HAL open science

Deriving map images of generalised mountain roads with generative adversarial networks

Azelle Courtial, Guillaume Touya, Xiang Zhang

To cite this version:

Azelle Courtial, Guillaume Touya, Xiang Zhang. Deriving map images of generalised mountain roads with generative adversarial networks. International Journal of Geographical Information Science, 2023, 37 (3), pp.499-528. 10.1080/13658816.2022.2123488 . hal-03825042

HAL Id: hal-03825042

https://hal.science/hal-03825042v1

Submitted on 21 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

DERIVING MAP IMAGES OF GENERALISED MOUNTAIN ROADS WITH GENERATIVE ADVERSARIAL NETWORKS.

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/10.1080/13658816.2022.2123488

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

October 21, 2022

ABSTRACT

Map generalisation is a process that transforms geographic information for a cartographic at a specific scale. The goal is to produce legible and informative maps even at small scales from a detailed dataset. The potential of deep learning to help in this task is still unknown. This article examines the use case of mountain road generalisation, to explore the potential of a specific deep learning approach: generative adversarial networks (GAN). Our goal is to generate images that depict road maps generalised at the 1:250k scale, from images that depict road maps of the same area using un-generalised 1:25k data. This paper not only shows the potential of deep learning to generate generalised mountain roads, but also analyses how the process of deep learning generalisation works, compares supervised and unsupervised learning, and explores possible improvements. With this experiment we have exhibited an unsupervised model that is able to generate generalised maps evaluated as good as the reference and reviewed some possible improvements for deep learning-based generalisation, including training set management, and the definition of a new road connectivity loss. All our results are evaluated visually using a four questions process and validated by a user test conducted on 113 individuals.

Keywords Deep learning · Mountain Road · Map Generalisation · Generative Adversarial Networks

1 Introduction

Map generalisation seeks to transform a detailed geographic dataset in order to display it as a small-scale map. This process allows a reduction in the amount of data acquisition, as maps at smaller scales can be derived from the detailed data used to create the maps at large scales. Automated map generalisation has been explored for decades in research and today some medium or small-scale maps can be automatically generalised [Duchêne et al., 2014]. But many challenges remain as maps can have very diverse source data, styles, and output scales, and current automatic methods seem to be insufficient to deal with this diversity.

Deep learning techniques have succeeded in resolving more and more computer vision tasks in several domains, such as increasing the resolution of an image [Ledig et al., 2017], or colorizing a black and white image [Nazeri et al., 2018]. Currently, most deep neural networks deal with image samples, and the more common tasks are image classification, image segmentation, and image generation. The definition of map generalisation as a raster problem should permit the

use of deep learning networks. [Touya et al., 2019] explain that: (1) images of maps capture most of the constraints that guide generalisation (minimum size, granularity, shape preservation, etc.); (2) the human decisions to generalize a map are very complex to model explicitly; (3) with all the existing maps generalized manually or semi-automatically, there is a large corpus to create massive training examples. However, the use of these techniques in map generalisation raises some new research challenges:

1. Deep neural networks learn to reproduce properties from samples, consequently input and target images used to train them have to be designed carefully. The images have to reflect the target generalisation, but also show the information necessary for generalisation. The identification of the required knowledge and its representation is a central point in machine learning-based generalisation [Weibel et al., 1995, Sester, 2000]. For example in road bend series generalisation, all bends have to be considered to know which one(s) should be removed [Lecordix et al., 1997]. Consequently, the segmentation of the study area into image examples is also a key issue [Du et al., 2021]. Finally, the diversity, quality, and quantity of situations and examples are decisive factors in deep learning.
2. What is the most adapted learning task? Deep learning can resolve several kinds of problems and the objective formulation can lead to different results. For example, the problem of end-to-end map generalisation could be expressed as a style transfer problem or as an image generation task [Isola et al., 2017]. And even intermediary generalisation tasks can be resolved using several tasks, e.g. the recognition of highway interchanges is a useful pre-process for generalisation and can be achieved using a network that classifies images, or using an image segmentation network that finds the pixels covering the interchange [Touya and Lokhat, 2018].
3. Which network architectures and parameters give the best results, and how to improve them regarding a generalisation problem? The deep learning state of the art is really rich but there are few studies conducted using deep learning in cartography. Moreover, a better network on a specific task would not necessarily be adapted for another problem. We have to define the characteristics of a network architecture adapted for map generalisation. For example, a network designed for style transfer is a good option, as the output image is created by altering the input image. We can note two kinds of approaches for the choice of network architectures: designing a specialized network for the use case [Ganguli et al., 2019] or using a generic network (that could have shown some expected characteristics on other use cases and/or transferability) and eventually adapt it to geographic data [Fu et al., 2019].
4. How to evaluate a deep learning prediction? Evaluation is an important question in generalisation. Map generalisation should preserve the information while increasing the legibility to respond to a predefined function. However, with a deep learning approach, a third point has to be assessed: the result should look like a real map. Moreover, most of the existing measures are calculated on vector data and a conversion back to vector is unsuitable in an iterative process such as the deep neural network refinement [Courtial et al., 2022].
5. How to integrate deep learning results in a generalisation framework or how could we use it as an end-to-end process? The conversion back to vector format can be described as the first step through integration [Du et al., 2021].
6. Is the raster representation of information relevant for map generalisation? Raster-based generalisation has been very rarely studied compared to vector-based generalisation. Nowadays, the input data are mainly stored and shared using vector formats. The conversion from vector to raster is not cost-less [Peuquet, 1979]. The most important challenge of our research would be induced by this data format and the fact that a raster representation of a map hides important information [Touya et al., 2019].
7. Can deep learning generalisation preserve the topology of linear objects? Relations between map objects are central for a proper map use and the preservation of these relations is mandatory for map generalisation. For example, roads are a part of a network and network connectedness is necessary to follow a route on the map. Raster-based approaches do not model topology, consequently defining a model that preserves it with images is a challenge.

The goal of this article is to explore the potential of generative adversarial networks (GAN) for a classical use case in map generalisation: mountain road generalisation. Our strategy is to adapt a generic deep neural network to generate images of generalised roads. Mountain roads are often very sinuous, which causes symbol coalescence when scale decreases. This generalisation involves simplification, caricature, typification, displacement, and smoothing operations [Roth et al., 2011], and frequently requires the combination of several algorithms. Generative adversarial networks were already used to deal with comics sketch simplification [Simo-Serra et al., 2016], map style transfer [Kang et al., 2019], and coastline generalisation [Du et al., 2021]. Moreover, the main problem of previous attempts to generalise mountain roads with deep learning, was the realism of generated images [Courtial et al., 2020] (e.g. producing geometries that are not smooth or straight enough to be real cartographic features), and GANs are specialized in realistic image generation.

In this article, we first review related research. Then, we present our experiments, especially the creation of our training images set, its improvement, the choice of the network architecture, and loss function. In the fourth part, we present and evaluate our results. Finally, we discuss the potential and limits of deep learning for mountain road generalisation.

2 Related work

2.1 Mountain road generalisation

Mountain road generalisation is a classical road generalisation problem, where the key issue is the symbol coalescence due to road sinuosity. Consequently, mountain road generalisation is often computed using vector data format rather than raster data, and is composed of two steps: first, road selection allows the elimination of insignificant roads, then graphical generalisation removes symbol coalescence and smooths the geometries. Road selection is based on shape characteristics [Thomson and Richardson, 1999], semantics features (importance, status, width, etc.), and the role played in the road network [Garcia-Balboa and Ariza-López, 2008]. Most of these elements are not accessible in a small image of the map, showing only an extract of the network, so we believe that convolutional neural networks and especially image representation are not really adapted for this step [Touya et al., 2019]. Moreover, many methods succeed to combine these factors and properly select roads (e.g. [Touya, 2010, Benz and Weibel, 2013] , etc.). Then, graphical generalisation adapts road line shape to be legible at a target scale. We distinguish two kinds of road simplification processes, (1) a global approach that tries to find the optimal shape [Bader and Barrault, 2001, Sester, 2005]; (2) iterative processes that separate roads in homogeneous sections, apply several algorithms [Lecordix et al., 1997] to each of them, and merges them together [Mustiere, 1998, Duchêne, 2014]. A deep learning approach would be similar to the global approach as the model is learned from multiple situations and aims to fit a wider set of situations. Finally, traditional methods for mountain road generalisation are already efficient at generalizing mountain roads. However, as roads fulfil a major role in maps, their generalisation is always the first step for map generalisation and we argue that it would be impossible to learn more complex generalisation without learning first how to generalise roads.

2.2 Raster-based map generalisation

In the early years of research on map generalisation, raster-based and vector-based approaches were developed in parallel as both approaches were equally promising [Peuquet, 1979, Monmonier, 1983, McMaster and Monmonier, 1989, Li, 1994, Li and Su, 1995]. In the '90s, vector-based generalisation finally out-performed raster-based generalisation [Peter and Weibel, 1999, Daley et al., 1997]. As generalisation was more and more monitored by constraints on objects of the map, the vector-based methods appeared to be more practical to use. In recent years, new raster-based generalisation algorithms were very rare [Pantazis et al., 2009], besides terrain generalisation where raster-based methods are prominent [Raposo and Samsonov, 2014, Guilbert et al., 2014].

2.3 Deep learning and map generalisation

The potential of this approach is theoretically explored in [Touya et al., 2019]. In this domain, first tests explored deep convolutional networks designed for image segmentation called U-Nets [Ronneberger et al., 2015] for the segmentation of generalised buildings [Feng et al., 2019], generalised mountain roads [Courtial et al., 2020] or for the generation of the relief shading of a topographic map [Jenny et al., 2020]. The principle is to use an image of the detailed shape of the road, to predict an image of the same size where each pixel value represents the probability to belong to the generalised shape. These attempts have shown the potential of deep learning to produce abstract representations from cartographic data and for graphical generalisation, but they are not adapted for the generalisation of multiple objects and the results for mountain road generalisation suffer from a lack of realism. When we use the concept of realism in this paper, we do not mean that a realistic image looks like the ground truth, as usual in cartography where realism is the opposite of abstraction. Here, a realistic image generated by deep learning looks like a real map crafted with traditional tools.

Generative adversarial networks (GAN) are deep learning models designed to generate photo-realistic pictures. They combine a classification network called discriminator to an image transformation network called a generator. Both these networks work in opposition: the discriminator trains to distinguish a fake generated image from real target images, while the generator tries to produce an image that fools the discriminator. Consequently, GANs are designed to create an image that looks like an image of the target domain, called prediction, from an input image of a different domain. For example, [Simo-Serra et al., 2018] have illustrated the ability of GAN for a problem that has similarities with map generalisation: the transformation of a comic sketch into comic clean image, which involves the selection, simplification and smoothing of drawing lines. Moreover, there were experiments of map style transfer using a GAN, where the GAN generates maps in the style of Google Maps from images of OSM vector data [Kang et al., 2019]. The authors noted

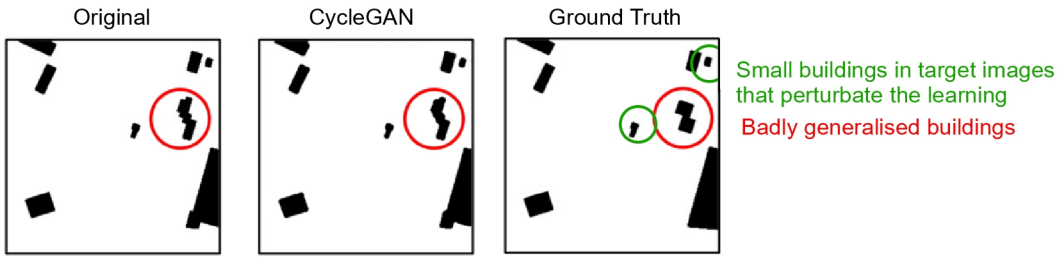


Figure 1: The GAN generalisation of building shapes failed. Results from [Kang et al., 2017].

