International Cartographic Association
Association Cartographique Internationale

PolygonTranslator - learning to simplify building footprints from one scale to another

Yu Feng [a,*], Ning Qian [b], Cheng Fu [c], Zhiyong Zhou [c], Monika Sester [b]

[a] *Chair of Cartography and Visual Analytics, Technical University of Munich, Germany, y.feng@tum.de*

[b] *Institute of Cartography and Geoinformatics, Leibniz University Hannover, Germany, ning.qian@stud.uni-hannover.de, monika.sester@ikg.uni-hannover.de*

[c] *Department of Geography, University of Zurich, Zurich, Switzerland, cheng.fu@geo.uzh.ch, zhiyong.zhou@geo.uzh.ch*

* Corresponding author

Keywords: map generalization, deep learning, transformer, simplification

Abstract:

Generalization of building maps is a classic task in cartography, which aims at generating smaller-scale representations from large-scale spatial data. Simplification is one of the most frequent operators that poses interesting challenges to be automated. In recent years, deep learning models have been introduced into the research of cartographic generalization. Instead of defining complicated rules and many thresholds, the learning-based approach is more flexible and can reduce the interplay when applying different operators, while exploiting the rich amount of training data from existing map series. Sester et al. (2018) proposed to apply conventional deep convolutional neural networks for image segmentation on rasterized building footprints to generalize building shapes. Feng et al. (2019) further experimented with different network architectures and showed that the models could achieve the expected generalization effects, however, they could not preserve linear boundaries and right angles very well. Courtial et al. (2022) applied similar models on road maps. Still, road connectivity is a major challenge for raster-based approaches.

Vector-based deep learning is thus more desirable. GCN (Graph neural networks) has already been studied a lot with purposes such as shape recognition and encoding (Yan et al., 2021). More recently, Zhou et al. (2022) have used the GCN to predict the actions needed at each building vertex, including the binary decision of remove, and the translation with an (x, y) offset. However, the experimental results showed that it often does not always present desired simplification actions as expected.

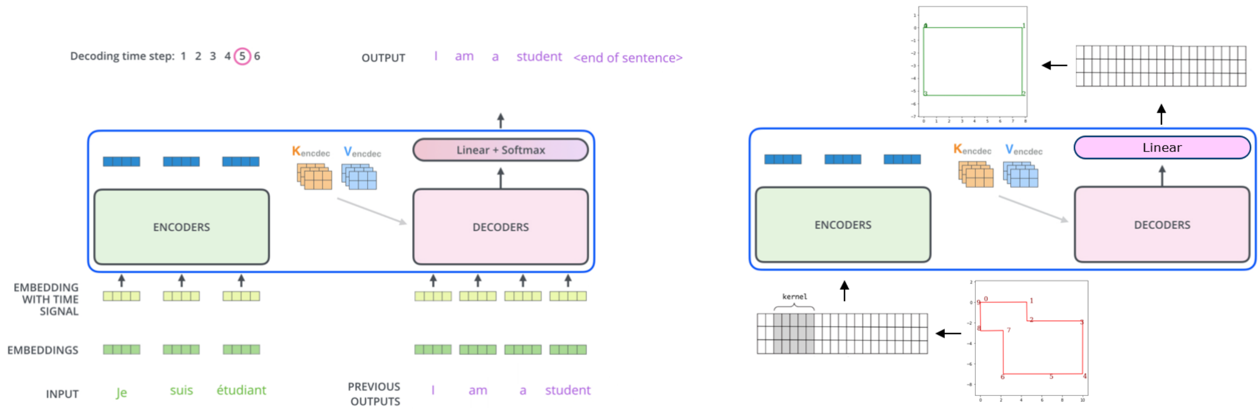

Figure 1. Illustration of the Transformer proposed in this work (right) in comparison with the network for sentence translation (left)[1].

Therefore, in this work, instead of using a graph to encode the polygon, we regarded the polygon as a data sequence and try to learn this simplification process with a transformer network. The transformer model is a neural network architecture that was initially designed for natural language processing (NLP) tasks. The self-attention mechanism has been applied, which processes the entire input at once, and each node receives the context information of the entire sequence. Therefore, it is ideally suited for GIS vector data. Simply borrowing these ideas, we aim to "translate" the building vectors from one map scale to another, e.g., from 1:10,000 to 1:25,000 in this work. The data from Feng et al. (2019) is used for learning this model.

In sentence translation, the semantics of the words can be captured by word embedding techniques. However, this does not exist naturally for the GIS vector data. Therefore, at the input stage of the network, 1D convolution with residual connections has been applied for the sequences of (x,y) coordinates. A second difference to the conventional transformer

1. The illustration is adapted based on the figure from https://jalammar.github.io/illustrated-transformer/ under CC BY-NC-SA 4.0.

Abstracts of the International Cartographic Association, 6, 60, 2023.
31st International Cartographic Conference (ICC 2023), 13–18 August 2023, Cape Town, South Africa.
https://doi.org/10.5194/ica-abs-6-60-2023

is that in the prediction step, an iterative process is deployed to predict the next word based on the previous words. However, very short, and not closed polygon segments worsen the result. Therefore, the encoded feature maps are directly decoded with the entire sequence as output. The model uses L1 loss to optimize the Euclidean distance between the target and predicted vertices.

