

Generic Semantic Segmentation of Historical Maps

Rémi Petitpierre¹, Frédéric Kaplan² and Isabella di Lenardo¹

¹*Institute for Area and Global Studies, EPFL, Lausanne, Switzerland*

²*Digital Humanities Laboratory, EPFL, Lausanne, Switzerland*

Abstract

Research in automatic map processing is largely focused on homogeneous corpora or even individual maps, leading to inflexible models. Based on two new corpora, the first one centered on maps of Paris and the second one gathering maps of cities from all over the world, we present a method for computing the figurative diversity of cartographic collections. In a second step, we discuss the actual opportunities for CNN-based semantic segmentation of historical city maps. Through several experiments, we analyze the impact of figurative and cultural diversity on the segmentation performance. Finally, we highlight the potential for large-scale and generic algorithms. Training data and code of the described algorithms are made open-source and published with this article.

Keywords

historical map processing, neural networks, semantic segmentation, computer vision, topology

1. Introduction

The creation of large digital databases on urban development is a strategic challenge, which could open new perspectives in urban planning, environmental sciences, sociology, economics, and analysis of urban ecosystems in general [16, 22]. Digital geohistorical data can also be valorized by cultural institutions, for instance in the form of 3D/4D models [28, 29]. When the data are of good quality, and when large and homogeneous corpora are considered, it is possible to obtain excellent segmentation results with traditional computer vision and decision algorithms [37, 38]. However, these algorithms are very specific. They are based on a detailed and exact knowledge of the processed map and its figuration [10]. This implies that the process potential of these inflexible methods lies only in large and very homogeneous cartographic corpora. Consequently, map vectorization is still largely manual, despite being extremely time-consuming. In order to process the immense and diverse cartographic collections hosted by heritage institutions around the world, the development of generic and automatic tools is required.

This research intends to set the foundations for a generic approach of the semantic segmentation of historical maps. The ambition is to design a system capable of processing map corpora characterized by both graphical and content heterogeneity. Recent progress on convolutional neural networks (CNN) tend to support the idea that genericity can be achieved, in particular for segmentation tasks [42]. However, the challenges of generic semantic interpretation (or

CHR 2021: Computational Humanities Research Conference, November 17–19, 2021, Amsterdam, The Netherlands

✉ remi.petitpierre@epfl.ch (R. Petitpierre); frederic.kaplan@epfl.ch (F. Kaplan); isabella.dilenardo@epfl.ch (I.d. Lenardo)

🆔 0000-0001-9138-6727 (R. Petitpierre); 0000-0002-6991-5730 (F. Kaplan); 0000-0002-1747-9164 (I.d. Lenardo)



© 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

representational flexibility [24]) of graphical objects produced in diverse technical and cultural contexts remain largely unsolved. In the following sections, we experimentally map the design space for such a generic processing pipeline and present first working prototypes tested on two large map corpora.

1.1. Previous Works

Classical segmentation algorithms are based on the specific knowledge of a map collection. Despite color appearing relatively late in map printing processes [47], this graphical component is frequently used. Some specialists even consider color-based pre-processing an “essential” task [36]. The simplest paradigm in this regard is color thresholding [12, 15]. Other studies add a morphological approach, using region growing algorithms [9, 31, 30], or rely on human feedback [11]. In deep learning methods, region growing is also used to flood fill and add a semantic layer to extracted polygons using watershed [41]. Another salient graphical component is texture. Hatched areas are particularly targeted [56, 4, 39]. Other methods focus on texture energy [34]. These approaches in general have met some success on textured maps [36, 51, 55]. Unlike color, however, textures have many degrees of freedom, such as size and rotation. Their use for segmentation therefore requires fine parametrization.

Beyond pure graphical marks, some researches focus on detecting morphological features, such as lines [34], or closed polygons [35]. These approaches are generally confronted with the problem of incomplete lines, due to the degradation of the document or graphical choices (e.g. dashed lines) and therefore require the development of reconstruction algorithms [33, 2, 25, 49]. Moreover, the extraction of the geometries is sensitive to information overlay, which is extremely frequent in cartography. Many specific methods were therefore developed to detect and eventually remove every disruption, such as the map grid [26], the background texture [34], the text [1], or the symbols [57]. The detection of these interfering elements generally requires precise knowledge of their nature and their visual characteristics: size, shape, texture, color, etc [19]. To conclude, the extraction of the morphology is still unable to add a semantic layer to the map content. For instance, a rectangle might well be a building, but it can equally well be a courtyard or a basin.

