 Framework for the operator classification approach.

5.2 Data balancing and sampling

As evident in Figure 16, the training dataset is imbalanced with respect to the generalization operators that were applied to the samples. To assess the level of imbalance in a multi-label dataset, Charte et al. (2013) propose various quantitative measures. Given a multi-label dataset with m samples and n labels, the dataset can be represented as an $m \times n$ matrix Y, where element $y_{ij} \in \{0, 1\}$ indicates the absence (0) or presence (1) of label j in sample i. Based on this matrix, the imbalance ratio IR can be calculated for every label j as the ratio between the number of samples with the most common label and the number of samples with the label j, as displayed in Equation (1).

$$IR_j = \frac{\max\limits_{1 \leq l \leq n} \left(\sum\limits_{i=1}^{m} y_{il} \right)}{\sum\limits_{i=1}^{m} y_{ij}} \tag{1}$$

Using the imbalance ratio per label, the mean label imbalance ratio $MeanIR$ can be computed according to Equation (2), representing the average level of imbalance within the dataset.

$$MeanIR = \frac{1}{n} \sum_{j=1}^{n} IR_j. \tag{2}$$

The coefficient of variation $CVIR$ can be calculated as illustrated in Equation (3), indicating whether imbalance is uniform or varies significantly among the labels (Charte et al. 2015).

$$CVIR = \frac{IR\sigma}{MeanIR}, \quad IR\sigma = \sqrt{\sum_{j=1}^{n} \frac{(IR_j - MeanIR)^2}{n-1}}. \tag{3}$$

According to Charte et al. (2015), any multi-label datasets whose $MeanIR$ and $CVIR$ values exceed 1.5 and 0.2, respectively, can be considered imbalanced. The building training database with annotated generalization operators displays $MeanIR$ and $CVIR$ values of 2.27 and 0.67 with respect to the investigated contextual operators, reinforcing concerns that imbalance is an issue and that the database could benefit from the application of tailored resampling techniques. Conducting DL based on imbalanced datasets can lead the models to adopt biases towards the majority classes while exhibiting poor generalization ability for the underrepresented labels (Mathews & Hari 2018). In the context of the operator classification task, this implies that the models may struggle to learn rarely applied operators. In fact, initial experiments conducted by training the DL models presented in Section 5.3 on the original imbalanced dataset revealed their tendency to overwhelmingly classify training samples according to the majority labels. Therefore, the techniques employed to balance the datasets used to train the elimination and multi-operator models are outlined in Section 5.2.1 and Section 5.2.2, respectively.

5.2.1 Elimination model

According to Figure 16, elimination is the most imbalanced operator for the scale transition between 1:25,000 and 1:50,000 when each generalization operator is considered independently. As the elimination model is concerned with performing single-label binary classification regarding whether or not a given focal building should be retained, resampling is trivial. The balanced dataset for training the elimination model is constructed through random undersampling (RUS) by indiscriminately discarding training samples, until an equal number of eliminated and retained buildings remains (Tahir et al. 2009). Due to the abundance of samples in the database, a sufficiently large balanced training dataset can be constructed without resorting to random oversampling (ROS) techniques that consist of replicating examples from the minority class at random and therefore can increase the risk of overfitting and lead to poorer generalization (Tarawneh et al. 2022).

5.2.2 Multi-operator model

In contrast to the elimination model, the multi-operator model is tasked with predicting the presence of multiple generalization operators for selected buildings simultaneously, thus requiring training on a multi-label dataset. In the context of multi-label datasets, a *labelset* refers to a distinct combination of labels assigned to a training sample (Charte et al. 2015). With respect to the dataset described in Chapter 4, each unique combination of generalization operators applied to a building can be considered a labelset. Naturally, some labelsets tend to be more common than others. The use of conventional resampling techniques as typically applied to single-label classification problems cannot be directly transferred to multi-label classification, as the label imbalance may be present within labels, between labels, and among labelsets. Therefore, datasets in which samples and their corresponding labels are unevenly distributed across the data space pose a major challenge for multi-label classification, establishing the need for specialized techniques to balance training data (Haixiang et al. 2017, Tarekegn et al. 2021).

The label powerset (LP) transformation is a specialized resampling technique designed to address multi-label classification problems (Boutell et al. 2004, Charte et al. 2015). It consists of mapping each unique combination of labels into a single labelset, essentially transforming a multi-label dataset into a multi-class dataset (Tarekegn et al. 2021). Every observation with the same combination of present labels is assigned a single common label during the LP transformation. Figure 19 illustrates the LP transformation based on arbitrary combinations of contextual generalization operators applied to selected buildings. For the multi-operator model, there are 4 generalization operators that are either present or absent within the training data. This implies the existence of $2^4 = 16$ possible labelsets, of which in practice only 14 were annotated during the generalization operator annotation workflow. The LP transformation has the potential to facilitate data balancing, as resampling with respect to a single, multi-class label is easier compared to resampling in a multi-label context. The number of buildings within each labelset is illustrated in Figure 20.

Aggregation	Typification	Displacement	Enlargement
0	1	0	1
1	1	0	0
1	1	1	1
0	0	1	1

→ LP transform

LP-transformed operator label
Typification, Enlargement
Aggregation, Typification
Aggregation, Typification, Displacement, Enlargement
Displacement, Enlargement

Figure 19 LP transformation of generalization operators applied to selected buildings.

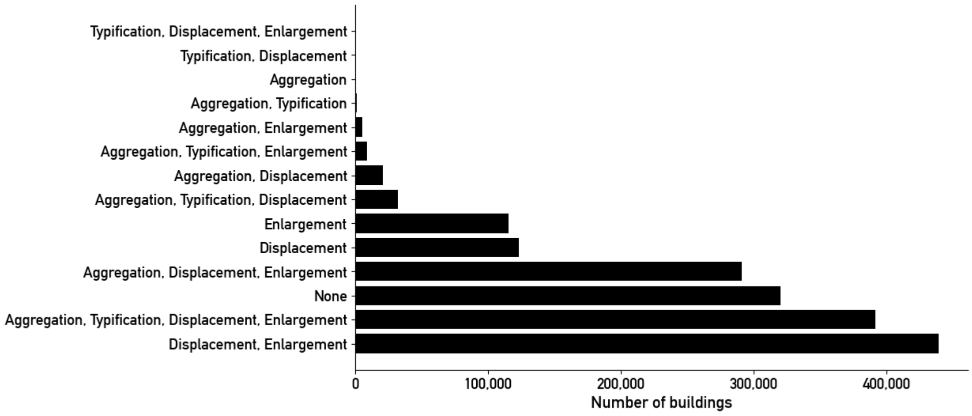


Figure 20 Number of buildings per generalization operator labelset.

Based on the LP transformation, Charte et al. (2013) propose the application of ROS and RUS to observations of underrepresented and overrepresented labelsets, respectively, to obtain a balanced dataset. LP-based resampling techniques have proven to be effective for addressing a variety of applications where imbalanced data pose a problem, ranging from scientific document classification (Hafeez et al. 2023) to movie genre prediction (Kumar et al. 2023). The application of LP-RUS and LP-ROS is particularly effective in cases where the ratio of distinct labelsets over examples is sufficiently small (Sechidis et al. 2011). Furthermore, there should be clearly discernible minority and majority labelsets (Charte et al. 2015). Both of these properties are present for the multi-label generalization operator database, establishing an LP-based hybrid resampling technique incorporating both LP-RUS and LP-ROS as a sensible approach to address the imbalance in the multi-label dataset. Based on the work of Charte et al. (2013, 2015), the proposed approach is implemented according to a stratified sampling technique. Given a target size for the resulting balanced dataset, the goal is to oversample all minority labelsets and undersample all majority labelsets until they contain the same number of elements given by the ratio between the target size and the number of unique labelsets. The exact implementation of the LP-based stratified balancing approach is outlined in Algorithm 1.

The distribution of the individual operators prior to and following the application of the LP-based stratified balancing approach is illustrated in Figure 21. Evidently, the sampling approach involving stratification according to the labelsets produces a balanced dataset, with each operator appearing in roughly 50% of the training samples. After applying the tailored resampling technique, the balanced dataset displays $MeanIR$ and $CVIR$ values of 1.12 and 0.14, respectively, indicating that the proposed hybrid resampling approach manages to effectively address the imbalance present in the dataset.

Algorithm 1 LP-based stratified balancing approach

function LPResample (buildings, targetSize)
Input: imbalanced building dataset, target size of the balanced dataset
Output: balanced building dataset

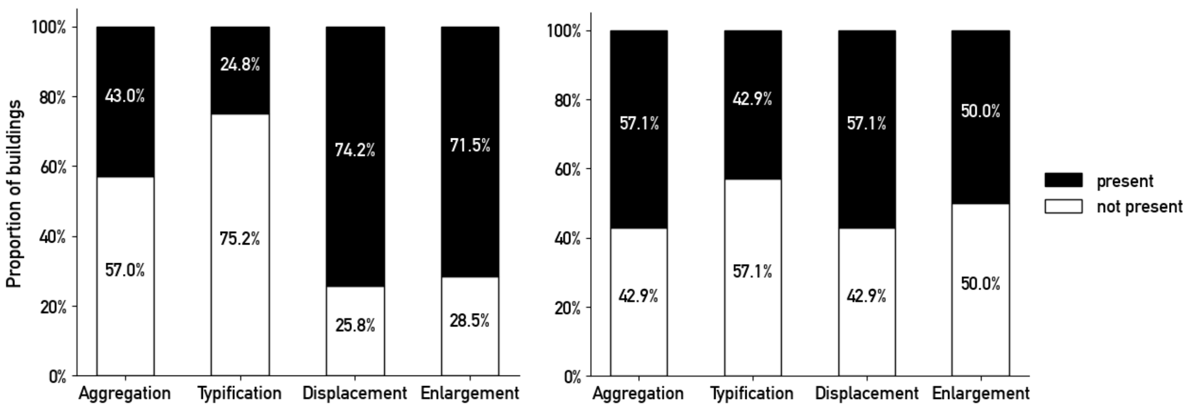
labelsets ← getLabelsets(buildings) ▷ Obtain labelsets

for $i = 1 \to$ |labelsets| **do** ▷ Group buildings into bags according to their labelsets
  labelsetBags$_i$ ← buildingsWithLabelset(i)
end for

targetLabelsetBagSize ← ⌈targetSize / |labelsets|⌉ ▷ Compute required labelset bag size

for $i = 1 \to$ |labelsetBags| **do**
  if |labelsetBags$_i$| < targetLabelsetBagSize **then** ▷ Oversample minority labelset bags
    labelsetBags$_i$ ← randomly oversample labelsetBags$_i$ to targetLabelsetBagSize
  end if
  if |labelsetBags$_i$| > targetLabelsetBagSize **then** ▷ Undersample majority labelset bags
    labelsetBags$_i$ ← randomly undersample labelsetBags$_i$ to targetLabelsetBagSize
  end if
end for

balancedBuildings ← reconstructBuildingsFrom(labelsetBags)
return balancedBuildings



(a) Original imbalanced dataset **(b)** Balanced dataset

Figure 21 Multi-operator dataset distribution before and after data balancing.

5.2.3 Sampling

Effective development, training, and evaluation of the DL models presented throughout the remainder of this thesis requires the creation of various disjoint subsets from the annotated operator database outlined in Chapter 4 for the elimination and multi-operator models. The training of the models is conducted based on balanced datasets containing 125,000 samples obtained by applying the resampling techniques introduced in Sections 5.2.1 and 5.2.2. The dataset used for training the elimination models is balanced with respect to elimination, whereas the dataset employed for fitting the multi-operator models is balanced with respect to aggregation, typification, displacement, and enlargement and only includes non-eliminated buildings. To assess the performance of DL models and to prevent overfitting, they should be evaluated on a set of examples not seen during the training process (Xu & Goodacre 2018). Therefore, the two balanced datasets are subsequently shuffled and randomly partitioned into a training and validation set according to an 80/20 split such that both the training and validation sets exhibit the same distribution, leaving 100,000 samples for training and 25,000 for validation. Since the training and validation sets stem from the same underlying balanced distribution, the validation set can be reliably used to assess overfitting and provide unbiased performance estimates during the learning process (Goodfellow et al. 2016).

The final evaluation of model performance is conducted on separate, imbalanced test sets derived by randomly drawing 25,000 samples from the original database not used during the training process of the respective model, which better reflects real-world conditions. For the test set of the multi-operator model, only non-eliminated buildings are considered. Therefore, the test sets can be used to assess whether the models can transfer the knowledge obtained by training on artificially balanced data to existing generalized maps that tend to be highly imbalanced with respect to the operators applied during the generalization process (Regnauld & McMaster 2007). The focal buildings in the training, validation, and test sets are subsequently subjected to the rasterization and graph transformation procedures outlined in Section 5.3.4 and Section 5.3.5 to serve as suitable inputs for the respective models. The sampling approach applied to derive the training, validation, and test sets for the elimination and multi-operator models is illustrated in Figure 22.

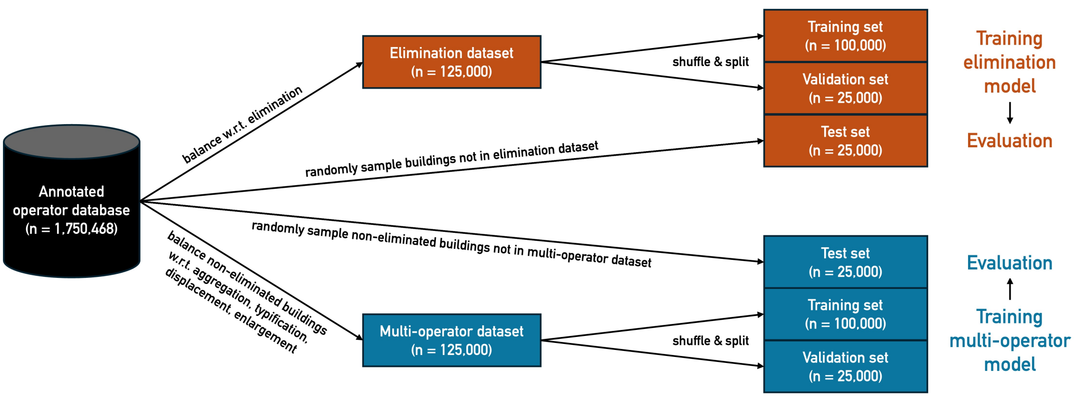


Figure 22 Sampling approach for deriving the training, validation, and test sets.

5.3 Deep learning models

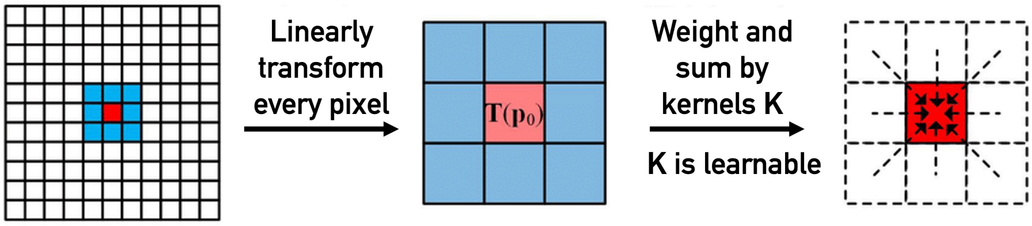
5.3.1 Technical setup

The CPU-heavy tasks such as the transformation of the training data to suitable inputs for the DL models were executed locally on a MacBook Pro (M1 Max, 32 GB RAM). Since the training dataset was neatly partitioned into street blocks, the problem can be considered embarrassingly parallel, enabling the use of multiprocessing to effectively handle the large number of samples (Watkinson et al. 2019). Apart from small-scale testing and prototyping, training of the DL models was carried out on the UZH S3IT ScienceCluster[3] equipped with powerful CUDA[4]-enabled NVIDIA Tesla V100 GPUs and RAM configurations of up to 48 GB, which facilitated model training and significantly expedited processing times. The DL models were implemented in Python using PyTorch[5] (Paszke et al. 2019), a widely adopted open-source ML library. The development of the vector-based models additionally involved the use of PyG[6] (Fey & Lenssen 2019), a library built on PyTorch for training GNNs. The code and all packages necessary for reproducing the findings are available at github.com/jorissenn/genops.

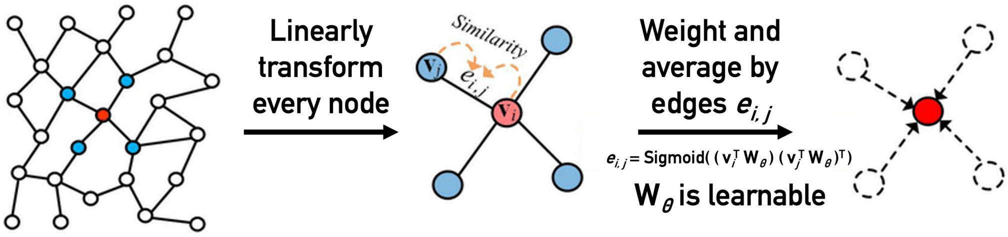
5.3.2 Model architectures

In order to conduct the operator classification task, two related and established model architectures are to be investigated for each of the raster-based and vector-based approaches. In both cases, the first type of architecture is a simple network (in terms of number of parameters) that relies on convolutions to process training data efficiently. The convolution operation involves sliding a kernel over the input data to generate a feature map that highlights important features, such as edges, since local groups of values tend to be highly correlated (LeCun et al. 2015, Albawi et al. 2017). Figure 23 shows the results of applying convolution to images and graphs. The convolution operation can be used to learn representations of pixels and nodes based on their neighborhood by aggregating information from their respective neighbors to detect local topological features and relationships while conserving correlations within the neighborhood (Bei et al. 2019, Yan et al. 2021, Harrie et al. 2024). Therefore, it naturally exploits the principles inherent in cartography such as the first law of geography (Tobler 1970) and spatial association (Anselin 1995). In the context of DL-based cartographic generalization, this property can be beneficial, since generalization operators such as typification are commonly applied to groups of buildings (Regnauld & McMaster 2007, Stanislawski et al. 2014), which implies that their application patterns tend to be spatially autocorrelated. From a conceptual point of view, this establishes model architectures such as CNNs and GCNNs as especially intriguing for DL-driven cartographic generalization, providing a way of encoding cartographic domain knowledge in the generalization process through the appropriate choice of model architecture (Ai 2022).

[3] docs.s3it.uzh.ch
[4] developer.nvidia.com/cuda-toolkit
[5] pytorch.org
[6] pyg.org, formerly called *PyTorch Geometric*.



(a) Image convolution



(b) Graph convolution

Figure 23 Convolution applied to images and graphs (Liu et al. 2022).

The second type of architecture investigated for each modality is based on the transformer architecture that incorporates a self-attention mechanism, allowing each part of the input data to interact with every other part to better capture their relationships (Vaswani et al. 2017). Within transformers, the concept of self-attention is implemented through multiple attention heads that independently process data, allowing the model to simultaneously focus on different parts of the provided input (Voita et al. 2019). In the context of DL, attention is a mechanism inspired by the human perception system that allows models to focus on distinctive, hidden parts of the input data that are more relevant to make predictions as opposed to treating all parts equally (Niu et al. 2021). Due to their success in capturing long-range dependencies, attention mechanisms have been gradually replacing convolutional layers in state-of-the-art DL approaches and have even shown the ability to learn to perform convolution (Ramachandran et al. 2019, Cordonnier et al. 2020).

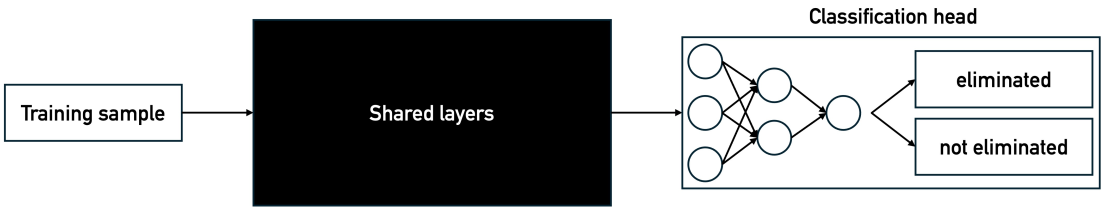
Against the backdrop of DL-driven cartographic generalization, the inclusion of attention mechanisms has the potential to facilitate better predictions, e.g., by focusing the attention on context buildings in close proximity to the focal building that may be more relevant to the decision of the operators that should be applied in a given scenario. Therefore, the incorporation of attention mechanisms into DL-driven cartographic building generalization has been subject to investigation in both raster-based (e.g., Winkler 2023 or Fu, Zhou, Feng & Weibel 2024) and vector-based (e.g., Zhou et al. 2023) approaches, establishing such architectures as auspicious candidates for the present operator classification task. All attention-based models presented in the remainder of this thesis are trained using eight attention heads according to the original transformer architecture proposed by Vaswani et al. (2017). The remaining parameters for the DL architectures were chosen empirically, starting with sensible defaults employed in existing approaches and iteratively refining them such that model performance is improved.

5.3.3 Model structure

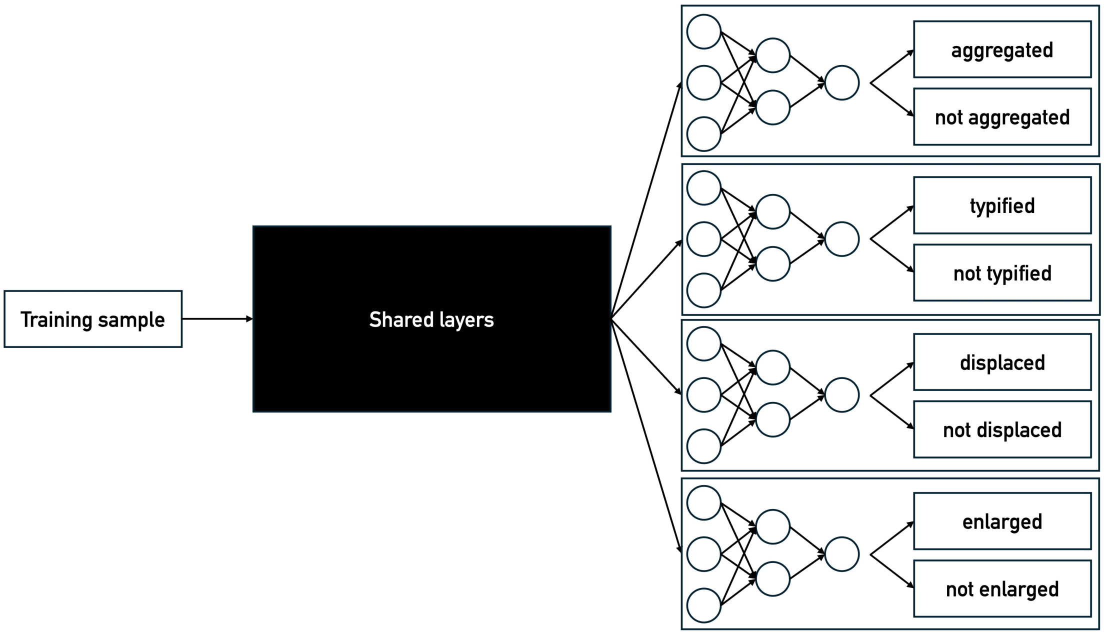
The multi-operator model implemented to predict the presence of aggregation, typification, displacement, and enlargement needs to be able to perform multi-label classification, as several generalization operators may be present simultaneously. To perform multi-label classification, an algorithm adaptation approach is adopted that involves modifying existing off-the-shelf architectures for the multi-label classification problem (Zhang & Zhou 2014), accounting for high-order relationships among generalization operators (Ji et al. 2008). Since the prediction of multiple generalization operators can be considered a process consisting of several connected tasks, an inductive multi-task learning framework is proposed to implement the classification of multiple generalization operators concurrently. Multi-task learning involves using a unified model to simultaneously learn multiple distinct tasks, leveraging shared parameters to exploit common features and employing task-specific parameters to handle differences, which has been shown to significantly enhance the performance and generalization ability of DL models applied to multi-label classification problems (Chapelle et al. 2010).

In the context of the operator classification task, it is reasonable to assume that some of the features and properties extracted by the models may be leveraged to facilitate the classification of multiple operators, e.g., the spatial relationships between buildings presumably play a crucial role for determining the presence of every operator. Compared to the construction of individual models for each operator (akin to the elimination model), the idea behind the multi-task classification framework is that such low-level features can be learned by the initial shared layers in the network, while the exact decision on whether or not an operator should be applied is based on operator-specific parameters, as has been shown in other studies (Ruder 2017, Crawshaw 2020, Vandenhende et al. 2022). Therefore, the multi-task learning framework is implemented by appending the model architectures presented throughout Sections 5.3.4 to 5.3.6 with operator-specific hidden layers that act as separate classification heads for each operator. Each classification head is charged with solving a binary classification problem using the information propagated by the shared previous layers. Consequently, the multi-operator model can be conceived as a multi-task model with four classification heads, whereas the elimination model is implemented as a single-task model with a single classification head.

The classification heads consist of a sequence of fully-connected layers. The first linear layer contains a number of neurons equal to the number of features returned by the last layer of the respective model architecture. The inputs are subsequently mapped through the remaining linear layers to a single output neuron, whose activation determines the probability that the respective generalization operator is present. Between the linear layers, the inputs are passed through a rectified linear unit (ReLU) activation function that helps mitigate the vanishing gradient problem, allowing for faster and more effective training (Picchiotti & Gori 2021). In addition, dropout layers are introduced between the linear layers to prevent overfitting by randomly disabling a subset of neurons during training (Srivastava et al. 2014). A schematic overview of the structure of the final layers of the single-task elimination and multi-task multi-operator models is provided in Figure 24.



(a) Single-task elimination model



(b) Multi-task multi-operator model

Figure 24 Structure of the single-task and multi-task models.

5.3.4 Raster models

Training sample generation

The strategy employed to derive training samples for the raster-based models is outlined in the following. In the first step, the vector geometries of the focal building, context buildings, and context roads are extracted and stored separately. The bounding box (BBOX) of the focal building's street block is subsequently used to calculate the pixel resolution necessary for the street block to fit into a raster with a side length of 256 pixels according to Equation (4).

$$\text{resolution} = \frac{\max(\text{BBOX width}, \text{BBOX height})}{256}. \tag{4}$$

Existing studies concerned with raster-based DL approaches usually adopt a fixed resolution to rasterize vector data (e.g., Sester et al. 2018, Feng et al. 2019, Kang et al. 2020, Courtial et al. 2021*b*, Fu, Zhou, Feng & Weibel 2024). However, the varying size of the street blocks coupled with the fact that established DL models for image processing require considerable adjustments in order to handle images with differing input shapes (Zhang et al. 2020) necessitates the choice of an adaptive raster resolution to generate the training samples. Therefore, the features are rasterized using the computed resolution to generate three binary rasters indicating the presence or absence of the focal building, context buildings, and context road features, respectively, at a given pixel. The created rasters are subsequently padded in the direction of the shorter side of the street block BBOX until they reach a common target shape of 256 x 256 pixels. The extent of all computed rasters corresponds to the extent of the street block. The raster shape of 256 x 256 pixels is empirically chosen given the map scale according to the suggestions provided by Touya et al. (2019) as it offers a good compromise between sharpness of building boundaries and computational cost.

In a last step, the three computed rasters are stacked according to the layered data representation model proposed by Courtial et al. (2022*b*) and Fu, Zhou, Feng & Weibel (2024), thus storing the focal building, the context buildings, and the context roads in separate image channels similarly to the structure of an RGB image. The proposed three-layer structure is favored compared to a two-layer structure storing the context buildings and roads in the same tensor, as the two-layer structure may be associated with a loss of information due to potential overlaps between the features induced by the different source scales (buildings at 1:25,000, roads at 1:50,000). The stacked rasters are stored as a tensor together with the information regarding the operators that were applied to generalize the focal building. The training sample generation procedure for the raster-based models is illustrated in Figure 25.

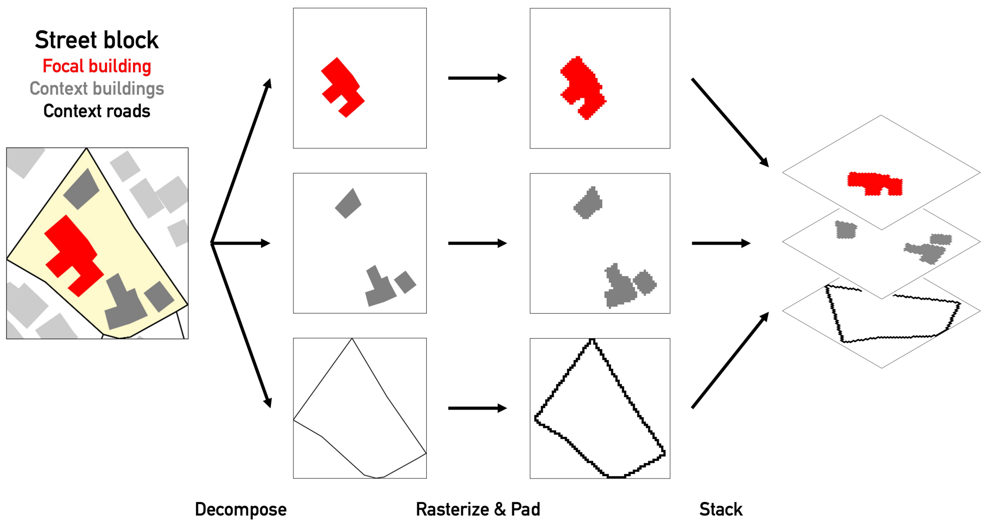


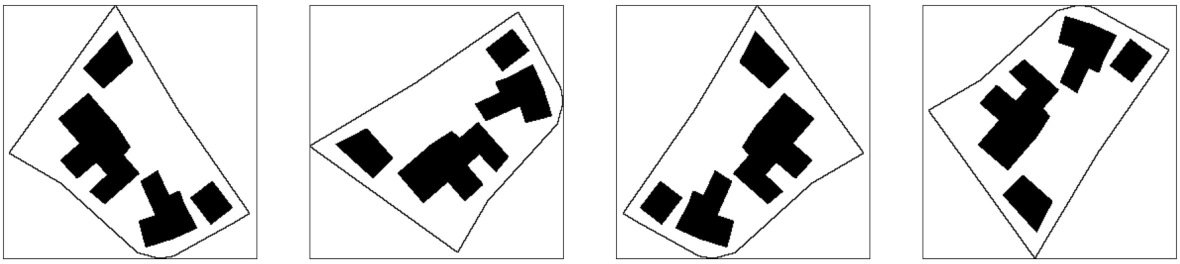
Figure 25 Procedure for deriving the layered raster representation (map data © swisstopo).[7]

Kang et al. (2020) hypothesize that cartographic generalization is a process that is invariant to direction. This implies that Approximation (5) holds given a geometric transformation t and a function g that generalizes a map m (Fu et al. 2019, Courtial 2023).

$$t(g(m)) \approx g(t(m)). \tag{5}$$

[7] A coarse raster resolution is deliberately chosen for illustrative purposes.