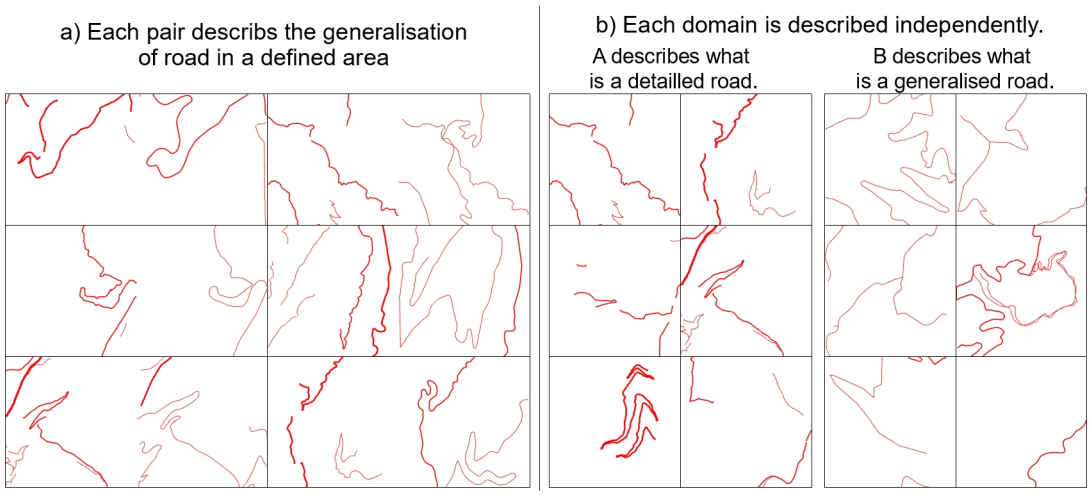


Figure 2: Example of tiles from paired (a) and unpaired (b) datasets for roads generalisation.

that the generated maps did not contain symbol coalescence as if the following generalisation operators were applied: selection, enhancement, and typification. Finally, the same authors also explored the improvement brought by a GAN architecture for building generalisation [Kang et al., 2017]. Their experiment did not succeed in improving the quality of the predicted images compared to the segmentation approach: some fuzzy/unrealistic shapes appeared and several buildings were not simplified enough (circled in red in Figure 1). Moreover, some buildings with edges smaller than the target minimum detail size remain in the target image of this experiment (circled in green in Figure 1). Consequently, the network seems unable to learn how to simplify buildings that have very small edges.

Then, the fuzzy borders are a problem for buildings that should have, most of the time, regular shapes. However, a smooth output geometry is expected for a mountain road, so this tendency to smooth regular shapes might not be a problem for the generalisation of roads. We believe that a GAN architecture could address the realism problem encountered with mountain road generalisation with U-Nets. Generative adversarial networks have also proved to be useful for coastline generalisation [Du et al., 2021].

GAN can deal with paired images [Isola et al., 2017] in a supervised way, or with unpaired images [1] [Zhu et al., 2017, Yi et al., 2017, Gan et al., 2017, Benaim and Wolf, 2017]. Using paired images for training means that you pair an input image with the ideal output image while using unpaired images means that you train the model with a set of good input images and a set of good output images (Figure 2).

Learning with paired images often permits to learn some features that unpaired models do not when input and target are similar, but when they are very different, or when it is difficult to find such paired data, unsupervised learning is the most adapted. Finally, semi-supervised learning represents a key issue for GAN improvement. It consists in combining learning on few paired and many unpaired examples. This approach succeeds to improve models for the aerial image to map transformations [Chen et al., 2020]. However, GANs still tend to produce a distortion of geographic objects' geometry. To solve this problem, [Fu et al., 2019] propose to constrain the geometric consistency of the output image: the prediction of a transformation of an image has to look like the transformation of the prediction.

[1] GAN designed for unpaired images are often called unsupervised in the deep learning literature.

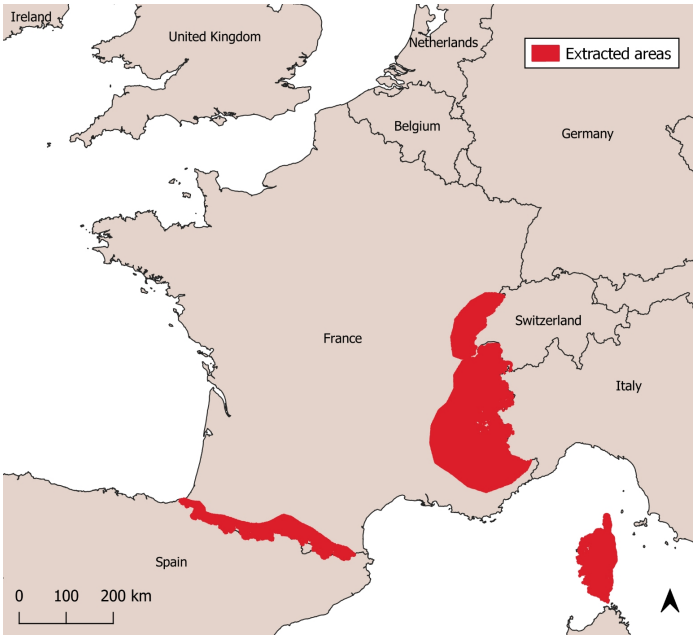


Figure 3: Situation map of the extracted areas, covering a large part of the mountain roads in France.

3 Methodology

In order to address our research questions, we set-up several experiment. The material for these experiments is presented in the following section.

3.1 Training set

We created a training set composed of the roads from the French national map agency (IGN) in maps at 1:25,000 and 1:250,000 scales. Figure 3 presents the different geographic areas selected during our experiment; it includes French high mountain areas where the most sinuous roads can be found (Corsica, the Alps, the Jura, and the Pyrenees). Roads are stored in cartographic databases [Grünreich, 1985] as vector data. The less detailed dataset has been derived from the detailed one many years ago using generalisation algorithms [Mustiere, 1998] and manual corrections, and then they had been separately updated (there are inconsistencies between both scales). As we only want to learn the graphical generalisation (excluding the road selection step), we decided to pre-process the detailed data and focus only on roads that are kept in the generalised dataset. We used a multi-criteria data matching algorithm to match both scales [Olteanu-Raimond et al., 2015], using topology, shape, and distance criteria. Then, the automated results were refined manually.

There are three possible methods to split the study area into small sample images. The object-based method consists in representing one entire road on each tile; this method produces images at different scales and it is not really optimal for map generalisation [Courtial et al., 2020]. The second method is a variation of the first one [Du et al., 2021]: tiles are created around segmented small portions of the line objects and are rotated following the main orientation of the line portion. This method seems to be adapted for coastlines but it does not fit the mountain road use case: contrary to coastlines, roads are parts of a network with many intersections, and it is not really possible to find tiles containing only one road, particularly at intersections. Furthermore, in road generalisation, it is important to have a broad view of the context around the road, to make sure the distortion caused by the simplification of the road does not create an overlap with a neighbouring road. As this second method produces very small tiles, the view of the context is not broad enough. We employed a third method that creates tiles using a fixed-size square window sliding on the study area, with overlaps between two tiles. We choose a tile size of 2.5*2.5 kilometres, and we draw red roads on a white background. The width of the road symbol is proportional to the road importance. We detail the effect of scale and resolution in next section.

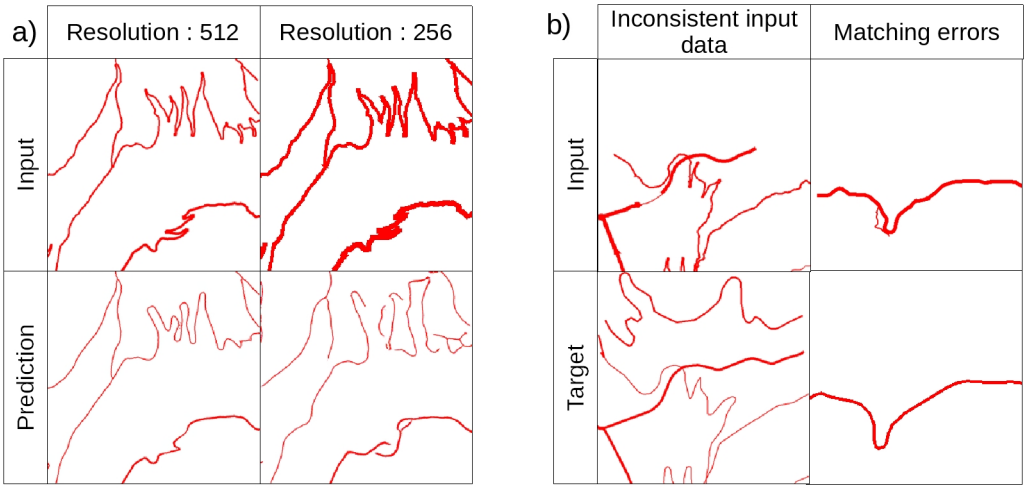


Figure 4: Example of quality issues: a) Effect of input image resolution on the prediction, b) example of different errors in our dataset. All tiles represent a part of the map of 2.5km².

3.2 Quality and quantity of training examples

It traditionally is assumed that the quality of the learning is positively influenced by the quantity and diversity of the learning examples, and the more complex is the task the more examples are needed. One standard way to improve deep learning is data augmentation, which increases the number of training examples by transforming the training images. However, experiments with a segmentation network [Courtial et al., 2020] have shown that these techniques are not adapted for our use case. Therefore, we rather decided to explore the effect of image quality on generalisation.

The first point for image quality is legibility. The network is unable to learn from unclear images, so we need to minimize coalescence even in the input images. With a specified tile size, there are two ways to improve legibility: reducing the road symbol width, or increasing the image resolution. We tested both options. In the default training dataset, the road symbol width is between 1 and 6 pixels in accordance with their importance (the roads can have 6 values of importance). Hence, roads with a medium to large importance, and a high sinuosity have low legibility. Therefore, we tested to reduce the width of roads between 1 and 3 pixels. We faced the two following problems: (1) the importance of the road is a key contextual element for generalisation learning [Courtial et al., 2020] and we were unable to generate a good generalisation without this information; (2) in case of an insufficient quantity of road pixels on images (due to few roads), the network learns to generate only background pixels, i.e., white images. Resolution represents details an image holds and can be described by pixel count in images of a fixed size. Most of the deep learning models are adapted for square images with a number of pixels equal to a power of two, in order to ease convolutions. Our experiments show that the use of 512*512 pixels in the image, instead of 256*256 improves the quality of results, especially to increase legibility when the input is complex (Figure 4.a.).