More recently, CNNs opened up new perspectives for the resolution of semantic segmentation problems [32]. In 2019, a first successful transfer was presented with the segmentation of parcels from “Napoleonic” cadastral maps of Venice [41], using a UNet architecture with a ResNet encoder. In 2020, further results were obtained on the extraction of the railway network from some USGS maps, using a modified PSPNet as encoder [8]. Heitzler and Hurni also presented a research on the extraction of building footprints from the Swiss Siegfried national maps (1872-1949) [21]. Other works were focusing on extracting specific elements, such as places of archaeological interest [17], or road types [6]. This research question being topical, the ICDAR 2021 conference dedicated a competition to the segmentation of building blocks on a corpus of maps from the Bibliothèque historique de la Ville de Paris [7]. One solution in particular stood out by proposing to use a DenseNet [23] architecture.

On the other hand, [54] proposed a method to operationalize cartographic figuration based on color-histogram based moments. The vector projection of these descriptors allows one to efficiently visualize the figurative conventions of a map or a corpus of maps. Subsequently, a texture extraction method has also been proposed [53]. Other efforts to operationalize cartographic figuration have also been carried out concerning graphic map load, i.e. the visual density of cartographic content [3, 52]. These researches are part of the field of information

theory.

1.2. Research approach

In this research, we seek to better understand the impact of figuration and figurative diversity on the learning of neural networks, in the context of the semantic segmentation of historical city maps. Thus, our research questions are the following: 1) How to measure the figurative diversity in a map corpus? 2) How robust are CNNs when facing high figurative diversity?

To answer these research questions, we first present a method to operationalize the cartographic figuration and measure the figurative diversity of a map corpus. We demonstrate the significant variability of our data, in comparison with other map corpora commonly found in the literature. Then, we propose an effective processing pipeline involving pre-segmentation of the map frame and semantic segmentation of the map itself. We highlight the potentials and the limitations of neural networks for solving generic semantic segmentation problems. Finally, we conduct a set of experiments that allow to investigate the learning mechanisms deployed by neural networks and to explore the design space of map semantic segmentation. Specifically, we seek to qualify the impact of learning transfer on each class. We then aim to evaluate the perspectives for an educated corpus constitution, through examination of cross-cultural performance and confidence prediction. Finally, we challenge the importance of graphical cues, in comparison to non-graphical concepts.

1.3. Dataset

To create an experimental context that challenges the genericity of the semantic segmentation, we contrast two corpora that present a different kind of variability. The first corpus gathers 330 maps of Paris from the collections of the Bibliothèque nationale de France (BnF) and the Bibliothèque historique de la Ville de Paris. Most of the maps were published between 1800 and 1950 in the Paris region, and their scale is generally comprised between 1:25'000 and 1:2'000. The second corpus gathers 256 maps of cities from all over the world, including numerous reference maps. They come from 32 different collections, the main ones being: the BnF, the Library of Congress, the Harvard Library, the David Rumsey Collection, the University of Bordeaux, the British Library, the Boston Public Library and the Institut Cartogràfic i Geològic de Catalunya. In total, the corpus represents 182 different cities in 90 countries. The distribution across the various regions is balanced, except for Oceania (8 maps only). The regions with the most maps are Eastern Europe & Central Asia (34), Western Europe (34), and East Asia (30). Conversely, the regions accounting for the fewest maps are Oceania, Subsaharian Africa (15), and Middle East (17). The urban form of each map in the World corpus was manually classified into three categories: regular (95 maps), irregular (68), or mixed (93). Most maps were published between 1720 and 1950. However, due to historical reasons, most non-Western publications occurred after 1800.

