CONSTRAINT-BASED EVALUATION OF MAP IMAGES GENERALIZED BY DEEP LEARNING

This version of the article has been accepted for publication, after peer review but is not the Version of Record. The Version of Record is available online at: https://doi.org/10.1007/s41651-022-00104-2

Azelle Courtial
LASTIG
University Gustave Eiffel, ENSG, IGN,
F-94160 Saint-Mande, France
azelle.courtial@ign.fr

Guillaume Touya
LASTIG
University Gustave Eiffel, ENSG, IGN,
F-94160 Saint-Mande, France
guillaume.touya@ign.fr

Xiang Zhang
School of Geospatial Engineering and Sciences,
Sun Yat-Sen University,
Guangzhou 510275, China

May 17, 2022

ABSTRACT

Deep learning techniques have recently been experimented for map generalization. Although promising, these experiments raise new problems regarding the evaluation of the output images. Traditional map generalization evaluation cannot directly be applied to the results in a raster format. Additionally, the internal evaluation used by deep learning models is mostly based on the realism of images and the accuracy of pixels, and none of these criteria is sufficient to evaluate a generalization process. Finally, deep learning processes tend to hide the causal mechanisms and do not always guarantee a result that follows cartographic principles. In this article, we propose a method to adapt constraint-based evaluation to the images generated by deep learning models. We focus on the use case of mountain road generalization, and detail seven raster-based constraints, namely, clutter, coalescence reduction, smoothness, position preservation, road connectivity preservation, noise absence, and color realism constraints. These constraints can contribute to current studies on deep learning-based map generalization, as they can help guide the learning process, compare different models, validate these models, and identify remaining problems in the output images. They can also be used to assess the quality of training examples.

Keywords Generalization · Evaluation · Constraints · Raster · Deep learning

1 Introduction

Deep learning is now widely used and is successfully applied in various tasks in computer vision and natural language processing. The potential of this approach for map generalization has recently been demonstrated [1, 2, 3].However, to bring the deep learning approach to map generalization to the next level, the question of how good the generalization results are must be addressed, whether it is during or after the generalization process.

Automated evaluation has multiple roles in the map generalization process due to the iterative nature of generalization [4]: (i) to evaluate the results while tuning the process and its parameters, (ii) to control the iterations of the process to make sure the map is improved by the generalization algorithms, (iii) to assess the quality of the final output map. For all these different types of evaluation, a balance between the legibility of the map and the preservation of important structures (e.g., topology and spatial distribution) is targeted [5, 6, 7], while respecting the fitness for use of the map

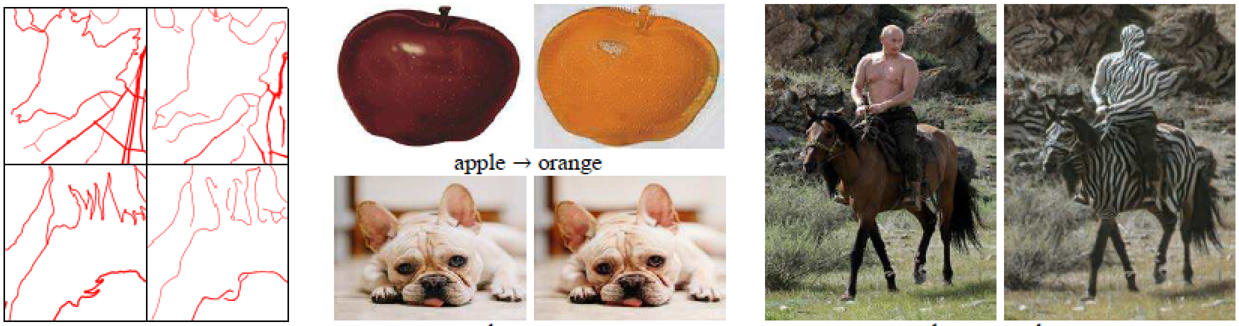

Figure 1: Deep learning image generation failure, roads images are from our experiments, while other are from [13].

(you do not want to disconnect your roads if your map are to be used to visualize trip itineraries). For most of these evaluation tasks, researchers found that the constraint-based approach provides a useful framework [8, 9, 10, 11, 12].

However, the use of deep learning implies some other challenges for the evaluation, mainly due to the use of raster (images) instead of vector format:

- The credibility, noise, and realism of the predicted images have to be evaluated. Indeed, deep learning generates a new image by choosing pixel values. The main constraint for the choice of the pixel values is that the prediction has to look like the target (i.e., ground truth) from a set of training examples. Figure 1 illustrates some unrealistic images of road map produced using deep learning, and some other failure cases by generative adversarial networks (GAN) [13]. So there is no guarantee that the results will follow cartographic rules, represent the same geographic information or be a functioning map. A good neural network should produce maps that respect these objectives. Nevertheless, as most deep learning models do not explicitly incorporate these considerations, it is more reasonable to examine if and how the map objectives are satisfied with an independent evaluation process.
- Most deep learning techniques are raster-based and the raster format is a limitation for using the traditional map generalization evaluation methods. Although converting to a vector format is possible, it can complicate the whole end-to-end process. Sometimes it is even not suitable. For example, when the evaluation is used to guide the learning process during generalization, real-time conversion can be demanding.

The aim of this article is to explore how constraint-based evaluation could contribute to addressing these challenges. To be specific, we focus on the evaluation of generalized raster images that should reflect the content of the map. To train the deep network for map generalization, our models take as input, images of the detailed map data, and try to predict images in the domain of generalized maps. The model predictions (i.e., generated images) have to look like a generalized map. Following previous research on deep neural network for road generalization in raster format.

The The following section is a literature review about map generalization evaluation and the challenges of deep learning approaches for map generalization. Then, we present our use case in 3 section. In 4 section, we propose seven raster-based constraints and their measurements for our use case. Finally, we present some concrete examples of the application of these constraints in our evaluation 5 section and some discussions about map generalization using deep learning quality management 6 section.

2 Literature Review

2.1 Deep Learning for Map Generalization

Deep learning methods are machine learning techniques mainly applied to images, point clouds, text, and graphs. Deep learning models are neural networks made of numerous abstraction layers. Each abstraction layer encodes the input until obtaining a prediction in the shape of the target, then the target and prediction are compared with a loss function. Finally, the weights of the abstraction layer are backpropagated in order to minimize the loss. Thus, the deep neural network iteratively learns to reproduce the target from the examples, and probably for most similar input situations.

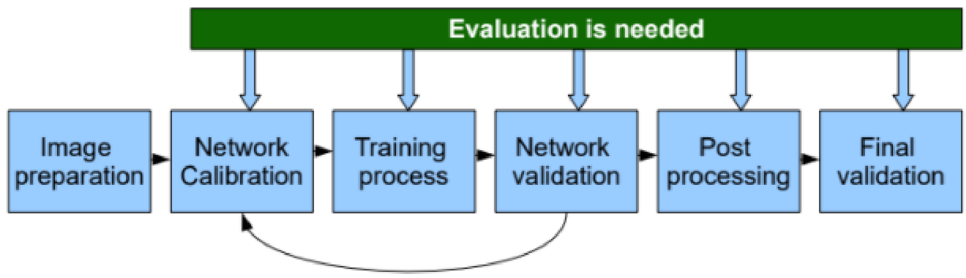

Figure 2: Steps where evaluation is required in a deep learning process.

Figure 2 presents the main steps where an evaluation would be required in the setup of a deep learning process. First of all, the evaluation of the training examples at the training set preparation step would permit to make sure the examples are good enough, or to filter examples of lower quality; and the quality of the deep learning model is mostly conditioned to the training set quality [14]. Then, evaluation can be used to choose the most adapted configuration of a network or guide the post-processing (e.g., identifying a noise to eliminate). Finally, the main interest of an evaluation measure calculated using images is to guide the learning process, i.e., including a particular evaluation measure in the loss function of the neural network.

Deep learning techniques are new at dealing with cartographic information, and very few models are used for map generalization. Among the possible kinds of deep learning models, three have already been employed to resolve tasks in the field of map generalization: classification models, segmentation models, and image generation models. The next paragraphs detail the interest of these techniques for map generalization and the relative need for evaluation.

Classification models are models that classify a whole image into a category. They can be used to obtain information about some geographic objects and contribute to information enrichment before its generalization. This approach has been used to classify the building shapes [15], and to detect some complex objects like highway interchanges [16]. In these examples, there is specific evaluation needed neither regarding the nature of the object (geographic information) nor regarding the nature of the task (generalization). The result of the classification is a class, and the evaluation consists in determining if the predicted classes correspond to the expected one, using a confusion matrix. However, in some cases the target classification is fuzzy and the relevance of global classification is more important than just the value of each element, e.g., for road section classification between the classes "selected at target scale" or "eliminated at target scale" [17]. Here, the global reduction of edge density while maintaining the structure and connectivity of the road network is more important than the exact accuracy of the selected sections.

