A NOVEL DATA AUGMENTATION METHOD TO ENHANCE THE TRAINING DATASET FOR ROAD EXTRACTION FROM SWISS HISTORICAL MAPS

Chenjing Jiao*, Magnus Heitzler, Lorenz Hurni

Institute of Cartography and Geoinformation, ETH Zürich, Zürich, Switzerland - (cjiao, hmagnus, lhurni) @ethz.ch

KEY WORDS: Road extraction, Historical maps, Deep learning, Data augmentation, Cartography, Image transformations.

ABSTRACT:

Long-term retrospective road data are required for various analyses (e.g., investigation of urban sprawl, analysis of road network evolution). Yet, it is challenging to extract roads from scanned historical maps due to their dissatisfying quality. Although deep learning has been exerting its superiority in image segmentation, its application to road extraction from historical maps is rarely seen in existing studies. Deep learning usually requires quite large amounts of training data, which is time-consuming and tedious to label. Data augmentation can to some extent solve this issue. The existing data augmentation techniques vary each training sample as a whole (e.g., rotation, flipping). But some features or symbols on maps will never occur in practice when they are rotated or flipped (e.g., numbers, labels). To solve this problem and to further improve the diversity of training samples, we propose a novel data augmentation method, which varies the target features instead of the whole training sample. The method is validated by applying it to road extraction from the historical Swiss Siegfried map. The experiment results show the effectiveness of the proposed method.

1. INTRODUCTION

Historical maps contain valuable retrospective spatial information that can be rarely found elsewhere. Many historical map series have been scanned into raster format and made widely accessible (Tsorlini et al., 2014). Long-term road network data are used to analyze the evolution of the road networks (Strano et al., 2012; Zhao et al., 2015) and to realistically reconstruct streetscapes of the past for education, entertainment and research purposes¹. The wide applications of road data and the image processing challenges due to the poor quality of historical maps (e.g., bleaching, paper distortion, blurring) (Leyk et al., 2005) induce an urgent demand for efficient methods to extract roads from historical maps.

Recently, deep learning has become a research hotspot and has been utilized widely owing to its generalisability. Specifically for image processing tasks, convolutional neural networks (CNN) have become the default choice. Convolutional layers in deep learning architectures take input image (patches) of any size and operate on local input regions based on relative spatial coordinates, unlike fully connected networks, which have fixed dimensions and do not explicitly exploit the spatial characteristics. Thus, Long et al. (2015) propose Fully Convolutional Network (FCN) by converting conventional fully connected layers to convolutional layers and supplementing the convolutional layers by successive deconvolution layers for upsampling. Apart from this, skip connections are added to combine finer scale predictions and coarser ones. The spatial informative output of FCNs make them a natural choice for endto-end dense prediction tasks like image segmentation (Buslaev et al., 2018). An improvement to the original FCN has been introduced by Ronneberger et al. (2015) in the form of the U-Net architecture. Compared with the original FCN, one important modification in U-Net is that the upsampling part also has a large number of feature channels corresponding to the downsampling part, which allows the network to propagate context information to higher resolution layers. Consequently, the upsampling part is more or less symmetric to the downsampling part. The downsampling steps gradually generate increasingly abstract feature maps of the input image, while the upsampling steps progressively reobtain the dimensions of the input and enable precise localization (Ronneberger et al., 2015). Concatenation operations are used to copy the feature maps of an intermediate step in the downsampling path to the corresponding step in the upsampling path, which empowers the network to combine low-level and high-level feature representations.

Saeedimoghaddam and Stepinski (2020) employ deep CNNs for road intersection extraction from USGS historical maps. Although with this method, road intersections represented as both single lines and double lines can be successfully extracted, road branches cannot be extracted, which are essential to the analysis of road network growth and urban sprawl (Masucci et al., 2014). Chiang et al. (2020) report a set of experiments for railroad extraction from USGS historical maps to investigate the impact of deep CNN architectures on feature extraction accuracy. Despite of the rapid development and the superiority in image segmentation and feature recognition of deep CNNs, their application to road extraction from historical maps is to some extent limited up to now (Jiao et al., 2021).