Both corpora present specific difficulties. For the Paris corpus (Fig. 1), one of the most complex issue to apprehend is the information cluttering or overlay. In particular, the superimposition of information concerning mobility, such as road or underground network (Fig. 1 A.2-3), but also on the water system, the catacombs, or more simply on administrative divisions (Fig. 1 B.3-4). For Paris, this intensive use of the city map as a tool for planning urban works was likely caused by the lack of a proper cadastre before the late 19th century. Low-contrast may also make the map images difficult to read (Fig. 1 B.3, C.4). For the World

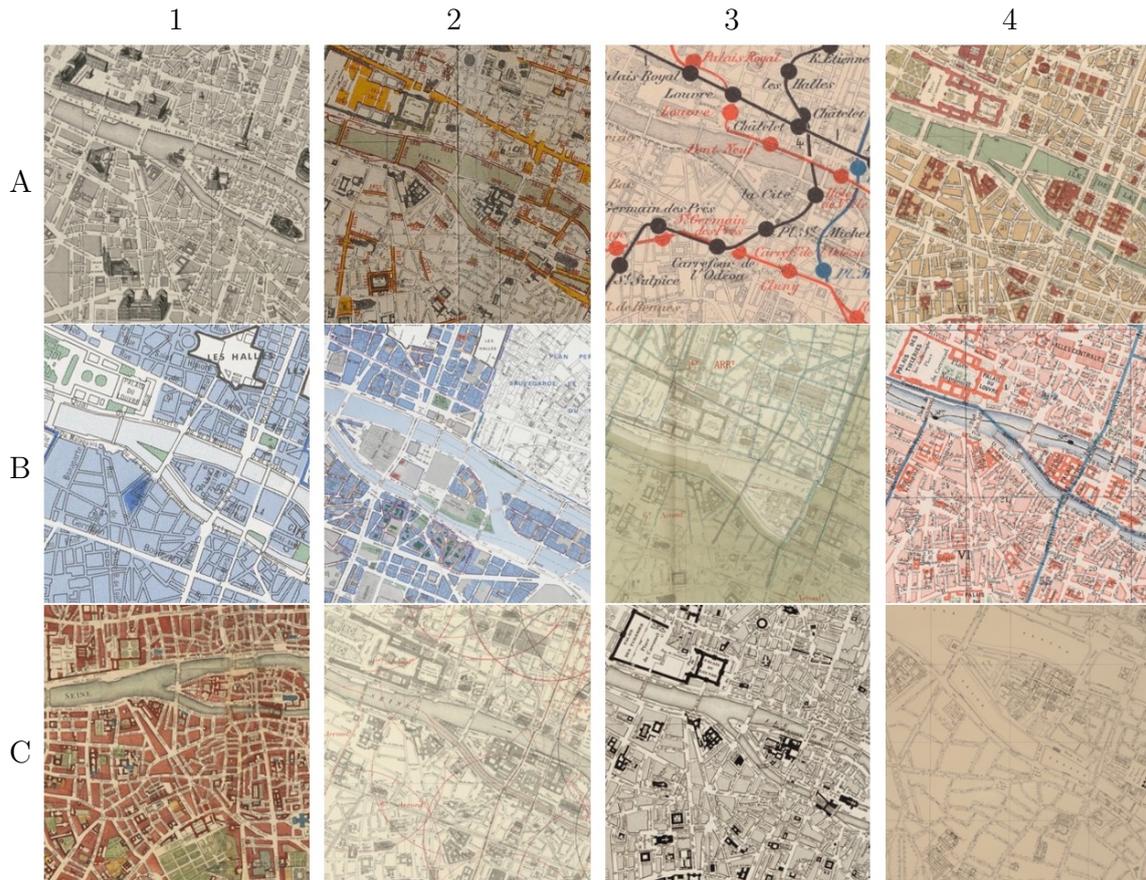


Figure 1: Map samples from the Paris corpus.

corpus, besides the extreme cultural diversity, the complexity is due to the differences in scale (Fig. 2 A.1-2), in formality (Fig. 2 A.3-4, B.1), in the graphical emphasis of monuments (Fig. 2 B.2-4), in the use of writing to represent areas or zones (Fig. 2 C.1-2), to differences in the regularity of urban form (Fig. 2 C.3-4, D.1), to figurative specificities, such as field shadowing (Fig. 2 D.2) or blueprint-style (Fig. 2 D.3), to information overlay (Fig. 2 D.4), to digitization imperfections or poor scanning quality (Fig. 2 E.1-2), and to material alterations (Fig. 2 E.3-4).

To summarize, we will be comparing two corpora. The first one, centered on Paris, is culturally and geographically homogeneous, but figuratively diverse. While the second, which will be called the World corpus, is culturally, geographically, and figuratively heterogeneous.