Segmentation models are used to classify the pixels that belong to a target object on an image. In map generalization, this method can be used to detect some objects and enrich geographic information as well as (or even better than) classification models [16]. But segmentation models can also be used to predict the shape of a generalized object from an image of the detailed shape. The potential of this technique for building [2], roads [18], and coastline [3] generalization has been explored. The result of segmentation is a mask of the generalized objects. The common evaluation measure in the segmentation literature is the intersection over union. This measure compares the predicted mask with the target one. However, in map generalization, the target is one possible generalization and a different shape can represent a good generalization too. Moreover, this measure is highly affected by the position and size of map objects and very lightly affected by local changes. Consequently, a map object (road or building) that is well generalized but displaced of two pixels to the left (e.g., to avoid the overlap with a river) would be evaluated as a bad result, while a road with a small disconnection or bend coalescence will be evaluated as correct. The issue of evaluating map generalization is to ensure the information preservation and the legibility of the output rather to compare with a reference.

Generation and style transfer approaches aim to generate a new image from an input one. These approaches are based on the combination of a generator network that produces an image and a discriminator that evaluates if this image is false or if it is a real target image. The generator learns to fool the discriminator while the discriminator is trained to improve its classification accuracy. Such an adversarial strategy would eventually come to a balance where the generator outputs images that look realistic (i.e., they look like real maps when the target is a map), and the discriminator cannot tell which one is real. Various such networks have been proposed to produce images of a map from aerial images [19], to produce an image of a map in the style of Google Maps from an image of OSM data in a raw style [20], and to produce an image of a generalized map from an image of detailed one [14]. For all these experiments, the evaluation is still challenging, as the network learns to reproduce a target domain that is not clearly defined. In the literature of generative model evaluation, one of the most important issues is the realism of the image. This question reflects the credibility of the output. In map generalization, however, information preservation and legibility are more important. The internal measures of the network are not sufficient to deal with both constraints simultaneously.

2.2 Map Generalization Evaluation

Map generalization evaluation is essential to control, assess, and validate the generalization process. In practice, map generalization is often evaluated visually by cartographers [7]. Hence, automated evaluation is necessary, and developing evaluation criteria is still challenging [4, 6, 7].

Constraints aim to formalize map specifications [8]. Thus, they are essential for many tasks in cartography including map generalization evaluation [21, 22]. Constraints are associated with an object and have a specific function. For example, the constraint “buildings must be greater than a certain size” aims at improving the legibility of buildings on the map. They can be classified according to the object they apply to [23], their nature or function [24], or both of them [10]. Evaluation is usually based on at least some constraints that aim to ensure information preservation and some other constraints to improve the legibility of the map objects [5].

Evaluating a map or a generalization process from a set of constraints raises the following issues:

1. The identification of relevant constraints/desired characteristics of the map. Both the objectives (legibility and preservation) need to have similar weights in the evaluation to avoid caricatured results [5]. Blank images are legible but do not preserve any information, and on the contrary, the identity transformation is a perfect preservation but does not improve legibility at a smaller map scale.
2. The identification of the relevant level of analysis for each constraint. This is a key issue in constraint definition: a constraint on each road section is different from those on a complete road, or on the whole road network. Ruas identifies three levels of constraints [23]: *micro* that relate to an individual object of the database used to make the map; *meso* that relates to a group of objects; and *macro* that relates to a population (e.g., all roads of the map).
3. The definition of a measurement method. The complexity of the measure depends on the constraint. For instance, constraints on object size can be easily measured while others require advanced spatial analysis, e.g., alignment preservation [6] and line position accuracy [25].
4. The computation of a level of satisfaction. The goal of this step is to determine if the result of the measure is acceptable. To do this, we can define an evaluation function [9]. For example, for the constraint “the road position must be preserved,” the measure could be the distance between two points from the previous and the new position, and the evaluation function would evaluate how the constraint is respected or violated given the measured distance.
5. The fusion of all the constraints into a synthetic value would permit the comparison of map generalization methods. However, generalization constraints often express conflicting ideas and cannot be totally satisfied at the same time: in a dense area, the legibility constraint demands information selection and/or object displacement which violates the preservation constraints. Consequently, several strategies can be used for this step [12]: minimize the value of worst constraint satisfaction, reduce the amount of critical measures, reduce the number of conflict areas, or increase the mean satisfying of all constraints, etc.

3 Use Case

3.1 Mountain Roads Generated by Deep Learning

In this article, we present how raster-based constraints can be defined, set up, and applied to the specific case of mountain roads. We focus on evaluating mountain road graphic generalization for a display at the 1:250,000 scale. The goal is to be able to evaluate the images generated by three different deep learning models, i.e., CycleGAN (Zhu et al. 2017), Pix2Pix [19], and U-Net [26] with a reference generalization (obtained by a combination of vector generalization algorithms and manual corrections). We choose to compare these three models as they have already been experimented for road map generalization, and the lack of evaluation methods make their comparison difficult. To make the comparison as fair as possible, we train the networks with the same amount of training examples, and tiles constructed similarly. We based our experiments on images of roads in the Alps extracted from IGN (the French national mapping agency) maps at 1:250,000 (target) and 1:25,000 (input) scale [27]. Our images (input and output)

Table 1: Table of possible levels of analysis for raster and vector data.

Level of analyse	Vector data	Raster data
Micro	Roads	Pixel value
Meso	Road network faces	Set of background pixel
	Stroke	
	Other set of road	Set of road pixel
Macro	Road network	Image

represent 2.5 km2 in 256×256 pixels, so the resolution is around 10 m. The width of the road symbols on these images corresponds to road importance and varies between 1 and 6 pixels. The reference images represent the expected generalization.

The main challenge of this use case is the raster format of maps. In vector map generalization, road segments, faces, and networks are used to set up constraints. On the contrary, a raster is a grid of valued pixels, and this format does not permit to define the limit of each object. Consequently, the constraints should not hold on to the notion of object shape or limits, and other levels of analysis have to be explored, such as the pixel, a set of pixels, or the image. These levels are presented in Table 1. We can construct a pixel set using the following rule: two pixels that are 4-connected and have the same color should belong to the same pixel set. Consequently, a road map is composed of road and background pixel sets.

3.2 Choosing the Constraints

Most of our constraints are directly derived from the usual requirements of mountain road map generalization, and road map generalization [28, 29]. We expect the mountain road generalization to reduce the coalescence in bend series and to smooth the road shape while preserving the position, connectivity, and structure of the network. Displacements are tolerated to avoid coalescence with other cartographic themes (rivers, for example), as well as bend removals sometimes, but the position has to be preserved as much as possible. We consequently decided to observe and measure the following elements:

1. The smoothness, the generalized shapes have to be smoother, less detailed than the initial ones;
2. The coalescence reduction, there should be no coalescence left;
3. The position preservation, the distances between input and generalized roads have to remain as small as possible;
4. The structure preservation, the road network structure and connectivity have to be preserved.

All these constraints are classical, and we only need to adapt the way they are measured. In addition, the observation of the example images of our use case validates our hypothesis that specific measures for deep learning results are required. Figure 3 describes an extract of our testing set. A visual evaluation of these images would guide the constraint definition and let us make the following observations.

U-Net prediction is promising but artifacts and disconnections impact both the information preservation and the legibility of maps. Pix2pix images are good generalizations for most situations, but bend series and complex situations are often not well generalized, and coalescence sometimes remains. CycleGAN gives a better prediction, and most of the situations are correctly generalized; some cases are too smooth, and the image is sometimes blurred. Finally, the images have very few visible noise pixels in the form of regular small blue spots. We observe that some predictions are unrealistic, blurred, or noisy. Consequently, we decided to add some realism constraints that aim at checking that the model will produce images that look like real maps.

3.3 Validation Set

In this section, we present the images used to set up and validate the proposed constraint measures. We propose to constitute what we call the *validation set* with both “real data” (images extracted from our deep learning results presented above), and “fake data”: hand-drawn images of interesting configurations of roads. For example, Table 2 lists interesting fake situations to construct by modifying the shape of a random road.

