International Cartographic Association
Association Cartographique Internationale

# Constraint Based Evaluation of Generalized Images Generated by Deep Learning.

**Azelle Courtial [a]*, Guillaume Touya [a], Xiang Zhang [b],**

[a] *LASTIG, Univ Gustave Eiffel, ENSG, IGN, F-94160 Saint-Mandé, France, azelle.courtial@ign.fr, guillaume.touya@ign.fr,*
[b] *School of Resource and Environmental Sciences, Wuhan University, Wuhan 430079, China, xiang.zhang@whu.edu.cn*

* Corresponding Author

**Abstract:**
The use of deep learning techniques for map generalisation raises new problems regarding the evaluation of the results: (1) images are used as input/output instead of vector data; (2) the deep learning processes do not guarantee results that follow cartographic principles; (3) the deep learning models are black boxes that hide the causal mechanisms. Also, deep learning intern evaluation is mostly based on the realism of the images and the pixel classification accuracy, and none of these criteria is sufficient to evaluate a generalisation process. In this article, we propose an adaptation of the constraint-based evaluation to the images generated by deep learning. Six raster-based constraints are proposed for a mountain road generalisation use case.

**Keywords:** Generalization, Evaluation, Deep learning, Raster, Constraints

## 1. Introduction

Map generalisation recently gave in to the trendy deep learning techniques. These techniques seek to generate images of generalised maps from images of ungeneralised maps, using examples of maps already generalised to train the model. (Feng et al., 2019) used a supervised learning network to learn the segmentation of buildings generalised shape. Then, a similar approach was proposed for mountain road generalisation (Courtial et al., 2020). Otherwise, an approach by image generation seems to be able to give promising result for map generalisation, as shown by the experiments to generate a map from aerial images (Isola et al., 2017) and by the experiments about style transfer on maps (Kang et al., 2019). Among the challenges raised by these new techniques, the evaluation is particularly important (Touya et al., 2019). The current techniques to evaluate map generalisation (Mackaness and Ruas, 2007, Stoter et al., 2014) do not really apply to the output image of a deep learning generalisation model: (1) images are used as input/output instead of vectors, and even if the conversion to vector before evaluation is possible it is not adapted for an evaluation that aims to guide the learning process; (2) the deep learning processes do not guarantee results that follow cartographic principles, so additional realism constraint are needed; (3) the deep learning models are black boxes that hide the causal mechanisms, which makes the evaluation moreover interesting to identify the weakness of each network. Also, deep learning evaluation is mostly based on the realism of the images and the pixel classification accuracy, and none of these criteria are sufficient to evaluate a generalisation process in practice.

## 2. Use case

In this short paper, we propose an adaptation of the constraint-based evaluation to the images generated by deep learning, with a use case on images of generalized mountain roads for a display at the 1:250,000 scale. We implemented all our constraints in Python. We based our experiments on images of roads from the Alps extracted from IGN (the French national mapping agency) maps at 1:250,000 (reference of generalization) and 1:25,000 (input) scale. Our images (input and output) represent 2.5 km2 in 256*256 pixels, so the resolution is around 10 metres. The width of the road symbols on these images corresponds to their importance and varies between 1 and 6 pixels. We evaluate the images generated by three different deep learning models, i.e., CycleGAN (Zhu et al., 2017), Pix2Pix (Isola et al., 2017), and U-net (Courtial et al., 2020). Except for the colour constraint that uses the real prediction output (red roads on a white background), constraints are all adapted to square images post-processed in black (background) and white (roads).

## 3. Workflow

We draw our inspiration from the traditional process of automatic evaluation that aims to provide a measure that is both understandable by a human and a computer (Mackaness and Ruas, 2007). We try to adapt classical constraints to guarantee the realism of the image, the preservation of the road initial characteristics, and, of course, the legibility of the mountain roads. But the challenge is to compute the constraints satisfaction in raster mode. We also try to keep our constraints balanced between these tree objectives in order to ensure the quality of the evaluation (Zhang, 2012). As we only deal with pixels and

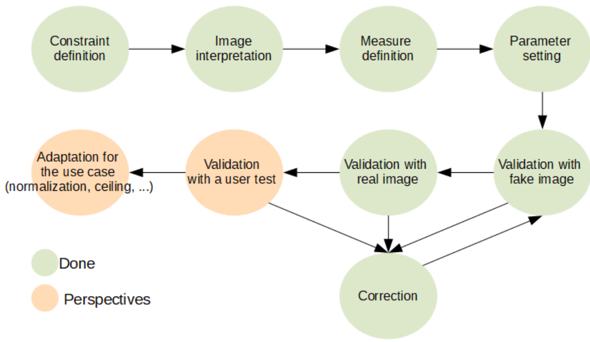


Figure 1. Workflow for constraint definition.