Then, the quality and consistency of images (input and target) are important for supervised learning because errors in the training set may perturb the learning process and be reproduced in predictions. With unpaired models, errors impact the domain characterization. In our dataset, the two main sources of inconsistencies are update differences in the data, and errors in pre-processing matching; two examples are presented in Figure 4.b. As a preliminary experiment, in order to validate the quantity and quality of our training set, we compared several subsets of increasing size. We also compared the impact of adding image tiles produced without the data matching manual refinement (called low-quality images in Figure 5). We observe that the quality of the final result varies more according to the quality of the data matching than according to the number of images. We conclude that both 1 000 and 2 000 images training sets are sufficient for our target task and that adding low-quality images cannot improve the prediction.

Finally, we could handle the quality of images by filtering the image set, thus eliminating examples that are not relevant for learning, with errors or with very few roads, or with a high difference between input and target due to matching or update inconsistencies.

All these points show that the improvement of the training set is a key first step for deep learning-based map generalisation. It involves improving data pre-processing, data representation, tile segmentation, image resolution and image selection.

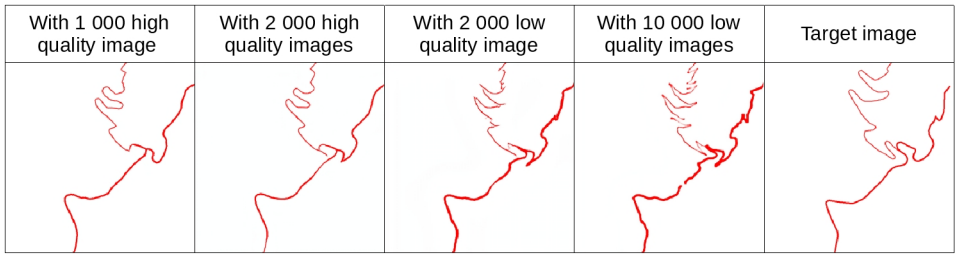


Figure 5: Predictions made by models trained with different quantities and qualities of image. All tiles represent a part of the map of 2.5km².

	Ratio of road pixel (%)	Sinuosity (%)
Relative standard deviation for Model A	92.08	17.75
Relative standard deviation for Model B	85.07	12.72

Table 1: Some diversity measures for the images used in the training sets.

3.3 Diversity and quantity of training examples

The question of example diversity is another important point in deep learning. The training set needs to cover the complex configurations of mountain roads with enough examples to correctly learn how to generalise them. Our preliminary experiments, not detailed in this manuscript, show that the worst results are often generalisations of uncommon configurations of roads. This can explain why using basic data augmentation techniques such as translation or rotation did not proved successful in the past [Courtial et al., 2020]. On the other hand, we propose to experiment diversity augmentation using a different geographic area. More complex road configurations are added with this new area, which should improve the results for these complex configurations without decreasing result quality for the most common situations. However, it is not so easy to heavily augment our training dataset as the number of complex mountain roads in a country is limited.

We also propose some metrics to evaluate the diversity of our images. We can verify the diversity of image characteristics (ratio of road pixels over background pixels, spatial distribution of road pixels, amount of coalescence, etc.) and the diversity of roads in the training area using the vector data (sinuosity, importance, etc.). These measures helped us to choose the images to add to the training set, by removing the redundant ones with a few straight lines. Table 1 presents some of these values related to diversity for the training sets of models A and B.

3.4 Supervised learning

In this section and the following one, we present the paired and unpaired techniques we experimented.

Pix2pix [Isola et al., 2017] is a supervised generative adversarial network that has been designed for a generic image to image translation with paired training data. It was employed for tasks such as night to day photograph conversion, colorization of images, or the conversion from aerial photographs to map images. This neural network opposes a generator network to a discriminator network, in order to generate realistic images. The discriminator is a convolutional neural network that classifies images to be realistic for the target domain, or not (in our case, the target domain is an image of roads generalised for the 1:250k scale). The generator is also a convolutional neural network that generates an image in the target domain (generalised roads) from an image of the input domain. As they are trained together, the generator learns to generate images that are classified as realistic images of the target domain by the discriminator.

Both the generator and discriminator are made of series of convolutions. The generator (G) is a U-net-like network [Ronneberger et al., 2015] that learns a mapping from an input image x and a random noise vector z to output an image y by reducing the information dimension of the input image through a series of convolutions, and then recreates an image using up-convolution. The discriminator (D) is a classifier that contains down convolutions and indicates the probability for a pair to be true.

Figure 6 represents one iteration of the learning process for the mountain road generalisation use case. Initially, the generator applies the random noise z to the input and produces a prediction. Then the discriminator examines if the pairs x-prediction and x-y are real. The correctness of the first prediction enables the adjustment of the generator weights, while the first and second predictions are used to adjust the discriminator weights.

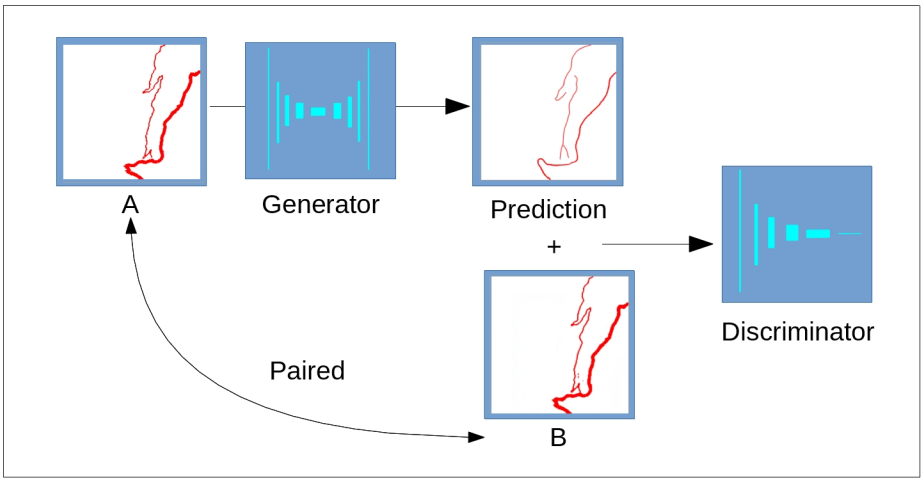


Figure 6: A learning iteration for the pix2pix network. The Generator convolutional network generates an image of domain Y, from an image of domain X, while the Discriminator convolutional network classifies the prediction generated by the Generator as realistic for domain Y, or not. Illustrated with a tile couple of 2.5km² from the training set.

In this manner, the objective of the network is to maximise the generator ability to generate images the discriminator classifies as realistic $log(1 - D(x, G(x, z)))$ while minimizing the discriminator error $log(D(x, y))$. Thus, the network conditional loss measures the following expectations:

$$L_{cGAN}(G, D) = E_{x,y}[logD(x, y)] + E_{x,z}[log(1 - D(x, G(x, z))] \tag{1}$$

It additionally includes a loss term that measures the expectation that the prediction $G(x, z)$ looks like the ground truth y according to the L1 distance:

$$L_{L1}(G) = E_{x,y,z}[\|y - G(x, z)\|_1] \tag{2}$$

Thus, the global objective $G*$ to find the generator that minimizes both losses while the discriminator maximizes the network conditional loss, is the following:

$$G* = arg\ min_G\ max_D\ L_{cGAN}(G, D) + \lambda L_{L1}(G). \tag{3}$$

3.5 Unpaired learning

We also tested an unpaired GAN, using an unpaired training set constructed on the same area as for paired learning. This strategy is usually better when the domains of the input and output images are significantly different (e.g. generating a painting from a photograph, or a zebra image from a horse image) or when the paired images are difficult to obtain. We choose to use the CycleGAN network introduced by [Zhu et al., 2017]. Its principle is to learn together the conversion from domain Y to domain X and the one from domain X to domain Y. Consequently, the prediction is not compared to a reference but converted back in the initial domain and the reconstruction is compared to the initial image. Figure 7 is an illustration of one iteration of the learning process. Each image is translated twice using both generators (F and G), then the following elements are evaluated: the probability of the first prediction to be fake, the probability of the reconstruction to be true and the proximity between the output image and the initial image. Consequently, the network loss is the sum of the two adversarial losses (4) and a cycle consistency loss (5). The adversarial losses oppose G the generator from X to Y domain to D_Y the discriminator from Y domain (and respectively, F from Y to X and D_X for X domain). The cycle consistency loss represents the expectation of the reconstruction to be similar to the initial image.

$$L_{GAN}(G, D_Y, X, Y) = E_y[logD_Y(y)] + E_x[log(1 - D_Y(G(x))] \tag{4}$$

$$L_{cyc}(G, F) = E_x[\|F(G(x)) - x\|_1] + E_y[\|G(F(y)) - y\|_1] \tag{5}$$

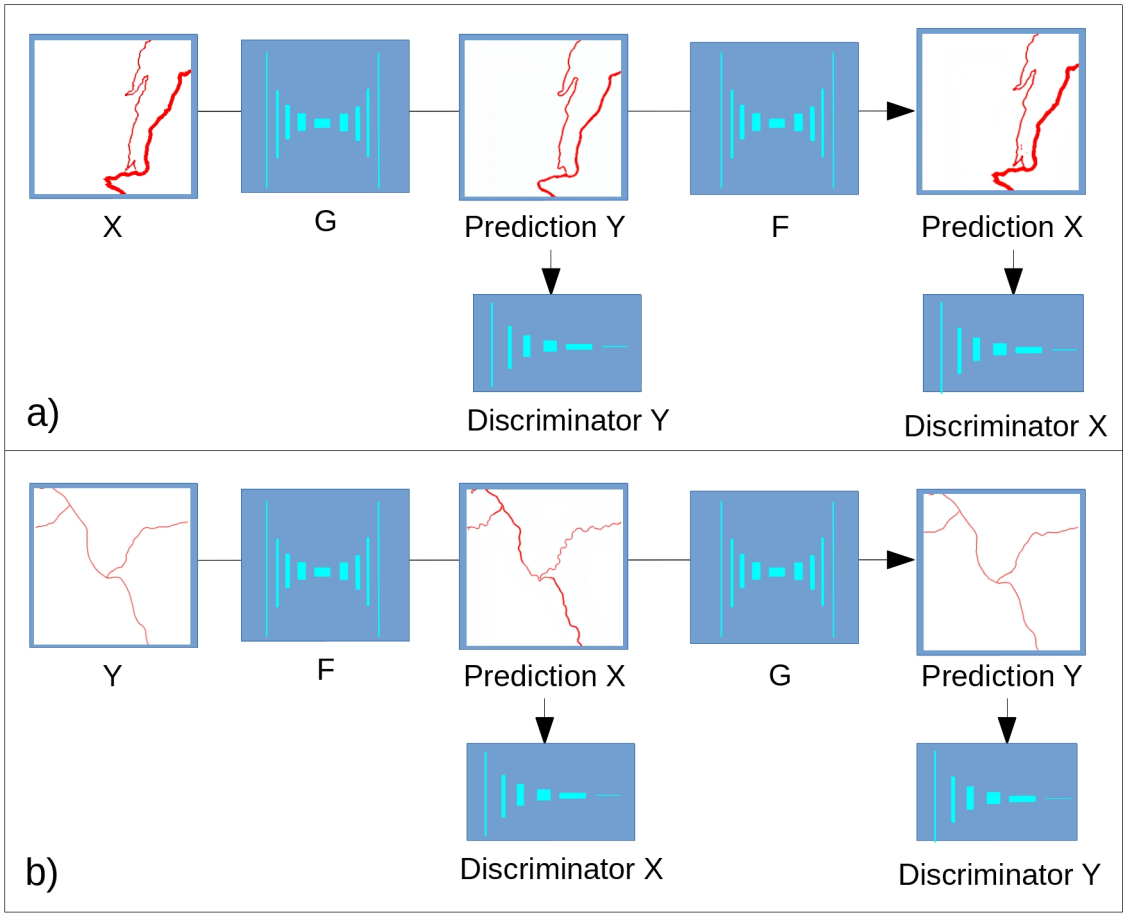


Figure 7: One learning iteration for CycleGAN. a) from X to Y. b) from Y to X.