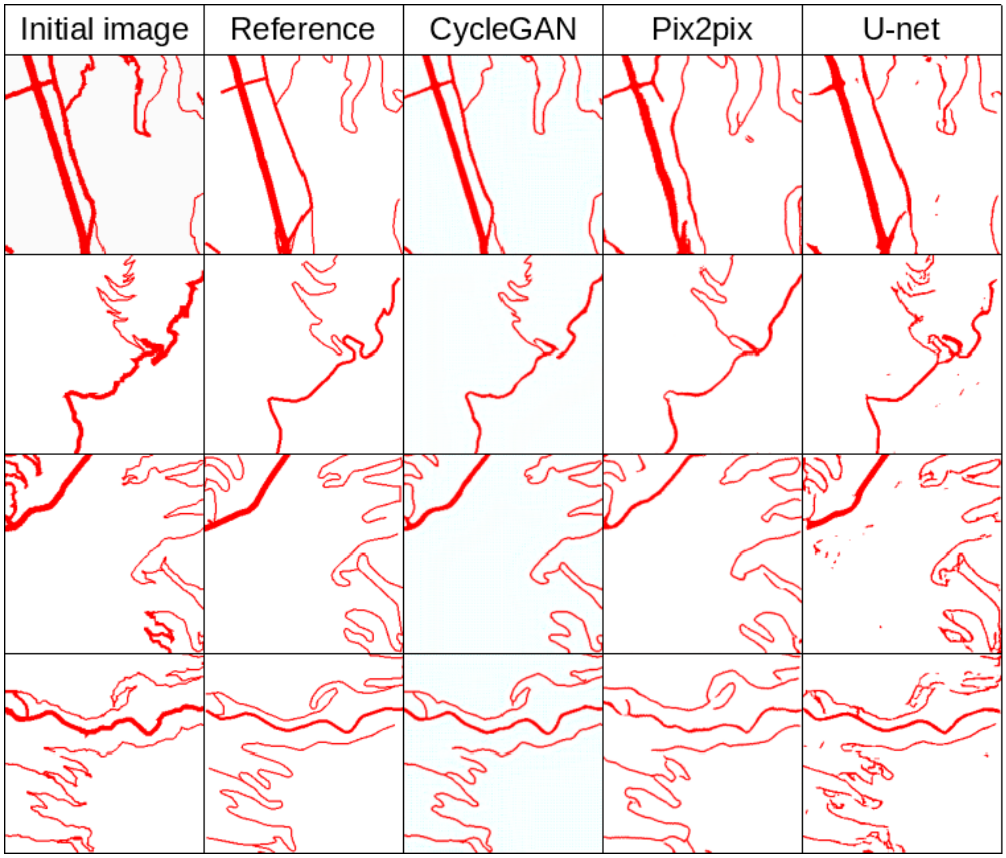

Figure 3: Examples of image in the test set.

Table 2: Alterations applied to create fake test images.

Legibility alteration	**Information alteration**
Enlarged bends	Appearance of a road
Enlarged road width	Noise in the image
Small displacement	Big structure alteration
Loop creation	Continuity alteration
Blurred road	Continuity alteration at border
Only background in the image	Only roads

4 Proposed Constraints ans Their Measurements

In this section, we present all the proposed constraints, associated measures, validation, and satisfaction functions. Most of these constraints are extensions of the measures first proposed in [30]. We implemented all our measures in Python using the open-source libraries PIL and skimage. The code should run on any square image, but most measures are only relevant for images constituted of roads (pixels close to red) and backgrounds (pixel close to white). We choose to define a qualitative level of constraint satisfaction: “very good,” “good,” “neutral,” “bad,” “very bad.” For our use case, we propose the definition of a threshold for each satisfaction level, based on the examination of the validation set of our experiment.

4.1 Clutter Reduction

Definition Clutter can be defined as too much and disorganized information in an image, and it can be seen as a proxy for image complexity. It has been shown that clutter makes the use of the images more difficult [31]. This is why we introduce a constraint to avoid that clutter raises during generalization. This constraint may be adapted for images of generalized maps in general, not only for images of generalized mountain roads.

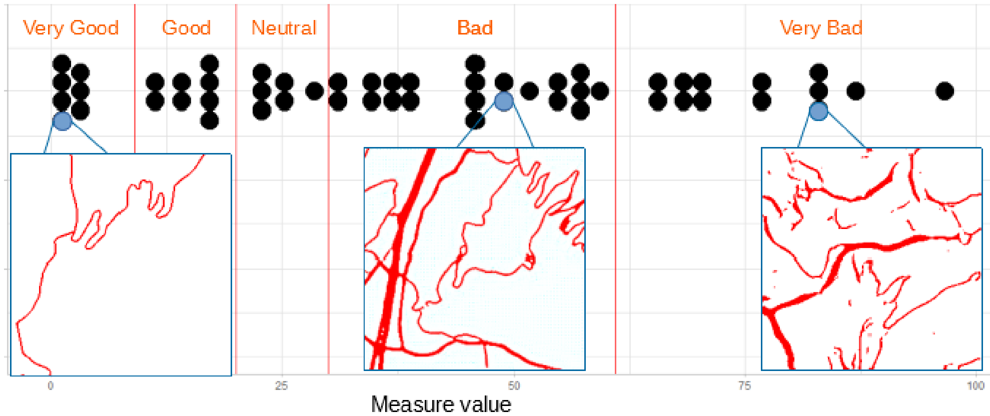

Figure 4: Distribution and examples of the value for edge-density clutter measure. Each point represents an image from our testing set, the corresponding level of satisfaction of the constraint is indicated in orange.

Table 3: Clutter level of satisfaction derived from the measure.

Very good	Good	Neutral	Bad	Very Bad
$0\% \leq x < 10\%$	$10\% \leq x < 20\%$	$20\% \leq x < 30\%$	$30\% \leq x < 60\%$	$x \geq 60\%$

Measure Clutter measures are image-level measures that aim to evaluate the complexity of an image. Their adaptation for map legibility evaluation is still a question [32]. We have tested four different measures for estimating the legibility of our road images:

- The edge density [31],
- The sub-band entropy [31], that is similar to JPEG compression,
- The quad-tree method, measures the number of homogeneous square cells in the image [33],
- The object segmentation-based measure [34].

Most of these measures represent a quantity; for more clarity, we propose to convert these quantities into a ratio using the formula 1

$$clutter\_generalisation = \frac{|clutter_{initial} - clutter_{prediction}|}{clutter_{initial}} \tag{1}$$

Validation and Limits Figure 4 presents the value of the edge density measure for test images (the images presented in Figure 3). This measure gives similar values for each image except the CycleGAN one. Indeed, these images include pixel noise that highly impacts the measure. This observation is similar for all tested clutter calculation methods, this is why we decide to use the edge density to measure clutter. The measure succeeds in detecting a complexity increase in the images from this method. However, if we look at the U-Net generalization, while a human could judge it as quite more complex than the others, the clutter measure only indicates a small difference.

Constraint Satisfaction We propose the satisfaction values associated to the clutter measure presented in Table 3. The thresholds were calibrated with our validation set of images (Figure 3).

4.2 Smoothness

Definition The smoothness constraint is a legibility measure that should constrain the generalized road to be smoother than the input road. The more the road is smooth, the fewer irregularities are visible, the more the shape is legible. This constraint is designed for images of generalized roads in general.

Measure To measure smoothness, we have to detect and measure the number of irregularities around the roads. We detect irregularities at the pixel level using mathematical morphology. We simplify the problem by transforming the image from colored (RGB) to a black and white map that represents the presence or absence of a road. A closing

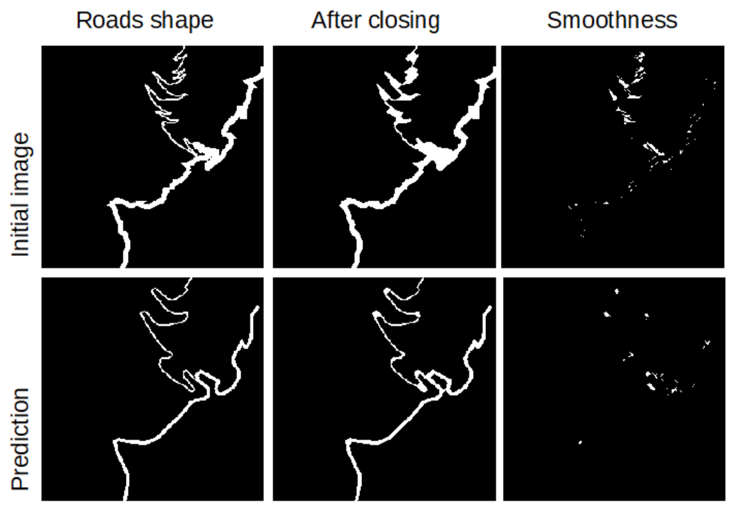

Figure 5: Effect of the smoothing operation on a road and its generalization.