To facilitate the encoding of this information in DL models, data augmentation approaches in the form of geometric transformations are commonly employed (Khosla & Saini 2020). In the context of cartographic generalization, the following geometric transformations are identified as suitable: random rotation, vertical flip, and horizontal flip (Kang et al. 2020). Unlike other augmentation techniques that involve resizing or cropping images, these transformations are not associated with the loss of potentially important parts of the image that may be vital for the operator prediction. During each training epoch, a given image is randomly rotated by 0°, 90°, 180°, or 270° and additionally flipped vertically or horizontally with a probability of 50%, respectively, yielding 16 possible configurations for each sample that can be fed to the models. Figure 26 illustrates the transformations applied to the street blocks during the training process. Transformations can be used to enhance the diversity of the training dataset through artificial means, ideally counteracting the loss of diversity induced by oversampling performed as part of the data balancing process. Furthermore, the application of data augmentation techniques in the form of geometric transformations has been proven to prevent overfitting, improve generalization, and improve the robustness of DL models (Shorten & Khoshgoftaar 2019).



(a) Original **(b)** Rotation **(c)** Horizontal flip **(d)** Vertical flip

Figure 26 Transformations applied to the map features during training (map data © swisstopo).

Network architecture

The first ANN architecture identified for performing raster-based generalization operator classification is a conventional CNN as introduced in Section 3.2.1. CNNs have demonstrated powerful capacities in a wide array of image classification tasks (LeCun et al. 2015), evoking the assumption that they are well suited to tackle the operator classification problem. AlexNet is a significant CNN architecture developed by Krizhevsky et al. (2017), having gained prominence for its exceptional performance in various conventional image classification tasks. Therefore, it is worth investigating to what extent this CNN architecture can be used to classify the generalization operators that should be applied to a given building. The architecture of the CNN was adapted from the AlexNet architecture implemented in PyTorch[8] by extending it with the classification heads outlined in Section 5.3.3. Since the operator classification task is assumed to be significantly different from the conventional image classification tasks that AlexNet is commonly trained on, no pre-trained weights were used, and the network is trained from scratch. A schematic overview of the implemented CNN architecture is provided in Figure 27.

[8]github.com/pytorch/vision/blob/main/torchvision/models/alexnet.py

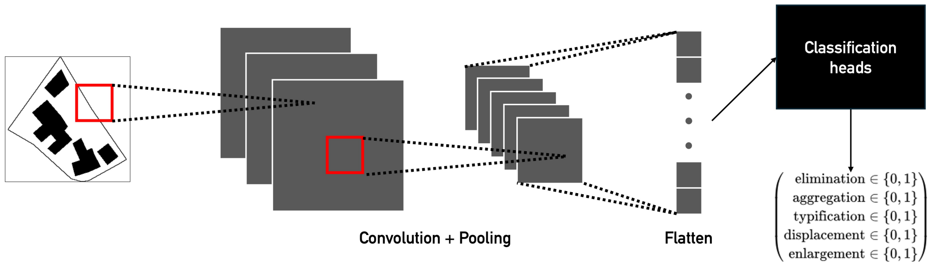


Figure 27 CNN architecture (map data © swisstopo).

In addition to the proposed CNN, a more sophisticated DL architecture in the form of Vision Transformer (ViT, Dosovitskiy et al. 2021) is additionally explored. ViT incorporates a transformer architecture to process images by dividing them into patches of a fixed size and subsequently generating linear embeddings of each patch. In conjunction with their positional encodings, these embeddings are fed into a standard transformer encoder that uses self-attention mechanisms to model relationships between different patches. The output of the transformer encoder is finally passed through a classification head that enables image classification based on the learned patch representations (Dosovitskiy et al. 2021). A schematic overview of the architecture of ViT is provided in Figure 28. ViT has been shown to excel in classifying, detecting, and segmenting remote sensing scenes (Bazi et al. 2021, Aleissaee et al. 2023) and has recently also been introduced into cartography, such as for the segmentation of features in historical maps (Wu et al. 2023, Xia et al. 2023). Moreover, experiments have shown the capabilities of architectures based on ViT for end-to-end cartographic generalization (Winkler 2023), establishing ViT as an intriguing candidate model architecture for raster-based classification of generalization operators. The architecture for ViT was adopted from an existing implementation[9]. An overview of the two implemented raster-based architectures is provided in Table 4.

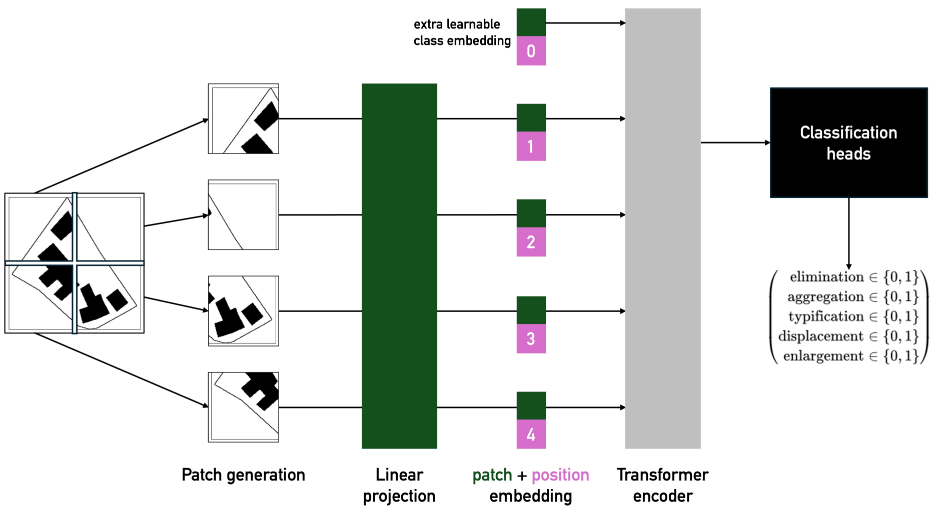


Figure 28 ViT architecture (Dosovitskiy et al. 2021, map data © swisstopo).

[9] github.com/lucidrains/vit-pytorch/blob/main/vit_pytorch/vit.py

Table 4 Properties of the raster-based architectures.

Architecture	Model	Parameters	Source
CNN	Elimination	4,075,585	Krizhevsky et al. 2017
	Multi-operator	8,893,252	
ViT	Elimination	20,586,241	Dosovitskiy et al. 2021
	Multi-operator	20,783,620	

5.3.5 Vector models

Training sample generation

The application of GNNs requires the transformation of the raw training data into graphs. To model the focal building and the context buildings and roads within a street block in a graph analogously to the raster-based approach, a heterogeneous graph structure is proposed. Heterogeneous graphs are graphs that contain multiple types of nodes and edges (Zhang, Song, Huang, Swami & Chawla 2019) and are therefore especially suited to capture the relationships between buildings and roads within a street block from a conceptual point of view. Heterogeneous graphs have been used to learn representations of groups of building polygons (Yu et al. 2024), reinforcing the assumption that they are suitable for modeling the operator classification task. Therefore, the goal is to transform the map features within a street block into an undirected heterogeneous graph with a single focal building node and multiple context building and road nodes.

To formalize the contextual and spatial relationships among the buildings in a street block, the approach outlined by Yan et al. (2019, 2020) is followed. It involves the creation of a graph, where each node represents a building (positioned at its centroid) and each edge denotes a spatial relationship between the buildings. Various proximity graphs such as nearest neighbor graphs, MST, relative neighborhood graphs, and Gabriel graphs are commonly employed to capture the spatial relationships among groups of neighboring buildings as a precursor step to generalization (Regnauld 2001, Cetinkaya et al. 2015, Deng et al. 2018, Wei et al. 2018). All of the aforementioned proximity graphs are subgraphs of the Delaunay triangulation (DT), which occupies the highest position in the hierarchy of frequently applied neighborhood graphs, ensuring that no important relationships among the buildings are omitted (Anders & Sester 2000). DT has been shown to efficiently capture proximity relationships to support a variety of conventional and DL-based tasks related to building generalization, such as building grouping (Yan et al. 2008), building pattern classification (Yan et al. 2019, Zhao et al. 2020), cartographic conflict detection (Ai et al. 2015), and generalization operator implementation (Ai et al. 2019, Xiao et al. 2024). Therefore, DT is chosen to construct the proximity graph among the building nodes within a street block.

Due to the identified importance of the road network for building generalization, structural information associated with the roads surrounding a street block is to be explicitly modeled in the graph. Domingo et al. (2019) propose the construction of a heterogeneous graph for the structural analysis of road and building layouts by modeling road segments as additional nodes in the graph and connecting them to building nodes based on urban morphology. This approach serves as inspiration for extending the proximity graph constructed between buildings through DT by additionally incorporating road nodes. The logic for determining the road nodes is described in the following and illustrated in Figure 29. For each building in a street block, the Voronoi polygon based on its polygon vertices is constructed, representing its region of influence with respect to the other buildings in the street block (Yan et al. 2019). If the Voronoi polygon generated from the building touches the road network, the intersection between the Voronoi polygon and the road network determines the road segment associated with the respective building. Road nodes are determined by computing the intersection points between the road segment and the major and minor axes of the minimum bounding rectangle (MBR) of the building polygon (Kong et al. 2024). The road nodes are subsequently connected to the building by an edge, ensuring that the orientation of the building is implicitly preserved in the graph structure. Road nodes are not connected to each other, as the information associated with roads is only expected to be propagated to buildings during the learning process, rather than among the roads themselves.

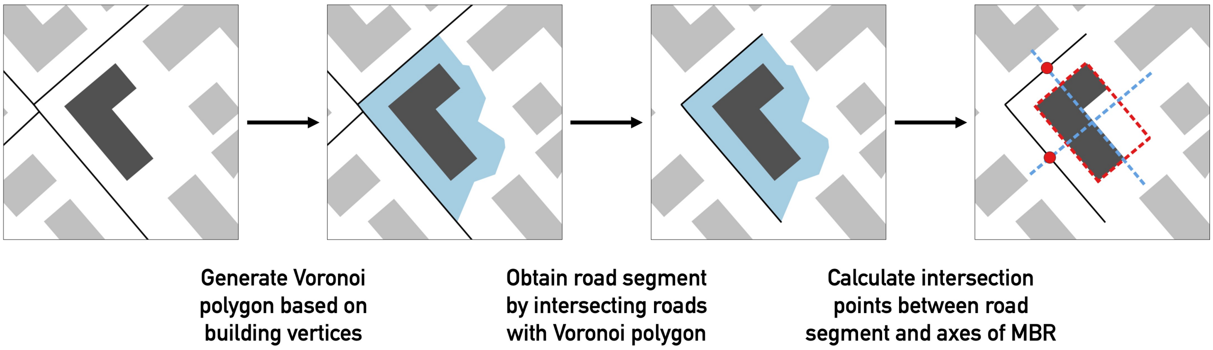


Figure 29 Procedure for identifying the road nodes (map data © swisstopo).

The heterogeneous street block graph is constructed by assigning focal building, context building, and road node types to the respective nodes according to the features they were derived from, such that they can be distinguished analogously to the multi-layered approach devised for the raster models. Based on the three node types, the edges are also provided with a type depending on the features they connect, leading to four different edge types (focal building ↔ context building, context building ↔ context building, focal building ↔ road, context building ↔ road). The heterogeneous street block graphs are stored together with the information regarding the operators that were applied to generalize the focal building. To the best of the author's knowledge, the present thesis constitutes the first attempt at explicitly representing the road network in a heterogeneous graph for DL-driven building generalization. The entire procedure for generating the street block graphs is illustrated in Figure 30. Unlike the raster-based representation, the graph only encodes relative positions between the map features. Therefore, it is refrained from applying similar geometrical transformations to the ones displayed in Figure 26 to the graph during training.

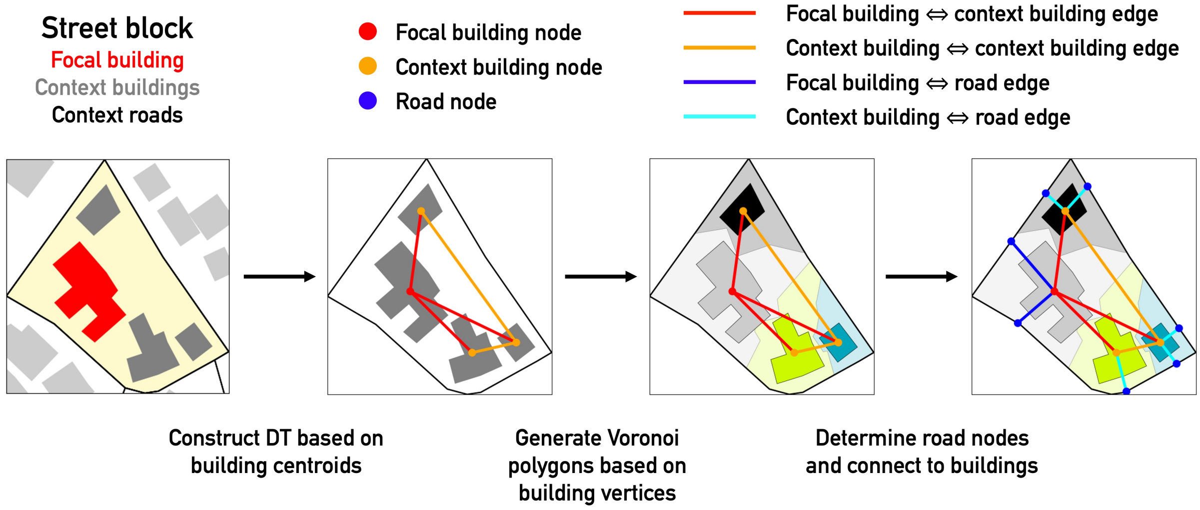


Figure 30 Procedure for deriving the heterogeneous graph representation (map data © swisstopo).

Feature extraction

In the context of graph-based DL, feature extraction refers to the process of attaching additional information to nodes and edges in the form of features derived from the raw data prior to model training that can be leveraged to guide the learning process (Acharya & Zhang 2020). The extraction of features enables the incorporation of domain knowledge into the DL-based generalization process and provides a way of encoding the rich information contained in the vector data lost during the graph construction procedure, potentially facilitating the classification of the presence or absence of generalization operators during the learning process (Xiao et al. 2024, Knura 2024). Table 5 provides an overview of the identified properties: distance, size, shape, density, orientation, and position. The properties were determined by studying relevant constraint-based and ML-based approaches for implementing the operators and identifying the structural measures utilized in their design.

Most properties have their roots in Gestalt principles (Wertheimer 1923), which state that humans instinctively perceive objects in structured arrangements. These principles are subconsciously applied by cartographers during manual generalization (Weibel 1996, Deng et al. 2018). Hence, Gestalt principles have been widely investigated in conventional generalization approaches such as for preliminary building grouping and pattern recognition (Yan et al. 2008, He et al. 2018, Deng et al. 2018, Yan et al. 2019, 2020) or for the implementation of generalization operators (Regnauld 2001, Li et al. 2004, Gong & Wu 2016). As DL-based cartographic generalization approaches are mainly concerned with replicating human decision making, it is reasonable to implicitly infuse these principles into the DL process. Based on the properties identified in Table 5, various features are determined using relevant literature from both conventional and DL-driven generalization approaches. These features are computed and attached to the building nodes in the heterogeneous street block graph to facilitate model training. The features are summarized in Table 6. For road nodes, only the features describing position were attached.

Table 5 Building properties exploited in the literature for operator implementation.

Operator	Property	Sources
Elimination	Size Distance Density	Li et al. 2004, Steiniger et al. 2010, Lee et al. 2017, Wang et al. 2017, Xiao et al. 2024
Aggregation	Size Distance Shape	Li et al. 2004, Allouche & Moulin 2005, Sester 2005, Lee et al. 2017
Typification	Size Distance Density Shape Orientation	Regnauld 2001, Li et al. 2004, 2005, Sester 2005, Burghardt & Cecconi 2007 Gong & Wu 2016, Shen et al. 2022, Xiao et al. 2024
Displacement	Distance Density Orientation	Ruas 1998, Lonergan & Jones 2001, Bader et al. 2005, Sester 2005
Enlargement	Size Distance	Steiniger et al. 2010

Table 6 Features attached to the heterogeneous street block graph.

Property	Feature	Calculation	Normalized w.r.t.	Source
Distance	Shortest Euclidean distance Hausdorff distance	see Figure 31a	max in block	Yan et al. 2019 Blana & Tsoulos 2022
Size	Building area Building perimeter	A_{building} P_{building}	max in block	Yan et al. 2019
Shape	Convexity Equivalent rectangular index	$\frac{A_{\text{building}}}{A_{\text{CH}}}$ $\sqrt{\frac{A_{\text{building}}}{A_{\text{MBR}}}} \cdot \frac{P_{\text{MBR}}}{P_{\text{building}}}$	normalized	Basaraner & Cetinkaya 2017
Density	Area of Voronoi polygon Impact area ratio	$\frac{A_{\text{Voronoi}}}{A_{\text{block}}}$ $\frac{A_{\text{building}}}{A_{\text{Voronoi}}}$	normalized	Xiao et al. 2024 Zhang et al. 2008
Orientation	Orientation of MBR Wall average	see Figure 31b	π	Duchêne et al. 2003 Hangouët 1998
Position	Centroid x-coordinate Centroid y-coordinate	$\frac{1}{N}\sum_{i=1}^{N} x_i$ $\frac{1}{N}\sum_{i=1}^{N} y_i$	block bounds	Knura 2024

To characterize the spatial proximity between two buildings for graph-based building pattern classification, Yan et al. (2019) propose the use of the shortest Euclidean distance $d(a,b) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}$ between the closest points $a = (a_1, a_2)$ and $b = (b_1, b_2)$ on two building outlines as weight for the edge incident on adjacent buildings. However, relying solely on Euclidean distances may be insufficient to capture the spatial relationships between polygonal buildings. Since polygons can be regarded as ordered sets of points, the Hausdorff distance (Hausdorff 1914) can be used to quantify the spatial distance between two polygons. Given two sets of points $A = \{a_1, \ldots, a_n\}$ and $B = \{b_1, \ldots, b_n\}$, the directed Hausdorff distance $h(A, B)$ is defined according to Equation (6) as the maximum, over all points in A, of the minimum Euclidean distance from each point in A to its closest point in B (Rote 1991).

$$h(A, B) = \max_{a \in A}\{\min_{b \in B} d(a, b)\}. \tag{6}$$

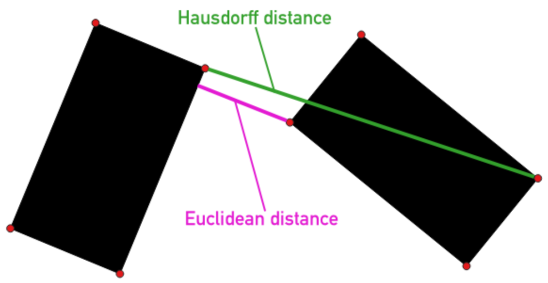
To obtain the undirected Hausdorff distance $H(A, B)$, the directed measures are commonly combined according to Equation (7) (Dubuisson & Jain 1994).

$$H(A, B) = \max\{h(A, B), h(B, A)\}. \tag{7}$$

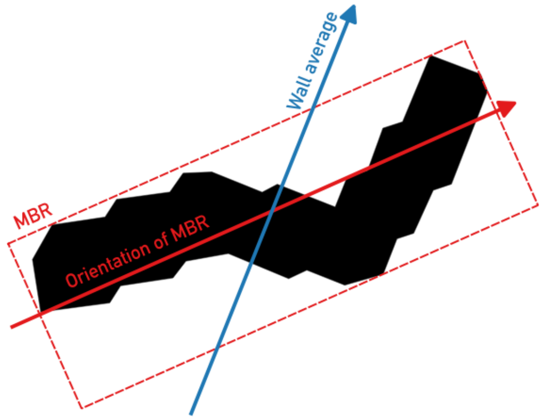
Compared to the shortest Euclidean distance between two objects, the Hausdorff distance manages to account for the shape of the objects to some degree (Filippovska et al. 2008, Blana & Tsoulos 2022). Moreover, the Hausdorff distance provides an approximation of the distance in cases where buildings overlap with the roads, causing the Euclidean distance to become zero. In addition to the Euclidean distance, the Hausdorff distance is therefore also considered as an edge weight between adjacent features. The two distance measures attached to the proximity graphs as edge weights are illustrated in Figure 31a.

According to Duchêne et al. (2003), buildings are commonly characterized by their position, shape, density, size, and orientation. Therefore, the remaining measures in Table 6 are attached to the building nodes in the heterogeneous street block graph. To characterize the size of a building, its area and perimeter are calculated (Yan et al. 2019). According to Basaraner & Cetinkaya (2017), convexity, the ratio between building area and the area of its convex hull (CH), and equivalent rectangular index (ERI), the deviation of a polygon from an equivalent rectangle, are efficient indices for describing the shape of building polygons.

Building density is measured with respect to the Voronoi tessellation created from building polygons within the street block, analogously to the tessellation created for the graph transformation procedure illustrated in Figure 30. The Voronoi polygon associated with a building can be considered its impact area, capturing the overall layout of a group of buildings within a street block and how they compete for space (Yan et al. 2019). The Voronoi tessellation allows for the calculation of two density measures for each building in the form of the ratio between its area and its impact area (Zhang et al. 2008) and the ratio between its impact area and the area of its street block (Xiao et al. 2024).



(a) Distance



(b) Orientation

Figure 31 Extracted building features (map data © swisstopo).

Encapsulating the orientation of a polygonal building in a single measure is considered a difficult task, especially for buildings with complex shapes that feature multiple characteristic orientations (Ma et al. 2023). According to Duchêne et al. (2003), building orientation can either refer to a building's general orientation characterizing its elongation or to the orientation of its walls. They identify the orientation of a building's MBR as an appropriate descriptor for the general orientation. Furthermore, Hangouët (1998) proposes the wall average to describe the orientation of a building in terms of the average orientation of its edges weighted by their length. The two building orientation measures are illustrated in Figure 31b. Finally, since the graph structure only encodes the relative position between nodes, the centroid coordinates of the respective building are assigned to all building nodes and the coordinates of the determined road points to all road nodes in the street block graph (Yan et al. 2019, Knura 2024).

In established DL approaches, extracted features are usually normalized to a common range. Feature normalization stabilizes and accelerates training and makes the model more robust, ultimately improving generalization (Ioffe & Szegedy 2015, Zhou et al. 2018). Therefore, the extracted features are normalized to the interval $[0, 1]$ with respect to the other map features within the same street block. The respective column in Table 6 summarizes the normalization strategy chosen for each feature. Some features, such as the measures extracted to capture shape and density, are normalized by design. The distance and size measures are normalized with regard to the respective maximum value encountered within the street block. The measures relating to building orientation are confined to the interval $[0, \pi]$. Therefore, they are normalized with respect to π. The coordinates describing node positions are normalized to the interval $[0, 1]$ with respect to the bounds of the street block enclosing the map features. A detailed sensitivity analysis regarding the relevance of the extracted features for operator prediction is provided in Section 6.1.2 in the form of an ablation study.

Network architecture

Heterogeneous graph neural networks (HGNN; Zhang, Song, Huang, Swami & Chawla 2019) are specialized GNNs designed to handle heterogeneous graphs. While homogeneous GNNs typically employ the same transformation function for every node, heterogeneous GNNs operate by leveraging specialized node-specific and edge-specific transformation functions, allowing them to learn representations that respect the semantic meaning of each node and edge type. Heterogeneous graph transformers (HGT, Hu et al. 2020) extend HGNNs by incorporating a transformer architecture utilizing specialized attention mechanisms for different node and edge types to discern intricate patterns (Yu et al. 2024). Hence, HGNN and HGT are sensible network architecture choices to conduct the operator classification task. Given a heterogeneous street block graph and the operators applied to the focal building, the models are trained to perform node-level classification for the focal building node, predicting the generalization operators that should be applied.

The architectures for HGNN[10] and HGT[11] were adopted from examples provided in the PyG documentation. HGNN uses a series of GraphSAGE convolutional layers specific to each edge type that perform message passing to gather information from neighboring nodes and combines it in a way that respects the unique roles of the different node and edge types in the heterogeneous street block graph. HGT operates based on a similar premise but uses HGT convolutional layers specifically designed for heterogeneous graphs that apply transformer-based convolution using multi-head attention across different node and edge types, allowing the model to weigh the importance of different node connections dynamically during the training process. Due to the inclusion of attention mechanisms, HGT is generally more complex, but may also offer higher expressiveness than HGNN, which relies on standard message passing. Finally, after the node embeddings have been constructed by the respective heterogeneous GNN, the updated focal building node features are passed through the classification heads outlined in Section 5.3.3 to perform the generalization operator prediction for the focal building. The architectures are illustrated in Figure 32 and their important properties are summarized in Table 7.

(a) HGNN

(b) HGT

Figure 32 Heterogeneous GNN architectures (map data © swisstopo).

[10] github.com/pyg-team/pytorch_geometric/blob/master/examples/hetero/hetero_conv_dblp.py
[11] github.com/pyg-team/pytorch_geometric/blob/master/examples/hetero/hgt_dblp.py

Table 7 Properties of the graph-based architectures.

Architecture	Model	Parameters	Source
HGNN	Elimination	481,665	Zhang, Song, Huang, Swami & Chawla 2019
	Multi-operator	540,548	
HGT	Elimination	700,466	Hu et al. 2020
	Multi-operator	750,389	

5.3.6 Multimodal models

Multimodal DL involves using multiple forms of data (*modalities*) to train DL models (Ngiam et al. 2011). The primary incentive to utilize multimodal data lies in the potential to extract complementary information from each modality involved in a specific learning task (Baltrušaitis et al. 2019). Multimodal DL aims to create a more comprehensive representation, potentially leading to significantly improved performance compared to what can be achieved through the use of a single modality alone (Ramachandram & Taylor 2017). Approaches based on multimodal DL have shown promising capabilities in a variety of fields, such as medical imaging (Huang et al. 2020) and speech recognition (Mroueh et al. 2015). Due to its success in other disciplines, Lafon et al. (2023) claim that multimodal models have the potential to overcome the existing limitations of models solely relying on individual modalities for classifying or segmenting cartographic data. Previous DL-based cartographic generalization approaches have shown increased success by integrating raster and vector data for improved results (Du, Wu, Zhu, Liu & Wang 2022, Jiang et al. 2023).

In the context of the operator classification task, multimodal DL models can be constructed that are capable of processing the raster and graph representations outlined in Sections 5.3.4 and 5.3.5 simultaneously, using the complementary information provided by the two modalities to make more informed decisions with respect to the operators that should be applied to generalize a focal building. The implementation of the multimodal models is based on a late fusion approach. Late fusion involves the integration of the outputs of separate models previously trained on individual modalities in order to construct an ensemble model that is capable of processing both modalities concurrently (Gadzicki et al. 2020). Therefore, the multimodal models are constructed by unifying the best-performing previously trained vector and raster models in a new model. Since two architectures were investigated for the raster and vector models, respectively, they can be combined into four multimodal architectures. Table 8 provides an overview of the possible combinations and their complexity. The two models are integrated by stripping them of their classification heads and adding a new linear layer that concatenates the high-level embeddings extracted by the individual models. The constructed multimodal models are subsequently appended with one and four classification heads for the elimination and multi-operator model, respectively. During model training, the parameters of the vector and raster models are frozen, as it is assumed that they have been sufficiently trained beforehand. Therefore, only the weights of the last few linear layers and the classification heads are adjusted during training. A schematic overview of the architecture of the proposed multimodal models is provided in Figure 33.

Table 8 Properties of the multimodal architectures.

Architecture	Model	Parameters
CNN + HGNN	Elimination	4,556,993 (1,622,273 trainable)
	Multi-operator	9,432,772 (6,489,092 trainable)
CNN + HGT	Elimination	4,775,794 (1,622,273 trainable)
	Multi-operator	9,642,613 (6,489,092 trainable)
ViT + HGNN	Elimination	21,067,649 (82,177 trainable)
	Multi-operator	21,323,140 (328,708 trainable)
ViT + HGT	Elimination	21,286,450 (82,177 trainable)
	Multi-operator	21,532,981 (328,708 trainable)

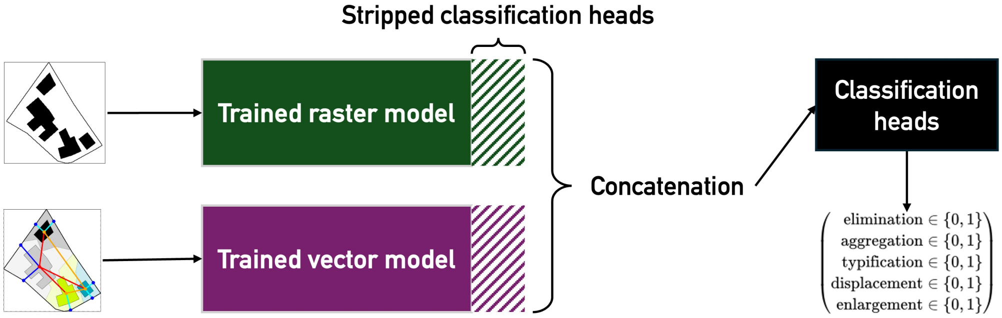


Figure 33 Multimodal model architecture (map data © swisstopo).