3.6 Loss function

Besides the enhancement of the training data, the second way to improve the results is to optimise the neural network architecture. An additional term to the loss function is proposed to improve road connectivity preservation, which is maybe the most important constraint of road generalisation. Loss functions are the functions the neural networks optimise during their training process. A new loss term can improve the overall output or constrain the predictions to avoid a specific error. We propose to focus on the second option, with a loss function term dealing with road connectivity preservation. Connectivity alterations include different types of errors in the structure of the network: the creation of loops, the connection or fusion of adjacent roads, the disconnection of several different roads, the separation of one road into several parts (internal disconnection), and the disconnection at the borders of the image. We notice that all these problems would change the number of sets of connected road pixels and/or background pixels (Figure 8).

The number of sets of connected background pixels increases when a loop is created, and the number of sets of connected road pixels increases when there is a disconnection. We can measure the network structure alteration by comparing these two values in the input and predicted images (Figure 9).

We combine the number of sets of connected background pixels $nback$ and the number of sets of connected road pixels $nroad$ for both input and prediction images (Eq. 6). We use the absolute value to avoid the creation and disappearance errors to make up for each other, and we divide the sum of differences by the number of different sets in the initial image in order to normalise the measure, as images sometimes include a large number of sets of connected pixels, and sometimes a very small number. This measure is only adapted for two-colour images, so with more complex images, this loss function should be adapted.

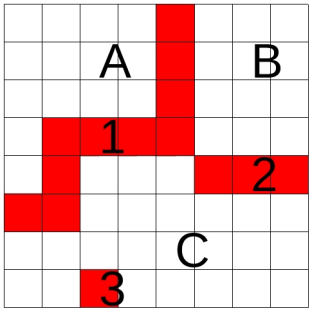


Figure 8: Example of road (enumerated from 1 to 3) and background (enumerated from A to C) pixel sets in a 8*8 pixels image.

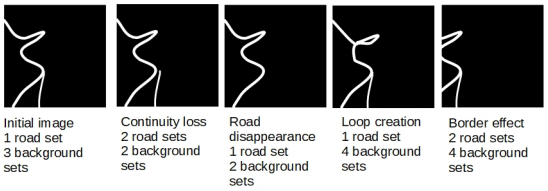


Figure 9: Example of road network structure alteration and the number of background and road set in an image not extracted from the training set.

$$L_{connectivity}(X,G) = E_x \frac{nback(x) - nback(G(x)) + nroad(x) - nroad(G(x))}{nback(x) + nroad(x)} \tag{6}$$

Then, this value is weighted by an importance value (λ_2) and added to the generator loss value of the network presented in equation 3.6. We tested different values show for λ_2, and the empirical best value is 10. L(G,F, $D_X, D_Y) = L_{GAN}(G, D_Y, X, Y) + L_{GAN}(F, D_X, Y, X) + \lambda_1 \ L_{cyc}(G, F)$
$+ \lambda_2 \ L_{connectivity}(X, G)$

3.7 Experiment

We carried out and compared the experiments presented in Table 2. Each model have been implemented in Python, with the PyTorch library, and trained over 200 epochs with default parameters. This number of epochs gives a good level of satisfaction for both the loss function values and the visual observation of validation data. The visual analysis of the results shows that below this number of epochs, the results are still unsatisfying while they show no more progress after this value for both paired and unpaired models.

4 Results

This section describes the results of the experiments presented in Section 3. The results are evaluated according to three methods. First, results are described using a visual evaluation. Then, an user test allows a more formal validation of the quality of the generated images in the Alps test area. In a third part, we describe how the generated images can be quantitatively evaluated. Finally, we use this quantitative evaluation to explore the potential of our model in different French regions.

Name	Strategy	Model	Sample image
Model A	Unsupervised	CycleGAN	1223 (from Alps only)
Model B	Unsupervised	CycleGAN	2271 (from several mountain regions)
Model C	Unsupervised	CycleGAN	1223 (+ connectivity loss)
Model D	Supervised	Pix2pix	1223 (from Alps only)

Table 2: List of the experiments presented in the paper.

4.1 Visual evaluation

Generalisation evaluation is traditionally conducted using expert visual evaluation or by measuring constraint violations on vector data [Mackaness and Ruas, 2007]. Consequently, as the second option is not really possible, we decided to evaluate our result employing visual analyses, focusing on a list of questions.

- **Q1**: Does the image look like a map and especially a road map? This question refers to the realism of the prediction. We trained the network with images of maps that follow cartographic rules, and the network should learn to generate images that look like a map.
- **Q2**: Are the geographic object credible? [Goodchild and Li, 2012] develop this idea in case of maps produced using volunteer geographic information (VGI) where the expertise of contributors is unknown and propose to control spatial relations between objects. This point is also important in the case of deep learning where the prediction process is a "black box"[Touya et al., 2019]. The key criteria for a road to be credible are its connectivity in the network and its linear and rather smooth shape.
- **Q3**: Is the image generalised enough to be legible at the target scale? In the use case of mountain roads, we have to take care of the smoothing level, and the reduction of symbol coalescence.
- **Q4**: Is the information preserved enough and are there errors created during the generalisation process? For instance, the GAN models tend to create fake loops in the road network (see results in Section 5).

Figure 10 presents the results for each method for 18 selected situations. Regarding the realism question (Q1), most of the images look like mountain road images, and for each strategy, the network seems to produce realistic mountain road images. However, the paired method (Model D) does not seem to have learnt how a mountain road on a generalised image really looks like: the shape of the road is not always credible. Regarding (Q2) for pix2pix, loops and disconnected roads are rare in a mountain road network, but abundant in the prediction of this method. Thus, these tiles are not functional for being part of a road map where the user wants to follow a route. Moreover, these elements affect the preservation of information (Q4): disconnections alter the connectivity of the road network, which is a key element in road map. Then, for the legibility question (Q3), in general, smoothing and bend enlargement seem to have correctly been applied, but for some complex situations, coalescence remains for all techniques, including the reference (e.g. images at lines 7, 8 and 14); in some cases, roads can be over-simplified (e.g. images at lines 13 and 18). Finally, legibility is troubled by the inability of the network to perform displacements in the case of close parallel roads (e.g., images at lines 8, 14, 15, and 17).

The benefit of adding a connectivity loss for cycleGAN can be observed by comparing Models A and C in Figure 10. At first sight, the new loss does not impact the final result, and we can even notice some visible disconnections (e.g., line 15), and some new connections (e.g., line 8). Moreover, the loop in line 7 has not been corrected and is balanced by a disconnection, so the number of sets of background pixels remains unchanged. However, the other evaluation criteria have not been affected and the measured connectivity was improved considering all test images, which is confirmed by the values of the loss function. This difference between measure and visual perception is explained by corrections that are hardly visible (e.g., small, at borders, not visually significant) and by a compensation phenomenon. For example, in the images at line 1, a very small set of background pixels on the left is better preserved with the new loss.

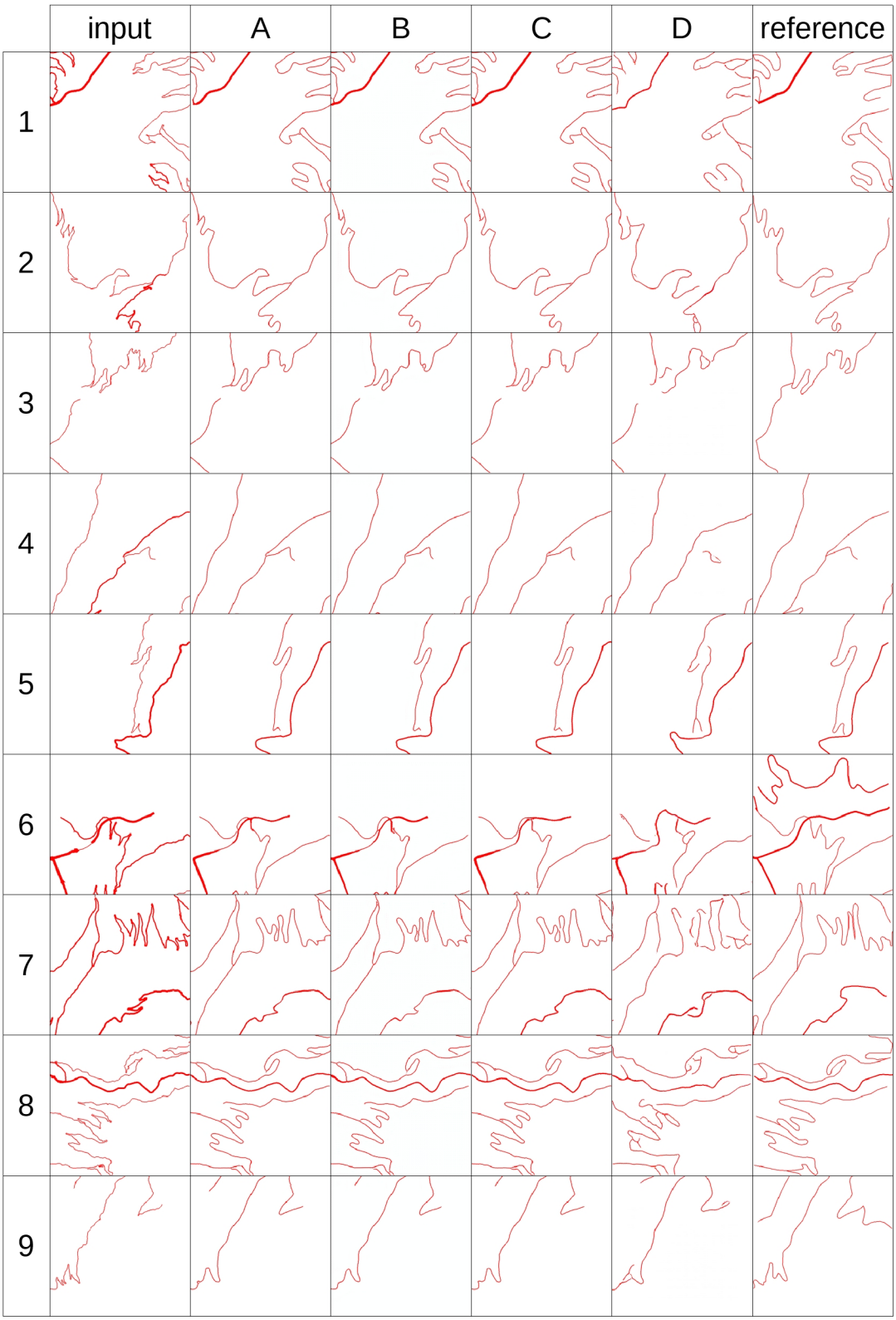
The benefit of adding training examples is not clear, as Models A and B have very similar results. We suppose that the differences in prediction are probably due to random effects during the learning process.