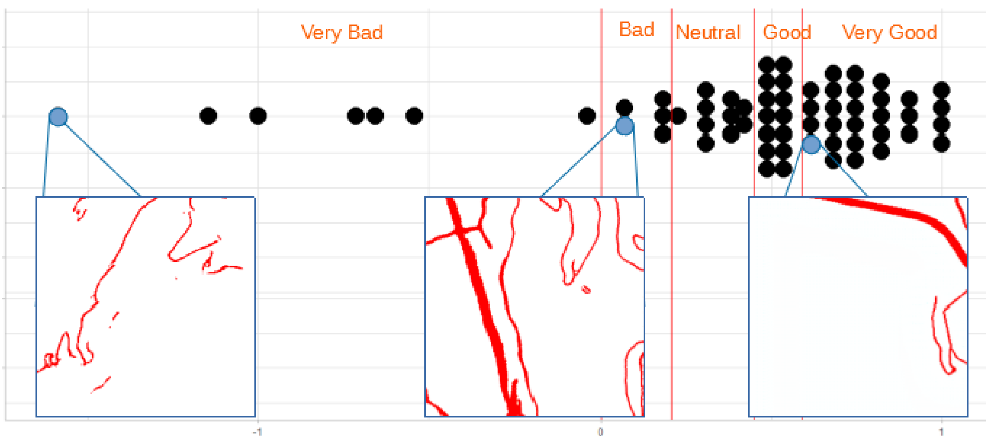

Figure 6: Distribution and examples of values for the smoothness measure. Each point represents an image from our testing set, the corresponding level of satisfaction of the constraint is indicated in orange.

(dilation followed by erosion) of road pixels on the image should permit, with an adapted size to fill in the concave irregularities of the road shape. Then, the deletion of the road shape $(irregularpixels = closingpixels - roadpixels)$ isolates the irregular pixels near the road as illustrated in Figure 5. Then we also propose a measure at an image level: the ratio of the initial roughness that is removed by generalization (2).

$$smoothness = \frac{irregular_{initial} - irregular_{prediction}}{irregular_{initial}} \tag{2}$$

Parameters The size of the closing determines the size of the biggest irregularity that should be smoothed by generalization; 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 resolution and scale of our use case with a closing value of three pixels (30 m). Our experiment has shown that closing with less than three pixels detects very few irregularities while closing with more pixels fills the bends and tiny regular background areas.

Validation and Limits Figure 6 represents the distribution and some example values of the smoothness measure. This measure is globally well correlated with our visual perception of smoothness.

However, we observed the following limitations:

Figure 7: Different kinds of pixels detected as irregular: a) concave only irregularity, b) convex only irregularity, c) symmetric irregularity.

Table 4: Smoothness level of satisfaction (at image level).

Very good	Good	Neutral	Bad	Very Bad
$x > 60\%$	$60\% \geq x > 40\%$	$40\% \geq x > 20\%$	$20\% \geq x > 0\%$	$x \leq 0\%$

- The accuracy of the irregularity detection has to be questioned; depending on the threshold so irregularities can be missed and/or some bends can be wrongly identified as smoothness problems.
- The measure depends on the resolution and scale of the road; other use cases might require different parameters.
- The measure is quite sensitive to the initial situation: roads with very concave shapes, even well generalized, will be evaluated as less smooth compared to straighter roads. Moreover, convex irregularities are not identified. In most of the situations where the road is concave at the right, it is convex at the left (and conversely), but in some really rare cases where the road shape is large, the left and right sides of the shape are not correlated (illustrated in Figure 7). In these cases, the convex irregularities are not correctly measured: (a) is considered as not smooth, while (b) and (c) as similar.
- If the initial image has no irregularity the ratio measure cannot be computed.

Constraint Satisfaction We propose the satisfaction values associated with the measure of smoothness constraint presented in Table 4.

4.3 Coalescence Reduction

Definition The coalescence constraint is a legibility constraint. Coalescence is the fact that the symbol of a line overlaps itself in the interior of a sharp bend [28]. The generalized image should contain no symbol coalescence. This constraint is designed for generalized mountain roads, where coalescence might occur.

Measure To measure the coalescence, we also use mathematical morphology operations to detect pixels that belong to a coalescent part of the road. This measure is also based on black and white images. The measure is also at the pixel level if we consider the map of the coalescent pixels and aggregated at the image level as we consider the ratio of pixels in an image. First, a dilation of roads (of n pixels) simulates the coalescence occurring when the road symbol is enlarged due to scale change; then an erosion of size n + w (the maximal width of roads) erases the parts that are not coalescent; and finally, a dilation of w pixels permits to evaluate the number of pixels in the coalescent areas. This process is illustrated in Figure 8. Finally, similarly to the smoothness constraint, we propose to measure the evolution of coalescence instead of the number of coalescent pixels using Eq. 3.

$$Coalescence = \frac{coalescent\_pixel(initial) - coalescent\_pixel(prediction)}{coalescent\_pixel(initial)} \tag{3}$$

Figure 8: Steps for detecting coalescent pixels, from left to right: initial image, dilation, erosion, dilation.

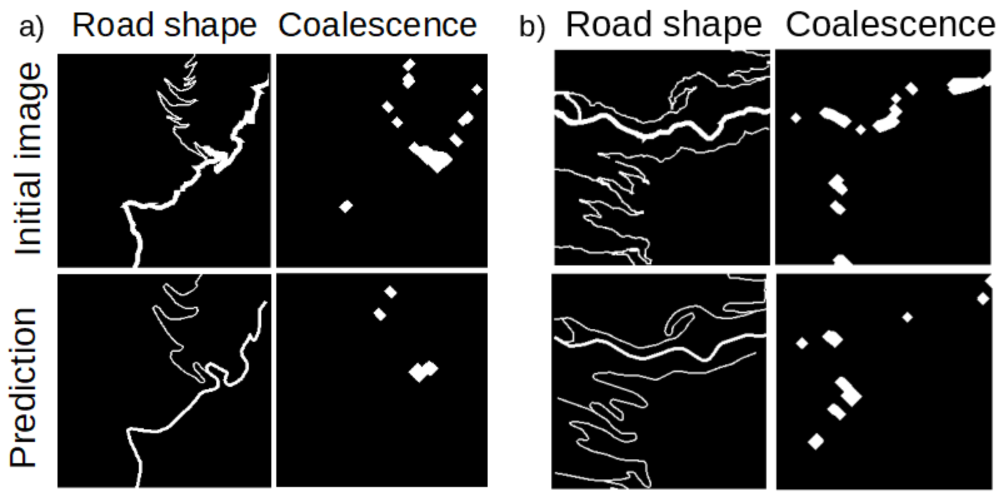

Figure 9: Effect of the coalescence detection on a road and its generalization. a) example with a few overdetection. b) example with an overdetection of coalescence in the generalized image.

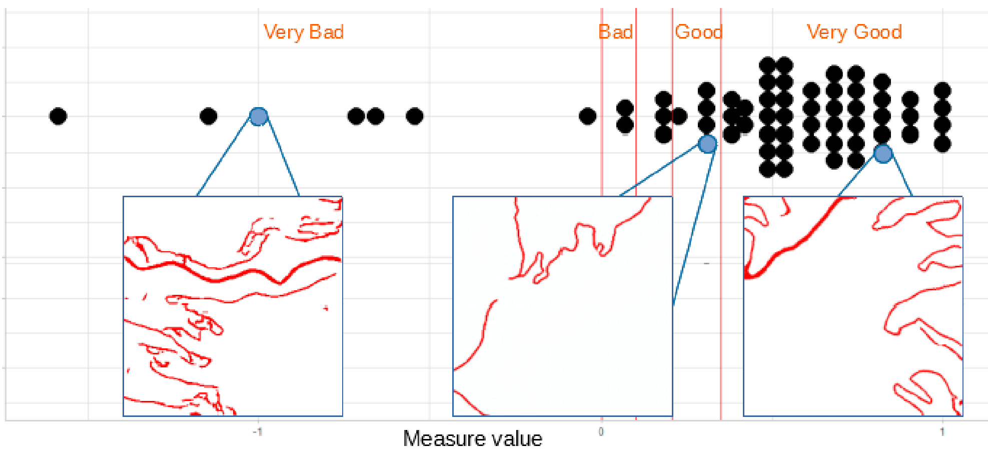

Figure 10: Distribution and examples of values for the coalescence reduction measure. Each point represents an image from our testing set, the corresponding level of satisfaction of the constraint is indicated in orange.

Parameter In this process, the size n represents the maximal distance between two coalescent parts of the road from the map specifications. For our use case (at the 1:250k scale), a threshold of five pixels is the most adapted, and it corresponds to a distance of 50 m. This value allows the detection of most of the coalescent areas. However, it also detects some un-coalescent areas as presented in Figure 9, while a thinner threshold would miss many coalescent areas.

Validation and Limits Figure 10 presents the distribution and some examples of values for the coalescence reduction measure. This measure is more questionable than the others. First of all, it does not exactly measure the coalescent parts but the pixels that could be coalescent with a larger symbol. This definition implies that the overlapping parts of the coalescent road are identified as a unique road and not detected as coalescent. Nevertheless, the observed ratio value is most of the time consistent with our visual evaluation of coalescence. But frequently, high coalescence values are explained by two parallel roads that are separated by fewer pixels than the threshold (detected as coalescent by the measure), and this is not coalescence as these are different roads, but just symbol overlap. These situations are limited to 5% of the cases in our real test set.