Unlike the easy availability of historical maps, it is timeconsuming and laborious to manually label the corresponding training data. However, deep learning usually requires large amounts of training data. One solution to this issue is data augmentation, which can be used to enhance the size and quality of training datasets so that performant machine learning models can be trained (Shorten and Khoshgoftaar, 2019). Conventional data augmentation methods are applied on the image patch level, which, for example, flip or rotate the image patch as a whole. This study proposes to use data augmentation on the feature level by rotating or flipping the target features only. It not only avoids the generation of possibly unrealistic training

^{*} Corresponding author

¹ https://ai.googleblog.com/2020/10/recreating-historicalstreetscapes.html

data resulting from rotating or flipping some map features (e.g., labels, numbers, triangulation points), but also improves the diversity of training samples, thereby empowering the deep learning network to learn invariant representations unique to target features (e.g., roads). The effectiveness of the novel data augmentation method is verified by applying it to road extraction from the Swiss Siegfried map.

2. DATA AND METHOD

2.1 Data

The Swiss Siegfried map is a comprehensive Swiss national map series published between 1872 and 1949 at the scales of 1:25,000 (Jura and Swiss plateau) and 1:50,000 (Alps) (Heitzler and Hurni, 2020; Jiao et al., 2020). The map series depicts various geographical features such as buildings, roads, railways, hydrological features, vegetation areas. The Siegfried map sheets are scanned into raster format by Swiss Federal Office of Topography, and georeferenced based on the map frame corner points and the coordinate grid lines (Heitzler et al., 2018). The size of each scanned map sheet is 7,000 pixels × 4,800 pixels. The resolution of map sheets used in this study is 1.25 m/pixel with a scale of 1:25,000. The map sheet has three color channels, namely RGB.

Roads are represented by six different symbols, namely single dashed line, single solid line, the combination of a solid line and a parallel dashed line, two parallel lines, a thin line together with a thicker line, and two parallel lines with short strokes in between, as marked by red arrows shown in Figure 1. The symbols correspond to different road grades.



Figure 1. Roads on Siegfried map. Geodata © Swisstopo

The labelled road data we have at hand only covers Zurich city. The red lines in Figure 2 show the labelled data, which are road centerlines. Figure 2(a) is an overview of the data overlaying the corresponding Siegfried map sheets, (b) a part of the data, and (c) buffers of roads in (b), as shown by white areas. The buffers are generated based on road width. For example, the width of roads represented by single solid lines and dashed lines is usually four meters, so the buffer size is two meters.

2.2 Sampling strategy

To get training samples from the input map sheet and to avoid the data imbalance issue, we adopt the following sampling strategy. First, "positive" points that are located close to roads are randomly generated in road buffers. "Negative" points that are located far from roads are also randomly generated. The positive points are randomly shifted by a small displacement within a neighbourhood of 13 pixels \times 13 pixels. Image tiles centered at these sampling points are cropped from the map sheet, which are sized 128 pixels \times 128 pixels. The positive points are shifted as roads will not always go through the center point of an image tile. The green dots in Figure 3 show the positive points and red dots the negative ones. The green embossed rectangle represents the map tile cropped centered at one positive point, and the red embossed rectangle the tile centered at one negative point. With this strategy we obtain sampling tiles with roads and without roads, so that the network can learn features of both road areas and non-road areas. In this study, the ratio of the positive samples to the negative samples is empirically set as about 5:1. Additionally, this sampling strategy allows for flexibly adding sampling points for a certain feature (road class in this use case). For example, if we see from the results that a certain road class is not well extracted, sampling points can be added specially for this road class.

2.3 A novel data augmentation method

Data augmentation is a data-space solution to the problem of over-fitting as well as limited training data, which are common issues in many applications of deep CNNs (Sun et al., 2019). It encompasses a suite of image transformations, such as scaling, rotation, flipping, color variation, noise injection, etc. (Shorten and Khoshgoftaar, 2019). Data augmentation enforces the network to learn and identify the desired invariance of feature representations. Specifically for our use case, the learned feature representations of roads should be invariant to variations in the map tiles that are irrelevant for the segmentation task (Dosovitskiy et al., 2014).