4.2 User test evaluation

We validate the visual evaluation by performing a user test. The goal is to compare the quality of the predictions according to the proposed models.

Task: The test is composed of two different tasks: an absolute evaluation (marking task) and a relative evaluation (ranking task). For both tasks, users do not know how each image has been produced. In the marking task evaluation, three questions are asked, and for each question, a Likert scale with four values from "very good" to "very bad" is proposed.

1. How well do you think the roads are generalised? The goal of this question is to evaluate if the generalisation operations have been applied properly.
2. How much do you think the generalised image seems realistic? This question evaluates if the images look like a map that has been traditionally produced.
3. How well do you think the generalised image is readable? This question evaluates the legibility of images.


input
A
B
C
D
reference
1
2
3
4
5
6
7
8
9

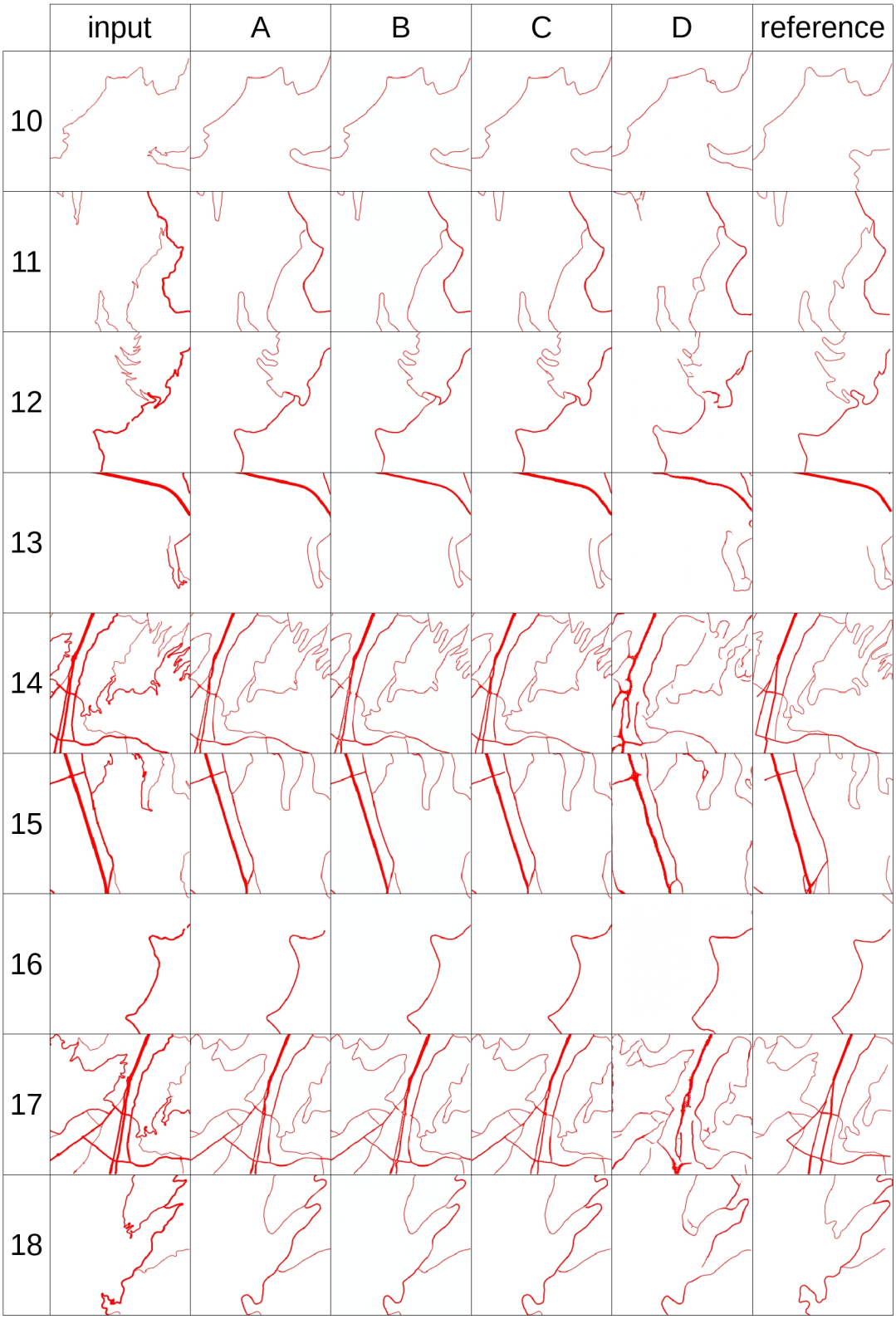


Figure 10: An extract of generated images for 18 situations from the testing set (unseen during the training) and Models A to D, Reference: state-of-the-art vector algorithms and manual corrections.) All tiles represent a part of the map of 2.5 km².

In the ranking task, participants have to drag and drop five images of generalisation ordering them from the most appreciated one on the left, to the less appreciated one on the right. This task is the most commonly used in GAN evaluation procedures [Wang et al., 2018], because it verifies if predictions cannot be distinguished from real images. The ordering task seems an easier task for people who are not familiar with the expected output, while the marking task evaluates the methods independently and assesses different aspects of the output quality separately. Finally, the combination of these two tasks makes the user test less repetitive for the participants.

Hypothesis: We make the following hypotheses:

- H1: Predictions should not be distinguished from reference.
- H2: The unpaired model (Model B) performs better than the paired model (Model D).
- H3: The predictions achieved with the model that includes connectivity loss (Model C) should be appreciated more.

Participants: We received 113 responses during the two weeks of the test from nonpaid volunteers. The participant population is made up of men at 60%; only 10% of the participants are aged under 25 and more than 40% are aged over 50. Most of the respondents use maps weekly or more often, but have never heard of generalisation. The "expert population" is composed of twelve generalisation experts, twelve geo-visualization or cartography experts, and 25 other informed people (who declared having already heard about map generalisation).

Procedure: The user test was designed as a web questionnaire; the link has been addressed to two different e-mail lists: one composed of generalisation experts and one composed of diverse employees from our university. The test is designed to take around 20 minutes. The first page of this survey gives information about the objective of the test, the images presented, and the generalisation of mountain roads. This introduction page can be reached again at any moment of the test if the user needs it. The last page asks some general information questions, such as their level of expertise on maps and generalisation. For the main part of the test, we selected 30 situations (24 for the first task and 6 for the second task), and we distributed the images using a latin square (for each situation and method) in 8 test configurations, which were randomly assigned to respondents. Each test is organised as follows. First, users are asked to complete 5 marking tasks. Then, they are asked to complete 1 ranking task. Then they repeat the process with 5 new marking tasks and one additional ranking task. The statistical analysis of the results is based on the marking task, Likert scale values being converted into marks between 0: "very bad" and 3: "very good". Then statistical tests are performed: M_r represents the mean realism rate (between 0 and 3) and M represents the mean of the sum of each mark (realism, generalisation and legibility, between 0 and 9).

Results: We obtained between 2 and 20 responses on each image. This difference is due to the random distribution of tests. All images with more than five responses are included in the evaluation. Figure 11 represents the mean mark given to all images for each method. These results are shown for all participants, but it is interesting to note that results are similar with the experts only. First of all, the unpaired model gives better quality results (M=6.43) than the paired one (M=3.32), an unilateral t-test allows to reject H0: the learning approach does not affect the generalisation (t(7) = 7.70, p $<<$ 0.01)) and confirms H2. Consequently, it is more adapted for our use case to learn "what is a generalised road?" instead of "how to generalise a road?". The ranking evaluation also confirms this idea; Model D is classified as the last one for 72.5% of the trials.

A second observation is that the CycleGAN method and its improvements (Models A, B, and C) are marked better than the reference image for the generalisation question (1) and at least as good as the reference for the realism (2) and legibility (3) questions. The mean realism mark of images from reference (M_r=1.90) and Model A (M_r=2.05) are similar. A paired t-test on realism mark means for each image shows that these marks are not significantly different (t(8) = 0.26, p= 0.79), which validates (H1). Moreover, the standard deviation is lower for CycleGAN images than for the reference images, i.e. there are more differences of quality perception between different situations when they are traditionally generalised than when they are generalised using deep learning. Deep learning seems to make map generalisation more uniform. The higher deviation for improved models (Models B and C) can be explained by the error correction on some images and some other images where errors are introduced compared to the standard CycleGAN. The mean marks of images from Models A (M=6.43) and C (M=6.51) are similar. A paired-unilateral t-test on marks shows that Model C is not significantly better than Model A (t(6) = 0.80, p= 0.77), which refutes (H3). The ranking evaluation gave similar results: the location of the reference in the ranking varies the most, and most of the time Models A, B, and C are well placed.

The study of some controversial images helps to explain this deviation more precisely. These images are grouped in Figure 12. Image 25 is interesting because it is the one where people less agree: 25% of the participants rank the reference as the best and 45% as the worst generalisation result. The left of this image is a very good generalisation, whereas on the right there is inconsistent information compared to the input image. This update inconsistency in our data is obviously avoided in the deep learning process because the image is generated from the input data. The same

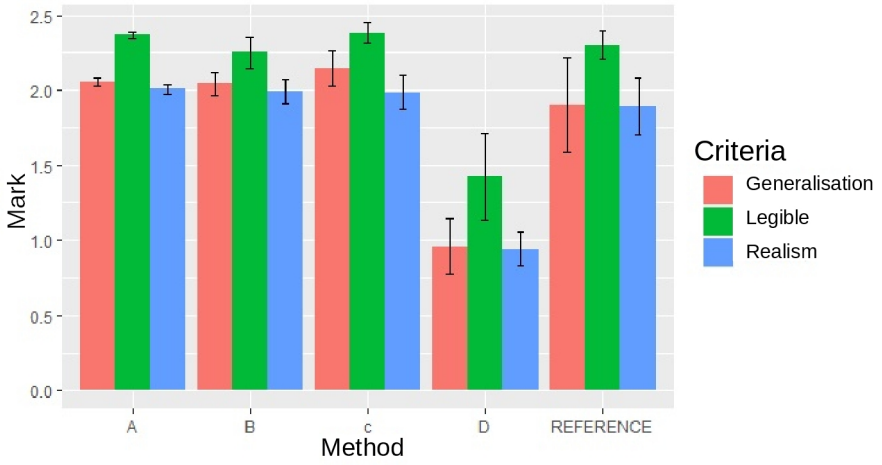


Figure 11: Mean evaluation of the different methods according to user marks.