Constraint Satisfaction We propose the satisfaction values associated to the measure of coalescence constraint presented in Table 5. As there is an over-detection of coalescence, we cannot really consider that any coalescence is bad, so we compute satisfaction based on the reduction of coalescent areas.

Table 5: Coalescence level of satisfaction (at image level).

Very good	Good	Neutral	Bad	Very Bad
$x > 30\%$	$30\% \geq x > 20\%$	$20\% \geq x > 10\%$	$10\% \geq x > 0\%$	$x \leq 0\%$

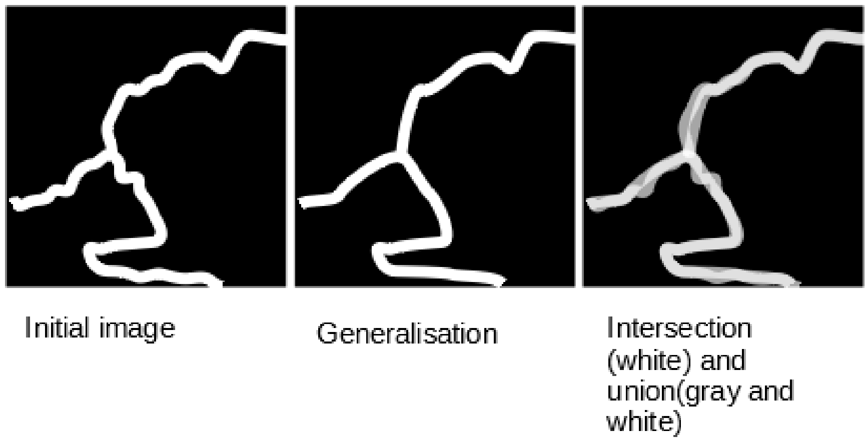

Figure 11: Intersection and union of roads and their generalization.

4.4 Position Preservation

Definition Then, this is one of the most basic constraints for map generalization, the generalized road has to be as much as possible at the same location as the initial one, which is as close as possible as the real position. This constraint is designed for an image of generalized roads in general.

Measure We measure this constraint using the intersection over union (IoU) of the road pixels of the initial and generalized images. Here, we also based our computation on transformed black and white images. The intersection between the road pixels in the initial image and the road pixels in the generalized image is computed and the area of the intersection is divided by the area of the union namely, the pixels that belong to generalized or the initial roads (Eq. 4). This is a classical measure to assess image segmentation and determine which parts of the road pixels are common in the initial image. Figure **??** represents the intersection and the union of a road and its generalization.

$$IoU = \frac{road\_pixel(initial) \cap road\_pixel(prediction)}{road\_pixel(initial) \cup road\_pixel(prediction)} \tag{4}$$

Parameters We propose to use a buffer of size n around roads instead of just the road pixels. This threshold would permit to model the area where the roads are allowed to move or be distorted without considering the position as too far. This threshold can be very large because the displacement offset can be large for mountain roads at this scale (1:250k). In our image set, the position is always visually preserved. However, the measured position accuracy is most of the time under 50%, so introducing this buffer with a size of 20 pixels (200 m) better reflects this preservation and the variation between two generalization methods: some cause more displacement than others.

Validation and Limits Figure 12 represents the distribution and some examples of values for the position measure. Despite a big dissimilarity of IoU values in our image set, we observe that all real images visually respect the position accuracy, even those with a threshold of IoU very low. Moreover, the measure is not regular: with a similar displacement, the IoU of two large objects will be more important than the one of two small objects. Another limit of this method is that it permits quantifying the displacement, but it does not identify if the displacement is relevant. Road displacements and distortions are permitted when they tend to increase legibility. Finally, the road position accuracy is a minor constraint: it can be seen as a soft constraint whose satisfaction does not need to be maximized.

Constraint Satisfaction We propose the satisfaction values associated with the measure of position constraint presented in Table 6.

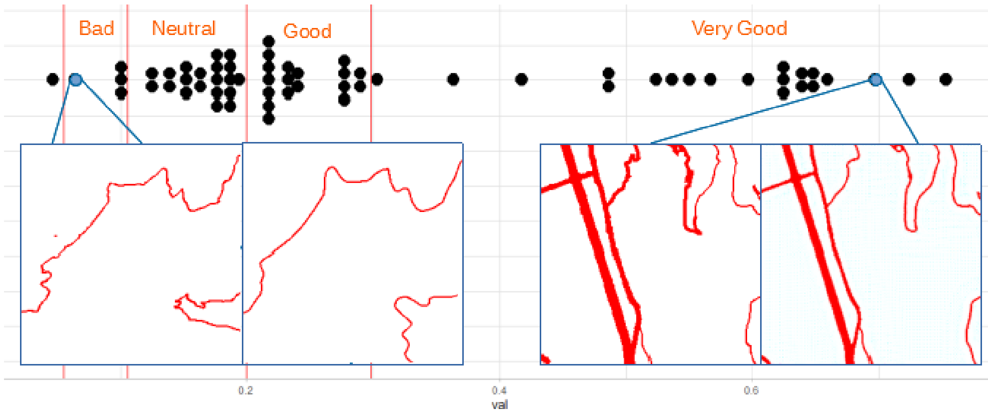

Figure 12: Distribution and examples of values for the position measure. Each point represents an image from our testing set, the corresponding level of satisfaction of the constraint is indicated in orange.

Table 6: Position's level of satisfaction (at image level).

Very good	**Good**	**Neutral**	**Bad**	**Very Bad**
$x > 30\%$	$20 < x < 30$ %	$10 < x < 20\%$	$5 < x < 10\%$	$x < 5\%$

4.5 Road Connectivity Preservation

Definition Then, we think that more than the position, the road network structure preservation is a key point for road maps. The connection between roads permits the creation and following an itinerary that is the most important function of such maps. This constraint is adapted for images of generalized roads in general.

Measure For the connectivity preservation constraint, we measure connections and disconnections using the number of sets of contiguous road pixels and contiguous background pixels. Here, we also based our calculation on transformed black and white images. Figure 13 illustrates how these numbers can reveal changes in the structure. A disconnection will reduce the number of background pixel sets and increase the number of road pixel sets. This measure also detects problems such as the creation of a loop, which deteriorates both the correctness and legibility of the information. Equation 5 shows how we combine these numbers of contiguous parts to measure continuity preservation. (x, y) is a pair of a detailed and a generalized image, $nroad$ is the number of road pixel sets in the image, and $nback$ is the number of background pixel sets.

$$structure\_alteration = \frac{|nback(x) - nback(y)| + |nroad(x) - nroad(y)|}{nback(x) + nroad(x)} \tag{5}$$

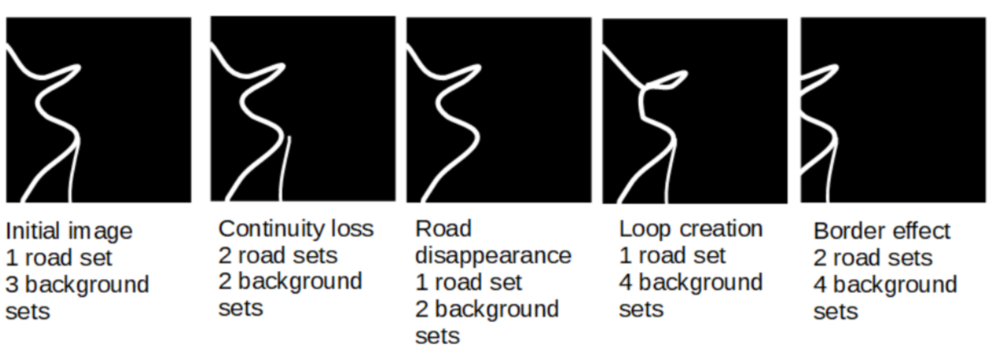

Figure 13: Different cases of disconnections or new connections in road images.

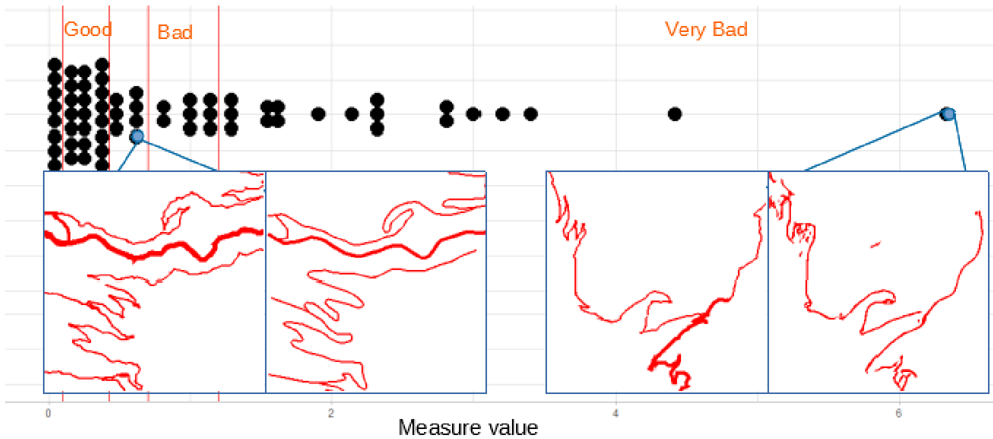

Figure 14: Distribution and examples of values for the connectivity alteration measure. Each point represents an image from our testing set, the corresponding level of satisfaction of the constraint is indicated in orange.