Color and scaling features are essential to road segmentation, so they should not be varied in data augmentation, and there is already much noise in the scanned Siegfried maps. Thus, we use the other two image transformations, namely rotation and flipping. Most of previous data augmentation methods rotate or flip the whole image or image patch. Siegfried maps, however, contain several features that only occur in certain ways. For example, numbers and triangulation points should not be rotated or flipped. Labels can only be rotated slightly and cannot be flipped, as large degree rotation (e.g., larger than 90°) and flipping are not character-preserving transformations (Shorten and Khoshgoftaar, 2019). Therefore, we rotate and flip only the road features, as we have road buffers as ground truth. Specifically, roads are first extracted from the original image patch based on the ground truth. They are randomly rotated or flipped. The remaining features on the patch are not rotated or flipped. The original road areas on the patch are replaced by pixels with the background color of the Siegfried map. Then, the rotated or flipped roads are overlaid on the patch, which produces an "augmented" patch. The ground truth is also rotated or flipped accordingly. Figure 4 shows two examples, where (a) and (c) respectively present the original map tiles cropped from Siegfried map, while (b) displays the result of rotating the roads in (a) by 270° anti-clockwise, and (d) vertical-axis flipping the roads in (c). Especially, the label in (d) is not transformed. These



Figure 2. The road training data and its buffers. Geodata © Swisstopo



Figure 3. Sampling points and the corresponding map tiles. Geodata © Swisstopo

road-only transformations empower the network to learn the features unique to roads, such as long slenderness, color, connectivity, topology, etc. Furthermore, the road-only rotation and flipping change the relative spatial relation between roads and non-roads, thereby adding more diversity to the training data than previous whole-image (patch) transformations.

2.4 Road extraction with U-Net

The road segmentation model in this study is developed based on a U-Net architecture. Specifically, the following parameters apply: The first convolution layer of the U-Net in this study has 16 channels. The bottleneck has 256 channels, as each downsampling step doubles the number of channels. The U-Net consists of four downsampling steps and four upsampling steps. Moreover, dropout is used at each downsampling and upsampling step with increasing dropout rates towards the bottleneck. It significantly prevents overfitting by avoiding the units co-adapting too much as well as enables to train and combine many different network architectures by randomly sampling a "thinned" network consisting of all the units that survive dropout (Srivastava et al., 2014; Jenny et al., 2020). The network is shown in Figure 5. For computing the probability of the



Figure 4. (a) a map tile cropped from Siegfried map, (b) rotating the roads in (a) by 270° anti-clockwise, (c) another map tile, (d) horizontal flipping the roads in (c). Geodata © Swisstopo

produced feature vector being road, a 1x1 convolution together with a sigmoid operation (Han and Moraga, 1995) is applied, as shown with a blue rectangle. The target prediction area is sized 64×64 pixels, as shown by a yellow embossed square. To enable the model to make precise predictions around the border of the target area, the input map tile is expanded by 32 pixels on each side. Thus, the input tile is 128×128 pixels. Furthermore, the weights in the filters are initialized with the method proposed by He et al. (2015), which helps with convergence of very deep networks trained directly from scratch. The sampled map tiles and their corresponding road buffers as ground truth are fed into the U-Net for training.



Figure 5. The U-Net architecture. Geodata © Swisstopo

3. EXPERIMENT

3.1 Training scenarios

We use Keras library to implement the experiment. We use Adam optimizer and initialize the learning rate as 0.001 (Kingma and Ba, 2014). Dice loss is used as the loss function (Dice, 1945; Milletari et al., 2016). Each model is trained with 100 epochs. The batch size is 64. To verify the effectiveness of the novel data augmentation method and the flexibility of improving results by adding samples of a certain road class, we implement three training scenarios, namely 1) training with 5000 original samples cropped from the Siegfried map, 2) training with 5000 original samples and additional 1400 samples produced with the novel data augmentation method, and 3) training with 5000 original samples, 1400 samples produced with the novel data augmentation method as well as 500 samples explicitly cropped from features of road class 1. As it is found that road class 1 is less well extracted than other classes, the latter case has been added to specifically improve the extraction capabilities of the model for road class 1, which is represented by dashed line.