5.3.7 Model training

The models are trained using the balanced training and validation sets outlined in Section 5.2.3 to eliminate the identified bias toward the majority classes. The model optimizes its parameters based on the training set, whereas the validation set is used during the training process to monitor the performance and assess the model's capability to generalize the knowledge to unseen data. Model training is carried out in batches, enabling more effective training by enhancing computational efficiency and stabilizing gradient estimates (Masters & Luschi 2018), resulting in better generalization (Keskar et al. 2016). Given the huge number of 100,000 training samples, a large batch size of 512 is chosen to train the models, which accelerates processing times. Early stopping is employed to determine the appropriate number of epochs for model training by identifying the epoch at which the model starts to exhibit overfitting with respect to the validation dataset (Prechelt 1998, Ying 2019).

Table 9 shows the training time per epoch required to process 100,000 training and 25,000 validation samples on a single NVIDIA V100 GPU for each investigated model architecture. Evidently, training the GNNs is much less computationally expensive compared to the raster models. This disparity stems from the inherent sparsity of graph data structures compared to images, which simplifies processing. In contrast, images are dense data representations requiring consideration of every pixel, which is vastly more computationally intensive. As a consequence of the time-consuming nature of the training procedure, performing more sophisticated model validation techniques, such as cross-validation, is considered unfeasible.

Table 9 Training times for the investigated model architectures.

Modality	**Architecture**	**Training time per epoch**
Raster	CNN	0.72 h
	ViT	1.43 h
Vector	HGNN	0.08 h
	HGT	0.22 h
Multimodal	CNN + HGNN	1.53 h
	CNN + HGT	1.79 h
	ViT + HGNN	2.87 h
	ViT + HGT	3.14 h

To iteratively update the network parameters based on the training data, the adaptive moment estimation (Adam; Kingma & Ba 2014) optimizer with a learning rate of $5 \cdot 10^{-4}$ was applied. Adam constitutes an extension to conventional stochastic gradient descent methods by maintaining adaptive learning rates per weight and incorporating a bias correction mechanism (Zaheer & Shaziya 2019). Furthermore, DL models require the specification of a loss function that measures the discrepancy between the predicted output of a model and the true target values, guiding the model during training to minimize this error by repeatedly applying the specified optimization algorithm (Cho et al. 2019). All models presented throughout this thesis are trained using binary cross-entropy (BCE) loss, which constitutes the default loss function choice for multi-label classification problems (Demirkaya et al. 2020, Ridnik et al. 2021). Given a set of n true binary labels $Y = \{y_1, \ldots, y_n\}$, where $y_i \in \{0, 1\}$, and its associated set of predicted probabilities $\hat{Y} = \{\hat{y}_1, \ldots, \hat{y}_n\}$, where $\hat{y}_i \in [0, 1]$, the BCE loss L_{BCE} is calculated as shown in Equation (8).

$$L_{BCE} = -\frac{1}{n}\sum_{i=1}^{n} y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i). \tag{8}$$

To determine the importance of cartographic context for the prediction of generalization operators and to assess the suitability of the proposed data model, all models and architectures are additionally trained by excluding the context roads from the input data, only feeding the data associated with the buildings to the models. For the raster-based approach, this involves supplying the models with a two-channel raster containing the focal building and the context buildings, respectively. To exclude the roads for the graph-based models, the road nodes and the associated edges connecting to the focal and context building nodes are removed. Thus, the resulting heterogeneous street block graph contains only two node types and two edge types (focal building $\leftrightarrow$ context building, context building $\leftrightarrow$ context building). The model architectures are adjusted accordingly to process the modified training samples. The results obtained using the adapted training samples and models can be used to ascertain if the developed data models can effectively capture the contextual nature of cartographic generalization and whether the inclusion of the road network is beneficial to model performance.

5.3.8 Model evaluation

After training for the appropriate number of epochs determined according to early stopping, the model can be used to conduct predictions regarding the operators that should be applied to generalize a given focal building. To obtain the prediction, the activation $x \in \mathbb{R}$ of the respective output neuron is mapped through a Sigmoid activation function σ as shown in Equation (9) to obtain values in $[0, 1]$. This value can subsequently be used to determine the presence or absence of the respective generalization operator using a threshold.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{9}$$

To evaluate the models, the predictions obtained from the models can be compared to the original annotated operators in the datasets. An effective tool for this comparison is the use of binary confusion matrices, which are contingency tables that help visualize the performance of DL models by showing the actual versus predicted operators. Each entry in the confusion matrix represents the number of predictions that fall into each category defined by the actual and predicted classes, specifically true positives, false positives, true negatives, and false negatives. Based on the confusion matrix, evaluation metrics such as precision, recall, false positive rate (FPR), overall accuracy, and F_1 score can be calculated to provide further insight into the performance of the models. These metrics are particularly useful for understanding how well the model performs across different operators and can help identify any biases or weaknesses in the predictive capabilities of the models (Luque et al. 2019). An overview of the metrics incorporated to evaluate trained models, including their calculation, is provided in Table 10.

To determine the optimal classification threshold for each operator, the receiver operating characteristic (ROC) curve can be used. The ROC curve plots FPR against the true positive rate (TPR, recall) at varying discrimination threshold values. In the case of the operator classification task, the optimal threshold is chosen as the threshold that minimizes FPR while maximizing TPR (Yang & Berdine 2017). However, it has been shown that for classification problems involving datasets with imbalanced classes, ROC curves may offer a misleadingly favorable view of model performance, since they are not sensitive to the proportion of positive and negative instances

Table 10 Metrics used to evaluate the models.

Actual \ Predicted	Not Present	Present	
Not Present	True Negative (TN)	False Positive (FP)	**FPR** $\frac{FP}{FP+TN}$
Present	False Negative (FN)	True Positive (TP)	**TPR / Recall** $\frac{TP}{TP+FN}$
	F_1 score $\frac{2TP}{2TP+FP+FN}$	**Precision** $\frac{TP}{TP+FP}$	**Accuracy** $\frac{TP}{TP+TN+FP+FN}$

(Davis & Goadrich 2006). As this is the case for the dataset containing annotated generalization operators, precision-recall (PR) curves that express precision as a function of recall can provide an effective alternative, since they do not incorporate true negatives (Saito & Rehmsmeier 2015). In the context of the PR curve, an optimal threshold maximizes both precision and recall.

To identify the optimal separation threshold for each operator, ROC and PR curves are generated for each model based on the validation set. The final threshold used to evaluate the performance of the models corresponds to the arithmetic mean between the optimal thresholds determined through both curves. After identifying the optimal threshold, the models are separately evaluated on their respective imbalanced test set. Assessment of model performance involves the calculation of the evaluation metrics in Table 10 for each operator. Additionally, the ROC and PR curves generated based on the test set can be used to compare different models for the same operator by calculating the area under the curve (AUC), whereby a larger AUC implies better performance (Yang & Berdine 2017).

In a first step, the previously outlined metrics and the ROC and PR curves are used to holistically evaluate model performance, allowing for the identification of the optimal model architecture for each modality. The best-performing architectures are subsequently used to conduct stratified performance evaluation with respect to the following grouping variables: street block area, urban-rural status, and operator combinations. The stratified evaluation is performed based on the test set not seen by the DL models during training to ensure unbiased results. After assigning the grouping variables, the aforementioned evaluation metrics can be calculated for every level of the variable to assess whether the models perform significantly better for certain subsets of the test set. The justification and methodology for implementing the stratification with respect to the different variables are described in the following.

Street block area

To construct training samples for the raster-based models, all buildings and roads belonging to the street block of a focal building are converted to a raster with 256 x 256 pixels by choosing the resolution according to Equation (4). Consequently, the resolution of the raster deteriorates with increasing street block area, as illustrated in Figure 34. The decrease in resolution has the potential to negatively affect the performance of the raster-based model for focal buildings that are part of large street blocks, as models may struggle to identify and distinguish buildings (Touya et al. 2019, Courtial et al. 2022*b*). In contrast to their raster-based counterparts, graph-based models are inherently less susceptible to this scale effect, as GNNs can process input graphs of variable size without loss of information (Zhou et al. 2023, Harrie et al. 2024, Knura 2024). To investigate the effect of street block area on generalization operator predictions, the focal buildings within the test set are stratified with respect to the area of their street blocks. The stratification involves dividing the samples in the test set into four distinct groups, each representing a respective quartile range of street block areas, as illustrated for all buildings in the training database in Figure 34. The first quartile contains the buildings belonging to the smallest 25% of the street blocks, whereas the largest 25% are contained in the fourth quartile.

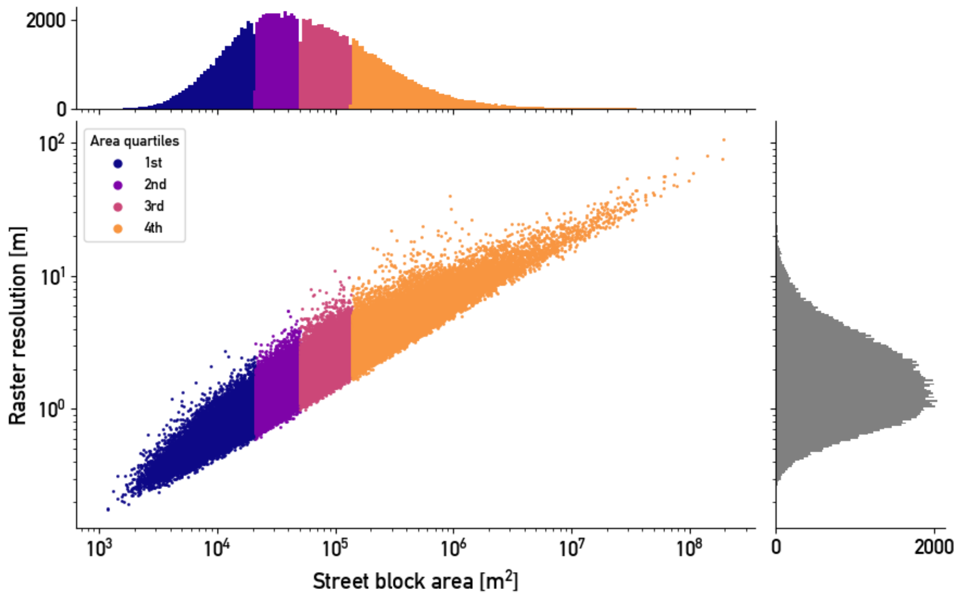


Figure 34 Relationship between street block area and raster resolution.

Urban-rural status

Compared to rural areas, urban areas on topographic maps exhibit a higher density of map features. Therefore, urban areas usually require a higher degree of generalization with an increased use of complex contextual generalization operators, such as typification or displacement, to address cartographic conflicts arising from densely packed map objects (Ruas & Mackaness 1997, Mustière & Moulin 2002, Spiess et al. 2005). The dependence of generalization criteria on geographic context implies that DL models may struggle to transfer their knowledge from urban to rural regions and vice versa (Zhou et al. 2023). In the context of the present thesis, training samples are randomly sampled from the original database without considering geographic context. Therefore, a stratification of the test set according to urban-rural status can provide information on how geographic context affects model performance.

To this end, the Swiss Land Use Statistics[12] maintained by the Federal Statistical Office are incorporated in the analysis. The land use statistics consist of 4.1 million sample points regularly spaced at 100 m intervals. Each sample point is assigned one of four main land use categories: settlement area, agricultural area, forested area, and unproductive area. The land use statistics points are reclassified into settlement area and non-settlement area and mapped onto a 100 x 100 m square grid to generate a seamless dataset of contiguous urban and rural grid cells. Based on the constructed grid, the street blocks are classified as urban or rural according to the predominant urban-rural status of their intersection area with the grid cells. The point-based reclassified land use statistics, the overlaid square grid, and the derived urban and rural street blocks are illustrated in Figure 35. To stratify the buildings in the test set with respect to their urban-rural status, every building is assigned the status previously computed for its street block.

[12]bfs.admin.ch/swiss-land-use-statistics

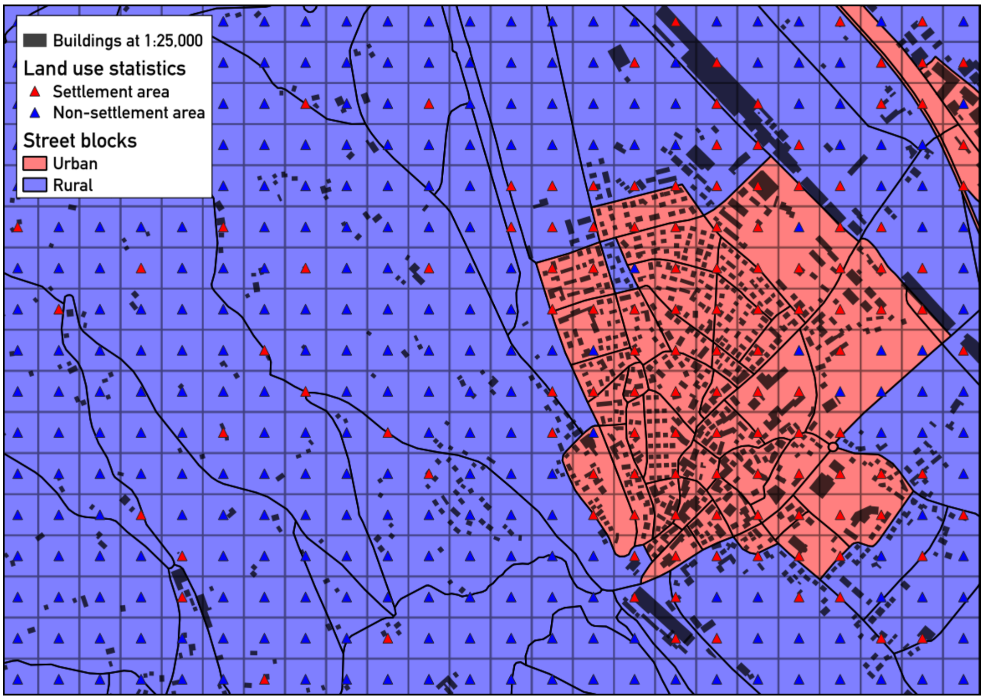


Figure 35 Classification of street blocks according to urban-rural status (map data © swisstopo, FSO).

Operator combinations

The DL models described in the previous sections operate by predicting individual generalization operators for a given focal building. However, during conventional generalization procedures, buildings are usually subject to the application of multiple operators in succession, as illustrated in Figure 20 (Regnauld & McMaster 2007). Therefore, an evaluation of the capacity of the multi-operator models to correctly predict combinations of generalization operators is of interest. To stratify the test set with respect to different operator combinations, a similar approach to the LP-transformation applied to balance the data is employed: The original multi-label classification problem is reformulated as a multi-class classification problem by considering each unique set of operators in the test set as a distinct class (Charte et al. 2015). The transformation is exemplified in Figure 36. Since the multi-operator model is tasked with predicting 4 generalization operators, there are $2^4 = 16$ possible sets of operators, of which only 8 are present within the test set. The predictions of the individual operators made by the models on the samples in the test set are analogously assigned to a set. Consequently, the true set of applied operators can be evaluated against the predicted set of operators to determine whether model performance differs significantly between operator combinations.

Aggregation	Typification	Displacement	Enlargement
0	1	0	1
1	1	0	0
1	1	1	1
0	0	1	1

LP transform

{0, 1, 0, 1}	{1, 1, 0, 0}	{1, 1, 1, 1}	{0, 0, 1, 1}
1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	1

Figure 36 Multi-label to multi-class transformation for selected operator combinations.

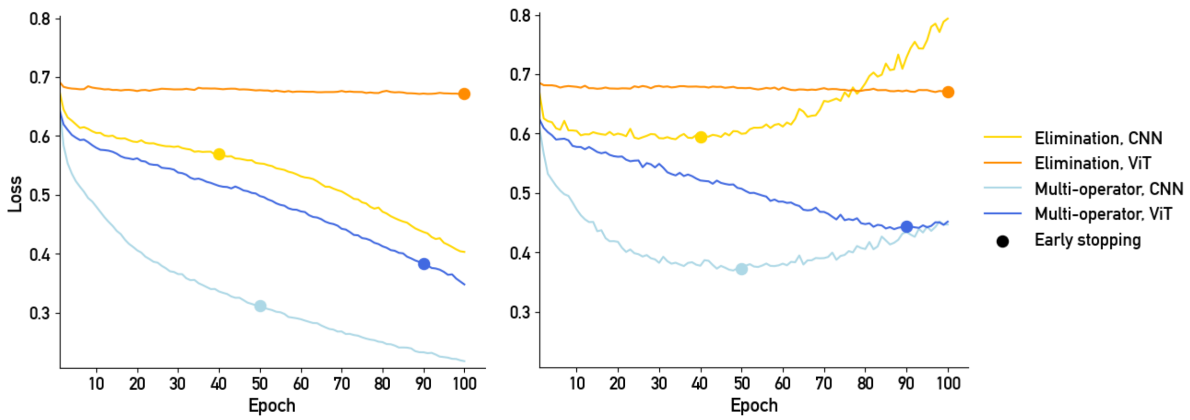
6
Results

6.1 Global evaluation

6.1.1 Raster models

Loss curves

Figure 37 shows the progression of training and validation loss as the raster models are trained over the course of the first 100 epochs. The number of epochs for training the final models chosen according to the early stopping criterion is indicated by points on the curve. As apparent by the non-decreasing training and validation loss, ViT struggles to learn the elimination operator, but also does not display any overfitting within the first 100 epochs. CNN manages to decrease the validation loss for the elimination model until about epoch 40, after which it exhibits overfitting, manifesting itself in an increase in validation loss. For the multi-operator model, both models achieve significantly lower training and validation loss values compared to the elimination model. CNN starts to exhibit overfitting past epoch 50, whereas training for 90 epochs is beneficial for ViT with respect to validation loss.



(a) Training loss **(b)** Validation loss

Figure 37 Loss curves for the raster models.

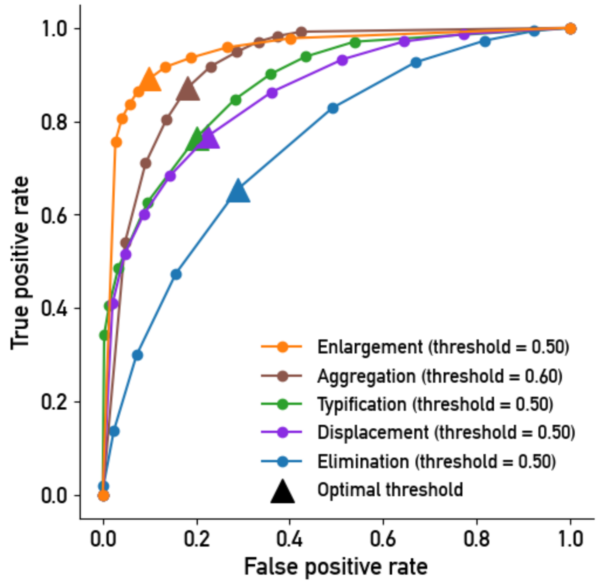
Evaluation metrics

The ROC and PR curves after training for the appropriate number of epochs for CNN and ViT are displayed in Figure 38 and Figure 39, respectively. When comparing the curves between the balanced validation set and the imbalanced test set, both architectures show similar patterns: The ROC values do not differ significantly between the two datasets, whereas the PR curves computed on the test set display smaller AUC. The effect is most pronounced for aggregation, typification, and elimination, raising potential overfitting concerns for these operators. With respect to AUC on the test set, both architectures excel at predicting enlargement. The large imbalance of samples for the displacement operator leads to a high performance of both architectures in classifying the operator on the test set with respect to precision and recall, whereas the ROC curve displays a significantly smaller AUC.

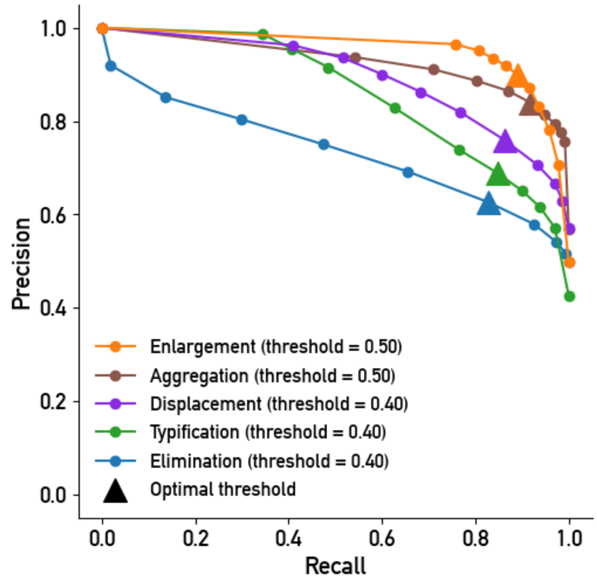
The evaluation metrics for the raster models using the optimal thresholds identified based on the ROC and PR curves constructed for the validation set are illustrated in Table 11. Evidently, enlargement is the easiest operator to predict, as both models achieve very high evaluation metrics. Enlargement is followed by aggregation and displacement, for which the two architectures still manage to exhibit acceptable evaluation metric values. For typification, a significant decrease in evaluation metrics can be observed. As opposed to the operators classified by the multi-operator model, the elimination model surprisingly exhibits by far the lowest evaluation metric values, even though the model is tasked with the supposedly easier classification task of predicting a single operator rather than four simultaneously. Using the threshold identified based on the validation set, both architectures achieve very low precision values. For CNN, the classification of the elimination operator only slightly outperforms a random guess, while ViT even displays an overall accuracy of below 0.5, implying that model performance can be improved by simply swapping the labels. Table 11 reveals that CNN outperforms ViT for every operator with respect to F_1 score, ROC AUC, and PR AUC. Therefore, CNN is identified as the optimal raster-based model architecture for further analysis.

Table 11 Evaluation metrics for the raster models.

Architecture	Metric	Operator				
		Elimination	**Aggregation**	**Typification**	**Displacement**	**Enlargement**
CNN	Accuracy	0.63	0.77	0.71	0.72	0.86
	Precision	0.23	0.73	0.51	0.95	0.97
	Recall	0.75	0.87	0.69	0.73	0.86
	F_1 score	0.35	0.79	0.58	0.82	0.91
	ROC AUC	0.74	0.85	0.76	0.76	0.91
	PR AUC	0.34	0.85	0.51	0.96	0.99
ViT	Accuracy	0.40	0.67	0.64	0.67	0.79
	Precision	0.16	0.63	0.44	0.94	0.95
	Recall	0.82	0.89	0.74	0.68	0.80
	F_1 score	0.27	0.73	0.55	0.79	0.87
	ROC AUC	0.60	0.76	0.71	0.66	0.85
	PR AUC	0.20	0.77	0.44	0.94	0.97

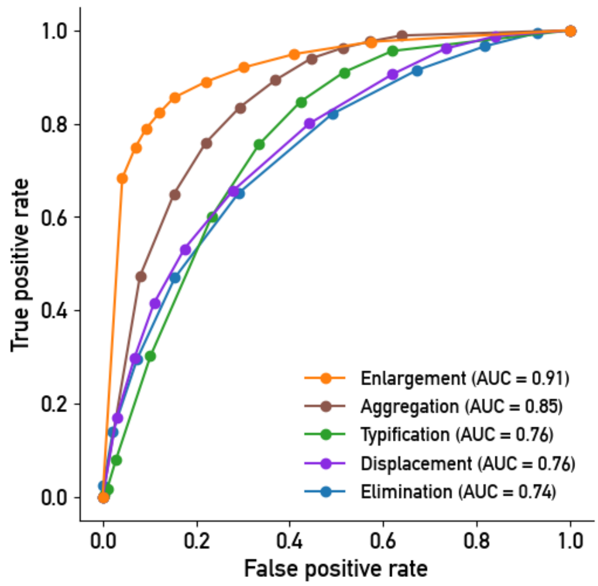


(ai) ROC

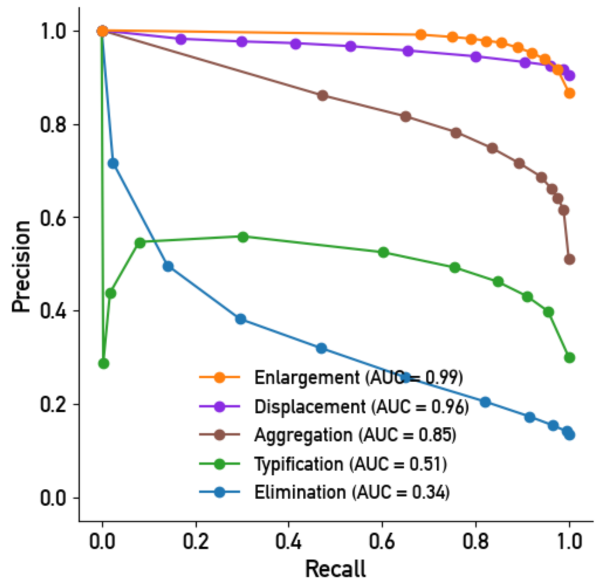


(aii) PR

(a) Validation set



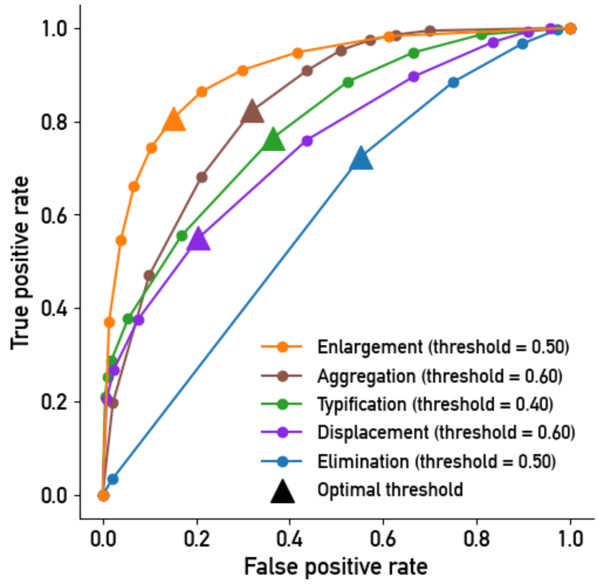
(bi) ROC



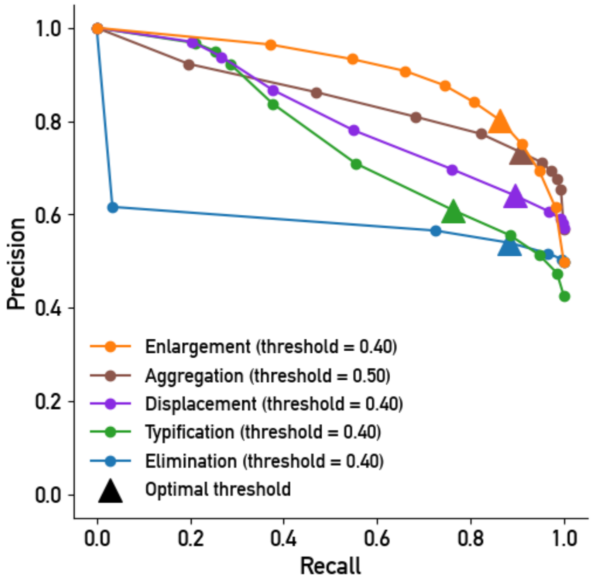
(bii) PR

(b) Test set

Figure 38 ROC and PR curves for CNN.

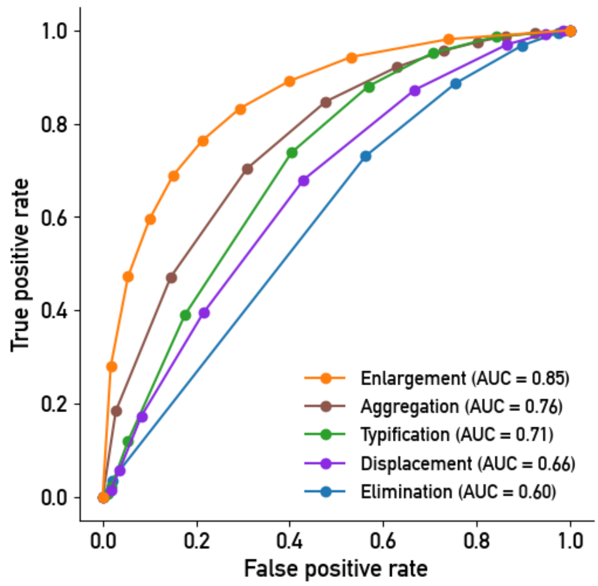


(ai) ROC

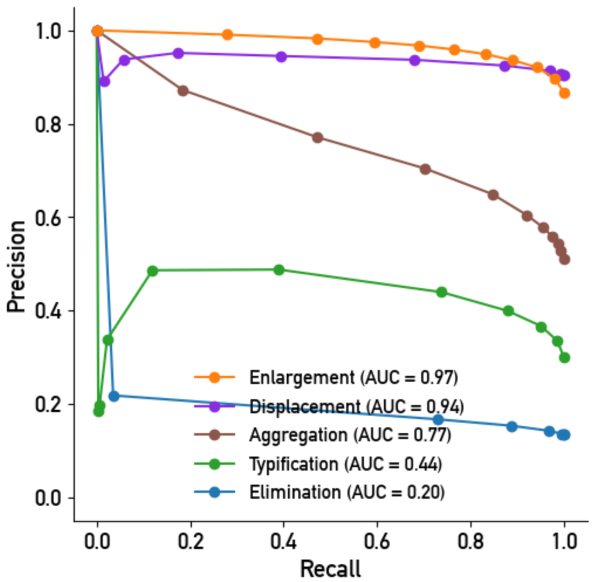


(aii) PR

(a) Validation set



(bi) ROC



(bii) PR

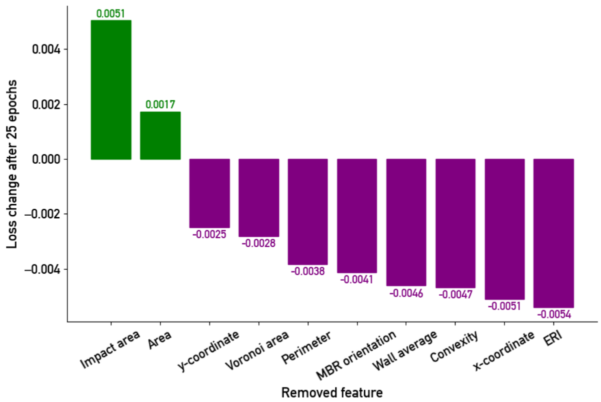
(b) Test set

Figure 39 ROC and PR curves for ViT.