Table 7: Road connectivity's level of satisfaction.

Very good	Good	Neutral	Bad	Very Bad
$x < 10\%$	$10\% \leq x < 40\%$	$40\% \leq x < 70\%$	$70\% \leq x < 100\%$	$\geq 100\%$

Parameters We choose to use the 4-connectivity to construct the sets of pixels; pixels of the same value connected by diagonal only, do not belong to the same group of pixels. We also include a threshold for set size: the groups composed of one or two pixels are not considered, consequently the appearance or disappearance of such group is not measured.

Validation and Limits Figure 14 gives the distribution and some examples of values for the connectivity measure, each pair represents an initial situation and the corresponding generalization. This measure is quite consistent with our observation but it seems to overvalue structure changes. This problem is explained by the fact that all changes have equal weight in the calculation, even the addition of very small sets or the disconnection of one pixel, which are not visible at the target scale. Then, this measure is sensitive to border effects (illustrated in the last column of Figure 13). 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.

Constraint Satisfaction We propose the satisfaction values associated to the road connectivity constraint presented in Table 7.

4.6 Color Realism

Definition The next two constraints are not adapted from conventional vector-based constraints but are specific to images generated by deep learning models. They aim at the quantification of errors that do not exist in traditional map generalization. The goal of the color constraint is to make sure that the generated images do not convey any unexpected information. In a map, a different color represents a different object. This constraint is adpated for all images of map generated using deep learning.

Measure We propose to measure if there is color noise: in our use case, it means pixels that are not roads (red) or background (white). We have decided to use an implementation of the CIEDE2000 distance between colors [35] because of its consistency with human color perception. Then, we counted the number of pixels that are visibly too distant from red or white.

Parameters We choose to consider a pixel as white or red when the distance with pure red (255,0,0) or pure white (255,255,255) is higher than 9.

Validation and Limits Figure 15 represents distribution and examples of values for the color measure. This measure has permitted quantifying the observed noise of CycleGAN images and verifying that images from other methods do not suffer from this problem.

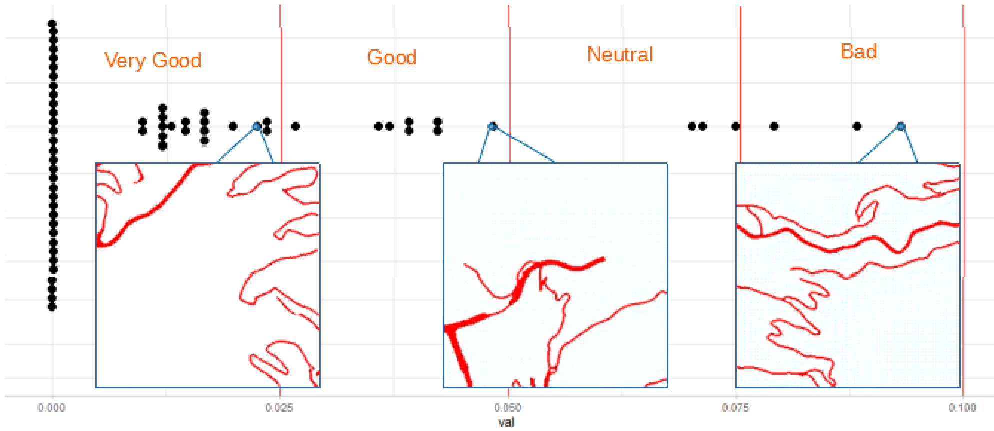

Figure 15: Distribution and examples of values for the colour measure. Each point represents an image from our testing set, the corresponding level of satisfaction of the constraint is indicated in orange.

Table 8: Color realism level of satisfaction (at image level).

Very good	Good	Neutral	Bad	Very Bad
$x < 2.5\%$	$2.5 \leq x < 5\%$	$5 \leq x < 7.5\%$	$7.5 \leq x < 10\%$	$\geq 10\%$

Constraint Satisfaction The satisfaction values of this constraint are computed with the threshold distances presented in Table 8.

4.7 Noise Absence

Definition We also found that some deep learning models can generate some noise pixels: e.g., isolated road pixels that are not roads. This is due to the fact that the loss function that is optimized in the training process is global and might not be fully optimized after all the iterations of the training. This constraint is designed for all images of maps generated using deep learning, not just generalized mountain roads.

Measure We just count the number of sets of contiguous road pixels that are too small to be real roads. Similarly to the road connectivity preservation constraint, we use the 4-connectivity to define sets of contiguous pixels.

Parameters The threshold for noise detection depends on the image resolution. For our use case, we have fixed a size below 6 pixels.

Validation and Limits Figure 16 presents the distribution and some examples of values for this measure. This measure is quite satisfying as it permits us to distinguish noise in both the fake images that we created with noise in our validation set, and in the images generated by a U-Net model, which often contain noise pixels.

Constraint Satisfaction The satisfaction values of the noise constraint are computed with the thresholds presented in Table 9.

5 How to Use the Constraints in Deep Learning Map Generalization

In this section, we present how our evaluation method can be used in three examples: comparing deep learning models, identifying and solving the quality issues of a deep learning model, and validating the results of a model. Our

Table 9: Noise level of satisfaction (at image level).

Very good	Good	Neutral	Bad	Very Bad
$x = 0$	$x < 3$	$3 \leq x < 6$	$6 \leq x < 9$	≥ 9

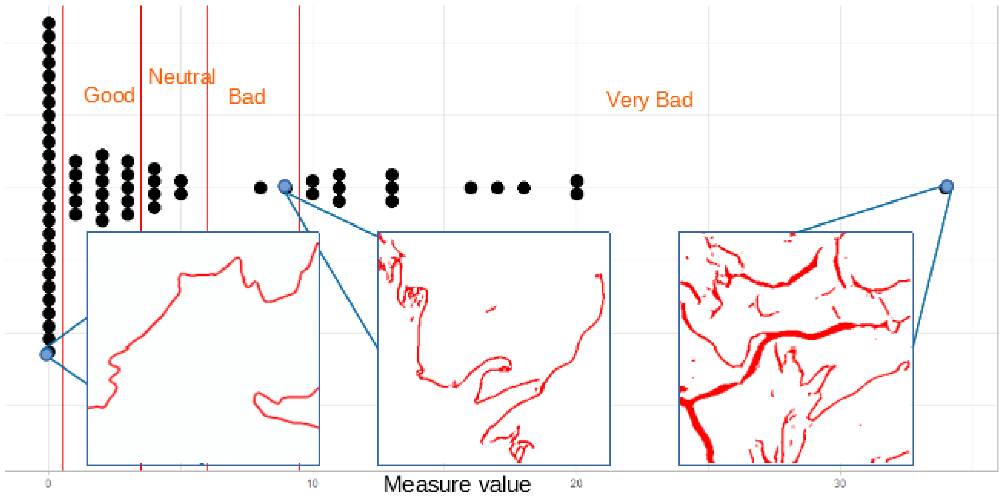

Figure 16: Distribution and examples of values for the noise measure. Each point represents an image from our testing set, the corresponding level of satisfaction of the constraint is indicated in orange.

observations are based on the evaluation of the results obtained with U-net, cycleGAN, and pix2pix models, trained and tested with default parameters on the same image set from the Alps.

5.1 Comparing Models

In this section, we propose to explain how the evaluation can be used to compare different deep learning models or similar models used with different parameters. A good generalization solution is a balance between several criteria, so it is not easy to determine if a solution is better than another. The constraints could contribute to evaluating which network configuration gives better results in an ambiguous case where it is visually hard to determine a preference. For comparison purposes, a good combination of different measures is needed to give a synthetic measure of the global quality of a tile image. For this combination, several methods are possible, but the method has to preserve and highlight the order and dissimilarity of quality. Moreover, the mean and worst values are interesting for this task. Instead of the worst value, we prefer the last decile value that excludes exceptionally bad values.

For our use case, the mean value allows a thinner comparison of the interest of CycleGan and pix2pix methods, which have quite visually similar results. Table **??** shows the mean and last decile values for the constraint measures and the associated level of satisfaction, for all the images from our test set. This table shows that unsupervised learning (CycleGAN) preserves more information while supervised learning (pix2pix) learns that distortions are tolerated. On the other hand, the comparison of the worst value permits highlighting the irregularities of the methods. The comparison shows that the reference images are sometimes less well generalized than deep learning predictions. This is due to defects in our data with roads displayed in the reference generalized image, but not in the initial data. Then, this analysis helped us in the identification of these badly generalized images, and excluding them from the example set would improve the learning.