3.2 Postprocessing

The trained models are applied to Siegfried map sheets that cover other areas than Zurich city. The pixel values in the raster prediction results indicate the probability of the pixel being a road, as shown with the white areas in Figure 6. Pixels with the probability greater than 0.5 are taken as roads. Subsequently, morphological operations are adopted to skeletonize the road areas. Then, the skeletons are vectorised and simplified as road centerlines by the "raster to polyline" tool in ArcGIS. Road extraction results from three typical areas, namely urban area, suburban area and rural area, are reported in Figure 7. Red lines represent road centerlines, which overlay the corresponding map images. The overlaid images are shown with 50% transparency to highlight the centerlines.

3.3 Evaluation

As shown in Figure 6, raster predictions with conventional and novel data augmentation have much fewer false positives, especially around streams and forest borders, which have very similar shape with roads. Predictions of novel data augmentation are more robust than those of conventional method, especially for double-line roads. In the rural area, the highly curved footpath is extracted with better continuity with the novel method than the conventional method. In addition, we use accuracy and F1 score for quantitatively evaluating raster road predictions, while correctness and completeness for vector road centerlines (Wegner et al., 2013). For raster predictions, the metric values are calculated with the number of correctly or wrongly predicted pixels, namely true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). To compute correctness and completeness, three-meter buffers are generated for the predicted centerlines and the ground truth. Correctness and completeness values are computed with the lengths of TP, TN, FP, FN.

Four Siegfried map sheets, including one covering an area in the northern part of Switzerland, which contains a city and hills (sheet number TA_017_1940), one covering rural area in the Swiss plateau (sheet number TA_199_1941), one covering rural area in the periphery of the Swiss Alps (sheet number TA 385 1941), and the last covering small cities/towns in the eastern part of Switzerland (sheet number TA 219 1944), are selected for evaluating and comparing results obtained by the models trained without data augmentation vs. with novel data augmentation vs. with both novel data augmentation and additional samples around road class 1. The average metrics values of the four sheets are reported in Table 1, Table 2 and Table 3, respectively. Accuracy, F1 and correctness obtained with novel data augmentation outperform those without data augmentation. Especially, correctness is largely improved because of the false positives reduced by applying the novel data augmentation method. Adding additional samples of road class 1 further improves the results, as shown in Table 3.

Metrics	Values
Accuracy	96.83%
F1 score	77.37%
Completeness	95.68%
Correctness	64.37%

 Table 1. Quantitative road extraction results (5000 original samples without data augmentation).

4. CONCLUSION

In recent years, deep learning techniques open an avenue to solve the challenge of extracting roads from historical maps. As a solution to the problem of limited training data, data augmentation is commonly used in deep learning applications. To

	Town	Rural area	
Original map image		Hohltaube /Stahl	
Raster predictions without data augmentation			
Raster predictions with conventional data augmentation			
Raster predictions with novel data augmentation			
Ground truth			

Figure 6. The comparison between raster road predictions obtained without data augmentation vs. with conventional data augmentation. Geodata © Swisstopo

Metrics	Values
Accuracy	97.20%
F1 score	79.63%
Completeness	95.24%
Correctness	66.66%

Table 2. Quantitative road extraction results (5000 original samples plus 1,400 augmented samples of novel data augmentation).

address the fallaciousness due to data augmentation applied to a whole image patch and to enhance the diversity of training samples, this study proposes a novel data augmentation method, which varies the target features instead of the whole image

Metrics	Values
Accuracy	97.76%
F1 score	83.05%
Completeness	96.65%
Correctness	67.35%

Table 3. Quantitative road extraction results (5000 original samples with additional 500 samples around road class 1, plus 1,400 augmented samples of novel data augmentation).

patch. The experiment results show the effectiveness of the proposed method. Especially, the method is very useful to reduce false positives. Although in this study we exemplarily apply the method to road extraction from historical maps, it can be

	Urban area	Suburban area	Rural area
Original map image			
Road centerlines			
Ground truth			

Figure 7. Vector road centerlines extracted by the model trained with novel data augmentation. Geodata © Swisstopo

generalizable to other features and data sources. Possible improvements and future work is to explore the optimal ratio of the augmented samples to the original ones.

ACKNOWLEDGEMENTS

We thank Ms. Monika Niederhuber for providing us the road training data of Zurich city as shown in Figure 2(a).