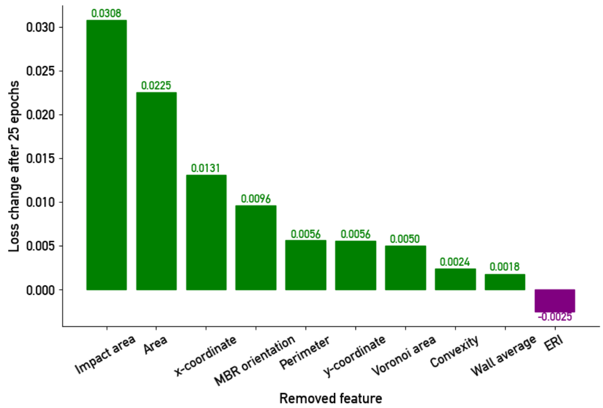
6.1.2 Vector models

Feature relevance

Following the approach of Zhou et al. (2023), an ablation study was performed to assess the relevance of the features identified in Table 6 for training the graph-based models. Feature ablation refers to the process of systematically retraining models after removing a single feature and observing the impact on performance metrics such as training loss compared to a model that uses all features (Meyes et al. 2019). If the removal of a feature has a positive effect on the evaluation metrics, it suggests that its inclusion is detrimental to the learning process. Therefore, this feature should be excluded during training. Figure 40 shows the change in training loss evoked by the omission of a single feature after training for 25 epochs compared to the loss obtained by training the respective model with all features. Features associated with a negative loss change imply that the removal of the feature decreases the training loss, which in turn is beneficial to model training. Therefore, features that induce a negative loss change can be considered irrelevant and should be excluded for the final training of the model.

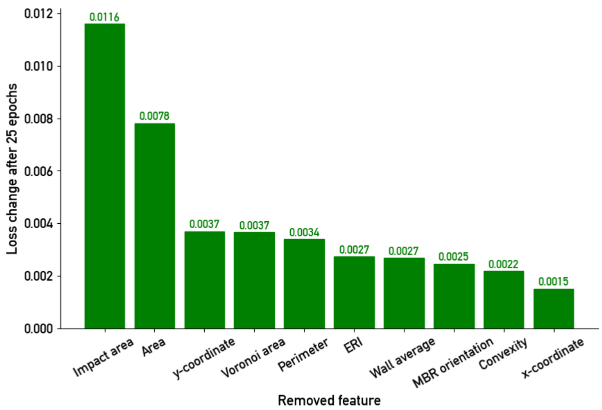


(ai) Elimination model

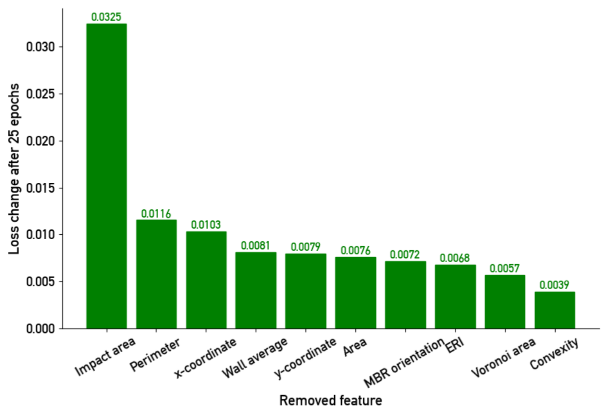


(aii) Multi-operator model

(a) HGNN



(bi) Elimination model



(bii) Multi-operator model

(b) HGT

Figure 40 Results of the feature relevance ablation study.

Based on Figure 40, the removal of all the features identified in Table 6 during the training process of HGT is associated with a loss increase, suggesting that all features are beneficial for training HGT on the operator prediction task. HGNN displays some features, whose exclusion leads to lower training loss, implying smoother training performance without these features. For the multi-operator model, the shape measure ERI has a negative effect on the training process, while all features apart from impact area and building area are detrimental to the elimination model. In fact, impact area (ratio between building area and the Voronoi polygon generated by the building with respect to its street block) and building area are the only features whose inclusion is beneficial for training every model.

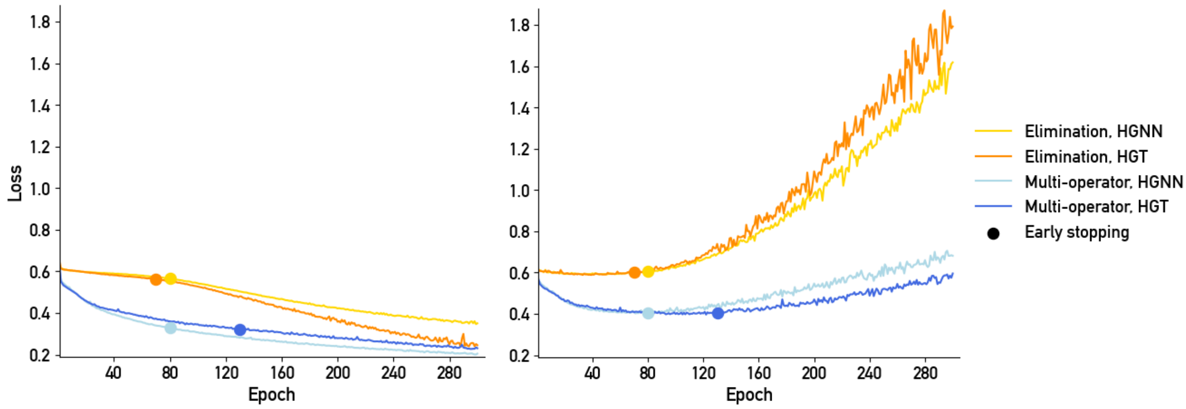
Across all models, impact area is the most relevant feature for predicting the generalization operators that should be applied to the focal building, as its exclusion is associated with the highest loss increase for every model. This suggests that building density within a street block is an important predictor for generalization operators. The shape measures convexity and ERI are the least relevant features for the task at hand, as their exclusion is associated either with large negative loss changes or small positive loss changes for all models. This is intuitive, as simplification is not subject to investigation and the application of the remaining contextual operators is not expected to be significantly affected by the shape of individual buildings. In the following, only the features whose exclusion negatively affects the training process, as summarized in Table 12, are incorporated to train the graph-based models.

Table 12 Features selected for training the graph-based models.

Architecture	**Model**	**Features**
HGNN	Elimination	Impact area, building area
	Multi-operator	All features in Table 6 except ERI
HGT	Elimination Multi-operator	All features in Table 6

Loss curves

Figure 41 illustrates the evolution of training and validation loss for the vector models over 300 epochs, whereby the epoch chosen for early stopping is indicated by a point on the respective curve. As opposed to the raster-based models, where significant differences can be observed between the investigated architectures, HGNN and HGT perform almost identically with respect to training and validation loss. Both architectures struggle to learn elimination, which is reflected in the inability of the models to significantly decrease the validation loss over the course of the first 100 epochs. Overfitting can be observed after 80 and 70 epochs for HGNN and HGT, respectively. Similarly to the raster-based models, the vector models are more adept at learning aggregation, typification, displacement, and enlargement, as both architectures manage to induce a substantial gradual decrease in validation loss for the multi-operator model. The validation loss only starts to increase after 80 and 130 epochs for HGNN and HGT, respectively, indicating that terminating model training at these epochs is appropriate to avoid overfitting.



(a) Training loss **(b)** Validation loss

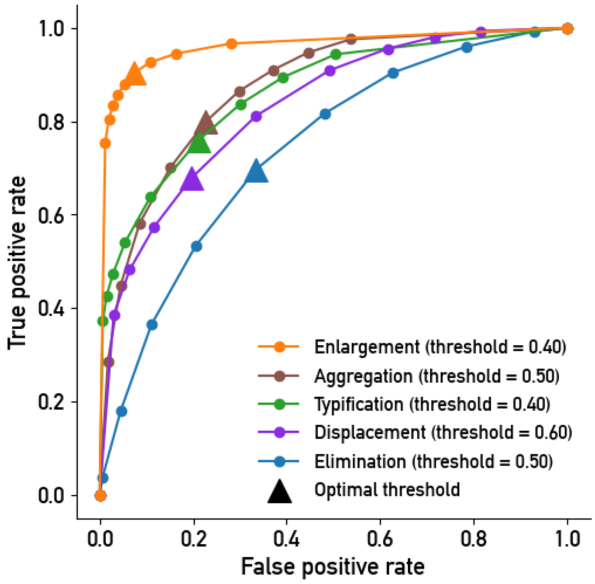
Figure 41 Loss curves for the vector models.

Evaluation metrics

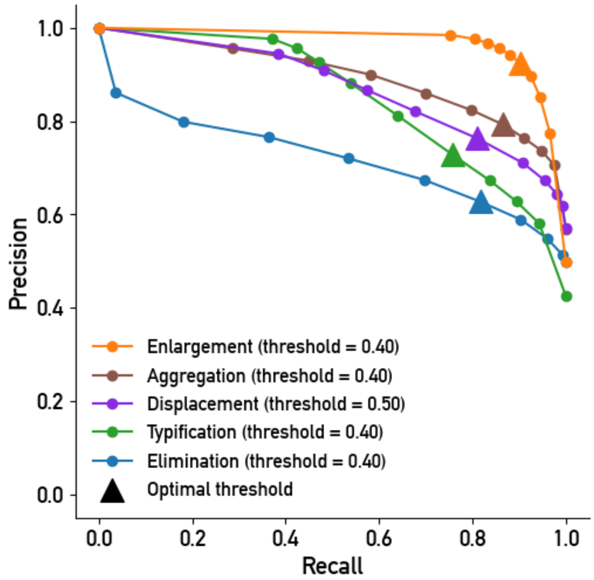
The ROC and PR curves for HGNN and HGT are displayed in Figure 42 and Figure 43, respectively. Interestingly, both architectures produce ROC and PR curves that are almost identical, implying that there is a negligible performance difference between the two. Furthermore, the curves also look very similar to those produced by the raster-based models, with the same observable difference between the validation and test set PR curves associated with the aggregation, typification, and elimination operators. The evaluation metrics for the vector models using the optimal discrimination threshold determined by the validation set ROC and PR curves are illustrated in Table 13. Analogously to the raster-based model, the vector-based multi-operator models yield better classification results compared to the elimination models. Since the difference between the evaluation metrics of the two architectures is insignificant, HGNN is chosen as the optimal vector-based architecture, as it is less complex and consequently less computationally expensive compared to HGT according to Tables 7 and 9.

Table 13 Evaluation metrics for the vector models.

Architecture	Metric	Operator				
		Elimination	Aggregation	Typification	Displacement	Enlargement
HGNN	Accuracy	0.61	0.71	0.72	0.58	0.88
	Precision	0.22	0.67	0.52	0.95	0.98
	Recall	0.78	0.84	0.64	0.56	0.88
	F_1 score	0.35	0.75	0.57	0.71	0.93
	ROC AUC	0.74	0.80	0.76	0.70	0.93
	PR AUC	0.31	0.81	0.52	0.95	0.99
HGT	Accuracy	0.62	0.72	0.71	0.61	0.89
	Precision	0.23	0.68	0.51	0.95	0.98
	Recall	0.77	0.85	0.69	0.60	0.89
	F_1 score	0.35	0.76	0.58	0.74	0.93
	ROC AUC	0.74	0.81	0.76	0.70	0.94
	PR AUC	0.31	0.82	0.51	0.95	0.99

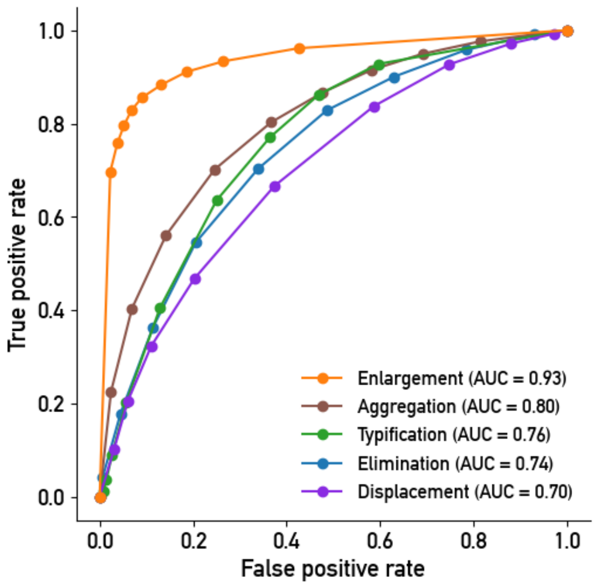


(ai) ROC

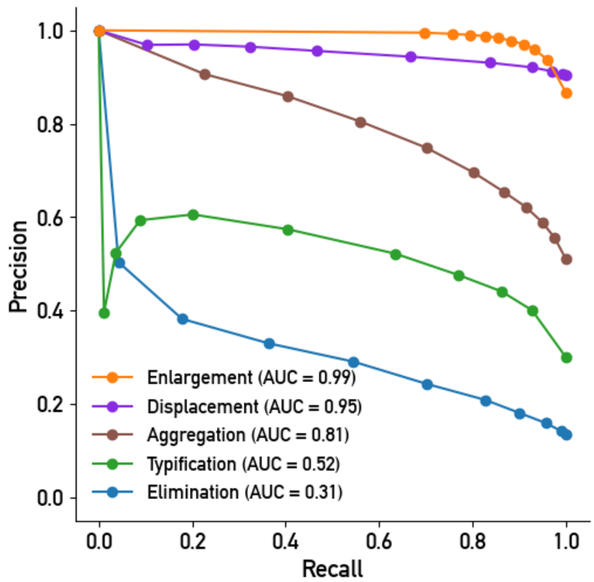


(aii) PR

(a) Validation set



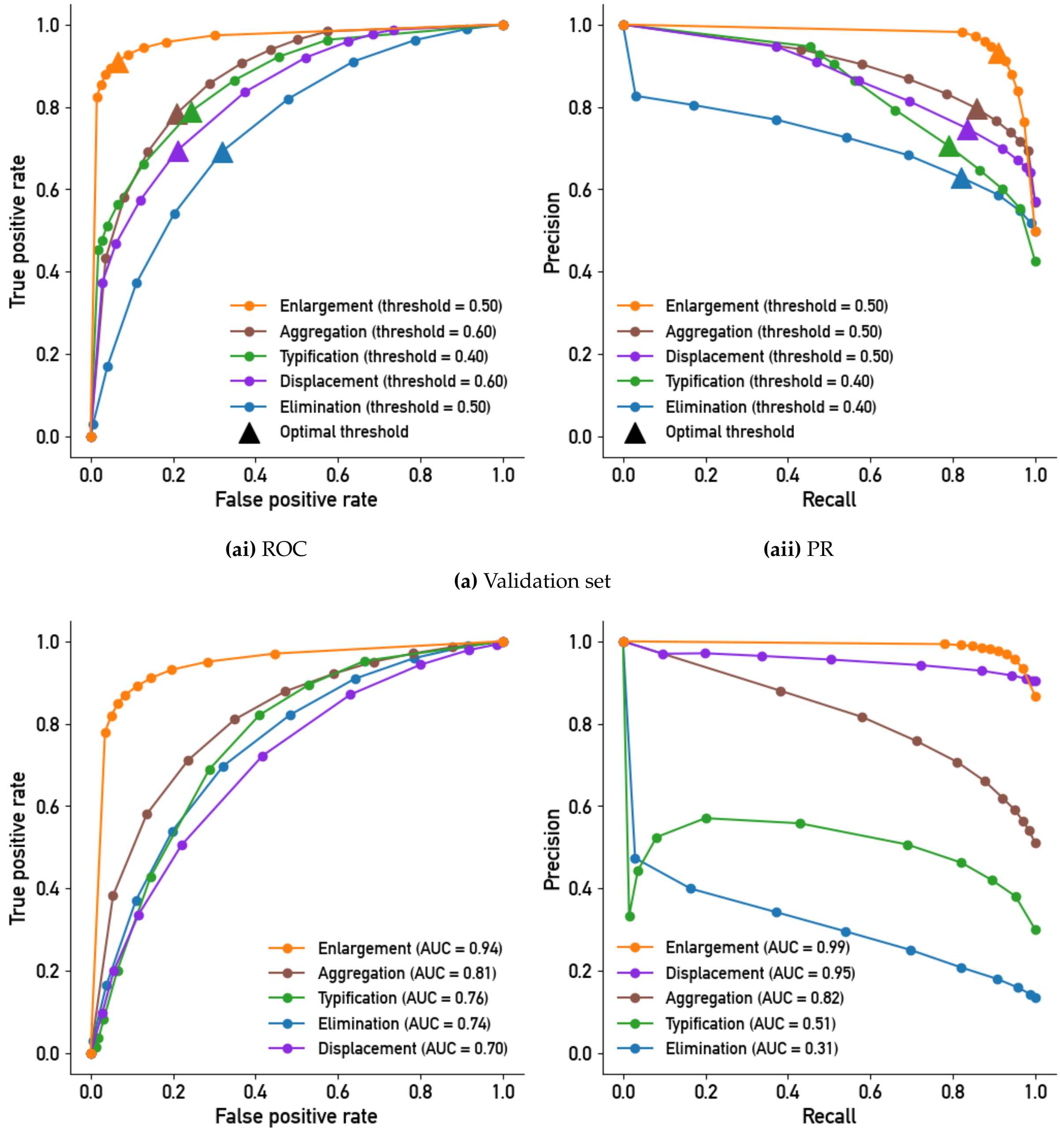
(bi) ROC



(bii) PR

(b) Test set

Figure 42 ROC and PR curves for HGNN.



(ai) ROC **(aii)** PR

(a) Validation set

(bi) ROC **(bii)** PR

(b) Test set

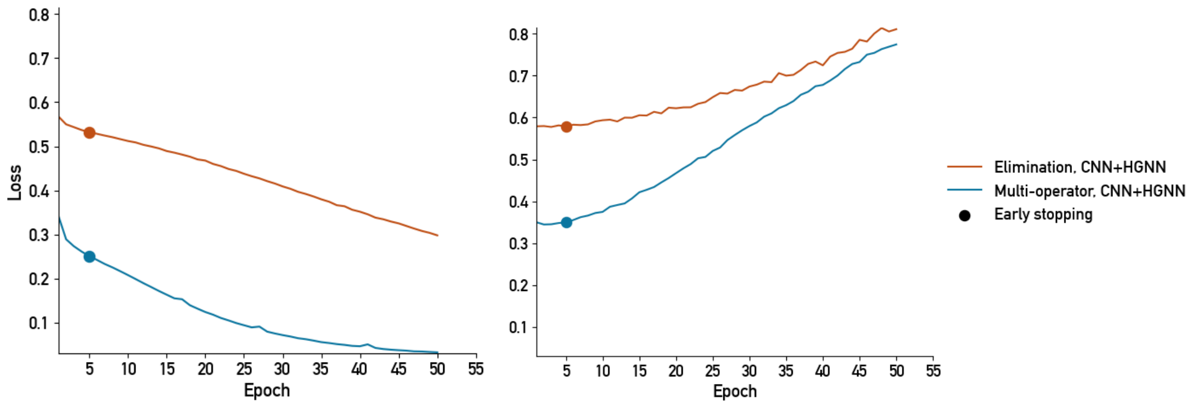
Figure 43 ROC and PR curves for HGT.

6.1.3 Multimodal models

Since CNN and HGNN were identified as the best-performing architectures for the models operating on individual modalities in Sections 6.1.1 and 6.1.2, respectively, they are chosen as the constituent architectures for the multimodal model.

Loss curves

Figure 44 shows the progression of training and validation loss as the multimodal models are trained over the course of the first 50 epochs. Judging from the validation loss curves, both the elimination and multi-operator models start to display excessive overfitting after just 5 epochs, which is therefore chosen as the number of epochs for early stopping.



(a) Training loss **(b)** Validation loss

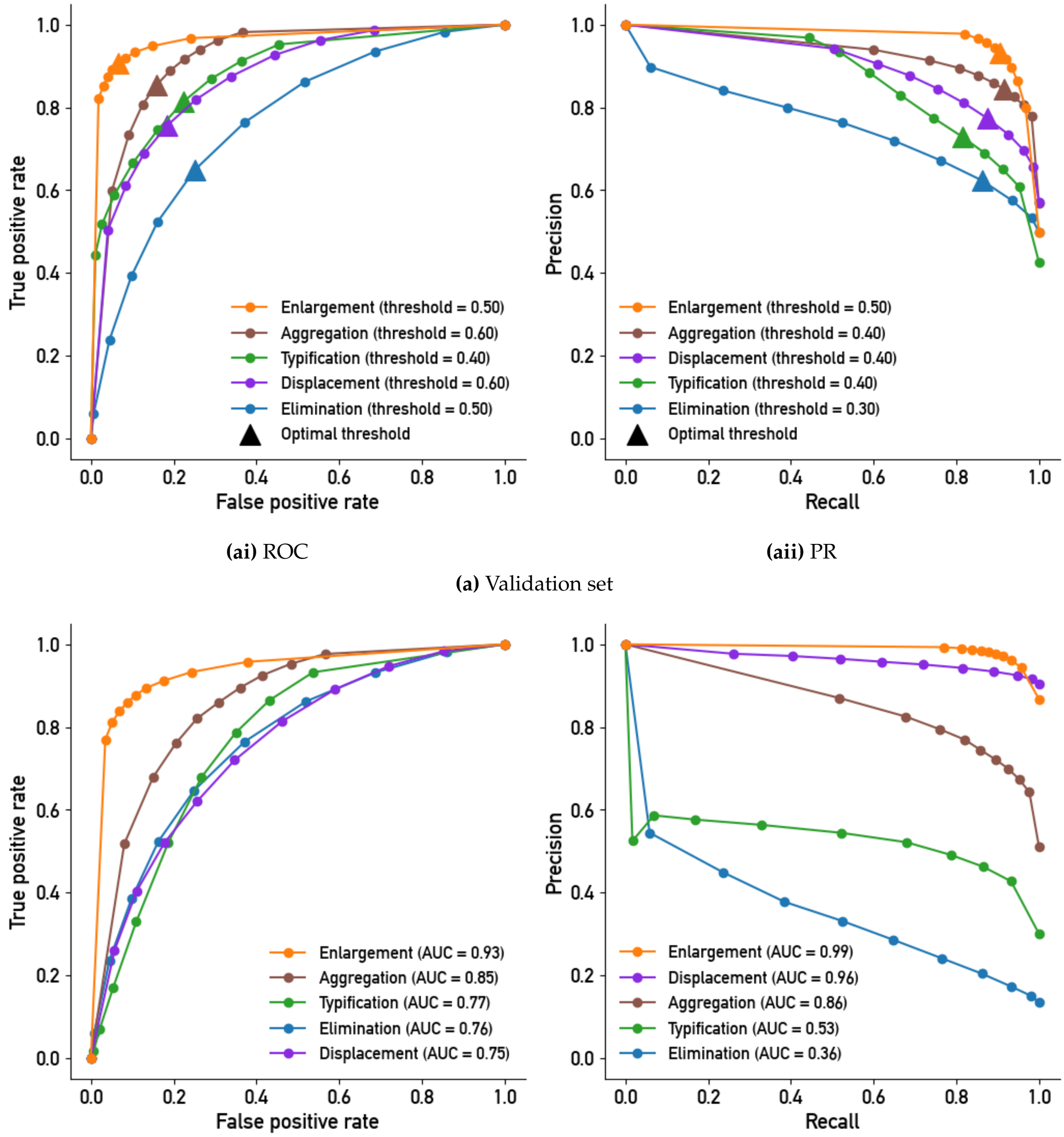
Figure 44 Loss curves for the multimodal models.

Evaluation metrics

The ROC and PR curves and the evaluation metrics for the multimodal model are shown in Figure 45 and Table 14, respectively. With respect to the ROC and PR curves, there is a notable decrease in AUC computed on the test set compared to the validation set for aggregation, typification, and elimination, raising some overfitting concerns. In general, the evaluation metrics follow a similar pattern compared to the raster and vector models.

Table 14 Evaluation metrics for the multimodal model.

Architecture	Metric	Operator				
		Elimination	Aggregation	Typification	Displacement	Enlargement
CNN + HGNN	Accuracy	0.65	0.78	0.72	0.71	0.88
	Precision	0.24	0.74	0.52	0.95	0.98
	Recall	0.77	0.86	0.68	0.72	0.88
	F_1 score	0.37	0.80	0.59	0.82	0.93
	ROC AUC	0.76	0.85	0.77	0.75	0.93
	PR AUC	0.36	0.86	0.53	0.96	0.99



(ai) ROC **(aii)** PR

(a) Validation set

(bi) ROC **(bii)** PR

(b) Test set

Figure 45 ROC and PR curves for the multimodal model (CNN + HGNN).

6.2 Stratified evaluation

6.2.1 Street block area

Figure 46 shows ROC curves for all modalities and operator combinations stratified by street block area quartile. For almost every combination of operator and modality, the first quartile corresponding to buildings located in the smallest street blocks ranks at the bottom with respect to AUC. This implies that all models, regardless of modality, tend to struggle with classifying operators for focal buildings located in small street blocks. On the other hand, focal buildings belonging to the highest quartile with respect to street block area are associated with the largest AUC for most of the modality and operator combinations. However, the magnitude of the difference in performance between the quartiles varies from operator to operator. For enlargement and elimination, the area of the street block only marginally affects performance, since all quartiles are associated with similar AUC. Displacement and aggregation are characterized by an apparent difference among modalities. Whereas the vector-based models display similar AUC for all quartiles, the raster and multimodal models exhibit larger AUC for larger street blocks. This effect is especially pronounced for typification, where a clear decrease in AUC with decreasing street block area can be observed for all three modalities. Generally, the raster-based and the multimodal models are shown to be more susceptible to the area of the street block of the focal buildings.

6.2.2 Urban-rural status

Figure 47 shows ROC curves for all modalities and operator combinations stratified by urban-rural status. With the exception of two cases (raster model for elimination and enlargement), all models display larger AUC for focal buildings in rural areas. The effect is most pronounced for typification, where all models display significantly better performance on the subset of rural buildings. The distinction between urban and rural buildings is clearest for the raster and multimodal models, whereas the performance is comparably similar for the vector model.

6.2.3 Operator combinations

The evaluation metrics stratified by operator combination set and modality computed on the test set using the multi-operator models are displayed in Table 15. Evidently, most of the operator sets display low evaluation metric values, since predicting the correct combination of operators that should be applied to a focal building constitutes a classification problem that is much harder to solve as opposed to classifying the operators individually. Generally, the performance of the multimodal model coincides with the raster model, whereas the vector models show some deviating evaluation metric values, such as lower precision for *enlargement* or lower recall for the set *displacement, enlargement*. Furthermore, the low evaluation metric values of the three modalities for ungeneralized focal buildings indicate that the constructed DL models tend to overestimate the necessary degree of generalization to be applied to the buildings.

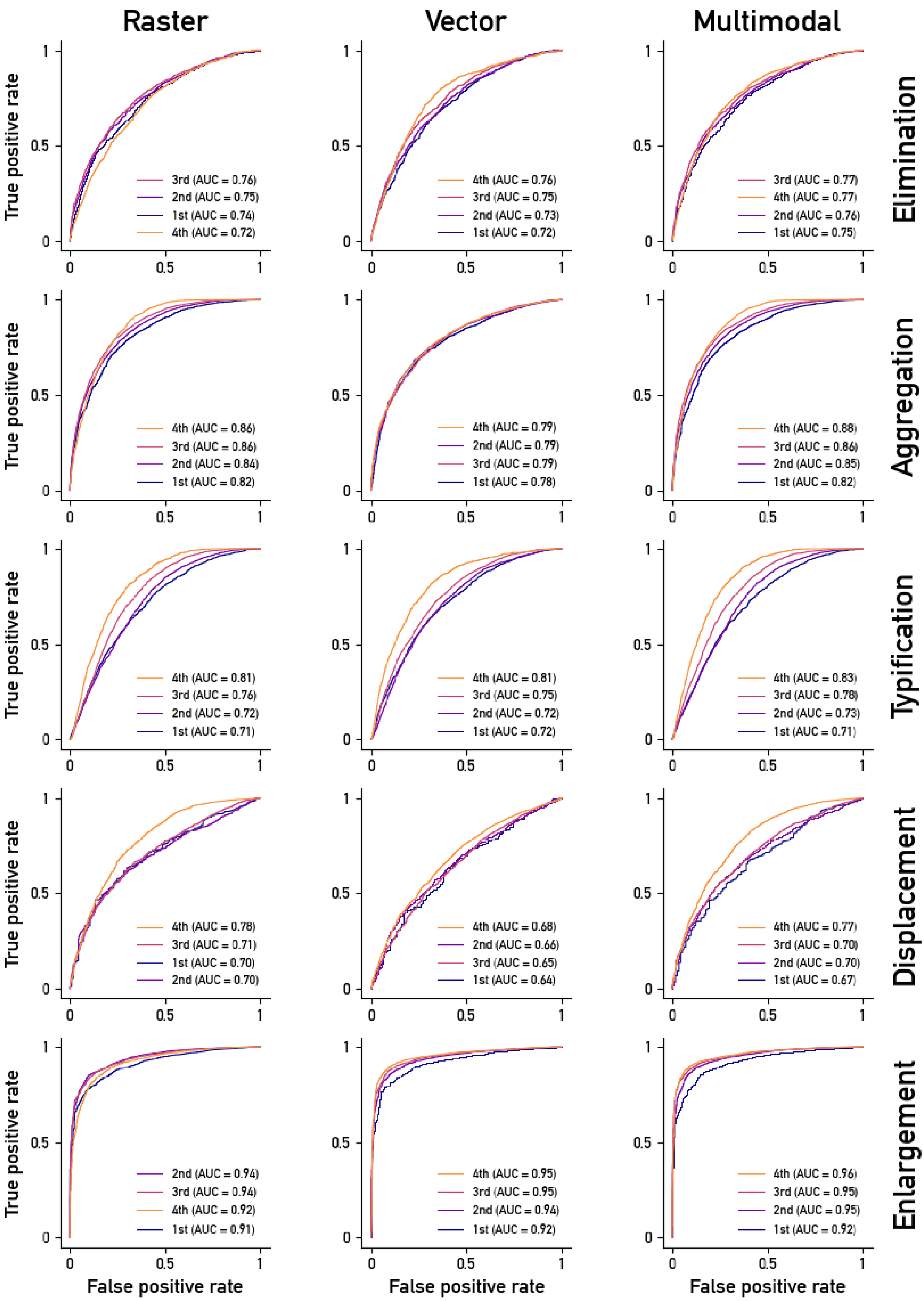


Figure 46 ROC curves by modality, operator, and street block area quartile.