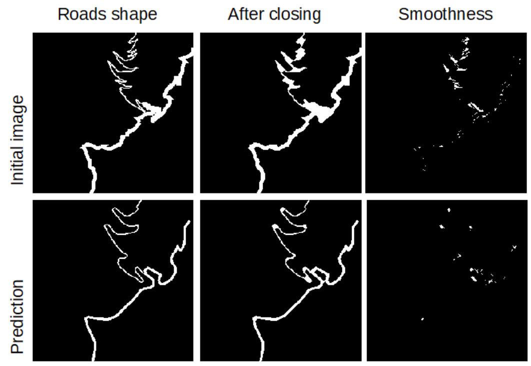


Figure 2. Effect of the closing on a road and its generalisation.

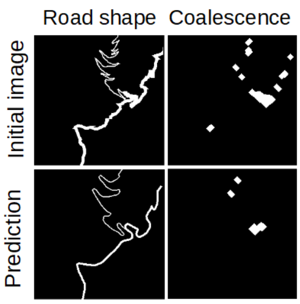


Figure 3. Effect of the coalescence measure, on a road and its generalisation.

images (eventually split in several sub-images) the level of the evaluation cannot be managed as for classical evaluation in term of micro, meso or macro objects, but there still are three levels: the evaluation is at the pixel level, at the level of a set of homogeneous pixels, or more generally, at the image level. We applied the method in Figure 1 in order to create and validate our constraints. After the constraint definition, we search for identifying the associated characteristics on the image and a way to measure them. Then, we tested the measures on real images in order to determine the optimal value of parameters. For the validation, we compare the evaluation value with our perceptual evaluation for some real and fake images, but we have not yet compared it with the evaluation from a user test.

## 4. Constraints

In the following paragraphs, we detail a proposition of six constraints. There are two legibility constraints: first, the generalized roads must be smoother than the initial one; the second legibility constraint deals with the coalescence problems frequently encountered with mountain roads. Then, we want the information to be preserved to a certain extent, especially the initial position and the structure of the road network. Finally, we introduce some additional constraints to avoid some artifacts we observed during our experiments with deep learning: noisy colours and shapes.

### 4.1 Smoothness

We propose to measure the smoothness of produced roads using a closing operation (dilation followed by an erosion) in mathematical morphology. It permits to fill in the concave irregularities of the road shape (second column of the Figure 2). Then, the deletion of the shape of the road only lets the number of necessary pixels to make the road smooth (white pixels on the third column of Figure 1). The size of the closing determines the size of the biggest irregularity we try to smooth; if we choose a threshold too large, concave portions that are not irregularities might be wrongly filled. This measure seems to work the best at the scale of our use case with a closing value of three pixels (30 meters). As the number of not smooth pixels is not easy to interpret (correlated with the number of roads in the image), we compute the ratio of the initial roughness that is removed by generalisation instead.

### 4.2 Coalescence

We propose to measure the coalescence of road symbols using the following mathematical morphology operations: first a dilation of roads of $n$ pixels, then an erosion of size $n + 6$ (the maximal width of roads in our images), and finally a dilation of 6 pixels. This process is illustrated on Figure 3. The size $n$ corresponds to the maximal distance between two coalescent parts of roads from map specifications. For our use case, a threshold of five pixels gives the best measure of coalescence.

### 4.3 Position accuracy

To measure position accuracy, the intersection between the road pixels in the initial and in the generalised images is computed. This is a classical measure to assess image segmentation and determine which parts of the roads pixels are common with the initial image. As displacement is tolerated, and sometimes necessary, we can use a buffer of size $n$ around roads instead of just road pixels. This threshold can be very large (around 20 pixels) because the displacement offset can be large for mountain roads.

### 4.4 Continuity

We based the measure of continuity (or connectivity) preservation on the number of sets of contiguous road pixels and contiguous background pixels. Figure 4 illustrate how these numbers can reveal changes in the structure.