References

- Buslaev, A., Seferbekov, S., Iglovikov, V., Shvets, A., 2018. Fully convolutional network for automatic road extraction from satellite imagery. *Proceedings of the IEEE conference* on computer vision and pattern recognition workshops, 207– 210.
- Chiang, Y.-Y., Duan, W., Leyk, S., Uhl, J. H., Knoblock, C. A. et al., 2020. Using historical maps in scientific studies. Springer,.
- Dice, L. R., 1945. Measures of the amount of ecologic association between species. *Ecology*, 26(3), 297–302.
- Dosovitskiy, A., Springenberg, J. T., Riedmiller, M., Brox, T., 2014. Discriminative unsupervised feature learning with convolutional neural networks. *Advances in neural information processing systems*, 27, 766–774.
- Han, J., Moraga, C., 1995. The influence of the sigmoid function parameters on the speed of backpropagation learning. *International workshop on artificial neural networks*, Springer, 195–201.
- He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *Proceedings of the IEEE international conference on computer vision*, 1026–1034.

- Heitzler, M., Gkonos, C., Tsorlini, A., Hurni, L., 2018. A modular process to improve the georeferencing of the Siegfried map. *e-Perimetron*, 13(2), 96–100.
- Heitzler, M., Hurni, L., 2020. Cartographic reconstruction of building footprints from historical maps: A study on the Swiss Siegfried map. *Transactions in GIS*, 24(2), 442–461.
- Jenny, B., Heitzler, M., Singh, D., Farmakis-Serebryakova, M., Liu, J. C., Hurni, L., 2020. Cartographic relief shading with neural networks. *IEEE Transactions on Visualization and Computer Graphics*, 27(2), 1225–1235.
- Jiao, C., Heitzler, M., Hurni, L., 2020. Extracting wetlands from swiss historical maps with convolutionalneural networks. Automatic Vectorisation of Historical Maps. International workshop organized by the ICA Commission on Cartographic Heritage into the Digital 13 March, 2020 Budapest. Proceedings, Department of Cartography and Geoinformatics, ELTE Eötvös Loránd University, 33–38.
- Jiao, C., Heitzler, M., Hurni, L., 2021. A survey of road feature extraction methods from raster maps. *Transactions in GIS*.
- Kingma, D. P., Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Leyk, S., Boesch, R., Weibel, R., 2005. A conceptual framework for uncertainty investigation in map-based land cover change modelling. *Transactions in GIS*, 9(3), 291–322.
- Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3431–3440.
- Masucci, A. P., Stanilov, K., Batty, M., 2014. Exploring the evolution of London's street network in the information space: A dual approach. *Physical Review E*, 89(1), 012805.

- Milletari, F., Navab, N., Ahmadi, S.-A., 2016. V-net: Fully convolutional neural networks for volumetric medical image segmentation. 2016 fourth international conference on 3D vision (3DV), IEEE, 565–571.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, Springer, 234–241.
- Saeedimoghaddam, M., Stepinski, T. F., 2020. Automatic extraction of road intersection points from USGS historical map series using deep convolutional neural networks. *International Journal of Geographical Information Science*, 34(5), 947–968.
- Shorten, C., Khoshgoftaar, T. M., 2019. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 1–48.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1), 1929–1958.
- Strano, E., Nicosia, V., Latora, V., Porta, S., Barthélemy, M., 2012. Elementary processes governing the evolution of road networks. *Scientific reports*, 2(1), 1–8.
- Sun, T., Di, Z., Che, P., Liu, C., Wang, Y., 2019. Leveraging crowdsourced gps data for road extraction from aerial imagery. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 7509–7518.
- Tsorlini, A., Iosifescu, I., Iosifescu, C., Hurni, L., 2014. A methodological framework for analyzing digitally historical maps using data from different sources through an online interactive platform. *e-Perimetron*, 9(4), 153–165.
- Wegner, J. D., Montoya-Zegarra, J. A., Schindler, K., 2013. A higher-order crf model for road network extraction. *Proceed*ings of the IEEE Conference on Computer Vision and Pattern Recognition, 1698–1705.
- Zhao, P., Jia, T., Qin, K., Shan, J., Jiao, C., 2015. Statistical analysis on the evolution of OpenStreetMap road networks in Beijing. *Physica A: Statistical Mechanics and its Applications*, 420, 59–72.