To constitute the training set, 330 random patches of 1000 pixels square are randomly cut out of each map in the Paris corpus. 30 are then selected for the validation set. For the World, 256 random patches are cut out of each map for the training set, and 49 additional patches are cut for the validation set. Four maps from the World corpus, whose resolution was too low, were upresolved by a factor 2 using CNN-based super-resolution [5]. The patches are then manually annotated using a raster-based software. The total annotation time is estimated to be 7 weeks of work. A simple 3-classes ontology and an extended 5-classes one are defined. The classes from the simple ontology correspond to: the road network, including railroads and bridges;



Figure 2: Map samples from the World corpus.

the frame, i.e. to non-geographic content of the document; and the map content, i.e. all the geographic content that do not belong to the road network. In the extended ontology, the map content was subdivided into 3 additional classes: the blocks, the water, and the non-built, which in fact includes all non-aquatic unbuilt land, except the road network, i.e. wasteland, meadows, crops, forests, but also parks, or inner courtyards. The Tab. 1 summarizes the distribution of classes in the different corpora and sets. The training patches are open-source and freely available online [45].

Table 1

Proportion of areas covered in the different training and validation sets.

Corpus	Set	Frame	Road netw.	Blocks	Water	Non-built
World	train	0.222	0.123	0.175	0.121	0.359
	val	0.185	0.114	0.178	0.073	0.450
Paris	train	0.236	0.208	0.338	0.026	0.192
	val	0.166	0.204	0.360	0.045	0.226
World	train	0.226	0.123	0.651		
	val	0.181	0.204	0.701		
Paris	train	0.236	0.206	0.559		
	val	0.166	0.201	0.633		

As in [7], we consider that the problem of the segmentation of the map frame is a different problem, mainly for scale reasons. This step is therefore carried out beforehand, and the pre-segmented background class is indicated as +1 in the next pages. The pre-segmentation simply takes the form of a mask applied on the input images

2. Methods

2.1. Operationalization of the figuration

To make cartographic figuration measurable, we extracted 3 sets of descriptors related to color, texture, and orientation. The color and texture features are based on previous works by Uhl et al. [54, 53]. First, mean and standard deviation are computed on the distribution of each color channel. Then, 256-bins histograms are extracted for each of the latter and the skewness and kurtosis in the distribution are computed. The mean is used to transcribe the hue and the color value, the standard deviation indicates the contrasts. The skewness is a descriptor of the asymmetry of the color distribution, while the kurtosis is a flattening coefficient of the curve and therefore also allows to describe the contrasts. Color standard deviation, kurtosis and skewness are summed for the 3 channels. Second, local binary patterns (LBP, [40]) are computed, using a radius of 2, on the Otsu-binarized images [43]. A 4-bins histogram is extracted on the LBP values. LBP are invariant to color value (i.e. brightness) and rotation. They can help differentiate between edges, corners or flat surfaces at a local scale and thus characterize the texture at a larger scale. Third, a 24-bins histogram of oriented gradient (HOG, [13]) is computed. Each bin corresponds to a certain orientation angle of the image local gradients. Therefore, they are ultimately grouped into 5 categories, summarizing the orientation of the gradients: vertical ($\pm\pi$), horizontal ($\pm\pi/2, \pm3\pi/2$), diagonal ($\pm\pi/4, \pm3\pi/4$), regular oblique ($\pm\pi/6, \pm2\pi/6, \pm4\pi/6, \pm5\pi/6$), and irregular oblique (all other orientations). At this point, there are 14 features in total: 6 (3+3) color descriptors, 4 LBP descriptors, and