A polygon can be represented with two strategies. The first is the *varying-length strategy*: A sequence of coordinates is used as input, and the distances between successive vertices can be arbitrarily long. The input sequences need to be zero-padded to the same number of points (i.e., 64 points) in order to be fed into the network. A one-hot encoded binary column is attached to the input sequence indicating the number of points, which needs to be learned with an extra binary cross-entropy term in the loss function (examples for strategy 1 are given in Figure 2 left red column). With this, the model also predicts the number of vertices of the output polygon in addition to the coordinates. The second strategy is the *fixed-length strategy*. It resamples the polygon of the different numbers of vertices into the same number of points, i.e., 64 points. Successive vertices have similar distances (examples in Figure 2 – right red column). Figure 2 presents several preliminary results of the two strategies, where the simplification outputs are compared qualitatively.

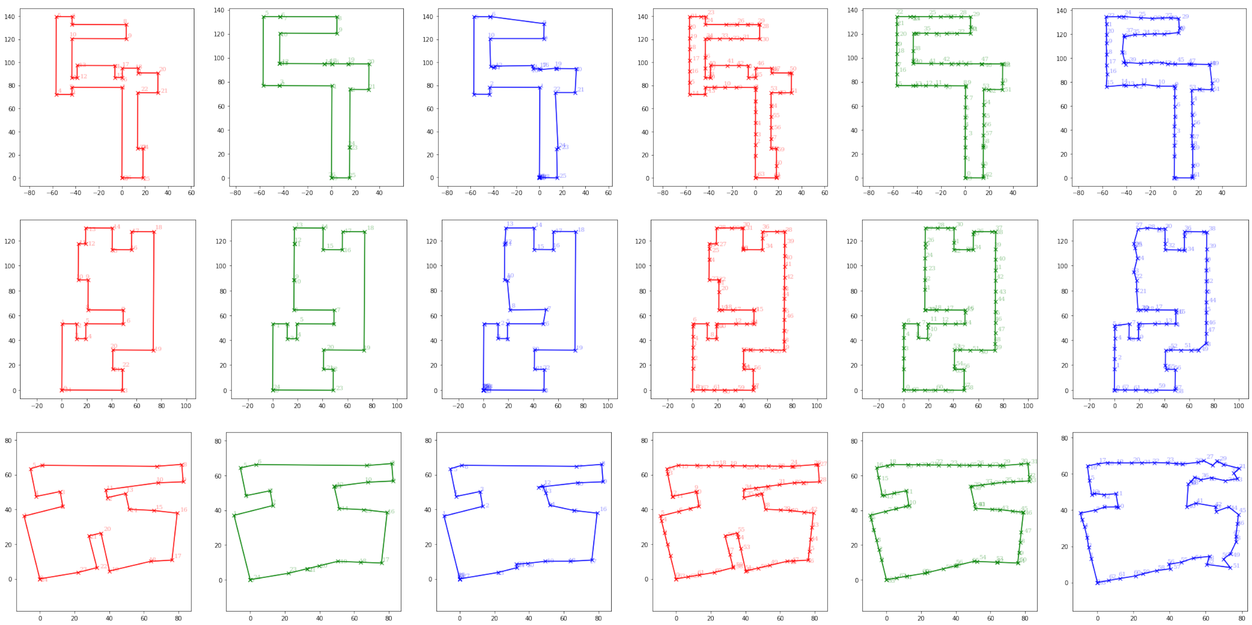

Figure 2. Qualitative comparison between models using the *varying-length strategy* (left 3 columns) and the *fixed-length strategy* (right 3 columns). The input polygons are in red, the target polygons are in green, and the predicted outputs are in blue. The vertex ids are placed on the right of each vertex.

The initial results show that both strategies are able to reconstruct simplified shapes as closed vectors. In general, the first strategy can provide more realistic building simplification results. However, the original positions of the vertices are often not precisely recovered for the cases using the first strategy. In this case, post-processing based on Least Squares Adjustment could be applied, which optimally adapts the simplified model to the original one (Sester, 2005). Nevertheless, it shows that the model is able to capture the simplification operation of the building shapes in most cases.

References

Courtial, A., Touya, G. and Zhang, X., 2022. Deriving map images of generalised mountain roads with generative adversarial networks. *International Journal of Geographical Information Science*, pp.1-30.

Feng, Y., Thiemann, F. and Sester, M., 2019. Learning cartographic building generalization with deep convolutional neural networks. *ISPRS International Journal of Geo-Information* 8(6), pp. 258.

Sester, M., Feng, Y. and Thiemann, F., 2018. Building generalization using deep learning. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-4 (2018)*, 42, pp.565-572.

Sester, M., 2005. Optimizing Approaches for Generalization and Data Abstraction, *International Journal of Geographic Information Science*, vol. 19 Nr. 8-9, pp. 871-897.

Yan, X., Ai, T., Yang, M. and 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), pp. 490–512.

Zhou, Z., Fu, C. and Weibel, R., 2022. Building simplification of vector maps using graph convolutional neural networks. *Abstracts of the ICA*, 5, p.86.

Abstracts of the International Cartographic Association, 6, 60, 2023.
31st International Cartographic Conference (ICC 2023), 13–18 August 2023, Cape Town, South Africa.
https://doi.org/10.5194/ica-abs-6-60-2023