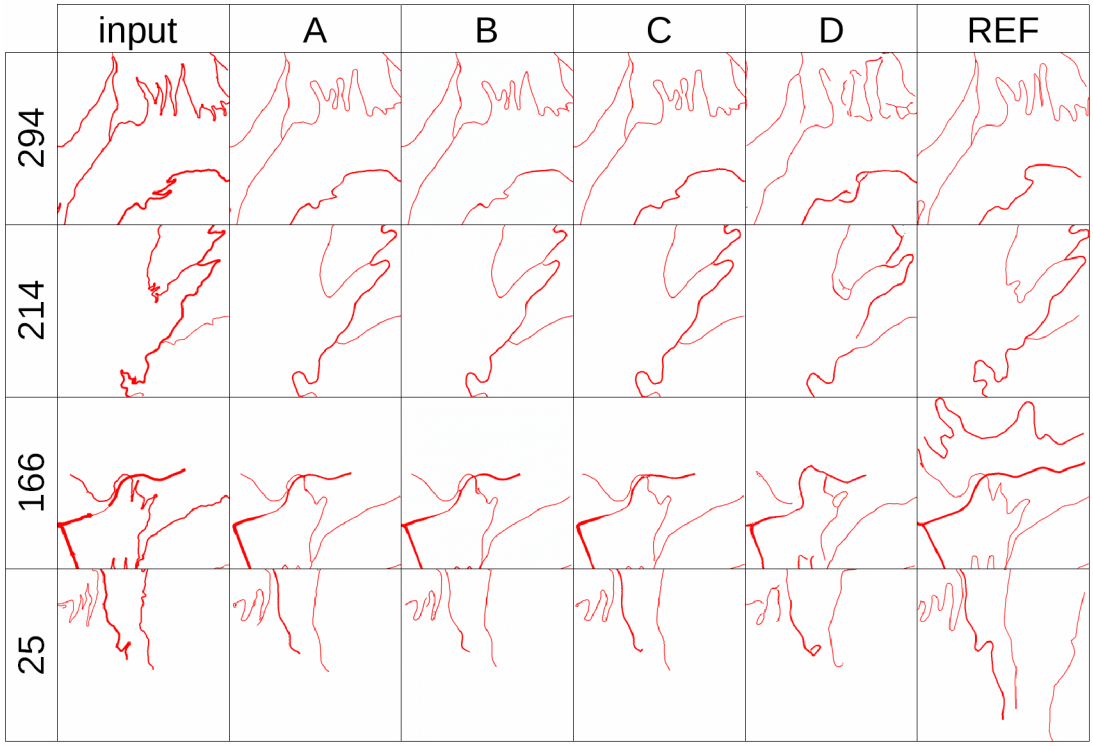


Figure 12: Example of images badly evaluated from the testing set. All tiles represent a part of the map of 2.5 km². Each line represents a situation and each column the learning method employed.

kind of problem is visible for Image 166. Then, regarding the realism question, there are some surprising answers: Image 294 has a very low realism for Model C only and Image 214 for all methods.

Finally, the user test reveals that Model B, the one trained with more examples of roads, seems to decrease the image legibility without really increasing the global quality. Consequently, the results of Model B show that adding examples does not necessarily improve the model when the diversity is not significantly increased.

4.3 Quantitative evaluation

The quantitative evaluation of GAN predictions is a complex issue [Shmelkov et al., 2018]: there is no unique correct solution to domain-to-domain transformation and the overall appearance of the generated image is more important than the pixel value similarity. If the evaluation could be based on the GAN loss function, none of them can indicate the general quality of generalisation. The adversarial loss (Eq. 4) describes the probability of the prediction to be seen as a real image and can contribute to evaluating the realism of the image. Moreover, the cycle loss (Eq. 5) evaluates the proximity between the initial and the generated images and can contribute to evaluating how much generalisation preserved the characteristics of the initial data.

Moreover, the statistical evaluation such as IoU or MSE are unsuitable to determine if a line is a good generalisation, as a road with almost no common pixel with the target can be a good generalisation, e.g. when a displacement occurs. Consequently, we choose to use constraints for a supplementary automated evaluation. Constraint-based evaluation aims to measure the violation of several constraints expected for good generalisation. [Courtial et al., 2022] proposed a set of constraints for mountain road generalisation that can be calculated on raster map images, which include constraints of realism and credibility for deep learning prediction. We decided to reuse the proposed following constraints:

- A smoothness constraint aims to ensure that the road shape in the prediction is smoother than in the initial image. It measures the quantity of un-smoothed pixels based on morphological operations on road pixels.
- A coalescence constraint evaluates the number of coalescent pixels using morphological operations on road pixels.
- A position accuracy constraint compares the relative position of roads in input and prediction images. It is calculated using the intersection over the union of road pixels.
- A connectivity constraint relies on the number of road and background pixel sets on input and prediction to measure disconnections.
- A colour constraint ensures the result is composed of red roads on a white background.
- A noise constraint measures the quantity of noise that perturbs the legibility of the prediction.

All these constraints are based on tiles and only evaluate a small extract of the map at a time. The region covered by one tile is not sufficient to evaluate the functional characteristics of the maps such as the ability to find and follow a route. However, the evaluation of the legibility, accuracy and credibility of each image is a necessary condition for the usability of the complete map. A grouping process that guarantees the correct combination of neighbouring tiles (especially regarding connection and topology) would give a usable map only if each tile is legible, credible and accurate. To show the interest of such constraints, they are used to assess how the unpaired model transfers to other mountainous areas in France.

The shape of roads is influenced by topographic elements (e.g. rivers and relief) and human factors (e.g. historical, social, and cultural factors). Consequently, mountain road generalisation can also vary in different contexts, because it is dependent on the initial shapes. The goal of this section is to analyse the ability of our model to adapt to situations in various geographical contexts.

We preprocessed and applied Model A to 30 images from five new areas: one in the south of the Alps (while the training images are from the north of the Alps), two in the Pyrenees, one in Corsica and one in the Jura mountains. Roads on images in Corsica and Jura are often thinner because of lesser importance, and the road density is smaller.

The generated images for all new areas are visually as good as in the initial area (Figure 13). All produced images look realistic (Q1) and are not noisy (see Table 4.3). The situations that are not credible (Q2) or where the information is not preserved (Q3), similarly to the initial test data, correspond to disconnections and loop creations (e.g. the Corsica image in Figure 13). The connectivity constraint is the one that varies the most. Produced images are legible (Q4) and smoothing is correctly transferred to all test areas and some simplifications and bend enlargements are correctly applied (e.g. at the bottom of Pyrenees West image in Figure 13), while some others do not (e.g. bottom of Alps south image). The coalescence and smoothness constraints are correctly satisfied during the transfer learning. Moreover, the constraint measures have a similar value for each of the test areas (Table 4.3).

5 Discussion

This experiment sheds light on some important issues about learning generalisation techniques that are detailed in this section.

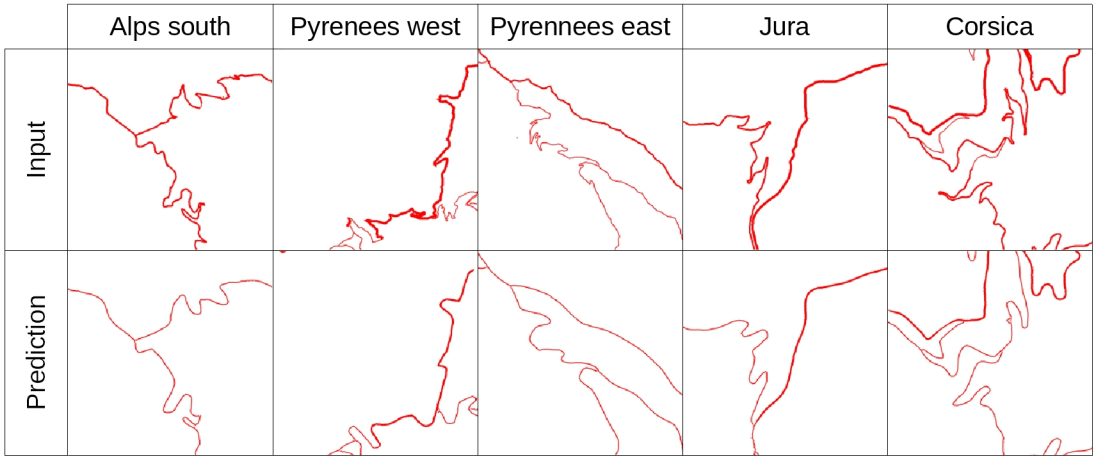


Figure 13: Examples of CycleGAN predictions in transfer areas. All tiles represent a part of the map of 2.5 km².

Mean constraint measure for each est area. The constraints used are the ones proposed by [Courtial et al., 2022]

	Noise	Connectivity	Position	Coalescence	Un-smoothness
Alps north	48	1.23	0.36	178.8	0.03
Alps South	31.19	2.92	0.36	23.09	0
Corsica	25.5	2.23	0.43	19.35	0
Jura	41.56	3.41	0.45	21.63	0
Pyrenees west	26.68	0.99	0.43	85.32	0
Pyrennees east	29.19	0.77	0.44	69.88	0

5.1 Learning strategy

First of all this article contributes to highlighting the comparative interest of several learning strategies: for our use case and unlike the building simplification use case [Feng et al., 2019], a GAN architecture gives way better results than a U-Net architecture [Courtial et al., 2020]. The discriminator forces the generated image to look like roads. The main difference between these two use cases is that buildings are mostly composed of angular shapes that GANs were unable to reproduce whereas, in the road use case, a smooth shape is desired. Nevertheless, GAN architectures generate images that are less constrained than U-Net architectures, and it causes for instance the appearance of small blue dots that are barely visible in the output images. This colour noise has to be post-processed.