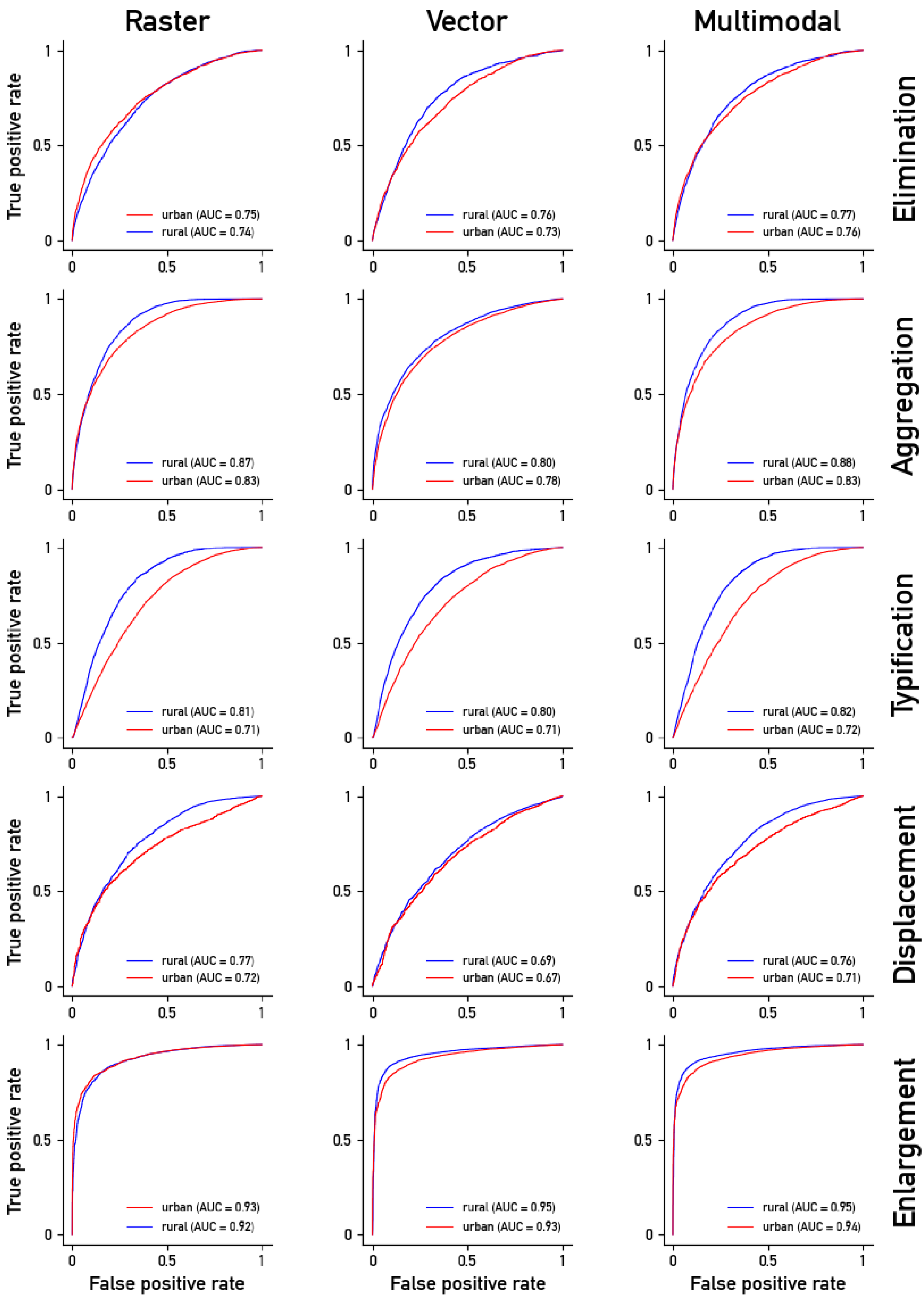


Figure 47 ROC curves by modality, operator, and urban-rural status.

Table 15 Evaluation metrics by operator combination and modality.

Operator set	Metric	Modality		
		Raster	Vector	Multimodal
None	Precision	0.20	0.15	0.18
	Recall	0.26	0.33	0.30
	F_1 score	0.22	0.21	0.23
Displacement	Precision	0.41	0.44	0.45
	Recall	0.40	0.27	0.39
	F_1 score	0.41	0.33	0.42
Enlargement	Precision	0.38	0.23	0.36
	Recall	0.53	0.52	0.54
	F_1 score	0.45	0.31	0.43
Displacement, Enlargement	Precision	0.67	0.57	0.68
	Recall	0.34	0.14	0.37
	F_1 score	0.45	0.22	0.48
Aggregation, Displacement	Precision	0.07	0.06	0.08
	Recall	0.31	0.29	0.34
	F_1 score	0.11	0.10	0.13
Aggregation, Displacement, Enlargement	Precision	0.36	0.32	0.39
	Recall	0.19	0.17	0.23
	F_1 score	0.25	0.23	0.29
Aggregation, Typification, Displacement	Precision	0.12	0.15	0.15
	Recall	0.36	0.36	0.39
	F_1 score	0.18	0.22	0.21
Aggregation, Typification, Displacement, Enlargement	Precision	0.55	0.58	0.57
	Recall	0.45	0.38	0.45
	F_1 score	0.50	0.46	0.50

6.3 Road network importance

To determine whether supplying the models with more contextual cartographic information in the form of the road network enclosing the street blocks is in fact beneficial for the operator classification task, the models are retrained for the number of epochs determined through early stopping after excluding the information associated with the roads from the training samples, which consequently only consist of the focal and context buildings. The ROC and PR AUC obtained by removing the roads are summarized in Table 16.

Table 16 ROC and PR AUC for the modalities when roads are removed.

Modality (Architecture)	Metric	Operator				
		Elimination	Aggregation	Typification	Displacement	Enlargement
Raster (CNN)	ROC AUC	0.72	0.85	0.77	0.69	0.91
	PR AUC	0.29	0.85	0.52	0.95	0.98
Vector (HGNN)	ROC AUC	0.74	0.79	0.75	0.70	0.94
	PR AUC	0.32	0.81	0.49	0.95	0.99
Multimodal (CNN + HGNN)	ROC AUC	0.75	0.85	0.77	0.69	0.93
	PR AUC	0.34	0.86	0.53	0.95	0.99

7 Discussion

7.1 Global performance

Table 17 presents a consolidation of the evaluation metrics obtained for each generalization operator through the best-performing architectures for the individual modalities.

7.1.1 Architectures and modalities

Courtial et al. (2021*b*, 2024) hypothesize that the implementation of DL architectures incorporating attention mechanisms has the potential to address some of the shortcomings that current DL-based approaches are faced with. However, according to the results obtained in Tables 11 and 13, the less complex architectures based solely on convolutions (CNN and HGNN) outperform the transformer-based architectures that additionally integrate attention mechanisms (ViT and HGT) with respect to evaluation metrics and training times for both the raster and vector models. This observation coincides with the results of existing studies, which declare similar performance (Fu, Zhou, Feng & Weibel 2024) or even a decrease in evaluation metrics (Zhou et al. 2023, Winkler 2023) when extending models with attention mechanisms. Compared to the approach chosen in the present thesis, Fu, Zhou, Feng & Weibel (2024) use a data model in which the focal building is always placed at the center of a patch and argue that this serves as a manual attention mechanism, citing this observation as a possible reason for the superior performance of convolution-based architectures. However, in the context of the present thesis, the results reveal that convolution-based models outperform architectures that integrate attention mechanisms even when the buildings are not centered.

Table 17 demonstrates that all three modalities exhibit similar capabilities with respect to the classification of the individual operators. None of the models are particularly adept at classifying any single operator compared to the others. The vector model generally performs slightly worse compared to the raster and multimodal models, although its processing times are also much lower. To better visualize the difference in performance, Table 18 illustrates the change in ROC and PR AUC using the vector model as a baseline. Evidently, the raster model outperforms the vector model with respect to all operators except for typification, despite the shifting focal

Table 17 Evaluation metrics for the best-performing model per modality.

Modality (Architecture)	Metric	Operator				
		Elimination	Aggregation	Typification	Displacement	Enlargement
Raster (CNN)	Accuracy	0.63	0.77	0.71	0.72	0.86
	Precision	0.23	0.73	0.51	0.85	0.97
	Recall	0.75	0.87	0.69	0.73	0.86
	F_1 score	0.35	0.79	0.58	0.82	0.91
	ROC AUC	0.74	0.85	0.76	0.76	0.91
	PR AUC	0.34	0.85	0.51	0.96	0.99
Vector (HGNN)	Accuracy	0.61	0.71	0.72	0.58	0.88
	Precision	0.22	0.67	0.52	0.95	0.98
	Recall	0.78	0.84	0.64	0.56	0.88
	F_1 score	0.35	0.75	0.57	0.71	0.93
	ROC AUC	0.74	0.80	0.76	0.70	0.93
	PR AUC	0.31	0.81	0.52	0.95	0.99
Multimodal (CNN + HGNN)	Accuracy	0.65	0.78	0.72	0.71	0.88
	Precision	0.24	0.74	0.52	0.95	0.98
	Recall	0.77	0.86	0.68	0.72	0.88
	F_1 score	0.37	0.80	0.59	0.82	0.93
	ROC AUC	0.76	0.85	0.77	0.75	0.93
	PR AUC	0.36	0.86	0.53	0.96	0.99

building and the scale effect. While the increase in performance is minor for elimination and enlargement, substantial increases in AUC values for aggregation and displacement can be observed. Having access to the exact outline of the buildings in the raster image is beneficial for identifying the presence of aggregation and displacement. As the building geometry is reduced to a node in the graph, the building outlines are missing in the vector models, potentially explaining the worse performance. The multimodal model shows similar performance increases over the vector model as the raster model, but additionally manages to boost performance for the operators elimination and typification. Therefore, the multimodal model outperforms the models based on the individual modalities despite the overfitting concerns identified in Figure 44, although only by a small margin.

Table 18 Change in ROC and PR AUC between modalities. The AUC values for the vector model are provided as a baseline, whereas the values associated with the raster and multimodal model represent the change in AUC from the vector model. Positive change implies better classification performance of the respective modality compared to the vector model.

Modality (Architecture)	Metric	Operator				
		Elimination	Aggregation	Typification	Displacement	Enlargement
Vector (HGNN)	ROC AUC	0.74	0.80	0.76	0.70	0.93
	PR AUC	0.31	0.81	0.52	0.95	0.99
Raster (CNN)	ROC AUC	+0.00	+0.05	+0.00	+0.06	+0.02
	PR AUC	+0.03	+0.04	-0.01	+0.01	+0.00
Multimodal (CNN + HGNN)	ROC AUC	+0.02	+0.05	+0.01	+0.05	+0.02
	PR AUC	+0.05	+0.05	+0.02	+0.01	+0.00

7.1.2 Generalization operators

The subsequent sections provide a discussion regarding the performance of the models on the individual operators. For visual investigation purposes, some classification results produced by the best-performing multimodal model are separately illustrated for each generalization operator in Figures 48 to 52. The examples are categorized into the following four cases by comparing the predicted operator with the operator that was annotated in the test dataset.

1. *True positive*, where the model correctly identifies the presence of an operator.
2. *True negative*, where the model correctly recognizes the absence of an operator.
3. *False negative*, where the model fails to identify an operator that is present.
4. *False positive*, where the model incorrectly detects an operator that is absent.

Elimination

According to Table 17, elimination evidently displays the worst classification evaluation metrics among all operators. This is unexpected, since the elimination model only has to optimize for a single operator, whereas the multi-operator model concerned with predicting the remaining operators must account for four operators simultaneously. Furthermore, buildings are commonly subjected to elimination once their area falls below a certain map scale threshold (Spiess et al. 2005). Consequently, the identification of elimination should be straightforward for the vector models, as they are explicitly supplied with features relating to building size. As depicted in Figure 16, individually considered, elimination is the most imbalanced generalization operator, as only 15% of buildings are subject to elimination. After being trained on a balanced dataset, the elimination models evidently cannot reproduce this imbalance using the determined threshold, as roughly 40% of buildings are classified as eliminated. This leads to high recall, low precision, and low PR AUC values. The resulting evaluation metrics are substantially lower when contrasted with a similar DL-based approach proposed by Xiao et al. (2024) that abstracts buildings as points and applies GCNN for selection, the complementary operator to elimination. The closest comparable ML-based approach proposed by Lee et al. (2017) also outperforms the results obtained for the elimination operator, achieving significantly higher evaluation metrics.

The classification samples displayed in Figure 48 show that the model addresses the imminent cartographic conflicts resulting from an enlargement of the road network by correctly classifying the small focal buildings in samples (a), (b), (c), and (d) as eliminated. On the other hand, it correctly recognizes that certain significant buildings should be retained where the conflicts can be solved by alternative operators, as seen in examples (m), (n), and (o) (by aggregation and typification) and (p) (by displacement and elimination of the insignificant surrounding buildings). However, the model struggles with the distinction between typification and elimination, as shown in examples (e) - (l). Example (h) shows a case where the focal building is annotated as eliminated, but ample space is available, which implies that the prediction generated by the model can be considered a valid alternative.

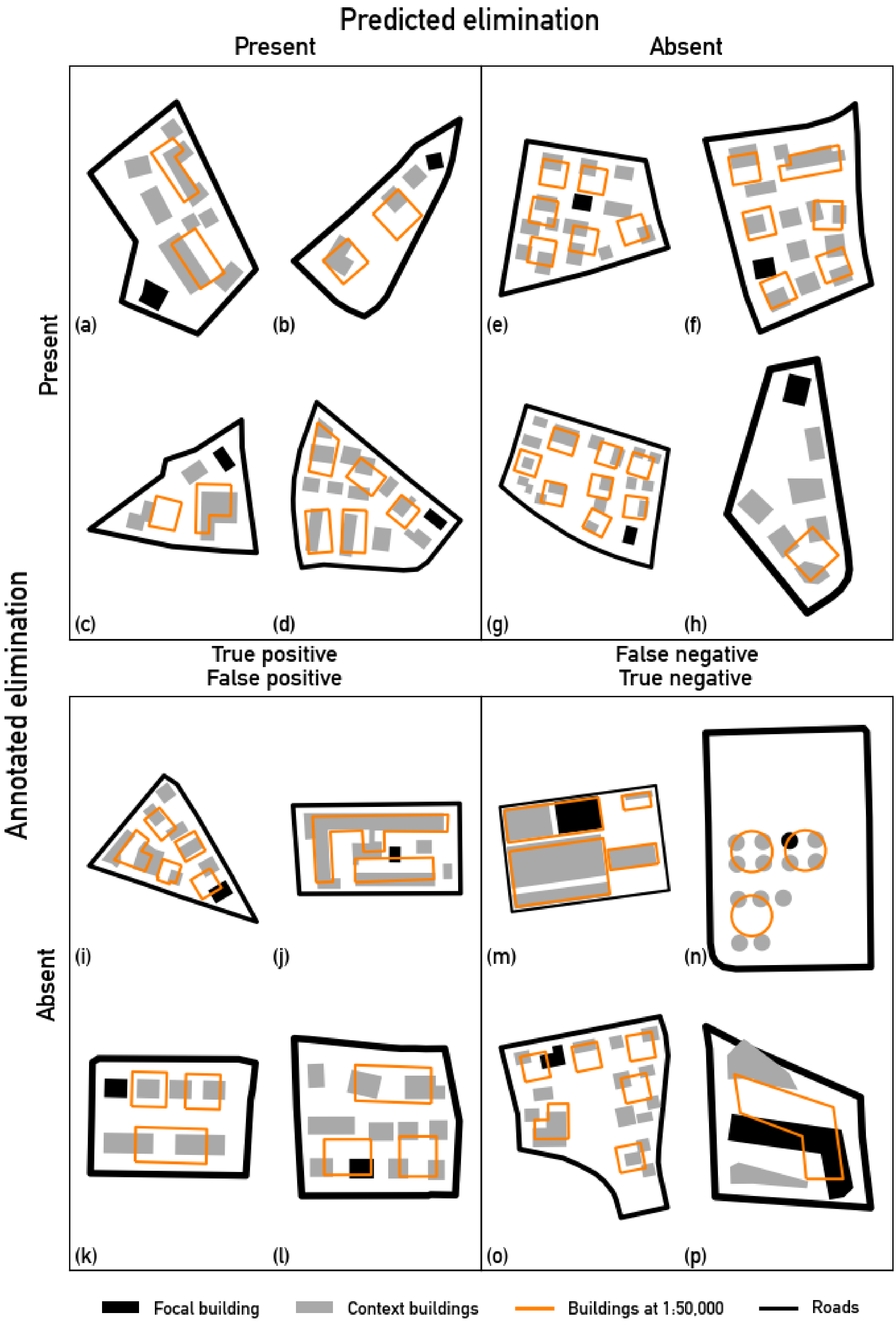


Figure 48 Visual evaluation of model performance on elimination (map data © swisstopo).

Aggregation

As illustrated in Table 17, the models achieve comparably high evaluation metrics of 0.8 for F_1 score and 0.85 for ROC and PR AUC, respectively, for classifying aggregation. For all three modalities, the models are able to leverage the provided features to obtain satisfactory classification evaluation metrics for determining whether buildings should be aggregated or not. In a comparable study, Lee et al. (2017) use ML techniques to classify aggregation, whereby their models struggle to predict the presence of aggregation, obtaining superior performance for the classification of elimination. Therefore, the DL-based approach proposed in the present thesis is ostensibly better suited to classify whether buildings should be aggregated or not.

Figure 49 allows for a visual evaluation of the performance of the model in predicting the presence or absence of aggregation. Examples (a) and (b) show that the DL model is able to correctly identify that small focal buildings forming part of larger, dense groups are at risk of becoming indistinguishable to the human eye at 1:50,000, for which the application of aggregation is proposed. Based on examples (g) and (h), the model is further able to correctly recognize focal buildings with access to abundant map space, where aggregation is not necessary and other operators such as displacement and enlargement are applied instead. Examples (c) and (d) illustrate two cases in which aggregation was applied in the underlying dataset, whereas the model chooses alternative operators to resolve the cartographic conflicts. In both cases, sufficient map space is available, which implies that the application of other operators such as displacement or enlargement may also provide satisfactory solutions. Finally, examples (e) and (f) illustrate instances where focal buildings are embedded within densely packed street blocks for which the model recommends aggregation, presumably due to the high building density. However, conflicts are resolved by applying alternative operators on the map.

Typification

Table 17 shows that the models perform slightly worse in classifying typification compared to aggregation, achieving F_1 scores of 0.6 and ROC and PR AUC values of 0.75 and 0.5, respectively. Given that the concept of typification has eluded cartographers for years (Gong & Wu 2016), the proposed DL approach produces promising evaluation metrics for predicting whether buildings should be subjected to typification or not.

Figure 50 illustrates the performance of the model in classifying typification. Examples (a) and (b) demonstrate that the model correctly recognizes that focal buildings in dense street blocks should be typified. On the other hand, examples (g) and (h) show that typification is fittingly classified as absent for significant buildings with characteristic shapes in dense street blocks, presumably because the preservation of these buildings is important for navigational purposes. However, examples (c) and (d) illustrate that the model fails to identify the presence of typification for small buildings in dense street blocks. Finally, examples (e) and (f) highlight that some aggregation operations are identified as typification, indicating that the model encounters difficulties in distinguishing between the two operators.

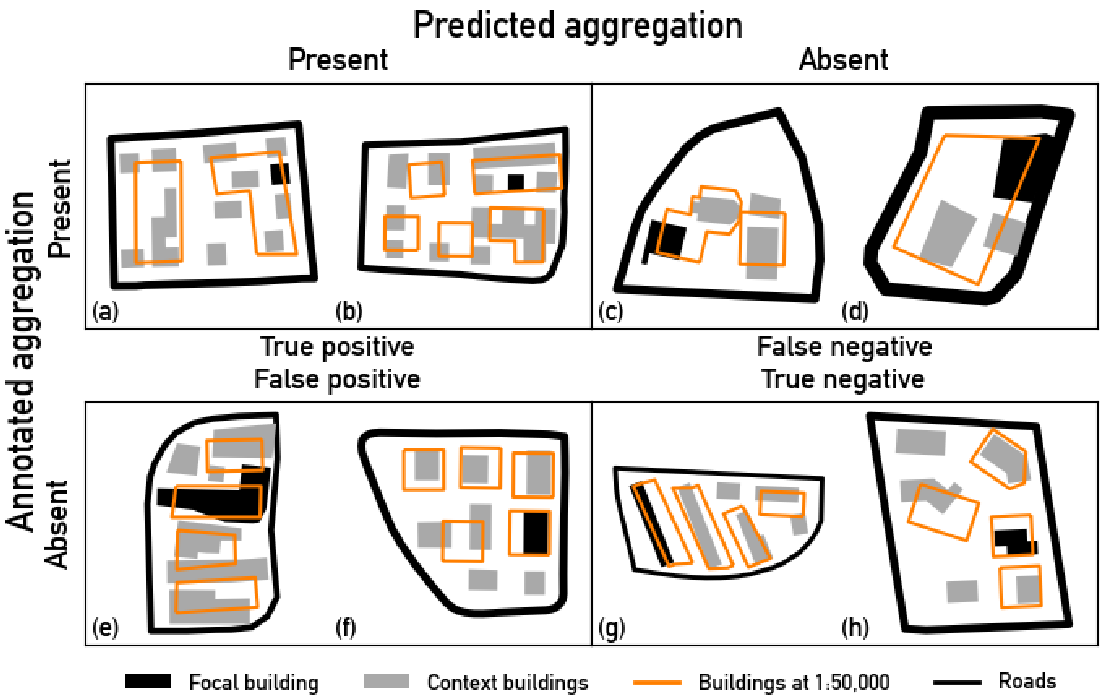


Figure 49 Visual evaluation of model performance on aggregation (map data © swisstopo).

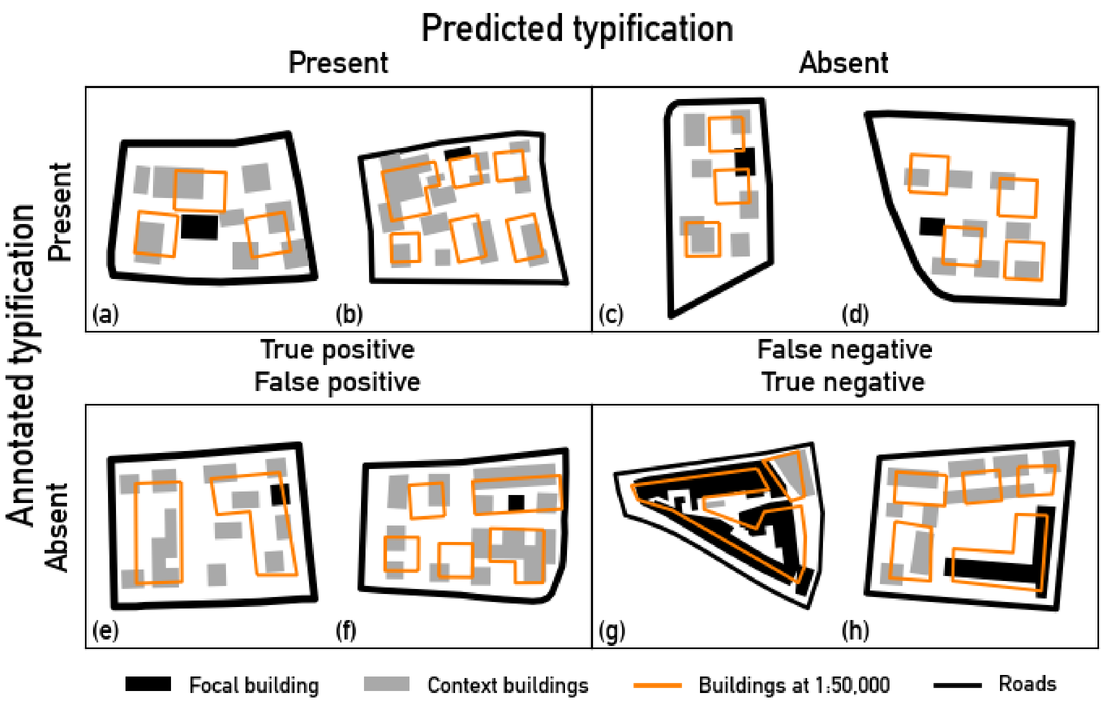


Figure 50 Visual evaluation of model performance on typification (map data © swisstopo).

Displacement

Table 17 illustrates that the models exhibit satisfactory performance in classifying displacement, achieving F_1 scores of 0.8 and ROC and PR AUC values of 0.75 and 0.95, respectively. Similarly to typification, displacement can be characterized as a highly contextual operator, for which conventional generalization approaches have struggled to conceive appropriate solutions (Ruas 2001, Regnauld & McMaster 2007). Therefore, the evaluation metrics obtained are a testament to the capacity of DL models for modeling the displacement of buildings.

Figure 51 provides a visual assessment of model performance in predicting the presence or absence of displacement. Examples (a) and (b) illustrate that the model correctly recognizes that focal buildings along roads should be displaced away from the road network, presumably because the road network is subject to significant enlargement when transitioning to 1:50,000. Furthermore, examples (g) and (h) demonstrate that the model appropriately classifies displacement as absent in cases where abundant space within the street block is available and displacement is evidently not necessary. On the other hand, example (c) shows an instance where the model proposes aggregation instead of displacement, and example (d) illustrates a case where the model fails to recognize the impending conflict with the road network. In some instances, as exemplified in (e) and (f), the model struggles to recognize that there is sufficient map space available, predicting displacement in cases where it is not necessary.

Enlargement

Based on Table 17, enlargement displays the highest evaluation metrics of all generalization operators, achieving F_1 scores of 0.9 and ROC and PR AUC values of 0.9 and 0.99, respectively. This is unsurprising, as buildings are routinely enlarged during the generalization process (Regnauld & McMaster 2007). Furthermore, enlargement can be considered the operator that relies least on contextual information, as its application is in large parts governed by the smallest recognizable building area at the next consecutive scale imposed by map specifications. Therefore, it is intuitive that the trained DL models are particularly adept at classifying the presence or absence of enlargement compared to the remaining contextual operators.

Figure 52 demonstrates how the model performs in classifying enlargement for exemplary focal buildings. Examples (a) and (b) illustrate cases in which small buildings should be retained to conserve the structure of the street block. The model correctly recognizes that these buildings have to be enlarged in order to comply with minimum dimensions stipulated by map specifications. Additionally, examples (g) and (h) demonstrate cases where the road network is significantly displaced to the point of intersecting the buildings, likely due to important map features in adjacent street blocks. In instances where the buildings extend beyond the street blocks, they should certainly not be enlarged, which the model correctly identifies. However, examples (c) and (d) show cases where the model fails to identify that thin buildings should be enlarged. Finally, examples (e) and (f) depict instances where building enlargement is unnecessary, but the model predicts enlargement regardless.

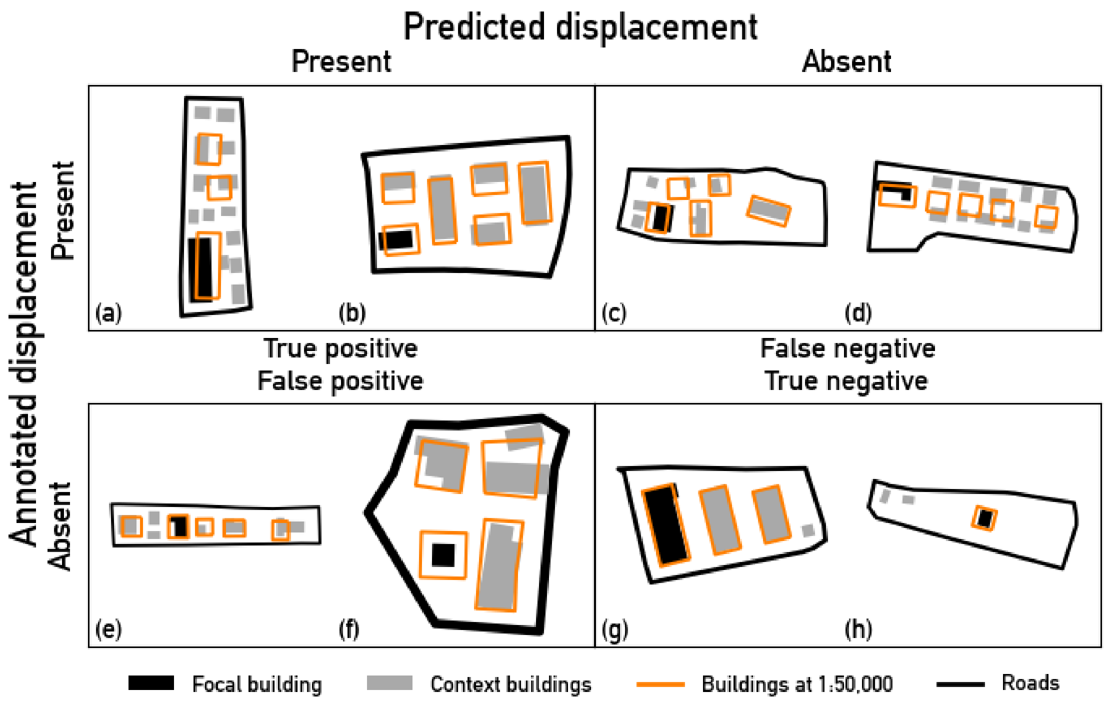


Figure 51 Visual evaluation of model performance on displacement (map data © swisstopo).