The equation 1 shows how we combine these numbers of


Figure 4. Number and roads and background pixel sets in some cases of continuity loss.

contiguous parts to measure continuity preservation:

$$\frac{nroad_{init} - nroad_{image} + nback_{init} - nback_{image}}{nroad_{init} + nback_{init}} \tag{1}$$

Where $nroad$ is the number of road sets in the image and $nback$ the number of sets of background pixels. This measure has two problems: first all changes are considered equal, even the addition of very small sets. To solve the first problem, we decided to count as a change only the appearance or the disappearance of sets larger than two pixels. Then, there is a border problem (illustrated in the last column of Figure 3), when a road has bends at the border of an image, these bends can be connected or disconnected by a small displacement, without a real change in the structure.

### 4.5 Color

We want to make sure that the generated images do not convey any unexpected information. In a map, a different colour may convey a different object. So we propose to measure if there is colour noise: pixel that are not roads (red) or background (white). We have decided to use an implementation of the CIEDE2000 distance between colours (Sharma et al., 2004), because of its consistency with human colour perception. Then, we counted the quantity of pixels that are visibly too distant from red or white ($d > 9$).

### 4.6 Noise

We also found that some deep learning models can generate some noise: isolated roads pixels that are not roads. We counted the number of sets of contiguous road pixels that are too small to be real roads (below 6 pixels).

## 5. How to use these constraints?

Depending on the objectives of the evaluation, the combination and analysis of these values will be different. For example, when the aim is to compare different deep learning models, a discretisation would be useful to derive constraint levels of satisfaction (e.g. very good, good, medium, bad, very bad); then it is easier to merge the constraint satisfaction values into one or several synthetic values. However, if we want to identify the weaknesses of a model, or detect if specific post-processes are required (removing colour noise for instance), the discretisation can be used but merging constraints is not necessary. Then, the evaluation can also be used to control a model: these constraints can be normalized and weighted to construct a loss function that controls how the deep learning internally assesses its results after each iteration. Finally, the evaluation can help to build a complete map from the several image tiles used in the deep learning models: as tiles cover similar portions of space, the constraints could be used to decide in which image the generalised output is the best one.

## 6. Conclusion

To conclude, we think that this work shows the applicability of constraint-based evaluation on images generated by deep learning. The next step would be to experiment the use of these constraint in different evaluation scenarios (comparison, triggering post-processes, monitoring the internal iterations…), with a particular focus on a loss function that is specific to map generalisation. We also plan to couple this constraint-based quantitative evaluation with a user survey to make sure that these measures do reflect the human perception of a "good map".

## References

Courtial, A., El Ayedi, A., Touya, G. and Zhang, X., 2020. Exploring the Potential of Deep Learning Segmentation for Mountain Roads Generalisation. *ISPRS International Journal of Geo-Information* 9(5), pp. 338. Number: 5 Publisher: Multidisciplinary Digital Publishing Institute.

Feng, Y., Thiemann, F. and Sester, M., 2019. Learning cartographic building generalization with deep convolutional neural networks. *International Journal of Geo-Information*.

Isola, P., Zhu, J.-Y., Zhou, T. and Efros, A. A., 2017. Image-to-image Translation with Conditional Adversarial Networks.

Kang, Y., Gao, S. and Roth, R., 2019. Transferring Multiscale Map Style Using Generative Adversarial Network.

Mackaness, W. A. and Ruas, A., 2007. Chapter 5 - Evaluation in the Map Generalisation Process. In: W. A. Mackaness, A. Ruas and L. T. Sarjakoski (eds), *Generalisation of Geographic Information*, International Cartographic Association, Elsevier Science B.V., Amsterdam, pp. 89–111.

Stoter, J., Zhang, X., Stigmar, H. and Harrie, L., 2014. Evaluation in Generalisation. In: D. Burghardt, C. Duchêne and W. Mackaness (eds), *Abstracting Geographic Information in a Data Rich World*, Springer International Publishing, Cham, pp. 259–297. Series Title: Lecture Notes in Geoinformation and Cartography.

Touya, G., Zhang, X. and Lokhat, I., 2019. Is deep learning the new agent for map generalization? *International Journal of Cartography* pp. 142–157.

Zhang, X., 2012. Automated evaluation of generalized topographic maps. PhD thesis.

Zhu, J.-Y., Park, T., Isola, P. and Efros, A. A., 2017. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. In: *Computer Vision (ICCV), 2017 IEEE International Conference on*.