Another surprising result in our experiment is that the paired model (pix2pix) is less effective than the unpaired (CycleGAN). In other words, it is easier to learn what a generalised map looks like than to learn how to generalise a map with mountain roads. The main flaw of images generated by the paired model, is that the network structure is hardly preserved: the multiple loop creations and disconnections make the generated images highly unrealistic. This flaw is common in deep learning approaches for line generalisation because the coalescent part of the bend is confused with a road intersection. It has already been observed in experiments about coastline generalisation with GANs [Du et al., 2021], where the issue is solved using a different method to cut the dataset into input images for the neural networks. In our case, the unpaired approach seems to resolve this problem. Both experiments help to understand why pix2pix learns to generate images with a disconnected network (or with loops). In the creation of the dataset using a fixed size square, several parts of roads can appear on each image, and the difference between two parts of independent roads on an image and disconnection is not big, so the discriminator may pass these alterations. However, unpaired data are less sensitive to disconnections and loop creation because of the cycle consistency loss that prevents generators to modify the information conveyed in the image. Nevertheless, paired learning has the advantage to be able to learn displacement, even if displacement is less applied than with a simple U-Net [Courtial et al., 2020]. It is not well applied mostly since the contextual necessary information is not presented: the road is often displaced in the reference because a river, not displayed in our images is close to the road. A mixed approach such as the one proposed in SMAPGAN [Chen et al., 2020], could help us get the advantages of both approaches.

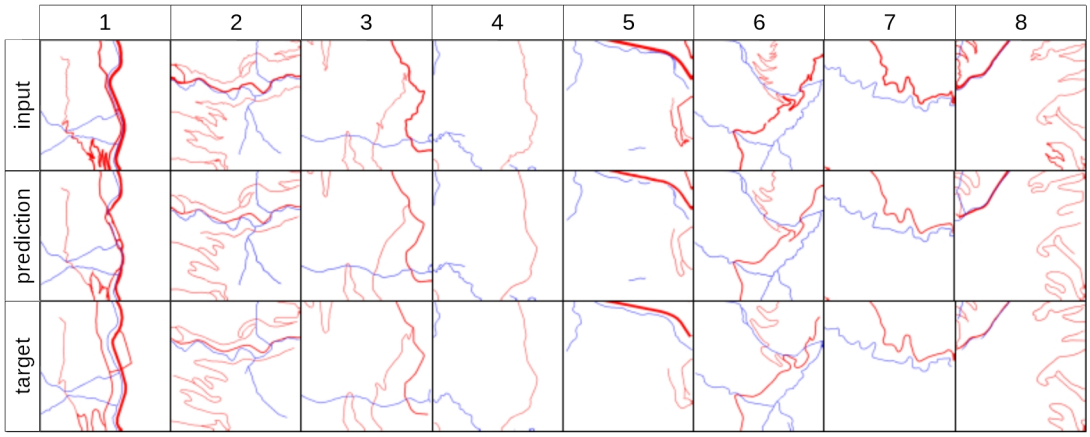


Figure 14: Example of CycleGAN predictions for images with roads and rivers. Roads are displayed in red, and rivers in blue, all tiles represent a part of the map of 2.5 km².

5.2 Creation of the training set

In this section, we discuss how to optimally create sample images from vector cartographic data.

The first point is about the quality, quantity, and diversity of images. On one side, the resolution experiment has shown that the network cannot learn from images that are not completely legible. Moreover, images containing errors and very few road pixels are also harmful to the quality of the predictions (see Sect. 3.2). On the other side, despite the relatively small number of images in our training dataset, CycleGAN is able to correctly generalise most of the situations because the most frequent cases are also the simplest to generalise, and only some specific road configurations are improved by the data augmentation. Consequently, adding new images to the training set only works if the images added have good quality and add diversity to the training set. The experiments about the number of images could explain the bad impact of data augmentation techniques such as rotating, cropping, or choosing a greater overlap between tiles for road generalisation as mentioned in the literature [Courtial et al., 2020]: these methods increase the number of images without adding new situations. Moreover, we observe that images with a small amount of information, even without errors, could be useless as they have almost always good evaluation because very few pixels are impacted by learning errors.

Then, the tile definition and the scale of the image have to be balanced between the quantity and the legibility of information: an overcut dataset is legible and generates much more images, but the entire road is sometimes not visible and the generalisation is not always optimal without seeing the entire shape of the road. On the contrary, with very large tiles the information available on each image is more important, but there is more symbol coalescence and there are fewer images. [Du et al., 2021] present a set of requirements and conditions for this issue, and then construct a segmentation method that maximizes the information in fixed-size tiles. This method is for now adapted to independent linear objects like coastlines, but it cannot really be adapted to networks such as roads and rivers, where lines often intersect. However, it shows the interest of an adapted tile segmentation.

Finally, it is important to remember that the network cannot learn something that is not in the training images. If other topographic features (e.g. rivers) are included in the context, the network might learn useful information to improve the results.

5.3 Generalisation of roads and rivers

The comparison with past experiments on buildings has shown that the appropriate deep learning architecture is highly related to the use case, and our method cannot be directly applied to other geographic objects. In order to assess the ability of the proposed method to generalise other features, we also experimented with the joint generalisation of roads and rivers. The geometric generalisation of both map themes has similarities but rivers naturally have smoother shapes. Adding rivers should provide more context in tiles, which should help, for example, to explain the displacement frequently applied to roads to avoid overlaps between roads and rivers. However, more context will produce more complex situations and make learning harder.

For this experiment, CycleGAN was trained with the same datasets, but the image tiles contained blue lines representing the rivers, under the road lines (red pixels are visible when roads and rivers intersect). Some results are presented in Figure 14. We observe that roads are globally as well generalised as in Model A. Then, rivers are really slightly generalised as expected; the model does learn that roads and rivers have to be generalised differently. We can note very few local errors due to a very coalescent input image, like the road bend that is not enlarged enough in situation 6, or the island creation in situation 4.

Then, the relation between roads and rivers seems to be a problem for the model. In Image 1, the enlargement to avoid symbol overlap is not applied, but some parts of the rivers are erased. Some deletions are also visible in Image 5. So, it seems that our model has learnt that roads and rivers do not have to overlap; however, it does not learnt how to avoid it. We can find the explanation for this problem in our input data: only half of our target training data present a significant displacement to avoid symbol overlap between roads and rivers (1,5,7), and the other half only propose a very small displacement (2), or no displacement at all (8). For these last cases, it can simply be due to the lack of overlapping problems: the roads and rivers only cross, or they do not follow the same valley for instance. But it can also be due to a lack of generalisation in our target dataset. In fact, due to the complexity of generalisation, the 1:250k map of IGN from which our target images are extracted have not always been generalised, and there are many situations where the road lines lie on top of the river lines. And the GAN logically learns to generate badly generalised maps from badly generalised training examples. A correction of the 1:250k dataset would be necessary to check if CycleGAN is able to learn how to generate images where roads and rivers have been displaced from each other. However, this use case has shown the ability of deep neural networks to learn the generalisation of several different objects at the same time.

Here, spatial relations are an important issue: the image representation seems insufficient to embed the relation between two close geographic objects (in our case the proximity between two roads or between a road and a river). We believe that with other knowledge representations, we could better deal with such spatial relations. For example, a graph representation has shown the ability to extract information about groups of map objects [Yan et al., 2019]. We believe that a mixed deep learning approach would benefit from image and graph information and generalise together several objects.

5.4 Integration into a larger generalisation framework

The user test shows that the generated images are considered as legitimate as the reference map. Nevertheless, these results are not really adapted for cartographic production for the following reasons:

- Some important loops and disconnections remain in some tiles.
- Only roads are generalised in the experiment and a complete topographic map is made of other topographic features. And these features (e.g., nearby buildings, relief, or rivers) may even influence how the roads are generalised.
- The output tiles are not a usable map, and post-processes are necessary to integrate these predictions in a map or in a more classical generalisation process.

In this section, we will discuss this last point. Integrating raster-based and vector-based procedures is not a new idea [Peter and Weibel, 1999]. A first methodology would be to use deep learning models and possibly a GAN architecture to generate an image of a generalised map containing all themes of the map. It would require in particular the introduction of more generalisation constraints into the loss functions of the neural networks of the GAN, as we did in this paper with the connectivity loss. However, more themes represented in the map means more generalisation constraints to satisfy for the generated image, and we think that we are not close to the proposition of a deep learning model able to generate a good and complete generalised map.

The integration of the generated road shape in a map generalised using a combination of several other algorithms is another solution but the conversion of the road back to vector would be necessary. The tiles recombination and the conversion raise other challenges, but we believe that they are less challenging than the first method for now. Problems to combine tiles would be similar to the ones occurring when partitioning is performed prior to generalisation [Touya et al., 2017]. Finally, the image generated by a deep learning model can be seen as a hint of the road location and shape, to guide another generalisation algorithm.

6 Conclusion

As a conclusion, this paper shows that it is possible to create images depicting a generalised map with mountain roads, using generative adversarial networks. The quality of the results has been judged by users, and considered as good as the reference map. The unpaired approach is preferred by the evaluators. The proposed improvement techniques

correct some specific errors. However, some important local issues remain and the generated images cannot be used as end-product topographic maps yet. We show that unpaired generative adversarial networks seem to be the most adapted method for mountain road generalisation. Two improvements to the default CycleGAN model are proposed: (1) the addition of a specific loss element that measures road connectivity and (2) the improvement of quality, diversity, and quantity of training example.

This work can still be improved by dataset and network improvement. Moreover, it can be extended by adding multiple geographic themes in the training datasets to predict a complete map. This perspective would have to deal with spatial relations in a dense context, such as urban areas and multiple representations of objects. In parallel, we observe the emergence of several promising deep learning models for map generalisation. Especially, we believe that an attention mechanism would be able to focus on some parts of a large image and resolve some context-related issues such as displacement. Moreover, graph convolutional networks could be employed to analyse the relations between objects inside a graph structure and model the spatial relationships between geographic entities. This is perhaps another important issue to which researchers should pay more attention.

Data and code availability statement

The data that support the findings of this study are openly available at https://doi.org/10.5281/zenodo.4682683. The codes for pix2pix and CycleGAN are from third-party authors who are not affiliated with this manuscript. The code for the model adapted from CycleGAN that contains the connectivity measure is available here: https://github.com/umrlastig/DeepMapGen.git.

Notes on contributors

Azelle Courtial is a PhD candidate in geographic information sciences. She is interested in exploring the potential of deep learning for map generalization. Contribution: Conceptualization, Methodology, Software, Formal analysis, Writing - Original Draft and Visualization.

Guillaume Touya is a senior researcher, at IGN France (the French mapping agency) and Univ Gustave Eiffel, holding a PhD and habilitation in GI science from Paris-Est University. His research interests focus on automated cartography, map generalization and volunteered geographic information, with a particular interest in research approaches to multi-scale cartography that mix automated cartography, spatial cognition and human–computer interaction issues. He is the principal investigator of the recent LostInZoom project, funded by the Europe Research Council (ERC), and the chair of the ICA (International Cartographic Association) commission on map generalization and multiple representation. Contribution: Conceptualization, Methodology, Validation, Data Curation, Writing - Review Editing, Supervision and Funding acquisition.

Xiang Zhang is an associate professor at School of Geospatial Science and Engineering, Sun Yat-Sen University. His research interests include map generalization, crowd-sourced geoinformation, and artificial intelligence for GIScience. Contribution: Conceptualization, Validation, Writing - Review Editing and Supervision.

References

[Bader and Barrault, 2001] Bader, M. and Barrault, M. (2001). Cartographic displacement in generalization: Introducing elastic beams. In *4th workshop on progress in automated map generalisation*. ICA.