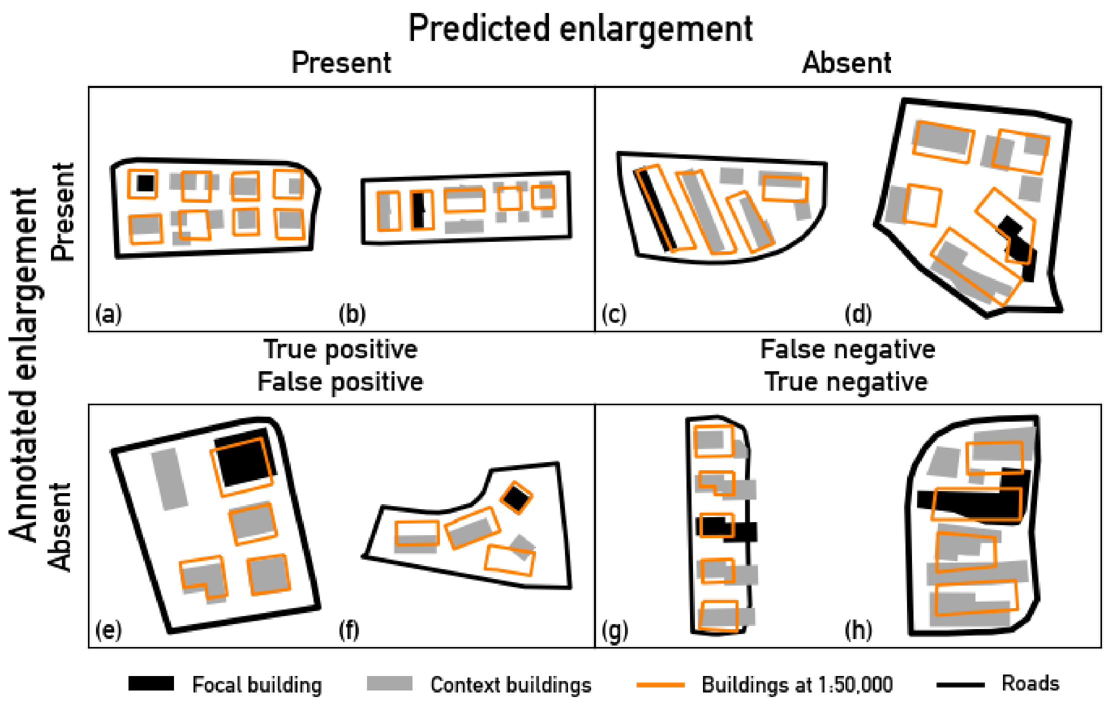


Figure 52 Visual evaluation of model performance on enlargement (map data © swisstopo).

7.2 Stratified performance

7.2.1 Street block area and urban-rural status

Table 19 summarizes the ROC AUC values depicted in Figures 46 and 47 when stratifying the test set with respect to street block area quartile and urban-rural status, respectively. Aside from the raster model on elimination and enlargement, all models exhibit an increase in performance with increasing street block area and on buildings in rural street blocks. This finding is intuitive, since buildings in rural areas tend to be part of street blocks with much larger areas, as illustrated in Figure 53. Furthermore, the findings can be considered representative since the training database evidently contains similar numbers of buildings in urban and rural contexts. In Section 5.3.8, it was hypothesized that especially the raster model may be susceptible to a scale effect, whereby performance would deteriorate with increasing street block area due to the associated decrease in resolution. However, the results presented in Table 19 reveal that the opposite is true, as classification performance improves with increasing street block area.

As there is no clear explanation for this inverse scale effect from a technical point of view, it is likely to be attributable to insights derived from conventional cartographic generalization practice. On the one hand, the higher performance observed for buildings within large street blocks can be linked to the fact that these blocks are predominantly situated in rural areas. Compared to urban contexts, the generalization of buildings in rural areas is generally considered an easier task, as the density of map objects is much lower (Ruas & Mackaness 1997, Mustière & Moulin 2002, Spiess et al. 2005). On the other hand, larger street blocks provide additional cartographic context that can facilitate the decision regarding the operators that should be applied. Transferring these insights to a DL context, it is logical that the models are more adept at performing the generalization operator prediction for buildings located in rural areas.

Table 19 ROC AUC upon stratification by street block area quartile and urban-rural status.

Operator	Modality	Street block area quartile				Urban-rural status	
		1st	2nd	3rd	4th	Urban	Rural
Elimination	Raster	0.74	0.75	0.76	0.72	0.75	0.74
	Vector	0.72	0.73	0.75	0.76	0.73	0.76
	Multimodal	0.75	0.76	0.77	0.77	0.76	0.77
Aggregation	Raster	0.82	0.84	0.86	0.86	0.83	0.87
	Vector	0.78	0.79	0.79	0.79	0.78	0.80
	Multimodal	0.82	0.85	0.86	0.88	0.83	0.88
Typification	Raster	0.71	0.72	0.76	0.81	0.71	0.81
	Vector	0.72	0.72	0.75	0.81	0.71	0.80
	Multimodal	0.71	0.73	0.78	0.83	0.72	0.82
Displacement	Raster	0.70	0.70	0.71	0.78	0.72	0.77
	Vector	0.64	0.66	0.65	0.68	0.67	0.69
	Multimodal	0.67	0.70	0.70	0.77	0.71	0.76
Enlargement	Raster	0.91	0.94	0.94	0.92	0.93	0.92
	Vector	0.92	0.94	0.95	0.95	0.93	0.95
	Multimodal	0.92	0.95	0.95	0.96	0.94	0.95

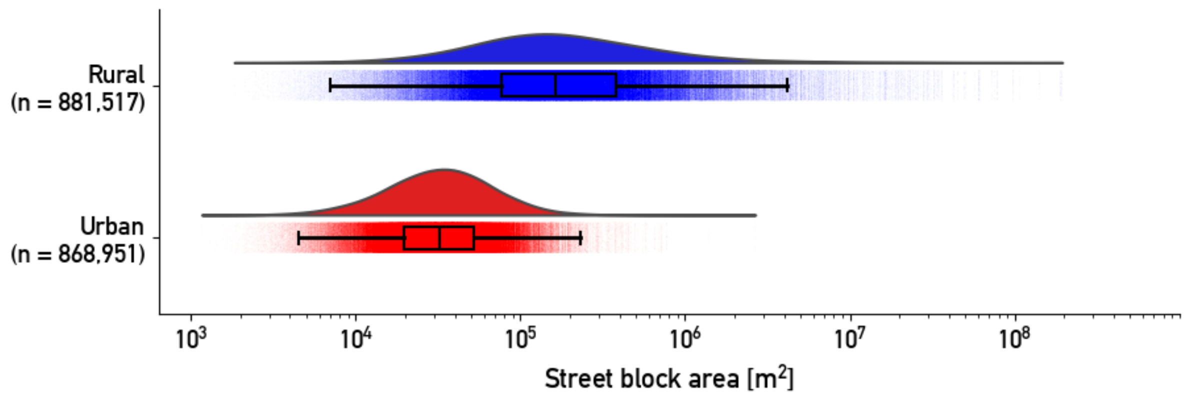


Figure 53 Street block area for buildings in urban and rural street blocks.

7.2.2 Operator combinations

Figure 54 shows the relationship between the number of occurrences of the respective operator combination in the test set and the F_1 score depicted in Table 15 obtained on the reformulated multi-class prediction task outlined in Section 5.3.8. Evidently, the models perform better for operator combinations that are frequently applied (such as *displacement, enlargement* and *aggregation, typification, displacement, enlargement*), while they struggle to correctly identify the operator combinations that appear rarely (such as *aggregation, displacement* and *aggregation, typification, displacement* and *none*). Furthermore, the models perform poorly on the common combination *aggregation, displacement, enlargement*. As the models were trained on a dataset containing samples that were balanced with respect to operator combinations, these results provide an indication that the balancing approach devised in Section 5.2.2 designed to avoid this problem is not particularly effective.

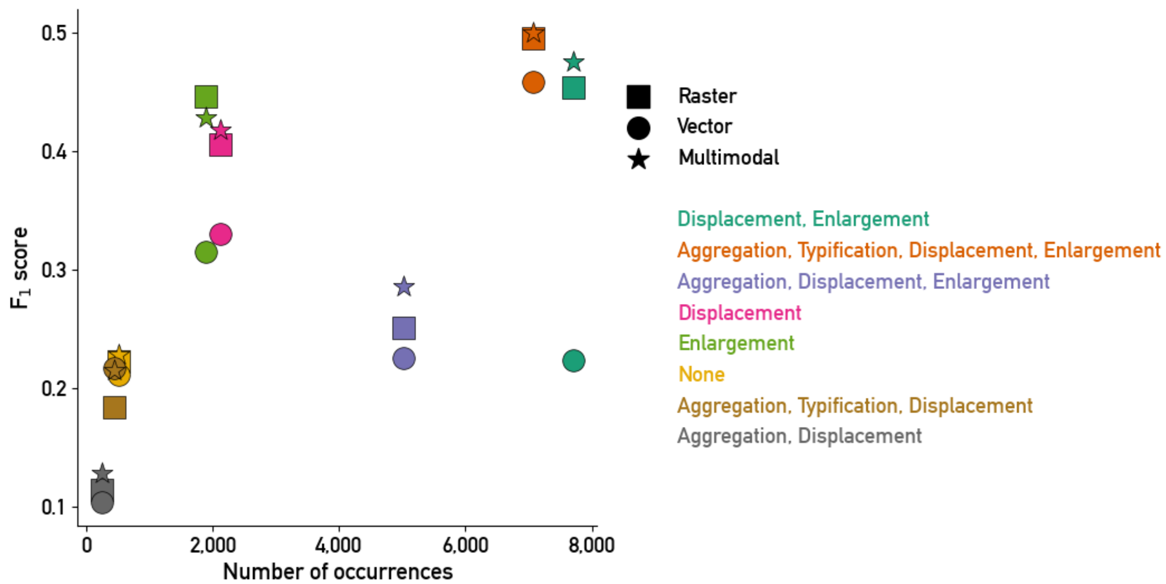


Figure 54 Relationship between operator combination prevalence and F_1 score.

7.3 Road network importance

Table 20 displays the change of ROC and PR AUC induced by the removal of the roads. The values in the table correspond to the difference between the respective metrics of Table 16 and Table 17. For all three modalities, the evaluation metrics for aggregation and enlargement are not significantly affected by the removal of the roads. This is consistent with conventional generalization approaches, as the application of these two operators is not expected to be considerably influenced by the layout of the road network. For the raster and multimodal models, the removal of the road network evidently leads to a substantial decrease in ROC and PR AUC for the operators elimination and displacement, implying better performance on these two operators if the roads are included for training and prediction. This observation coincides with conventional generalization practices, as the enlargement of the road network when transitioning to smaller scales commonly necessitates the elimination or displacement of buildings situated along roads. This phenomenon is illustrated in Figure 55, whereby the amount of available map space dictates the application of elimination or displacement. This finding indicates that providing the raster and multimodal models with more cartographic context enables them to learn and reproduce conventional cartographic knowledge. A similar effect can be observed for the vector-based model, which exhibits a slight decrease in performance for typification when roads are excluded. This is intuitive from the perspective of conventional generalization, as the layout of the road network constrains the amount of available map space which is critical for determining building density and therefore also the decision of whether typification should be applied or not (Lee 1996, Regnauld & McMaster 2007).

The lack of a decrease for displacement and elimination for the vector model when roads are excluded suggests that the developed graph construction method may not be the most optimal solution to encode structural knowledge associated with roads in the heterogeneous graph. Nonetheless, the obtained results demonstrate that including additional cartographic context in the form of the road network during the training process of the various DL models has the potential to increase the quality of the prediction regarding the generalization operators that should be applied to generalize a given focal building. The findings further suggest that DL models have the capacity to learn cartographic knowledge that is consistent with conventional and manual generalization approaches.

Table 20 Change in ROC and PR AUC induced by road removal. Negative change implies worse classification performance when the roads are excluded from model training.

Modality (Architecture)	Metric	Operator				
		Elimination	Aggregation	Typification	Displacement	Enlargement
Raster	ROC AUC	-0.02	0.00	0.01	-0.07	0.00
(CNN)	PR AUC	-0.05	0.00	0.01	-0.01	-0.01
Vector	ROC AUC	0.00	-0.01	-0.01	0.00	0.01
(HGNN)	PR AUC	0.01	0.00	-0.03	0.00	0.00
Multimodal	ROC AUC	-0.01	0.00	0.00	-0.06	0.00
(CNN + HGNN)	PR AUC	-0.02	0.00	0.00	-0.01	0.00

(a) Elimination **(b)** Displacement

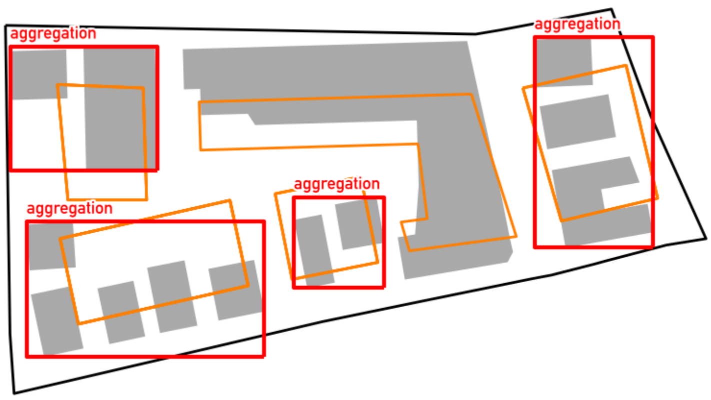
Figure 55 Operator application necessitated by the generalization of the roads (map data © swisstopo).

7.4 Limitations and further research

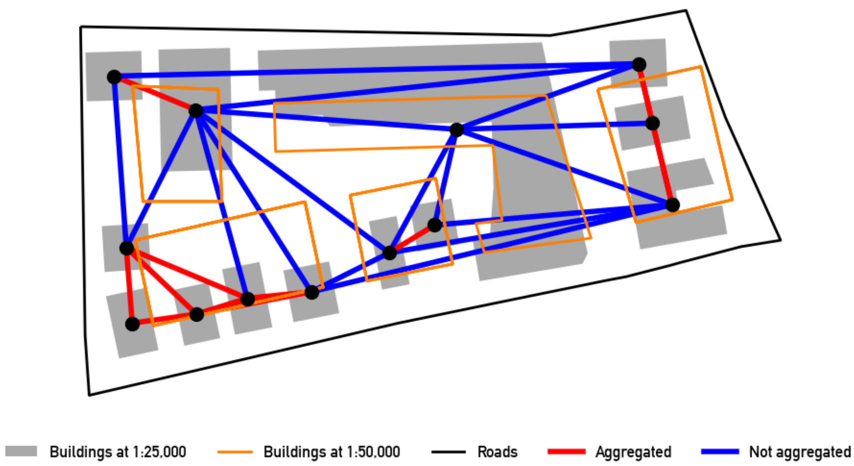
7.4.1 Conceptualization

According to Stanislawski et al. (2014), cartographic generalization extends beyond simply invoking a series of generalization operators. In light of this observation, the proposed approach is associated with various limitations from a conceptual point of view. An underlying assumption is that operator predictions are conducted separately for each focal building while keeping the surrounding context buildings constant, thereby neglecting that the context buildings are also subject to generalization. This can result in implausible predictions for the operators aggregation and typification that are applied to sets of buildings, whereby only a single building in a group of buildings is predicted as aggregated or typified. Moreover, the trained DL models are unable to identify and resolve potential conflicts at the street block level triggered by incompatible generalization operator predictions. Furthermore, the predictions supplied by the models do not include the sequence in which the operators are to be applied, which has been identified as a common problem with conventional approaches (Duchêne et al. 2018, Sester 2020). Additionally, the strict choice of street blocks as analysis units is not always suitable, as the distances between buildings within large street blocks in rural areas with low building densities become too vast to significantly influence each other with respect to generalization.

The limitations outlined previously could be addressed by alternatively formulating the classification problem. Instead of predicting multiple generalization operators simultaneously that should be applied to a single focal building given its cartographic context, separate models for each generalization operator may be constructed that predict a single operator concurrently for all buildings in a street block. For instance, the raster-based models could be extended from conventional image classification to object detection using a system such as YOLO (You Only Look Once, Redmon et al. 2015). YOLO conducts object recognition by directly predicting bounding boxes around objects and the associated class probabilities in a single evaluation. Figure 56a illustrates, how such a framework may be used to train models to identify sets of buildings that should be collectively subjected to aggregation. Similarly, the graph-based classification task could be reformulated from node-level to edge-level operator prediction. The operators predicted for the edges incident on buildings denote the generalization operators that should be applied between the buildings. Analogously to the raster-based case, Figure 56b illustrates how such a model could be used to recognize adjacent buildings that require collective aggregation. To improve the multimodal model, an approach based on early or joint fusion could be adopted, which is generally preferred over late fusion when the input modalities inherently complement each other (Huang et al. 2020), as is the case for the raster and vector representations of the same underlying street block employed throughout the thesis. Such an adapted multimodal approach also has the potential to alleviate the overfitting concerns identified in Section 6.1.3. Additionally, sparsely covered street blocks may be further partitioned based on measures such as k-nearest neighbor distance to serve as sensible analysis units.



(a) Raster-based model: object detection with YOLO



(b) Vector-based model: edge-level operator prediction

Figure 56 Illustration of potential avenues for future research (map data © swisstopo).

7.4.2 Sampling

As evident from the results presented throughout Chapter 6, the stratified sampling approach based on the LP-transformation devised in Section 5.2 shows limited effectiveness in addressing the operator imbalance present in the training database. Although the models manage to reproduce the operator distribution found in the imbalanced test dataset for the operators classified by the multi-operator model despite being trained on a balanced training dataset, the models fail to do so for predicting the presence or absence of elimination. Furthermore, the analysis conducted in Section 7.2.2 demonstrates that the models are more adept at classifying commonly applied operator combinations as opposed to combinations that rarely appear in the training dataset. The reason for this discrepancy can likely be attributed to the excessive oversampling applied to rare operator combinations, which was necessary to obtain a balanced dataset. Due to the oversampling, the models evidently struggle to generalize the knowledge acquired from samples with rare operator combinations during the training procedure, resulting in poor performance in classifying these operator combinations (Buda et al. 2018).

In light of the identified limitations, future approaches should focus on further developing techniques tailored to DL-based cartographic generalization to handle the inherent imbalance associated with generalization operators. In the present thesis, data augmentation in the form of rotation and flipping of the input rasters was confined to the training of the raster-based models. Therefore, future research should investigate the degree to which data augmentation techniques can be applied to graph-based cartographic generalization approaches, which have the potential to mitigate the negative impacts of oversampling minority labels (Zhao et al. 2021). Additionally, synthetic data generation techniques such as SMOTE (Synthetic Minority Oversampling Technique, Chawla et al. 2002) may be used to balance the datasets by creating artificial instances of cases where less frequent operator combinations were applied. To combat the effects of data imbalance, future approaches could additionally introduce weights into the BCE loss that increase the contribution of less frequent operator combinations to the loss calculated during training instead of resampling and balancing the data beforehand (Fernando & Tsokos 2022). Furthermore, alternative loss functions such as focal loss have been shown to outperform the standard BCE loss employed throughout this thesis on imbalanced classification problems (Lin et al. 2020). Finally, balancing a dataset with respect to a single label is substantially easier as opposed to balancing a dataset with respect to multiple labels. Therefore, future approaches should move towards implementing ensemble methods, whereby separate models are trained for individual generalization operators, whose predictions can subsequently be chained together while accounting for the correlations between the labels (Sun et al. 2015).

7.4.3 Explainability

As described in Section 3.1, the original motivation for introducing DL into the cartographic generalization domain was to bypass the knowledge acquisition bottleneck that has significantly hindered the development of conventional approaches by harnessing the implicit cartographic knowledge contained within existing maps (Weibel et al. 1995). Although the DL models presented throughout this thesis are evidently capable of effectively leveraging this implicit knowledge to learn the classification of generalization operators, it is difficult to ascertain why the models produce a specific prediction. Therefore, it can be argued that treating ANNs as black boxes (Rudin 2019) simply shifts this implicit knowledge from the data to the models during training, without making any meaningful contributions to explicit knowledge acquisition. The lack of explainability also inhibits the interpretation of the poor performance observed in classifying supposedly simple operators such as elimination.

Harrie et al. (2024) claim that the knowledge acquired by DL models during the training process may not adhere to any established principles of good cartographic practice. Hence, they emphasize the importance of providing DL models with explicit cartographic domain expertise in order to facilitate the understanding of the obtained knowledge. The approach outlined in the present thesis attempts to address this issue during the conceptualization phase by introducing established principles and constraints developed as part of conventional generalization approaches. For instance, street blocks are chosen as analysis units, surrounding roads are explicitly provided to the models, features supplied to GNNs are selected based on state-of-the-art generalization literature, and the multi-channel data model used for training the raster-based models emulates the procedures employed during manual generalization. Therefore, future research should seek to provide DL models with even more explicit procedural knowledge. For example, given the identified benefit of including roads to determine the presence or absence of certain operators, the proposed approach could be extended by introducing additional contextual map features that are assigned a higher priority in the generalization process compared to buildings, such as hydrographic or railway networks. Furthermore, semantic information associated with the road network may be included, since the influence of major roads, such as highways, on building generalization is considered to be more significant compared to less important roads (Spiess et al. 2005).

Hu et al. (2024) and Kang et al. (2024) argue that the explainability of DL models should be considered an important pillar of GeoAI. In the face of the counter-intuitive results presented throughout this thesis and in order to advance the field of DL-based map generalization, the introduction of techniques from the domain of explainable AI (XAI) is pivotal, providing human-interpretable insights into the decision-making process of the DL models (Fu et al. 2023, Gunning et al. 2019). Awareness regarding the cartographic knowledge acquired by DL models has the potential to help refine existing or even propose entirely novel DL-based approaches (Fu, Zhou, Xin & Weibel 2024). In the present thesis, attempts are made to facilitate explainability by stratifying training samples with respect to street block area quartile, urban-rural status, and operator combinations, thereby identifying cases or tasks that the models struggle to learn. Future research should strive to increasingly introduce techniques from the domain of XAI to DL-based map generalization. For example, Fu, Zhou, Xin & Weibel (2024) show how XAI tools may be used to augment raster-based DL approaches to determine the importance of individual pixels for the prediction. In the context of the DL models developed throughout this thesis, the raster-based CNN model could be subjected to gradient-weighted class activation mapping

(Grad-CAM, Selvaraju et al. 2019) to determine the parts of the input map that are considered important and to what degree the models rely on cartographic context in order to predict the generalization operators that should be applied to a given focal building. Analogously, XAI frameworks developed for heterogeneous GNNs could be applied to quantify the influence of certain nodes for arriving at the generalization operator predictions (Li et al. 2023).

7.4.4 Evaluation

A further limitation concerns the evaluation of the results, which represents a pivotal step in both conventional (Stoter et al. 2014) and DL-based generalization approaches (Courtial et al. 2022*a*). Due to the formalization of the operator classification task as a binary, multi-label classification problem, the approach assumes by design that the operators annotated within the training database represent the singular solution for generalizing the buildings from 1:25,000 to 1:50,000. However, generalization represents a process for which there are usually multiple satisfactory solutions, implying that it is difficult to formulate a global measure encapsulating the quality of a generalization (Touya 2012). This phenomenon is illustrated in Section 7.1.2, where the visual evaluation of the examples in Figures 48 to 52 has shown that the appropriate choice of generalization operators in a given situation is often ambiguous. Since the evaluation metrics presented throughout Chapter 6 are based on the assumption of the existence of a unique solution, they cannot be used to assess whether alternative operator combinations predicted by the models also provide acceptable results. Additionally, due to the fuzzy nature of the operator annotation process, the training database sporadically contains samples with mislabeled operators, which can hamper the learning and evaluation procedure. Furthermore, the implemented approach does not allow for the evaluation of whether the predictions produce satisfactory solutions at the street block level, which is generally considered more desirable compared to correctly predicting the operators for individual buildings (Courtial et al. 2021*b*).

In light of the outlined limitations, future research should aspire to incorporate alternative learning paradigms to the hitherto applied supervised and unsupervised frameworks that enable the models to explore a variety of suitable generalization operators, even if they are not explicitly annotated in the training dataset. For instance, the application of reinforcement learning has the potential to address this problem, as models are trained to maximize rewards formulated for different appropriate generalization operators (Arulkumaran et al. 2017). Evaluating the ambiguity of the solutions produced by such frameworks would require the development of a training database where multiple appropriate operators are annotated for each building. Based on such a dataset, techniques ranging from custom loss functions to soft labels (Collins et al. 2022) may be applied to assess prediction quality. Evaluation of the resulting generalization quality at the street block level constitutes a challenging task from a computational point of view. To address this issue, trained cartographers could be involved in the evaluation process to judge whether the predictions generated by the DL models are indeed suitable to generalize entire street blocks. Finally, additional research is required to evaluate the degree to which the developed models may be transferred to other settings and scale transitions.

8

Conclusion

Contributions

DL has emerged as an exciting candidate for a paradigm shift in automated cartographic generalization, exhibiting the potential to address many of the challenges encountered by conventional approaches. Against the backdrop of these developments, the present thesis makes the following contributions to the burgeoning field of DL-based cartographic generalization.

- Investigation of DL models for the **prediction of contextual generalization operators** (elimination, aggregation, typification, displacement, and enlargement) to be applied to a given building for the hitherto neglected scale transition from 1:25,000 to 1:50,000, which can become integral parts of more comprehensive DL-driven generalization workflows.
- Facilitation of operator classification through incorporation of an enriched **dataset annotated with generalization operators** based on the workflow proposed by Fu et al. (2025) containing building geometries generalized by expert cartographers.
- Implementation of a **data balancing approach** based on the annotated dataset designed to address the imbalance inherent in the distribution of generalization operators.
- Development and evaluation of convolution-based and attention-based models processing various **modalities**: a **raster** model operating on maps represented as multi-channel images, a **vector** model leveraging a novel heterogeneous graph structure, and a **multimodal** model that exploits both modalities concurrently to make predictions.
- Exploration of the role of additional **contextual information** in the form of surrounding buildings and the road network in improving predictions.

Findings

The findings reveal that DL models are capable of learning the prediction of the contextual generalization operators elimination, aggregation, typification, displacement, and enlargement. Furthermore, the results shed light on the importance of context and highlight the potential of multimodal approaches to automate cartographic generalization using DL. With respect to the research questions formulated in Section 3.4, the findings can be summarized as follows.

- **RQ1**: *To what extent can DL models be used to predict the generalization operators that should be applied to a given building?*

The trained DL models achieve satisfactory evaluation metrics for classifying operators whose application relies the least on contextual information, such as enlargement and aggregation, whereas they display worse classification performance in predicting highly contextual operators such as displacement and typification. Surprisingly, the worst performance is observed on elimination, likely due to the difficulty in distinguishing between elimination and typification. The models struggle to classify operator combinations that are rarely applied.

- **RQ2**: *To what degree can the inclusion of cartographic context enable more informed generalization operator predictions?*

The exclusion of the road network during model training leads to poorer performance in classifying elimination, typification, and displacement. The stratification shows an improvement in evaluation metrics when predicting the operators to be applied to buildings located within large street blocks in rural areas. Therefore, context matters for DL-based cartographic generalization.

- **RQ3**: *To what extent can a multimodal model integrating vector and raster representations outperform unimodal models based on the individual modalities?*

The experiments demonstrate that the models trained on the three investigated modalities exhibit similar classification performance. For all modalities, convolution-based architectures outperform architectures that incorporate attention mechanisms. Although the raster model outperforms the vector model with respect to the classification of aggregation and displacement, it does so at the expense of significantly longer processing times. The multimodal model additionally achieves marginal performance improvements over the raster model for predicting elimination and typification, while displaying similar evaluation metrics for the remaining operators. In conclusion, the multimodal model is shown to outperform the unimodal models.

Limitations

The proposed approach is limited from a conceptual point of view, as it assumes that buildings are generalized individually rather than collectively. Therefore, it cannot account for cartographic conflicts and inconsistencies evoked by the operator predictions. Additionally, the models are not able to specify the order in which the operators are to be applied to the buildings. From the perspective of the DL models, the experiments suggest that the multimodal model is prone to overfitting and that the encoding procedure of street blocks as heterogeneous graphs conceived for the vector model is likely too convoluted to be efficient. Moreover, the stratified sampling approach devised to address the imbalance present within the distribution of the generalization operators displays limited effectiveness due to excessive oversampling. Furthermore, the predictions generated by the model are not easily explainable, which restricts the contribution of the proposed approach to explicit knowledge acquisition. Finally, the approach is limited by its inability to evaluate the quality of alternative operator predictions.