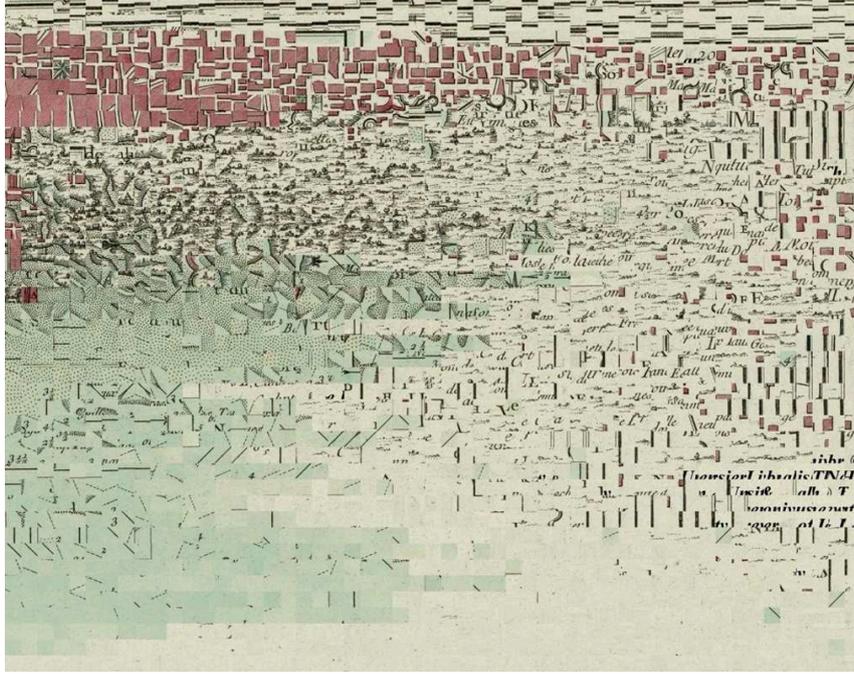


Figure 3: t-SNE projection of the figuration descriptive features of the Plan of the City of Moka.

5 HOG descriptors. Together, these descriptors can characterize the cartographic figuration, as can be seen in Fig. 3.

These visual features are computed, using 50x50 sub-patches, for each image of both corpora, as well as for the USGS [8], the Napoleonic cadaster [41], and the ICDAR21 dataset [7], for comparison purposes. On the one hand, the inter-corpora inter-class correlation is computed between the Paris and World datasets, as well as the intra-corpora inter-class correlation within each corpus.

On the other hand, a 32-bins histogram is computed on the distribution of each feature. As the distributions can be multimodal, due to repetitive homogeneous figuration, the modes are extracted by smoothing the histogram with a Savitzky-Golay filter [50] of width 3 and polynomial degree 1. The local minima are identified on a window of width 3 and the histogram is split between each mode. To investigate the first research question, we want to determine how much these features vary in a map corpus. To this extend, a κ -coefficient is defined as the proportionally weighted sum of the kurtosis on each mode of the distribution. In other terms, the κ -coefficient is a measure of the acuteness of the feature distribution in a map corpus, and can thus characterize the homogeneity of the figuration. As the value of the κ -coefficient can vary according to the size of the corpus, the bigger sample sets were randomly downsampled, without replacement, to the size of the smallest sample set (here the Napoleonic cadaster set). The κ -coefficient was then computed 5000 times for various downsampling schemes and for each feature, the median κ -coefficient was retained as an estimator of the real κ -coefficient. The bias of this recalibration is below $\pm 3.6\%$ for the World, and below $\pm 2.0\%$ for Paris, with a confidence of 95%. The code of the described algorithms is made open-source and published with this article [44].

2.2. Map segmentation

For the semantic segmentation, we are using a CNN with UNet architecture [48] and ResNet [20] as encoder, implemented in a Pytorch version of the open-source tool dhSegment [42, 27]. The batch size and learning rate parameters are optimized. The method and the results of the tests are detailed in [46]. The selected parameters are a batch size of 1, as in the original article by Long and Shelhamer [32], and a learning rate of $5 * 10^{-5}$. The decoder weights are initialized using Glorot and Bengio uniform method [18]. The training data are augmented by side flip and upside-down flip, and by rotation ($r \in [0, \pi]$). The loss used is cross entropy. The optimization relies on stochastic gradient descent (SGD). Unless otherwise specified, the encoder weights are first pretrained on ImageNet [14].

The second pretraining is done in a crossed way, Paris being pretrained on the World, and the World being pretrained on Paris, as described in the following subsection. The CNN is then trained successively during 150 epochs on the Paris (2+1 and 4+1) datasets and on the World (2+1 and 4+1) datasets. ResNet101 is used as encoder with LeakyReLU as activation.

Three metrics are used to quantify performance: intersection over union (IoU), precision, and recall. In a second step, the confusion matrices between the different classes are computed. They are normalized regarding the proportion of pixels belonging to the class, according to the ground-truth.