[Benaim and Wolf, 2017] Benaim, S. and Wolf, L. (2017). One-Sided Unsupervised Domain Mapping. *Advances in Neural Information Processing Systems*, 30.

[Benz and Weibel, 2013] Benz, S. and Weibel, R. (2013). Road network selection using an extended stroke-mesh combination algorithm. *16th ICA Workshop on Generalisation and Multiple Representation*.

[Chen et al., 2020] Chen, X., Chen, S., Xu, T., Yin, B., Peng, J., Mei, X., and Li, H. (2020). SMAPGAN: Generative Adversarial Network-Based Semisupervised Styled Map Tile Generation Method. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1–19.

[Courtial et al., 2020] 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):338. Number: 5 Publisher: Multidisciplinary Digital Publishing Institute.

[Courtial et al., 2022] Courtial, A., Touya, G., and Zhang, X. (2022). Constraint-Based Evaluation of Map Images Generalized by Deep Learning. *Journal of Geovisualization and Spatial Analysis*, 6.

[Daley et al., 1997] Daley, N., Goodenough, D., Bhogal, A. S., Bradley, Q., and Yin, Z. (1997). Comparing raster and object generalization. *IGARSS'97. 1997 IEEE International Geoscience and Remote Sensing Symposium Proceedings. Remote Sensing - A Scientific Vision for Sustainable Development*, 2:677–679 vol.2.

[Du et al., 2021] Du, J., Wu, F., Xing, R., Gong, X., and Yu, L. (2021). Segmentation and sampling method for complex polyline generalization based on a generative adversarial network. *Geocarto International*, 0(0):1–23.

[Duchêne, 2014] Duchêne, C. (2014). Individual road generalisation in the 1997-2000 AGENT european project. Technical report, IGN, COGIT lab.

[Duchêne et al., 2014] Duchêne, C., Baella, B., Brewer, C. A., Burghardt, D., Buttenfield, B. P., Gaffuri, J., Käuferle, D., Lecordix, F., Maugeais, E., Nijhuis, R., Pla, M., Post, M., Regnauld, N., Stanislawski, L. V., Stoter, J., Tóth, K., Urbanke, S., van Altena, V., and Wiedemann, A. (2014). Generalisation in practice within national mapping agencies. In Burghardt, D., Duchêne, C., and Mackaness, W., editors, *Abstracting Geographic Information in a Data Rich World*, Lecture Notes in Geoinformation and Cartography, pages 329–391. Springer International Publishing.

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

[Fu et al., 2019] Fu, H., Gong, M., Wang, C., Batmanghelich, K., Zhang, K., and Tao, D. (2019). Geometry-Consistent Generative Adversarial Networks for One-Sided Unsupervised Domain Mapping. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2422–2431, Long Beach, CA, USA. IEEE.

[Gan et al., 2017] Gan, Z., Chen, L., Wang, W., Pu, Y., Zhang, Y., Liu, H., Li, C., and Carin, L. (2017). Triangle generative adversarial networks. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems 30*, pages 5247–5256. Curran Associates, Inc.

[Ganguli et al., 2019] Ganguli, S., Garzon, P., and Glaser, N. (2019). GeoGAN: A conditional GAN with reconstruction and style loss to generate standard layer of maps from satellite images. *arXiv:1902.05611*.

[Garcia-Balboa and Ariza-López, 2008] Garcia-Balboa, J. L. and Ariza-López, F. J. (2008). Generalization-oriented road line classification by means of an artificial neural network. *Geoinformatica*.

[Goodchild and Li, 2012] Goodchild, M. F. and Li, L. (2012). Assuring the quality of volunteered geographic information. *Spatial Statistics*, 1:110–120.

[Grünreich, 1985] Grünreich, D. (1985). Ein vorschlag zum aufbau einer grossmassstabäbigen topographisch-kartographischen datenbank unter besonderer berücksichtigung der grundrissdatei des alk-systems. In *Nachrichten aus dem Karten- une Vermessungswesen*, pages 53–62. Series I, no. 85.

[Guilbert et al., 2014] Guilbert, E., Gaffuri, J., and Jenny, B. (2014). Terrain Generalisation. In Burghardt, D., Duchêne, C., and Mackaness, W. A., editors, *Abstracting Geographic Information in a Data Rich World*, Lecture Notes in Geoinformation and Cartography, pages 227–258. Springer International Publishing.

[Isola et al., 2017] Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[Jenny et al., 2020] Jenny, B., Heitzler, M., Singh, D., Farmakis-Serebryakova, M., Liu, J., and Hurni, L. (2020). Cartographic Relief Shading with Neural Networks. *IEEE Transactions on Visualization and Computer Graphics*.

[Kang et al., 2019] Kang, Y., Gao, S., and Roth, R. (2019). Transferring multiscale map style using generative adversarial network. *International Journal of Cartography*.

[Kang et al., 2017] Kang, Y., Rao, J., Wang, W., Peng, B., Gao, S., and Zhang, F. (2017). Towards cartographic knowledge encoding with deep learning. *CaGIS Journal*, page 6.

[Lecordix et al., 1997] Lecordix, F., Plazanet, C., and Lagrange, J.-P. (1997). A platform for research in generalization: Application to caricature. *GeoInformatica*, 1(2):161–182.

[Ledig et al., 2017] Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., and Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. *arXiv:1609.04802 [cs, stat]*.

[Li, 1994] Li, Z. (1994). Mathematical Morphology in Digital Generalization of Raster Map Data. *Cartography*, 23(1):1–10.

[Li and Su, 1995] Li, Z. and Su, B. (1995). Algebraic Models for Feature Displacement in the Generalization of Digital Map Data Using Morphological Techniques. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 32(3):39–56. Publisher: University of Toronto Press.

[Mackaness and Ruas, 2007] Mackaness, W. A. and Ruas, A. (2007). Chapter 5 - evaluation in the map generalisation process. In Mackaness, W. A., Ruas, A., and Sarjakoski, L. T., editors, *Generalisation of Geographic Information*, International Cartographic Association, pages 89 – 111. Elsevier Science B.V.

[McMaster and Monmonier, 1989] McMaster, R. B. and Monmonier, M. (1989). A conceptual framework for quantitative and qualitative raster-mode generalization. In *Proceedings of GIS/LIS'89*, pages 390–403, Orlando, Florida. American Society for Photogrammetry and Remote Sensing.

[Monmonier, 1983] Monmonier, M. S. (1983). Raster-mode area generalization for land use and land cover maps. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 20(4):65–91. Publisher: University of Toronto Press.

[Mustiere, 1998] Mustiere, S. (1998). GALBE : Adaptative generalization. the need for an adaptative process for automated generalisation an exemple on roads. *Procesding of 1st GIS PlaNet conference*.

[Nazeri et al., 2018] Nazeri, K., Ng, E., and Ebrahimi, M. (2018). *Image Colorization Using Generative Adversarial Networks*, pages 85–94.

[Olteanu-Raimond et al., 2015] Olteanu-Raimond, A.-M., Mustière, S., and Ruas, A. (2015). Knowledge formalization for vector data matching using belief theory. *Journal of Spatial Information Science*, 10:21–46.

[Pantazis et al., 2009] Pantazis, D., Karathanasis, B., Kassoli, M., Koukofikis, A., and Stratakis, P. (2009). Morphing techniques: Towards new methods for raster based cartographic generalization. In *Proceedings of the 24th International Cartographic Conference*, Santiago, Chile. ICA.

[Peter and Weibel, 1999] Peter, B. and Weibel, R. (1999). Using Vector and Raster-Based Techniques in Categorical Map Generalization. In *Proceedings of Third ICA Workshop on Progress in Automated Map Generalization*, Ottawa, Canada.

[Peuquet, 1979] Peuquet, D. J. (1979). Raster Processing: An Alternative Approach to Automated Cartographic Data Handling. *The American Cartographer*, 6(2):129–139.

[Raposo and Samsonov, 2014] Raposo, P. and Samsonov, T. (2014). Towards general theory of raster data generalization. In *Proceedings of 17th ICA Workshop on Generalisation and Multiple Representation*, Vienna, Austria.

[Ronneberger et al., 2015] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. volume abs/1505.04597, pages 234–241.

[Roth et al., 2011] Roth, R. E., Brewer, C. A., and Stryker, M. S. (2011). A typology of operators for maintaining legible map designs at multiple scales. *Cartographic Perspectives*, (68):29–64. Number: 68.

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

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

[Shmelkov et al., 2018] Shmelkov, K., Schmid, C., and Alahari, K. (2018). How good is my GAN? pages 213–229.

[Simo-Serra et al., 2018] Simo-Serra, E., Iizuka, S., and Ishikawa, H. (2018). Mastering sketching: Adversarial augmentation for structured prediction. *arXiv:1703.08966*.

[Simo-Serra et al., 2016] Simo-Serra, E., Iizuka, S., Sasaki, K., and Ishikawa, H. (2016). Learning to simplify: Fully convolutional networks for rough sketch cleanup. *ACM Trans. Graph.*, 35(4).

[Thomson and Richardson, 1999] Thomson, R. and Richardson, D. (1999). The 'good continuation' principle of perceptual organization applied to the generalization of road networks. *ICA Proceedings*.

[Touya, 2010] Touya, G. (2010). A road network selection process based on data enrichment and structure detection. *Transactions in GIS*, 14(5):595–614.

[Touya et al., 2017] Touya, G., Berli, J., Lokhat, I., and Regnauld, N. (2017). Experiments to distribute and parallelize map generalization processes. *The Cartographic Journal*, 54(4):322–332.

[Touya and Lokhat, 2018] Touya, G. and Lokhat, I. (2018). Deep learning for enrichment of vector spatial databases: application to highway interchange. *ACM Transactions on Spatial Algorithms and Systems*, page 21.

[Touya et al., 2019] Touya, G., Zhang, X., and Lokhat, I. (2019). Is deep learning the new agent for map generalization? *International Journal of Cartography*, 5(2-3):142–157. Publisher: Taylor & Francis.

[Wang et al., 2018] Wang, X., Girshick, R., Gupta, A., and He, K. (2018). Non-local neural networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7794–7803, Los Alamitos, CA, USA. IEEE Computer Society.

[Weibel et al., 1995] Weibel, R., Keller, S., and Reichenbacher, T. (1995). Overcoming the knowledge acquisition bottleneck in map generalization: The role of interactive systems and computational intelligence. In Frank, A. U. and Kuhn, W., editors, *Spatial Information Theory A Theoretical Basis for GIS*, volume 988 of *Lecture Notes in Computer Science*, pages 139–156. Springer.

[Yan et al., 2019] Yan, X., Ai, T., Yang, M., and Yin, H. (2019). A graph convolutional neural network for classification of building patterns using spatial vector data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 150:259–273.

[Yi et al., 2017] Yi, Z., Zhang, H., Tan, P., and Gong, M. (2017). DualGAN: Unsupervised dual learning for image-to-image translation. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2868–2876. IEEE.

[Zhu et al., 2017] Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2242–2251. IEEE.