Finally, this comparison allows the identification of parameters or models that are more appropriate for several road configurations. It could contribute to improving results in most cases by using the best model for each configuration. Thus, the output generalization of the study area would be a partition of tiles generalized by different deep learning models. The determination of the most appropriate methods for each configuration relies on a robust evaluation that permits to compare methods.

5.2 Identifying and Solving the Quality Issues of a Model

In this section, we explain how a model can be used to detect the quality issues of a given deep learning model and then eventually improve the model. On one hand, the constraints at the image level must provide a comprehensive indicator of quality for the constrained property, which allows the identification of a quality issue. On the other hand, constraints at the pixel level (e.g., coalescence and smoothness) permit the identification of conflicts and situations that are badly generalized within each tile.

Table 10: Table of value for comparing models.

Method	Constraint	Reference	CycleGAN	Pix2pix	U-Net
Mean value	Smoothness	0,85	0,92	0,82	-0,23
Satisfaction		Very good	Very good	Very Good	Very Bad
Worst value		0,54	0,80	0,50	-3,37
Satisfaction		Good	Very good	Good	Very bad
Mean value	Coalescence	0,47	0,62	0,63	-0,02
Satisfaction		Very good	Very good	Very Good	Very bad
Worst value		-0,20	0,38	0,42	-0,91
Satisfaction		Very bad	Very good	Good	Very bad
Mean value	Position	0,19	0,63	0,22	0,23
Satisfaction		Neutral	Very good	Good	Good
Worst value		0,08	0,54	0,13	0,16
Satisfaction		Bad	Very good	Neutral	Neutral
Mean value	Connectivity	0,32	0,39	0,74	2,65
Satisfaction		Good	Good	Bad	Very bad
Worst value		0,45	1,07	1,46	3,91
Satisfaction		Neutral	Bad	Very bad	Very bad
Mean value	Color	0	0,67	0,61	0
Satisfaction		Very good	Good	Very good	Very good
Worst value		0	0,96	1,01	0
Satisfaction		Very good	Bad	Good	Very good
Mean value	Noise	0,19	0,25	5,06	12,75
Satisfaction		Good	Good	Good	Very bad
Worst value		1	1	10,5	20
Satisfaction		Very good	Very good	Very bad	Very bad
Majority vote		Very good	Very good	Good	Very bad

If a characteristic is often badly generalized, it means that the model is not adapted to learn to reproduce this characteristic. For example, disconnections are important errors in map generalization, but they affect very few pixels, so they are not penalized in the learning process. Indeed, the learning process is based on the minimization of a loss function, and the default loss seems not to be adapted to reproduce structure preservation. Since this default has been identified, it can be solved by the exploration of adapted loss functions. For example, by modifying the weight of loss components and increasing the part that guarantees, the resemblance of the input and prediction images, or by including geometric consistency in the loss computation [36]. Another improvement could be to add the constraint violation measure to the loss function. Therefore, the model learns to minimize the sum of the loss function and the constraint violation (e.g., the connection preservation measure).

If a situation or a kind of situation is often badly generalized, it might reflect an imbalance in the training dataset. For example, in our experiments, the cases with two close parallel roads are under-represented. Here, the solution is to change the training set either by adding images that represent similar situations or by modifying example weights to give more impact in the optimization to the examples that are less represented than others. The calculation of the constraint satisfaction with the pair of images initial/reference from the training set instead of prediction/initial would permit the detection of such situations.

5.3 Validation of a Model

To validate a model, we propose to verify that predictions from this model give results that are under a defined threshold of satisfaction for each constraint. The definition of the threshold of validation is quite similar to the categorization for identifying quality issues. We can consider that the images with a “good” or “very good” value are valid, and then a method with a majority of valid images is valid [12]. These thresholds are defined by examination of the test data and especially the constraint satisfaction values obtained with the reference images, which should be good by definition and should be evaluated as “good” or “very good” for most situations and constraints.

6 Discussion

6.1 Do We Need a Meso Level of Analysis?

In this section, we will discuss the interest of using a meso level of analysis. First, the use of a sliding window over the image can create a new sub-image or a patch that can be better adapted for the calculation of some constraints. This process can permit locating problems in the image for constraints that are only measurable at the image level. With this method, we can also identify the images that suffer from border or size effects: indeed these images would have a remarkable distribution of a constraint value in a patch. For example, the connectivity constraint of an image that suffers from border effects would have a high value for a patch at the border but not for a patch in the middle of the tile.

Then, we believe it could be useful to define a meso constraint that synthesizes the satisfaction of several pixellevel constraints [23] for pixels contained in a meso object. For example, instead of measuring the ratio of the pixels that are not smooth, we could define that an image is smooth if all or most of its pixels are not irregular. This definition first designed for the agent models is also adapted to guide the deep learning processes where the loss is a statistical summary of the pixels or patches values.

6.2 Is it Possible to Automatically Evaluate a Generalization with Images Only?

Image representation allows a visual evaluation of generalization. The automation of this evaluation is a complex combination of multiple constraints. In this article, we describe how some constraints can be adapted for a calculation based on raster data. However, we were unable to prove that all existing constraints can be converted by this method. Especially, the relation preservation constraints seem to be challenging as the raster format does not permit to encode object boundaries and spatial relations.

Moreover, the evaluation of a map must relate to the map function. The automatic evaluation of map efficiency is still challenging for both raster and vector data, as it relates to human perception. A good value of evaluation is a necessary condition for the final map to be a functional map but not a sufficient one.

Finally, the proposed measure still suffers from multiple limitations. We think the raster evaluation can be useful at a stage where vector conversion is not ideal. However, the vectors remain more accurate and practical and permit a comparison with classical generalization approaches or validation.

6.3 Can These Methods be Adapted to a Multi-Theme Map?

Then, a challenge is to adapt this constraint-based evaluation to maps with several themes, not just roads. In these maps, the spatial relations are even more important and concern several different objects, e.g., buildings are aligned along a road, or a road is crossing a forest. And the preservation of these relations is an important point in information preservation as they contribute to the global structure and understanding of the map [37] and help to determine map location.

Moreover, the raster representation makes conflict detection harder as each pixel can have only one value even if there is a coalescence of several elements at this place. Finally, the identification of groups of objects can be done using a distance criterion using buffers, but the identification with similarity criteria could be more difficult. In Table **??**, we propose to estimate the difficulty of the raster adaptation of constraints from [11]. We identify four levels of adaptability of measure:

- Directly measured in the image: ++ ;
- May require image processing: + ;
- May require a new paradigm: -;
- No possible adaptation/may require vectorization: – .

7 Conclusion

Deep learning techniques are still recent but promising and their development and integration into the map generalization process would be conditioned to the existence of common and adapted evaluation methods. This work shows the applicability and the interest of constraint-based evaluation on images generated by deep learning.

Measuring constraint current values based on raster data can require image processing (segmentation, mathematical morphology statistical analysis, etc.) or a new level of analysis (e.g., considering a group of pixels instead of an object).

Table 11: Table of possible constraint conversion to raster.

Id	Type	Geometry	Property	Adaptation
1-1	Minimal dimension	polygon	area	++
1-4	Minimal dimension	polygon	area of any hole in a polygon	+
1-5	Minimal dimension	line/polygon	length of an edge/line	−−
1-13	Shape	polygon	squareness	−
1-18	Orientation	any	general orientation	−−
2-1	Minimal dimensions	any - any	minimal distance	+
2-3	Orientation	line/polygon - line/polygon	orientation	−−
2-9	Topology	any - any	topological consistency	+
3-4	Topology	line and polygon	intersection	−
3-5	Topology	line and polygon	connectivity	−
3-7	Topology	polygon	spatial distribution	−−
3-8	Topology	polygon	spatial distribution	−−

But sometimes, it seems not possible to measure a constraint value with an image. For the use case of mountain roads, we demonstrate that four of the most important constraints can be adapted. We note that the measures for our test set are often consistent with our visual evaluation. To go further in the validation of these constraints, it would be useful 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."

We also presented the interest of a new type of constraint designed for evaluating images generated using deep learning. We propose two examples of such constraints. Then, this article has demonstrated the interest in the constraint-based approach for comparing, improving, and validating deep learning models. Moreover, proposed measures are indicators of a prediction's quality and they can be used as a term of the loss function that the deep neural network aims to minimize.

However, the seven proposed constraints for the case of roads generalization are only partly sufficient for the validation of deep learning models, and the limitations remain important (size and border effects). Future research is necessary to improve the proposed constraints, develop new ones, and try to connect raster-based and vector-based constraint evaluation.

Data and Code availability Data and code presented in the submitted work are available in the DeepMapGen repository, https://github.com/ umrlastig/ DeepMapGen.