Outlook

To address the identified limitations, future studies should seek to explore novel DL architectures capable of simultaneously predicting the presence or absence of individual operators for all buildings in a street block. Additionally, the unsolved problems associated with operator imbalance call for the development of tailored resampling and data augmentation approaches, the exploration of alternative loss functions, and the implementation of ensemble methods. The encoding of additional procedural domain knowledge and the application of techniques from the domain of XAI have the potential to make meaningful contributions to knowledge acquisition and to the explainability of DL-based cartographic generalization approaches. Finally, future research should strive to investigate alternative learning frameworks leveraging training datasets with fuzzy labels and for the development of sophisticated evaluation strategies that better capture the ambiguities intrinsic to the generalization process. All of these developments are paramount to truly ushering in DL as the next paradigm for end-to-end automated cartographic generalization.

Bibliography

Acharya, D. B. & Zhang, H. (2020), Feature Selection and Extraction for Graph Neural Networks, *in* 'Proceedings of the 2020 ACM Southeast Conference', ACM SE '20, Association for Computing Machinery, New York, NY, USA, p. 252–255.

Ai, T. (2022), 'Some thoughts on deep learning empowering cartography', *Journal of Geography and Cartography* **5**(2), 25–40.

Ai, T., Yin, H., Shen, Y., Yang, M. & Wang, L. (2019), 'A formal model of neighborhood representation and applications in urban building aggregation supported by Delaunay triangulation', *PLOS ONE* **14**(7).

Ai, T., Zhang, X., Zhou, Q. & Yang, M. (2015), 'A vector field model to handle the displacement of multiple conflicts in building generalization', *International Journal of Geographical Information Science* **29**(8), 1310–1331.

Albawi, S., Mohammed, T. A. & Al-Zawi, S. (2017), Understanding of a convolutional neural network, *in* '2017 International Conference on Engineering and Technology (ICET)', Antalya, Turkey, pp. 1–6.

Aleissaee, A. A., Kumar, A., Anwer, R. M., Khan, S., Cholakkal, H., Xia, G.-S. & Khan, F. S. (2023), 'Transformers in Remote Sensing: A Survey', *Remote Sensing* **15**(7), 1860.

Allouche, M. K. & Moulin, B. (2005), 'Amalgamation in cartographic generalization using Kohonen's feature nets', *International Journal of Geographical Information Science* **19**(8–9), 899–914.

Alpaydin, E. (2020), *Introduction to Machine Learning*, 4 edn, MIT Press, London.

Anders, K.-H. & Sester, M. (2000), 'Parameter-free cluster detection in spatial databases and its application to typification', *International Archives of Photogrammetry and Remote Sensing* **33**, 75–83.

Anselin, L. (1995), 'Local Indicators of Spatial Association — LISA', *Geographical Analysis* **27**(2), 93–115.

Armstrong, M. P. (1991), Knowledge Classification and Organization, *in* B. Buttenfield & R. B. McMaster, eds, 'Map Generalization: Making Rules for Knowledge Representation', Addison-Wesley Longman Ltd, pp. 86–102.

Arulkumaran, K., Deisenroth, M. P., Brundage, M. & Bharath, A. A. (2017), 'Deep Reinforcement Learning: A Brief Survey', *IEEE Signal Processing Magazine* **34**(6), 26–38.

Bader, M., Barrault, M. & Weibel, R. (2005), 'Building displacement over a ductile truss', *International Journal of Geographical Information Science* **19**(8–9), 915–936.

Baltrušaitis, T., Ahuja, C. & Morency, L.-P. (2019), 'Multimodal Machine Learning: A Survey and Taxonomy', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **41**(2), 423–443.

Barrault, M., Regnauld, N., Duchêne, C., Haire, K., Baeijs, C., Demazeau, Y., Hardy, P., Mackaness, W. A., Ruas, A. & Weibel, R. (2001), Integrating multi-agent, object-oriented and algorithmic techniques for improved automated map generalisation, *in* 'Proceedings 20th International Cartographic Conference', number 1, pp. 2110–2116.

Basaraner, M. & Cetinkaya, S. (2017), 'Performance of shape indices and classification schemes for characterising perceptual shape complexity of building footprints in GIS', *International Journal of Geographical Information Science* **31**(10), 1952–1977.

Basaraner, M. & Selcuk, M. (2008), 'A Structure Recognition Technique in Contextual Generalisation of Buildings and Built-up Areas', *The Cartographic Journal* **45**(4), 274–285.

Bazi, Y., Bashmal, L., Rahhal, M. M. A., Dayil, R. A. & Ajlan, N. A. (2021), 'Vision Transformers for Remote Sensing Image Classification', *Remote Sensing* **13**(3), 516.

Beard, K. (1991), Constraints on Rule Formation, *in* B. Buttenfield & R. B. McMaster, eds, 'Map Generalization: Making Rules For Knowledge Representation', Addison-Wesley Longman Ltd, pp. 121–135.

Beglinger, N. (2023), 'Exploring Deep Learning for Deformative Operators in Vector-Based Cartographic Road Generalization'. MSc Thesis, Dept. of Geography, University of Zurich.

Bei, W., Guo, M. & Huang, Y. (2019), 'A Spatial Adaptive Algorithm Framework for Building Pattern Recognition Using Graph Convolutional Networks', *Sensors* **19**(24), 5518.

Bićanić, Z. & Solarić, R. (2017), 'Cartographic Generalization and Contributions to its Automation', *Geoadria* **7**(2), 5–21.

Blana, N. & Tsoulos, L. (2022), 'Generalization of Linear and Area Features Incorporating a Shape Measure', *ISPRS International Journal of Geo-Information* **11**(9), 489.

Borůvka, O. (1926), 'O jistém problému minimálním', *Práce Moravské přírodovědecké společnosti* **3**(3), 37–58.

Boutell, M. R., Luo, J., Shen, X. & Brown, C. M. (2004), 'Learning multi-label scene classification', *Pattern Recognition* **37**(9), 1757–1771.

Brassel, K. E. & Weibel, R. (1988), 'A review and conceptual framework of automated map generalization', *International Journal Of Geographical Information Systems* **2**(3), 229–244.

Brinkhoff, T., Kriegel, H.-P., Schneider, R. & Braun, A. (1995), Measuring the Complexity of Polygonal Objects, *in* 'ACM-GIS', Vol. 109.

Bronstein, M. M., Bruna, J., Cohen, T. & Veličković, P. (2021), 'Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges'. https://arxiv.org/abs/2104.13478.

Buda, M., Maki, A. & Mazurowski, M. A. (2018), 'A systematic study of the class imbalance problem in convolutional neural networks', *Neural Networks* **106**, 249–259.

Burghardt, D. & Cecconi, A. (2007), 'Mesh simplification for building typification', *International Journal of Geographical Information Science* **21**(3), 283–298.

Buttenfield, B. P. & McMaster, R. B. (1991), *Map Generalization: Making Rules for Knowledge Representation*, Longman Scientific & Technical London.

Cebrykow, P. (2017), 'Cartographic generalization yesterday and today', *Polish Cartographical Review* **49**(1), 5–15.

Cecconi, A., Weibel, R. & Barrault, M. (2002), Improving Automated Generalisation for On-Demand Web Mapping by Multiscale Databases, *in* D. E. Richardson & P. van Oosterom, eds, 'Advances in Spatial Data Handling', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 515–531.

Cetinkaya, S., Basaraner, M. & Burghardt, D. (2015), 'Proximity-based grouping of buildings in urban blocks: a comparison of four algorithms', *Geocarto International* **30**(6), 618–632.

Chapelle, O., Shivaswamy, P., Vadrevu, S., Weinberger, K., Zhang, Y. & Tseng, B. (2010), 'Boosted multi-task learning', *Machine Learning* **85**(1–2), 149–173.

Charte, F., Rivera, A., del Jesus, M. J. & Herrera, F. (2013), A First Approach to Deal with Imbalance in Multi-label Datasets, *in* J.-S. Pan, M. M. Polycarpou, M. Woźniak, A. C. P. L. F. de Carvalho, H. Quintián & E. Corchado, eds, 'Hybrid Artificial Intelligent Systems. HAIS 2013. Lecture Notes in Computer Science', Vol. 8073, Springer, Berlin, Heidelberg, pp. 150–160.

Charte, F., Rivera, A. J., del Jesus, M. J. & Herrera, F. (2015), 'Addressing imbalance in multilabel classification: Measures and random resampling algorithms', *Neurocomputing* **163**, 3–16.

Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. (2002), 'SMOTE: Synthetic Minority Over-sampling Technique', *Journal of Artificial Intelligence Research* **16**, 321–357.

Cheng, B., Liu, Q. & Li, X. (2015), 'Local Perception-Based Intelligent Building Outline Aggregation Approach with Back Propagation Neural Network', *Neural Processing Letters* **41**(2), 273–292.

Cheng, B., Liu, Q., Li, X. & Wang, Y. (2013), 'Building simplification using backpropagation neural networks: a combination of cartographers' expertise and raster-based local perception', *GIScience & Remote Sensing* **50**(5), 527–542.

Cho, K., Roh, J.-h., Kim, Y. & Cho, S. (2019), A Performance Comparison of Loss Functions, *in* '2019 International Conference on Information and Communication Technology Convergence (ICTC)', Jeju, South Korea, pp. 1146–1151.

Christophe, S., Mermet, S., Laurent, M. & Touya, G. (2022), 'Neural map style transfer exploration with GANs', *International Journal of Cartography* **8**(1), 18–36.

Collins, K. M., Bhatt, U. & Weller, A. (2022), 'Eliciting and Learning with Soft Labels from Every Annotator', *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* **10**(1), 40–52.

Cordonnier, J.-B., Loukas, A. & Jaggi, M. (2020), 'On the Relationship between Self-Attention and Convolutional Layers'. https://arxiv.org/abs/1911.03584.

Courtial, A. (2023), Exploring the potential of deep learning for map generalization, PhD thesis, Université Gustave Eiffel, Champs-sur-Marne, France.

Courtial, A., El Ayedi, A., Touya, G. & Zhang, X. (2020), 'Exploring the Potential of Deep Learning Segmentation for Mountain Roads Generalisation', *ISPRS International Journal of Geo-Information* **9**(5), 338.

Courtial, A., Touya, G. & Zhang, X. (2021*a*), 'Can Graph Convolution Networks Learn Spatial Relations?', *Abstracts of the ICA* **3**.

Courtial, A., Touya, G. & Zhang, X. (2021*b*), 'Generative adversarial networks to generalise urban areas in topographic maps', *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **43**(4), 15–22.

Courtial, A., Touya, G. & Zhang, X. (2022*a*), 'Constraint-Based Evaluation of Map Images Generalized by Deep Learning', *Journal of Geovisualization and Spatial Analysis* **6**(13).

Courtial, A., Touya, G. & Zhang, X. (2022*b*), 'Representing Vector Geographic Information as a Tensor for Deep Learning Based Map Generalisation', *AGILE: GIScience Series* **3**, 32.

Courtial, A., Touya, G. & Zhang, X. (2023), 'Deriving map images of generalised mountain roads with generative adversarial networks', *International Journal of Geographical Information Science* **37**(3), 499–528.

Courtial, A., Touya, G. & Zhang, X. (2024), 'DeepMapScaler: a workflow of deep neural networks for the generation of generalised maps', *Cartography and Geographic Information Science* **51**(1), 41–59.

Crawshaw, M. (2020), 'Multi-Task Learning with Deep Neural Networks: A Survey'. https://arxiv.org/abs/2009.09796.

Daigavane, A., Ravindran, B. & Aggarwal, G. (2021), 'Understanding Convolutions on Graphs', *Distill* . https://distill.pub/2021/understanding-gnns.

Danuser, G. (2011), 'Computer Vision in Cell Biology', *Cell* **147**(5), 973–978.

Davis, J. & Goadrich, M. (2006), The relationship between Precision-Recall and ROC curves, *in* 'Proceedings of the 23rd International Conference on Machine Learning', ICML '06, Association for Computing Machinery, New York, NY, USA, p. 233–240.

Delone, B. N. (1934), 'Sur la sphère vide', *Bulletin de l'Académie des Sciences de l'URSS* **6**, 793–800.

Demirkaya, A., Chen, J. & Oymak, S. (2020), Exploring the Role of Loss Functions in Multiclass Classification, *in* '2020 54th Annual Conference on Information Sciences and Systems (CISS)', Princeton, NJ, USA, pp. 1–5.

Deng, M., Tang, J., Liu, Q. & Wu, F. (2018), 'Recognizing building groups for generalization: a comparative study', *Cartography and Geographic Information Science* **45**(3), 187–204.

Domingo, M., Thibaud, R. & Claramunt, C. (2019), 'A graph-based approach for the structural analysis of road and building layouts', *Geo-spatial Information Science* **22**(1), 59–72.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J. & Houlsby, N. (2021), 'An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale'. https://arxiv.org/abs/2010.11929.

Douglas, D. H. & Peucker, T. K. (1973), 'Algorithms for the reduction of the number of points required to represent a digitized line or its caricature', *Cartographica: The International Journal for Geographic Information and Geovisualization* **10**(2), 112–122.

Du, J., Wu, F., Xing, R., Gong, X. & Yu, L. (2022), 'Segmentation and sampling method for complex polyline generalization based on a generative adversarial network', *Geocarto International* **37**(14), 4158–4180.

Du, J., Wu, F., Yin, J., Liu, C. & Gong, X. (2022), 'Polyline simplification based on the artificial neural network with constraints of generalization knowledge', *Cartography and Geographic Information Science* **49**(4), 313–337.

Du, J., Wu, F., Zhu, L., Liu, C. & Wang, A. (2022), 'An ensemble learning simplification approach based on multiple machine-learning algorithms with the fusion using of raster and vector data and a use case of coastline simplification', *Acta Geodaetica et Cartographica Sinica* **51**(3), 373–387.

Dubuisson, M.-P. & Jain, A. (1994), A modified Hausdorff distance for object matching, *in* 'Proceedings of 12th International Conference on Pattern Recognition', Vol. 1, Jerusalem, Israel, pp. 566–568.

Duchêne, C., Bard, S., Barillot, X., Ruas, A., Trevisan, J. & Holzapfel, F. (2003), Quantitative and qualitative description of building orientation, *in* 'Fifth Workshop on Progress in Automated Map Generalisation'.

Duchêne, C., Baella, B., Brewer, C. A., Burghardt, D., Buttenfield, B. P., Gaffuri, J., Käuferle, D., Lecordix, F., Maugeais, E., Nijhuis, R., Pla, M., Post, M., Regnauld, N., Stanislawski, L. V., Stoter, J., Tóth, K., Urbanke, S., van Altena, V. & Wiedemann, A. (2014), Generalisation in Practice Within National Mapping Agencies, *in* D. Burghardt, C. Duchêne & W. Mackaness, eds, 'Abstracting Geographic Information in a Data Rich World: Methodologies and Applications of Map Generalisation', Springer International Publishing, Cham, pp. 329–391.

Duchêne, C., Touya, G., Taillandier, P., Gaffuri, J., Ruas, A. & Renard, J. (2018), Multi-Agents Systems for Cartographic Generalization: Feedback from Past and On-going Research, Technical report, IGN France, LaSTIG - COGIT team.

Eckert, M. (1908), 'On the Nature of Maps and Map Logic', *Bulletin of the American Geographical Society* **40**(6), 344–351. Translated by W. Joerg.

Feng, Y. (2023), Prompt-aided Map Generalization with Diffusion Models, *in* '1st Workshop on CartoAI: AI for Cartography'.

Feng, Y., Qian, N., Fu, C., Zhou, Z. & Sester, M. (2023), 'PolygonTranslator - learning to simplify building footprints from one scale to another', *Abstracts of the ICA* **6**(60).

Feng, Y., Thiemann, F. & Sester, M. (2019), 'Learning Cartographic Building Generalization with Deep Convolutional Neural Networks', *ISPRS International Journal of Geo-Information* **8**(6), 258.

Fernando, K. R. M. & Tsokos, C. P. (2022), 'Dynamically Weighted Balanced Loss: Class Imbalanced Learning and Confidence Calibration of Deep Neural Networks', *IEEE Transactions on Neural Networks and Learning Systems* **33**(7), 2940–2951.

Fey, M. & Lenssen, J. E. (2019), 'Fast Graph Representation Learning with PyTorch Geometric'. https://arxiv.org/abs/1903.02428.

Filippovska, Y., Walter, V. & Fritsch, D. (2008), 'Quality evaluation of generalization algorithms', *ISPRS Commission II, WG II* **7**.

Fu, C., Senn, J., Winkler, J., Weibel, R. & Zhou, Z. (2025), An official map-sourced building data set for deep learning-based map generalization. unpublished.

Fu, C., Zhou, Z., Feng, Y. & Weibel, R. (2024), 'Keeping walls straight: data model and training set size matter for deep learning in building generalization', *Cartography and Geographic Information Science* **51**(1), 130–145.

Fu, C., Zhou, Z., Winkler, J., Beglinger, N. & Weibel, R. (2023), Progress in Constructing an Open Map Generalization Data Set for Deep Learning, *in* R. Beecham, J. A. Long, D. Smith, Q. Zhao & S. Wise, eds, '12th International Conference on Geographic Information Science (GIScience 2023)', Vol. 277 of *Leibniz International Proceedings in Informatics (LIPIcs)*, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, pp. 30:1–30:6.

Fu, C., Zhou, Z., Xin, Y. & Weibel, R. (2024), 'Reasoning cartographic knowledge in deep learning-based map generalization with explainable AI', *International Journal of Geographical Information Science* **38**(10), 2061–2082.

Fu, H., Gong, M., Wang, C., Batmanghelich, K., Zhang, K. & Tao, D. (2019), Geometry-Consistent Generative Adversarial Networks for One-Sided Unsupervised Domain Mapping, *in* '2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)', pp. 2422–2431.

Förster, T., Stoter, J. & Köbben, B. (2007), Towards a formal classification of generalization operators, *in* 'Proceedings of the 23rd International Cartographic Conference'.

Gadzicki, K., Khamsehashari, R. & Zetzsche, C. (2020), Early vs Late Fusion in Multimodal Convolutional Neural Networks, *in* '2020 IEEE 23rd International Conference on Information Fusion (FUSION)', Rustenburg, South Africa, pp. 1–6.

Gao, S., Hu, Y. & Li, W. (2023), *Handbook of Geospatial Artificial Intelligence*, 1 edn, CRC Press.

Gong, X. & Wu, F. (2016), 'A typification method for linear pattern in urban building generalisation', *Geocarto International* **33**(2), 189–207.

Goodchild, M. F. (2018), 'A GIScience Perspective on the Uncertainty of Context', *Annals of the American Association of Geographers* **108**(6), 1476–1481.

Goodfellow, I. J., Bengio, Y. & Courville, A. (2016), *Deep Learning*, MIT Press. http://www.deeplearningbook.org.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. & Bengio, Y. (2014), 'Generative Adversarial Networks'. https://arxiv.org/abs/1406.2661.

Grünreich, D. (1985), 'Computer-assisted generalization', *Papers CERCO-Cartography Course* .

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S. & Yang, G.-Z. (2019), 'XAI—Explainable artificial intelligence', *Science Robotics* **4**(37), eaay7120.

Hafeez, A., Ali, T., Nawaz, A., Rehman, S. U., Mudasir, A. I., Alsulami, A. A. & Alqahtani, A. (2023), 'Addressing Imbalance Problem for Multi Label Classification of Scholarly Articles', *IEEE Access* **11**, 74500–74516.

Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H. & Bing, G. (2017), 'Learning from class-imbalanced data: Review of methods and applications', *Expert Systems with Applications* **73**, 220–239.

Hamilton, W. L., Ying, R. & Leskovec, J. (2017), Inductive representation learning on large graphs, *in* I. Guyon, U. von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan & R. Garnett, eds, 'Proceedings of the 31st International Conference on Neural Information Processing Systems', NIPS'17, Curran Associates Inc., Red Hook, NY, USA, p. 1025–1035.

Hangouët, J.-F. (1998), Approches et méthodes pour l'automatisation de la généralisation cartographique - Application en bord de ville, PhD thesis, Université de Marne-la-Vallée, Champs-sur-Marne, France.

Harrie, L., Touya, G., Oucheikh, R., Ai, T., Courtial, A. & Richter, K.-F. (2024), 'Machine learning in cartography', *Cartography and Geographic Information Science* **51**(1), 1–19.

Harrie, L. & Weibel, R. (2007), Modelling the Overall Process of Generalisation, *in* W. A. Mackaness, A. Ruas & L. T. Sarjakoski, eds, 'Generalisation of Geographic Information: Cartographic Modelling and Applications', International Cartographic Association, Elsevier Science B.V., pp. 67–87.

Hausdorff, F. (1914), *Grundzüge der Mengenlehre*, Veit & Comp, Leipzig.

Hayes-Roth, F., Waterman, D. A. & Lenat, D. B. (1983), *Building Expert Systems*, Addison-Wesley Longman Publishing Co., Inc., Reading, MA.

He, K., Zhang, X., Ren, S. & Sun, J. (2015), 'Deep Residual Learning for Image Recognition'. https://arxiv.org/abs/1512.03385.

He, X., Zhang, X. & Xin, Q. (2018), 'Recognition of building group patterns in topographic maps based on graph partitioning and random forest', *ISPRS Journal of Photogrammetry and Remote Sensing* **136**, 26–40.

Hu, Y., Goodchild, M., Zhu, A.-X., Yuan, M., Aydin, O., Bhaduri, B., Gao, S., Li, W., Lunga, D. & Newsam, S. (2024), 'A five-year milestone: reflections on advances and limitations in GeoAI research', *Annals of GIS* **30**(1), 1–14.

Hu, Y., Liu, C., Li, Z., Xu, J., Han, Z. & Guo, J. (2022), 'Few-Shot Building Footprint Shape Classification with Relation Network', *ISPRS International Journal of Geo-Information* **11**(5), 311.

Hu, Z., Dong, Y., Wang, K. & Sun, Y. (2020), Heterogeneous Graph Transformer, *in* 'Proceedings of The Web Conference 2020', WWW '20, Association for Computing Machinery, New York, NY, USA, p. 2704–2710.

Huang, S.-C., Pareek, A., Seyyedi, S., Banerjee, I. & Lungren, M. P. (2020), 'Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines', *npj Digital Medicine* **3**(1), 136.

Ioffe, S. & Szegedy, C. (2015), 'Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift'. https://arxiv.org/abs/1502.03167.

Isola, P., Zhu, J.-Y., Zhou, T. & Efros, A. A. (2016), 'Image-to-Image Translation with Conditional Adversarial Networks'. https://arxiv.org/abs/1611.07004.

Janowicz, K., Gao, S., McKenzie, G., Hu, Y. & Bhaduri, B. (2020), 'GeoAI: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond', *International Journal of Geographical Information Science* **34**(4), 625–636.

Ji, S., Tang, L., Yu, S. & Ye, J. (2008), Extracting shared subspace for multi-label classification, *in* 'Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining', KDD '08, Association for Computing Machinery, New York, USA, p. 381–389.

Jiang, B., Xu, S. & Li, Z. (2023), 'Polyline simplification using a region proposal network integrating raster and vector features', *GIScience & Remote Sensing* **60**(1).

Jones, C. B. (2014), *Geographical Information Systems and Computer Cartography*, Routledge, London.

Kang, Y., Gao, S. & Roth, R. E. (2019), 'Transferring multiscale map styles using generative adversarial networks', *International Journal of Cartography* **5**(2–3), 115–141.

Kang, Y., Gao, S. & Roth, R. E. (2024), 'Artificial intelligence studies in cartography: a review and synthesis of methods, applications, and ethics', *Cartography and Geographic Information Science* **51**(4), 599–630.

Kang, Y., Rao, J., Wang, W., Peng, B., Gao, S. & Zhang, F. (2020), 'Towards cartographic knowledge encoding with deep learning: A case study of building generalization', *The 23rd International Research Symposium on Cartography* pp. 1–6.

Kavouras, M. & Kokla, M. (2007), *Theories of Geographic Concepts: Ontological Approaches to Semantic Integration*, 1 edn, CRC Press, Boca Raton.

Keskar, N. S., Mudigere, D., Nocedal, J., Smelyanskiy, M. & Tang, P. T. P. (2016), 'On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima'. https://arxiv.org/abs/1609.04836.

Khosla, C. & Saini, B. S. (2020), Enhancing Performance of Deep Learning Models with different Data Augmentation Techniques: A Survey, *in* '2020 International Conference on Intelligent Engineering and Management (ICIEM)', London, UK, pp. 79–85.

Kilpelainen, T. (1994), Updating multiple representation geo-databases by incremental generalization, *in* H. Ebner, C. Heipke & K. Eder, eds, 'ISPRS Commission III Symposium: Spatial Information from Digital Photogrammetry and Computer Vision', Vol. 2357, International Society for Optics and Photonics, SPIE, pp. 440–447.

Kingma, D. P. & Ba, J. (2014), 'Adam: A Method for Stochastic Optimization'. https://arxiv.org/abs/1412.6980.

Kipf, T. N. & Welling, M. (2016*a*), 'Semi-Supervised Classification with Graph Convolutional Networks'. https://arxiv.org/abs/1609.02907.

Kipf, T. N. & Welling, M. (2016*b*), 'Variational Graph Auto-Encoders'. https://arxiv.org/abs/1611.07308.

Knura, M. (2021), 'Deep Learning for Map Generalization: Towards a new Approach using Vector Data', *Abstracts of the ICA* **3**.

Knura, M. (2024), 'Learning from vector data: enhancing vector-based shape encoding and shape classification for map generalization purposes', *Cartography and Geographic Information Science* **51**(1), 146–167.

Kong, B., Ai, T., Zou, X., Yan, X. & Yang, M. (2024), 'A graph-based neural network approach to integrate multi-source data for urban building function classification', *Computers, Environment and Urban Systems* **110**, 102094.

Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2017), 'ImageNet classification with deep convolutional neural networks', *Communications of the ACM* **60**(6), 84–90.

Kumar, S., Kumar, N., Dev, A. & Naorem, S. (2023), 'Movie genre classification using binary relevance, label powerset, and machine learning classifiers', *Multimedia Tools and Applications* **82**(1), 945–968.

Lafon, E., Potié, Q. & Touya, G. (2023), Salient building detection using multimodal deep learning, *in* 'Abstracts from GIScience 2023 Workshop on CartoAI: AI for cartography'.

Lagrange, F., Landras, B. & Mustière, S. (2000), 'Machine learning techniques for determining parameters of cartographic generalisation algorithms', *International Archives of Photogrammetry and Remote Sensing* **33**(4), 718–725.

Lamy, S., Ruas, A., Demazeau, Y., Jackson, M., Mackaness, W. & Weibel, R. (1999), The Application of Agents in Automated Map Generalisation, *in* 'Proceedings 19th Int. Cartographic Conference', Ottawa, pp. 160–169.

LeCun, Y. & Bengio, Y. (1998), Convolutional networks for images, speech, and time series, *in* 'The Handbook of Brain Theory and Neural Networks', MIT Press, Cambridge, MA, USA, p. 255–258.

LeCun, Y., Bengio, Y. & Hinton, G. (2015), 'Deep learning', *Nature* **521**(7553), 436–444.

Lee, D. (1996), 'Automation of Map Generalization: The Cutting-Edge Technology', *ESRI White Paper Series* .

Lee, J., Jang, H., Yang, J. & Yu, K. (2017), 'Machine Learning Classification of Buildings for Map Generalization', *ISPRS International Journal of Geo-Information* **6**(10), 309.

Li, H., Guo, Q. & Liu, J. (2005), Rapid Algorithm of Building Typification in Web Mapping, *in* 'Proceedings of the International Symposium on Spatio-temporal Modelling, Spatial Reasoning, Analysis, Data Mining and Data Fusion', Beijing, China, pp. 27–29.

Li, P., Yan, H. & Lu, X. (2024), 'MultiLineStringNet: a deep neural network for linear feature set recognition', *Cartography and Geographic Information Science* **51**(1), 114–129.

Li, T., Deng, J., Shen, Y., Qiu, L., Yongxiang, H. & Cao, C. C. (2023), 'Towards Fine-Grained Explainability for Heterogeneous Graph Neural Network', *Proceedings of the AAAI Conference on Artificial Intelligence* **37**(7), 8640–8647.

Li, W. (2020), 'GeoAI: Where machine learning and big data converge in GIScience', *Journal of Spatial Information Science* **20**, 71–77.

Li, Y., Sakamoto, M., Shinohara, T. & Satoh, T. (2020), 'Automatic label placement of area-features using deep learning', *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **33**(4), 117–122.

Li, Z. (2007), 'Digital Map Generalization at the Age of Enlightenment: A Review of the First Forty Years', *The Cartographic Journal* **44**(1), 80–93.

Li, Z. & Openshaw, S. (1993), 'A Natural Principle for the Objective Generalization of Digital Maps', *Cartography and Geographic Information Systems* **20**(1), 19–29.

Li, Z. & Su, B. (1995), 'From phenomena to essence: envisioning the nature of digital map generalisation', *The Cartographic Journal* **32**(1), 45–47.

Li, Z., Yan, H., Ai, T. & Chen, J. (2004), 'Automated building generalization based on urban morphology and Gestalt theory', *International Journal of Geographical Information Science* **18**(5), 513–534.

Liao, S., Bai, Z. & Bai, Y. (2012), 'Errors prediction for vector-to-raster conversion based on map load and cell size', *Chinese Geographical Science* **22**(6), 695–704.

Lin, T.-Y., Goyal, P., Girshick, R., He, K. & Dollár, P. (2020), 'Focal Loss for Dense Object Detection', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **42**(2), 318–327.

Liu, C., Hu, Y., Li, Z., Xu, J., Han, Z. & Guo, J. (2021), 'TriangleConv: A Deep Point Convolutional Network for Recognizing Building Shapes in Map Space', *ISPRS International Journal of Geo-Information* **10**(10), 687.

Liu, Q., Xiao, L., Yang, J. & Wei, Z. (2022), 'Multilevel Superpixel Structured Graph U-Nets for Hyperspectral Image Classification', *IEEE Transactions on Geoscience and Remote Sensing* **60**, 1–15.

Lonergan, M. & Jones, C. B. (2001), 'An Iterative Displacement Method for Conflict Resolution in Map Generalization', *Algorithmica* **30**(2), 287–301.

Long, J., Shelhamer, E. & Darrell, T. (2015), Fully Convolutional Networks for Semantic Segmentation, *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)', IEEE Computer Society, Los Alamitos, CA, USA, pp. 3431–3440.

Luque, A., Carrasco, A., Martín, A. & de las Heras, A. (2019), 'The impact of class imbalance in classification performance metrics based on the binary confusion matrix', *Pattern Recognition* **91**, 216–231.

Ma, W., Wang, B., Liu, C., Li, Q., Yang, C., Pan, J., Zhou, B. & Wang, Y. (2023), 'Complex buildings orientation recognition and description based on vector reconstruction', *International Journal of Applied Earth Observation and Geoinformation* **123**, 103486.

Mackaness, W. A. (1995), A constraint based approach to human computer interaction in automated cartography, *in* 'Proceedings of the 17th International Cartographic Conference', Barcelona, pp. 1423–1432.

Mackaness, W., Burghardt, D. & Duchêne, C. (2014), Map Generalisation: Fundamental to the Modelling and Understanding of Geographic Space, *in* D. Burghardt, C. Duchêne & W. Mackaness, eds, 'Abstracting Geographic Information in a Data Rich World: Methodologies and Applications of Map Generalisation', Springer International Publishing, Cham, pp. 1–15.

Mackaness, W. & Edwards, G. (2002), The Importance of Modelling Pattern and Structure in Automated Map Generalisation, *in* 'Proceedings of the Joint ISPRS/ICA Workshop on Multi-Scale Representations of Spatial Data', Ottawa, Canada.

Mai, G., Janowicz, K., Hu, Y., Gao, S., Yan, B., Zhu, R., Cai, L. & Lao, N. (2022), 'A review of location encoding for GeoAI: methods and applications', *International Journal of Geographical Information Science* **36**(4), 639–673.

Mai, G., Jiang, C., Sun, W., Zhu, R., Xuan, Y., Cai, L., Janowicz, K., Ermon, S. & Lao, N. (2023), 'Towards general-purpose representation learning of polygonal geometries', *GeoInformatica* **27**(2), 289–340.

Masters, D. & Luschi, C. (2018), 'Revisiting Small Batch Training for Deep Neural Networks'. https://arxiv.org/abs/1804.07612.

Mathews, L. & Hari, S. (2018), Learning From Imbalanced Data, *in* M. Khosrow-Pour, ed., 'Encyclopedia of Information Science and Technology', 4 edn, IGI Global, pp. 1825–1834.

McMaster, R. B. & Mark, D. M. (1991), The design of a graphical user interface for knowledge acquisition in cartographic generalization, *in* 'Proceedings GIS/LIS', Atlanta, GA, pp. 311–320.

McMaster, R. B. & Shea, K. S. (1992), Generalization in Digital Cartography, Association of American Geographers, Washington, DC.

Meyes, R., Lu, M., Waubert de Puiseau, C. & Meisen, T. (2019), 'Ablation Studies in Artificial Neural Networks'. https://arxiv.org/abs/1901.08644.

Mitchell, T. (1997), *Machine Learning*, McGraw-Hill, New York.

Monmonier, M. (1986), 'Toward a practicable model of cartographic generalisation', *Nuclear Physics, Section A* pp. 257–266.

Mroueh, Y., Marcheret, E. & Goel, V. (2015), Deep multimodal learning for Audio-Visual Speech Recognition, *in* '2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)', South Brisbane, QLD, Australia, pp. 2130–2134.

Mustière, S. & Moulin, B. (2002), 'What is spatial context in cartographic generalization?', *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* **34**(4), 274–278.

Müller, J.-C. (1990), Rule Based Generalization: Potentials and Impediments, *in* 'Proceedings of the 4th International Symposium on Spatial Data Handling', Vol. 1, Zurich, Switzerland, pp. 317–334.

Müller, J.-C. (1991), Generalization of Spatial Databases, *in* D. J. Maguire, M. F. Goodchild & D. W. Rhind, eds, 'Geographical Information Systems: Principles and Applications', 1 edn, Longman, London, pp. 457–475.

Müller, J.-C., Weibel, R., Lagrange, J. P. & Salgé, F. (1995), Generalization: State of the art and issues, *in* J.-C. Müller, J. P. Lagrange & R. Weibel, eds, 'GIS And Generalisation: Methodology and Practice', Taylor & Francis, London, p. 3–17.

Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H. & Ng, A. Y. (2011), Multimodal Deep Learning, *in* 'Proceedings of the 28th International Conference on Machine Learning', Bellevue, WA, USA, pp. 689–696.

Niu, Z., Zhong, G. & Yu, H. (2021), 'A review on the attention mechanism of deep learning', *Neurocomputing* **452**, 48–62.

Nyerges, T. L. (1991), Representing Geographical Meaning, *in* B. Buttenfield & R. B. McMaster, eds, 'Map Generalization: Making Rules for Knowledge Representation', Addison-Wesley Longman Ltd, pp. 59–85.

Oktay, O., Schlemper, J., Le Folgoc, L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B. & Rueckert, D. (2018), 'Attention U-Net: Learning Where to Look for the Pancreas'. https://arxiv.org/abs/1804.03999.

Oucheikh, R. & Harrie, L. (2024), 'A feasibility study of applying generative deep learning models for map labeling', *Cartography and Geographic Information Science* **51**(1), 168–191.

Parsaye, K. & Chignell, M. (1988), *Expert Systems For Experts*, Wiley & Sons, New York.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J. & Chintala, S. (2019), 'PyTorch: An Imperative Style, High-Performance Deep Learning Library'. https://arxiv.org/abs/1912.01703.

Patricios, N. N. (2001), 'Urban design principles of the original neighbourhood concepts', *Urban Morphology* **6**(1), 21–32.

Peter, B. & Weibel, R. (1999), Using Vector and Raster-Based Techniques in Categorical Map Generalization, *in* 'Third ICA Workshop on Progress in Automated Map Generalization, Ottawa'.

Picchiotti, N. & Gori, M. (2021), 'Clustering-Based Interpretation of Deep ReLU Network'. https://arxiv.org/abs/2110.06593.

Powitz, B. M. (1993), 'Computer-assisted generalization - An important software tool in GIS', *International Archives of Photogrammetry and Remote Sensing* **29**, 664–672.

Prechelt, L. (1998), Early Stopping - But When?, *in* G. B. Orr & K.-R. Müller, eds, 'Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science', Vol. 1524, Springer, Berlin, Heidelberg, pp. 55–69.

Rainsford, D. & Mackaness, W. (2002), Template Matching in Support of Generalisation of Rural Buildings, *in* D. E. Richardson & P. van Oosterom, eds, 'Advances in Spatial Data Handling', Springer, Berlin, Heidelberg, pp. 137–151.

Ramachandram, D. & Taylor, G. W. (2017), 'Deep Multimodal Learning: A Survey on Recent Advances and Trends', *IEEE Signal Processing Magazine* **34**(6), 96–108.

Ramachandran, P., Parmar, N., Vaswani, A., Bello, I., Levskaya, A. & Shlens, J. (2019), 'Stand-Alone Self-Attention in Vision Models'. https://arxiv.org/abs/1906.05909.

Ratner, A., Bach, S. H., Ehrenberg, H., Fries, J., Wu, S. & Ré, C. (2020), 'Snorkel: rapid training data creation with weak supervision', *The VLDB Journal* **29**(2–3), 709–730.

Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. (2015), 'You Only Look Once: Unified, Real-Time Object Detection'. https://arxiv.org/abs/1506.02640.

Regnauld, N. (2001), 'Contextual Building Typification in Automated Map Generalization', *Algorithmica* **30**(2), 312–333.

Regnauld, N. & McMaster, R. B. (2007), A Synoptic View of Generalisation Operators, *in* W. A. Mackaness, A. Ruas & L. T. Sarjakoski, eds, 'Generalisation of Geographic Information: Cartographic Modelling and Applications', International Cartographic Association, Elsevier Science B.V., pp. 37–66.

Renard, J., Gaffuri, J., Duchêne, C. & Touya, G. (2011), Automated generalisation results using the agent-based platform CartAGen, *in* '25th International Cartographic Conference'.

Ridnik, T., Ben-Baruch, E., Zamir, N., Noy, A., Friedman, I., Protter, M. & Zelnik-Manor, L. (2021), Asymmetric Loss for Multi-Label Classification, *in* 'Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)', pp. 82–91.

Rieger, M. K. & Coulson, M. R. C. (1993), 'Consensus or Confusion: Cartographers' Knowledge of Generalization', *Cartographica: The International Journal for Geographic Information and Geovisualization* **30**(2–3), 69–80.

Ronneberger, O., Fischer, P. & Brox, T. (2015), U-Net: Convolutional Networks for Biomedical Image Segmentation, *in* N. Navab, J. Hornegger, W. M. Wells & A. F. Frangi, eds, 'Medical Image Computing and Computer-Assisted Intervention. MICCAI 2015. Lecture Notes in Computer Science', Vol. 9351, Springer, Cham, pp. 234–241.

Rote, G. (1991), 'Computing the minimum Hausdorff distance between two point sets on a line under translation', *Information Processing Letters* **38**(3), 123–127.

Roth, R. E., Brewer, C. A. & Stryker, M. S. (2011), 'A typology of operators for maintaining legible map designs at multiple scales', *Cartographic Perspectives* (68), 29–64.

Ruas, A. (1998), 'A method for building displacement in automated map generalisation', *International Journal of Geographical Information Science* **12**(8), 789–803.

Ruas, A. (1999), 'Modèle de généralisation de données urbaines à base de contraintes et d'autonomie', *Cybergeo: European Jornal of Geography* .

Ruas, A. (2001), Automatic Generalisation Project: Learning Process from Interactive Generalisation, Technical Report 39, European Organization for Experimental Photogrammetric Research.

Ruas, A. & Duchêne, C. (2007), A Prototype Generalisation System Based on the Multi-Agent System Paradigm, *in* W. A. Mackaness, A. Ruas & L. T. Sarjakoski, eds, 'Generalisation of Geographic Information: Cartographic Modelling and Applications', International Cartographic Association, Elsevier Science B.V., p. 269–284.

Ruas, A. & Mackaness, W. A. (1997), Strategies for Urban Map Generalization, *in* 'Proceedings of the 18th International Cartographic Conference', Vol. 3, Stockholm, Sweden, pp. 1387–1394.

Ruder, S. (2017), 'An Overview of Multi-Task Learning in Deep Neural Networks'. https://arxiv.org/abs/1706.05098.

Rudin, C. (2019), 'Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead', *Nature Machine Intelligence* **1**(5), 206–215.

Saito, T. & Rehmsmeier, M. (2015), 'The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets', *PLOS ONE* **10**(3), e0118432.

Sanchez-Lengeling, B., Reif, E., Pearce, A. & Wiltschko, A. B. (2021), 'A Gentle Introduction to Graph Neural Networks', *Distill* . https://distill.pub/2021/gnn-intro.

Sarjakoski, L. T. (2007), Conceptual Models of Generalisation and Multiple Representation, *in* W. A. Mackaness, A. Ruas & L. T. Sarjakoski, eds, 'Generalisation of Geographic Information: Cartographic Modelling and Applications', International Cartographic Association, Elsevier Science B.V., pp. 11–35.

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M. & Monfardini, G. (2009), 'The Graph Neural Network Model', *IEEE Transactions on Neural Networks* **20**(1), 61–80.

Schmidhuber, J. (2015), 'Deep learning in neural networks: An overview', *Neural Networks* **61**, 85–117.

Sechidis, K., Tsoumakas, G. & Vlahavas, I. (2011), On the Stratification of Multi-label Data, *in* D. Gunopulos, T. Hofmann, D. Malerba & M. Vazirgiannis, eds, 'Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2011. Lecture Notes in Computer Science', Vol. 6913, Springer, Berlin, Heidelberg, pp. 145–158.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. & Batra, D. (2019), 'Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization', *International Journal of Computer Vision* **128**(2), 336–359.

Senn, J., Fu, C., Zhou, Z. & Weibel, R. (2024), 'Efficient operator annotation for deep learning in cartographic building generalization', *Abstracts of the ICA* **7**, 149.

Sester, M. (2000), 'Knowledge acquisition for the automatic interpretation of spatial data', *International Journal of Geographical Information Science* **14**(1), 1–24.

Sester, M. (2005), 'Optimization approaches for generalization and data abstraction', *International Journal of Geographical Information Science* **19**(8–9), 871–897.

Sester, M. (2020), 'Cartographic generalization', *Journal of Spatial Information Science* **21**, 5–11.

Sester, M., Arsanjani, J. J., Klammer, R., Burghardt, D. & Haunert, J.-H. (2014), Integrating and Generalising Volunteered Geographic Information, *in* D. Burghardt, C. Duchêne & W. Mackaness, eds, 'Abstracting Geographic Information in a Data Rich World: Methodologies and Applications of Map Generalisation', Springer International Publishing, Cham, pp. 119–155.

Sester, M., Feng, Y. & Thiemann, F. (2018), 'Building Generalization Using Deep Learning', *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **42**(4), 565–572.

Shen, Y., Li, J., Wang, Z., Zhao, R. & Wang, L. (2022), 'A raster-based typification method for multiscale visualization of building features considering distribution patterns', *International Journal of Digital Earth* **15**(1), 249–275.

Shorten, C. & Khoshgoftaar, T. M. (2019), 'A survey on Image Data Augmentation for Deep Learning', *Journal of Big Data* **6**(1), 60.

Slocum, T. A., McMaster, R. B., Kessler, F. C. & Howard, H. H. (2022), *Thematic Cartography and Geovisualization*, 4 edn, CRC Press, Boca Raton.

Spiess, E., Baumgartner, U., Arn, S. & Vez, C. (2005), 'Topographic Maps - Map Graphics and Generalisation', *Cartographic Publication Series* **17**. Swiss Society of Cartography.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. (2014), 'Dropout: A Simple Way to Prevent Neural Networks from Overfitting', *Journal of Machine Learning Research* **15**(56), 1929–1958.

Stanislawski, L. V., Buttenfield, B. P., Bereuter, P., Savino, S. & Brewer, C. A. (2014), Generalisation Operators, *in* D. Burghardt, C. Duchêne & W. Mackaness, eds, 'Abstracting Geographic Information in a Data Rich World: Methodologies and Applications of Map Generalisation', Springer International Publishing, Cham, pp. 157–195.

Steiniger, S., Lange, T., Burghardt, D. & Weibel, R. (2008), 'An Approach for the Classification of Urban Building Structures Based on Discriminant Analysis Techniques', *Transactions in GIS* **12**(1), 31–59.

Steiniger, S., Taillandier, P. & Weibel, R. (2010), 'Utilising urban context recognition and machine learning to improve the generalisation of buildings', *International Journal of Geographical Information Science* **24**(2), 253–282.

Stoter, J., Zhang, X., Stigmar, H. & Harrie, L. (2014), Evaluation in Generalisation, *in* D. Burghardt, C. Duchêne & W. Mackaness, eds, 'Abstracting Geographic Information in a Data Rich World: Methodologies and Applications of Map Generalisation', Springer International Publishing, Cham, pp. 259–297.

Su, B. (1996), 'A Generalized Frame for Cartographic Knowledge Representation', *Cartography* **25**(2), 31–41.

Sug, H. (2018), 'Performance of Machine Learning Algorithms and Diversity in Data', *MATEC Web of Conferences* **210**(04019).

Sun, Z., Song, Q., Zhu, X., Sun, H., Xu, B. & Zhou, Y. (2015), 'A novel ensemble method for classifying imbalanced data', *Pattern Recognition* **48**(5), 1623–1637.

Tahir, M. A., Kittler, J., Mikolajczyk, K. & Yan, F. (2009), A multiple expert approach to the class imbalance problem using inverse random under sampling, *in* J. A. Benediktsson, J. Kittler & F. Roli, eds, 'Multiple Classifier Systems. MCS 2009. Lecture Notes in Computer Science', Vol. 5519, Springer, Berlin, Heidelberg, pp. 82–91.

Tarawneh, A. S., Hassanat, A. B., Altarawneh, G. A. & Almuhaimeed, A. (2022), 'Stop Oversampling for Class Imbalance Learning: A Review', *IEEE Access* **10**, 47643–47660.

Tarekegn, A. N., Giacobini, M. & Michalak, K. (2021), 'A review of methods for imbalanced multi-label classification', *Pattern Recognition* **118**, 107965.

Tobler, W. R. (1970), 'A Computer Movie Simulating Urban Growth in the Detroit Region', *Economic Geography* **46**, 234–240.

Touya, G. (2012), Social Welfare to Assess the Global Legibility of a Generalized Map, *in* N. Xiao, M.-P. Kwan, M. F. Goodchild & S. Shekhar, eds, 'Geographic Information Science. GIScience 2012. Lecture Notes in Computer Science', Vol. 7478, Springer, Berlin, Heidelberg, pp. 198–211.

Touya, G. & Lokhat, I. (2020), 'Deep Learning for Enrichment of Vector Spatial Databases: Application to Highway Interchange', *ACM Transactions on Spatial Algorithms and Systems* **6**(3), 1–21.

Touya, G., Potié, Q. & Mackaness, W. A. (2023), 'Incorporating ideas of structure and meaning in interactive multi scale mapping environments', *International Journal of Cartography* **9**(2), 342–372.

Touya, G., Zhang, X. & Lokhat, I. (2019), 'Is deep learning the new agent for map generalization?', *International Journal of Cartography* **5**(2–3), 142–157.

Tsoumakas, G. & Katakis, I. (2007), 'Multi-Label Classification: An Overview', *International Journal of Data Warehousing and Mining* **3**(3), 1–13.

Töpfer, F. & Pillewizer, W. (1966), 'The Principles of Selection', *The Cartographic Journal* **3**(1), 10–16.

Usery, E. L., Arundel, S. T., Shavers, E., Stanislawski, L., Thiem, P. & Varanka, D. (2022), 'GeoAI in the US Geological Survey for topographic mapping', *Transactions in GIS* **26**(1), 25–40.

van 't Veer, R., Bloem, P. & Folmer, E. (2018), 'Deep Learning for Classification Tasks on Geospatial Vector Polygons'. https://arxiv.org/abs/1806.03857.

Vandenhende, S., Georgoulis, S., Van Gansbeke, W., Proesmans, M., Dai, D. & Van Gool, L. (2022), 'Multi-Task Learning for Dense Prediction Tasks: A Survey', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **44**(7), 3614–3633.

Varanka, D. E. & Usery, E. L. (2018), 'The map as knowledge base', *International Journal of Cartography* **4**(2), 201–223.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. & Polosukhin, I. (2017), Attention is All you Need, *in* I. Guyon, U. von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan & R. Garnett, eds, 'Advances in Neural Information Processing Systems', Vol. 30, Curran Associates, Inc.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P. & Bengio, Y. (2017), 'Graph Attention Networks'. https://arxiv.org/abs/1710.10903.

Voita, E., Talbot, D., Moiseev, F., Sennrich, R. & Titov, I. (2019), 'Analyzing Multi-Head Self-Attention: Specialized Heads Do the Heavy Lifting, the Rest Can Be Pruned'. https://arxiv.org/abs/1905.09418.

Wang, L., Guo, Q., Liu, Y., Sun, Y. & Wei, Z. (2017), 'Contextual Building Selection Based on a Genetic Algorithm in Map Generalization', *ISPRS International Journal of Geo-Information* **6**(9), 271.

Ware, J. M., Jones, C. B. & Thomas, N. (2003), 'Automated map generalization with multiple operators: a simulated annealing approach', *International Journal of Geographical Information Science* **17**(8), 743–769.

Watkinson, N., Shivam, A., Nicolau, A. & Veidenbaum, A. (2019), Teaching Parallel Computing and Dependence Analysis with Python, *in* '2019 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)', Rio de Janeiro, Brazil, pp. 320–325.

Wei, Z., Guo, Q., Wang, L. & Yan, F. (2018), 'On the spatial distribution of buildings for map generalization', *Cartography and Geographic Information Science* **45**(6), 539–555.

Weibel, R. (1991), Amplified Intelligence and Rule-based Systems, *in* B. Buttenfield & R. B. McMaster, eds, 'Map Generalization: Making Rules For Knowledge Representation', Addison-Wesley Longman Ltd, pp. 172–186.

Weibel, R. (1995*a*), 'Map Generalization in the Context of Digital Systems', *Cartography and Geographic Information Systems* **22**(4), 259–263.

Weibel, R. (1995*b*), Three essential building blocks for automated generalization, *in* J.-C. Müller, J. P. Lagrange & R. Weibel, eds, 'GIS And Generalisation: Methodology and Practice', Taylor & Francis, London, p. 56–69.

Weibel, R. (1996), A Typology of Constraints to Line Simplification, *in* M. J. Kraak & M. Molenaar, eds, 'Proceedings of 7th International Symposium on Spatial Data Handling', London: Taylor & Francis, pp. 533–546.

Weibel, R. & Dutton, G. (1998), Constraint-based automated map generalization, *in* 'Proceedings of the 8th International Symposium on Spatial Data Handling', Vancouver: IGU-Geographic Information Science Study Group, pp. 214–224.

Weibel, R. & Dutton, G. (1999), Generalising spatial data and dealing with multiple representations, *in* P. A. Longley, M. F. Goodchild, D. J. Maguire & D. W. Rhind, eds, 'Geographical Information Systems: Principles, Techniques, Management and Applications', 2 edn, John Wiley & Sons, New York, pp. 125–155.

Weibel, R., Keller, S. & Reichenbacher, T. (1995), Overcoming the knowledge acquisition bottleneck in map generalization: The role of interactive systems and computational intelligence, *in* A. U. Frank & W. Kuhn, eds, 'Spatial Information Theory A Theoretical Basis for GIS. COSIT 1995. Lecture Notes in Computer Science', Vol. 988, Springer, Berlin, Heidelberg, pp. 139–156.

Wertheimer, M. (1923), 'Untersuchungen zur Lehre von der Gestalt', *Psychologische Forschung* **4**, 301–350.

Wilson, I. D., Ware, J. & Ware, J. (2003), 'A Genetic Algorithm approach to cartographic map generalisation', *Computers in Industry* **52**(3), 291–304.

Winkler, J. (2023), 'Exploring the Swin Transformer Architecture for the Generalization of Building Footprints in Binary Cartographic Maps'. MSc Thesis, Dept. of Geography, University of Zurich.

Wu, S., Chen, Y., Schindler, K. & Hurni, L. (2023), Cross-attention Spatio-temporal Context Transformer for Semantic Segmentation of Historical Maps, *in* 'Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems', SIGSPATIAL '23, Association for Computing Machinery, New York, NY, USA, pp. 1–9.

Xia, X., Jiao, C. & Hurni, L. (2023), Contrastive Pretraining for Railway Detection: Unveiling Historical Maps with Transformers, *in* 'Proceedings of the 6th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery', GeoAI '23, Association for Computing Machinery, New York, NY, USA, p. 30–33.

Xiao, T., Ai, T., Yu, H., Yang, M. & Liu, P. (2024), 'A point selection method in map generalization using graph convolutional network model', *Cartography and Geographic Information Science* **51**(1), 20–40.

Xu, Y. & Goodacre, R. (2018), 'On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning', *Journal of Analysis and Testing* **2**(3), 249–262.

Yan, H., Weibel, R. & Yang, B. (2008), 'A Multi-parameter Approach to Automated Building Grouping and Generalization', *GeoInformatica* **12**(1), 73–89.

Yan, X., Ai, T., Yang, M. & Tong, X. (2021), 'Graph convolutional autoencoder model for the shape coding and cognition of buildings in maps', *International Journal of Geographical Information Science* **35**(3), 490–512.

Yan, X., Ai, T., Yang, M., Tong, X. & Liu, Q. (2020), 'A graph deep learning approach for urban building grouping', *Geocarto International* **37**(10), 2944–2966.

Yan, X., Ai, T., Yang, M. & Yin, H. (2019), 'A graph convolutional neural network for classification of building patterns using spatial vector data', *ISPRS Journal of Photogrammetry and Remote Sensing* **150**, 259–273.

Yan, X., Ai, T. & Zhang, X. (2017), 'Template Matching and Simplification Method for Building Features Based on Shape Cognition', *ISPRS International Journal of Geo-Information* **6**(8), 250.

Yan, X. & Yang, M. (2022), 'A Comparative Study of Various Deep Learning Approaches to Shape Encoding of Planar Geospatial Objects', *ISPRS International Journal of Geo-Information* **11**(10), 527.

Yan, X. & Yang, M. (2024), 'A deep learning approach for polyline and building simplification based on graph autoencoder with flexible constraints', *Cartography and Geographic Information Science* **51**(1), 79–96.

Yang, M., Yuan, T., Yan, X., Ai, T. & Jiang, C. (2022), 'A hybrid approach to building simplification with an evaluator from a backpropagation neural network', *International Journal of Geographical Information Science* **36**(2), 280–309.

Yang, S. & Berdine, G. (2017), 'The receiver operating characteristic (ROC) curve', *The Southwest Respiratory and Critical Care Chronicles* **5**(19), 34–36.

Ying, X. (2019), 'An Overview of Overfitting and its Solutions', *Journal of Physics: Conference Series* **1168**(2), 022022.

Yu, D., Hu, Y., Li, Y. & Zhao, L. (2024), 'PolygonGNN: Representation Learning for Polygonal Geometries with Heterogeneous Visibility Graph'. https://arxiv.org/abs/2407.00742.

Yu, W. & Chen, Y. (2022), 'Data-driven polyline simplification using a stacked autoencoder-based deep neural network', *Transactions in GIS* **26**(5), 2302–2325.

Zaheer, R. & Shaziya, H. (2019), A Study of the Optimization Algorithms in Deep Learning, *in* '2019 Third International Conference on Inventive Systems and Control (ICISC)', Coimbatore, India, pp. 536–539.

Zhang, C., Song, D., Huang, C., Swami, A. & Chawla, N. V. (2019), Heterogeneous Graph Neural Network, *in* 'Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining', KDD '19, Association for Computing Machinery, New York, NY, USA, pp. 793–803.

Zhang, J., Lu, C., Wang, J., Yue, X.-G., Lim, S.-J., Al-Makhadmeh, Z. & Tolba, A. (2020), 'Training Convolutional Neural Networks with Multi-Size Images and Triplet Loss for Remote Sensing Scene Classification', *Sensors* **20**(4), 1188.

Zhang, M.-L. & Zhou, Z.-H. (2014), 'A Review on Multi-Label Learning Algorithms', *IEEE Transactions on Knowledge and Data Engineering* **26**(8), 1819–1837.

Zhang, X., Ai, T. & Stoter, J. (2008), The Evalutation of Spatial Distribution Density in Map Generalization, *in* 'Proceedings of the XXI Congress: Silk Road for Information from Imagery (ISPRS 2008)', Vol. XXXVII, International Society for Photogrammetry and Remote Sensing, Beijing, China, pp. 181–187.

Zhang, X., Ai, T., Stoter, J. & Zhao, X. (2014), 'Data matching of building polygons at multiple map scales improved by contextual information and relaxation', *ISPRS Journal of Photogrammetry and Remote Sensing* **92**, 147–163.

Zhang, X., Touya, G. & Meijers, M. (2024), 'Automated Map Generalization: Emerging Techniques and New Trends (Editorial)', *Journal of Geovisualization and Spatial Analysis* **8**(1), 11.

Zhang, Z., Chen, D., Wang, Z., Li, H., Bai, L. & Hancock, E. R. (2019), 'Depth-based subgraph convolutional auto-encoder for network representation learning', *Pattern Recognition* **90**, 363–376.

Zhang, Z., Liu, Q. & Wang, Y. (2018), 'Road Extraction by Deep Residual U-Net', *IEEE Geoscience and Remote Sensing Letters* **15**(5), 749–753.

Zhao, R., Ai, T., Yu, W., He, Y. & Shen, Y. (2020), 'Recognition of building group patterns using graph convolutional network', *Cartography and Geographic Information Science* **47**(5), 400–417.

Zhao, T., Liu, Y., Neves, L., Woodford, O., Jiang, M. & Shah, N. (2021), 'Data Augmentation for Graph Neural Networks', *Proceedings of the AAAI Conference on Artificial Intelligence* **35**(12), 11015–11023.

Zheng, J., Gao, Z., Ma, J., Shen, J. & Zhang, K. (2021), 'Deep Graph Convolutional Networks for Accurate Automatic Road Network Selection', *ISPRS International Journal of Geo-Information* **10**(11), 768.

Zhou, Y., Yuan, C., Zeng, F., Qian, J. & Wu, C. (2018), An Object Detection Algorithm for Deep Learning Based on Batch Normalization, *in* M. Qiu, ed., 'Smart Computing and Communication. SmartCom 2017. Lecture Notes in Computer Science', Vol. 10699, Springer, Cham, pp. 438–448.

Zhou, Z., Fu, C. & Weibel, R. (2023), 'Move and remove: Multi-task learning for building simplification in vector maps with a graph convolutional neural network', *ISPRS Journal of Photogrammetry and Remote Sensing* **202**, 205–218.

Zhou, Z., Fu, C. & Weibel, R. (2024), 'SpaGAN: A spatially-aware generative adversarial network for building generalization in image maps', *International Journal of Applied Earth Observation and Geoinformation* **135**, 104236.

Zhu, J.-Y., Park, T., Isola, P. & Efros, A. A. (2017), 'Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks'. https://arxiv.org/abs/1703.10593.

Personal Declaration

I hereby declare that the submitted thesis is the result of my own, independent work. All external sources are explicitly acknowledged in the thesis.

In accordance with the recommendations on the use of generative artificial intelligence at UZH, I further declare that the following generative AI tools were used during the development of the work: ChatGPT[13] for brainstorming, rephrasing, and coding, Consensus[14] for sourcing literature, DeepL[15] for translation, and Writefull[16] for grammar corrections. Nevertheless, I assume full responsibility for the content of this thesis.

Zurich, 28.01.2025

Joris Senn

[13] chatgpt.com
[14] consensus.app
[15] deepl.com
[16] writefull.com