Informed Consent This research involved no human participants.

Ethical Approval This research involved no human participants.

Conflict of Interest The authors declare no competing interests.

References

[1] Guillaume Touya, Xiang Zhang, and Imran Lokhat. Is deep learning the new agent for map generalization? *International Journal of Cartography*, 5(2-3):142–157, May 2019, doi:10.1080/23729333.2019.1613071. Publisher: Taylor & Francis.

[2] Yu Feng, Frank Thiemann, and Monika Sester. Learning cartographic building generalization with deep convolutional neural networks. *International Journal of Geo-Information*, May 2019.

[3] Jiawei Du, Fang Wu, Ruixing Xing, Xianyong Gong, and Linyi Yu. Segmentation and sampling method for complex polyline generalization based on a generative adversarial network. *Geocarto International*, 0(0):1–23, January 2021, doi:10.1080/10106049.2021.1878288. Publisher: Taylor & Francis.

[4] William A. Mackaness and Anne Ruas. Chapter 5 - Evaluation in the Map Generalisation Process. In William A. Mackaness, Anne Ruas, and L. Tiina Sarjakoski, editors, *Generalisation of Geographic Information*, International Cartographic Association, pages 89 – 111. Elsevier Science B.V., Amsterdam, 2007.

[5] Xiang Zhang. *Automated evaluation of generalized topographic maps*. PhD thesis, University of Twente, October 2012.

[6] Xiang Zhang, Jantien Stoter, Tinghua Ai, Menno-Jan Kraak, and Martien Molenaar. Automated evaluation of building alignments in generalized maps. *International Journal of Geographical Information Science*, 27(8):1550–1571, August 2013, doi:10.1080/13658816.2012.758264. Publisher: Taylor & Francis.

[7] Jantien Stoter, Xiang Zhang, Hanna Stigmar, and Lars Harrie. Evaluation in Generalisation. In Dirk Burghardt, Cécile Duchêne, and William Mackaness, editors, *Abstracting Geographic Information in a Data Rich World*, Lecture Notes in Geoinformation and Cartography, pages 259–297. Springer International Publishing, 2014.

[8] Kate M. Beard. Constraints on rule formation. In Barbara Buttenfield and Robert McMaster, editors, *Map Generalization*, pages 121–135. Longman Pages, 1991.

[9] Sylvain Bard. Quality Assessment of Cartographic Generalisation. *Transactions in GIS*, 8(1):63–81, 2004, doi:10.1111/j.1467-9671.2004.00168.x.

[10] D. Burghardt, S. Schmid, and J.E. Stoter. Investigations on cartographic constraint formalisation. In *ICC 2007 : proceedings of the workshop of the ICA commission on generalization and multiple representation, August 2-3 at the 23nd international cartographic conference ICC: Cartography for everyone and for you, 4-10 August 2007 Moscow, Russia. 16 p.* ICC, 2007.

[11] Jantien Stoter, Dirk Burghardt, Cécile Duchêne, Blanca Baella, Nico Bakker, Connie Blok, Maria Pla, Nicolas Regnauld, Guillaume Touya, and Stefan Schmid. Methodology for evaluating automated map generalization in commercial software. *Computers, Environment and Urban Systems*, 33(5):311–324, September 2009, doi:10.1016/j.compenvurbsys.2009.06.002.

[12] Guillamue Touya. Social Welfare to Assess the Global Legibility of a Generalized Map. *Geographic Information Science*, 2012.

[13] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2242–2251, Venice, October 2017. IEEE.

[14] A. Courtial, G. Touya, and X. Zhang. Generative Adversarial Networks to generalise Urban Area in Topographic Maps. In *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, volume XLIII-B4-2021, pages 15–22. Copernicus GmbH, June 2021. ISSN: 1682-1750.

[15] Xiongfeng Yan, Tinghua Ai, Min Yang, and Xiaohua Tong. Graph convolutional autoencoder model for the shape coding and cognition of buildings in maps. *International Journal of Geographical Information Science*, 0(0):1–23, May 2020, doi:10.1080/13658816.2020.1768260. Publisher: Taylor & Francis.

[16] Guillaume Touya and Imran Lokhat. Deep learning for enrichment of vector spatial databases: Application to highway interchange. *ACM Trans. Spatial Algorithms Syst.*, 6(3), apr 2020, doi:10.1145/3382080.

[17] Azelle Courtial, Guillaume Touya, and Xiang Zhang. Can Graph Convolution Networks Learn Spatial Relations? *Abstracts of the ICA*, 3:1–2, December 2021, doi:10.5194/ica-abs-3-60-2021.

[18] Azelle Courtial, Achraf El Ayedi, Guillaume Touya, and Xiang Zhang. Exploring the Potential of Deep Learning Segmentation for Mountain Roads Generalisation. *ISPRS International Journal of Geo-Information*, 9(5):338, May 2020, doi:10.3390/ijgi9050338. Number: 5 Publisher: Multidisciplinary Digital Publishing Institute.

[19] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. volume abs/1611.07004, 2016.

[20] Yuhao Kang, Song Gao, and Robert Roth. Transferring Multiscale Map Style Using Generative Adversarial Network. 2019.

[21] Stefan Werder. Formalization of Spatial Constraints. *In 12th AGILE International Conference on Geographic Information Science*, page 13, 2009.

[22] Guillaume Touya, Cécile Duchêne, and Anne Ruas. Collaborative Generalisation: Formalisation of Generalisation Knowledge to Orchestrate Different Cartographic Generalisation Processes. In Sara Irina Fabrikant, Tumasch Reichenbacher, Marc van Kreveld, and Christoph Schlieder, editors, *Geographic Information Science*, Lecture Notes in Computer Science, pages 264–278, Berlin, Heidelberg, 2010. Springer.

[23] Anne Ruas. *Modèle de généralisation de données géographiques à base de contraintes et d'autonomie*. phdthesis, Université de Marne la Vallée, 1999.

[24] Martin Galanda. Modelling constraints for polygon generalisation. In *Proceedings of 5th workshop on progress in automated map generalisation*. ICA, 2003. event-place: Paris, France.

[25] Andriani Skopeliti and Lysandros Tsoulos. A methodology for the assessment of generalization quality. In *Fourth Workshop on Progress in Automated Map Generalization*, pages 2–4, 2001.

[26] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. volume abs/1505.04597, pages 234–241, 2015.

[27] A. Courtial, G. Touya, and X. Zhang. Alpinebends – a benchmark for deep learning-based generalisation. In *Abstracts of the ICA 4:1–2.*, 2022.

[28] Sébastien Mustière. GALBE: Adaptive Generalisation - The need for an Adaptive Process for Automated Generalisation, an Example on Roads. In *1st GIS'PlaNet conference*, Lisbon, Portugal, 1998.

[29] Cécile Duchêne. Individual Road Generalisation in the 1997-2000 AGENT European project. Technical report, IGN, COGIT lab, Saint-Mandé, France, 2014.

[30] Azelle Courtial, Guillaume Touya, and Xiang Zhang. Constraint Based Evaluation of Generalized Images Generated by Deep Learning. In *23rd ICA Workshop on Map Generalisation and Multiple Representation*, Delft, Netherlands, November 2020. ICA.

[31] Ruth Rosenholtz, Yuanzhen Li, and Lisa Nakano. Measuring visual clutter. *Journal of vision*, 7:17.1–22, February 2007, doi:10.1167/7.2.17.

[32] Marion Dumont, Guillaume Touya, and Cécile Duchêne. Assessing the Variation of Visual Complexity in Multi-Scale Maps with Clutter Measures. *ICA Workshop on Generalisation and Multiple Representation*, 2016.

[33] Guillaume Touya, B. Decherf, M. Lalanne, and Marion Dumont. Comparing Image-Based Methods for Assessing Visual Clutter in Generalized Maps. *ISPRS Annals of Photogrammetry, Remote Sensing andSpatialInformation-Sciences*, II-3/W5, 2015, doi:10.5194/isprsannals-II-3-W5-227-2015.

[34] Mary J. Bravo and Hany Farid. A scale invariant measure of clutter. *Journal of Vision*, 8(1):23–23, January 2008, doi:10.1167/8.1.23. Publisher: The Association for Research in Vision and Ophthalmology.

[35] Gaurav Sharma, Wencheng Wu, and Edul N. Dalal. The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application*, 30(1):21–30, February 2005, doi:10.1002/col.20070.

[36] Huan Fu, Mingming Gong, Chaohui Wang, Kayhan Batmanghelich, Kun Zhang, and Dacheng Tao. Geometry-Consistent Generative Adversarial Networks for One-Sided Unsupervised Domain Mapping. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2422–2431, Long Beach, CA, USA, June 2019. IEEE.

[37] William A. Mackaness and Geoffrey Edwards. 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*, pages 7–8, July 2002.