2.3. Semantic segmentation experiments

2.3.1. Impact of learning transfer on each class

In a first experiment, the CNN with ResNet50 as encoder was trained during 100 epochs on the Paris 5-classes-dataset, respectively on the World 5-classes-dataset. After the first training, the network was re-trained on the Paris set, using the weights trained in the previous step on the World corpus as initialization. Reciprocally, the World set was re-trained by initializing this time the weights on the Paris corpus.

2.3.2. Analysis of cross-cultural performance and dataset design

To estimate the bias of the validation set, and to investigate cross-cultural performance, an 8-fold cross validation was performed on the World 4+1 dataset. Each time, the CNN was trained during 100 epochs as described in the subsection on the performance of semantic segmentation.

2.3.3. Perspectives on confidence prediction

In this experiment, we attempted to create an estimator of confidence at the patch scale. First, a 10-fold cross validation was performed on the Paris 3-classes-dataset in order to estimate the error on each patch of the training set. A simple ResNet50 encoder was used and trained during 60 epochs each time. The output predictions from the k th-fold are compared with the ground truth, and an accuracy map, in which the pixels take the value 1 if the prediction is correct and 0 if the prediction is wrong, is created.

A second network, identical to the first one, is then trained on 300 pairs of images and accuracy maps, and validated on another 30 pairs. Instead of segmenting, the aim of this network is to predict the accuracy map corresponding to the input image. The output of the CNN prediction is classified using a global threshold, which is set to meet the mean accuracy of the training set. The confidence index is defined as the patch accuracy, and the reference as the

accuracy previously measured by k-fold cross validation. In order to evaluate the confidence prediction performance, an 8-fold cross validation was performed on the World 4+1 dataset. Each time, the CNN was trained during 100 epochs with the same parameters as in 2.2.

2.3.4. Importance of graphical cues for learning

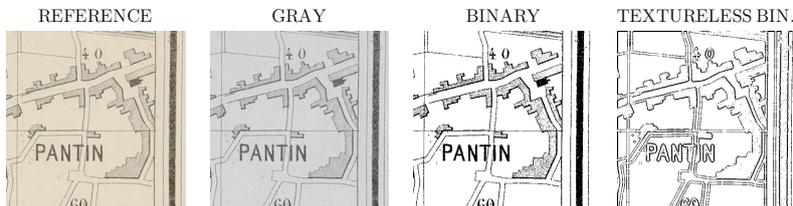


Figure 4: Images processed by phasing out visual characteristics.

This experiment aims to determine what role color, texture, and morphology take in the CNN performance, in contrast with non-graphical concepts. For this purpose, the images of the training and validation sets are subjected to 4 different treatments (Fig. 4): reference, gray, binary, and textureless binary. For the reference, the images are not modified in any way. The second treatment is the gray treatment. In the latter, the RGB color channels of the image are transformed into a grayscale. For the third treatment, the images are transformed into a grayscale, then binarized [43]. For the fourth treatment, the images are transformed into a grayscale, binarized, and texture is extracted with LBP [40] ($r = 3$). Finally, a second Otsu thresholding is applied. The CNN is trained 5 times for 60 epochs separately on each of the 4 datasets.

3. Results

3.1. Operationalization of the figuration

Table 2

Median overall κ for studied* and comparison† datasets. Lower values indicate greater figurative diversity.

	Paris*	World*	Napoleonic†	ICDAR21†	USGS†
median κ	1.97	1.99	4.36	5.04	29.1

The median overall κ was computed on each dataset (Tab. 2). The result is very close for the two studied corpora, while the Napoleonic cadaster and the ICDAR dataset are already further away. The USGS map is in a different order of magnitude.

Fig. 5 and Tab. 3 are the aggregated representations of the correlations of the figurative features between and within both corpora. The frame class seems to be represented very similarly in both corpora ($\rho = 0.96$, Fig. 5), while the non-built class is the most distant ($\rho = 0.81$, Fig. 5). In the Paris corpus, the blocks class seems to stand out clearly from the other classes ($\bar{\rho}_{Blocks} = 0.775$, Tab. 3, and Fig. 5), while in the World corpus, it is rather the road network that stands out ($\bar{\rho}_{RoadNetwork} = 0.888$, Tab. 3, and Fig. 5). In general, however, all classes are more distinct in the Paris corpus ($\bar{\rho}_{Mean} = 0.842$ for Paris, $\bar{\rho}_{Mean} = 0.917$ for the World, Tab. 3).

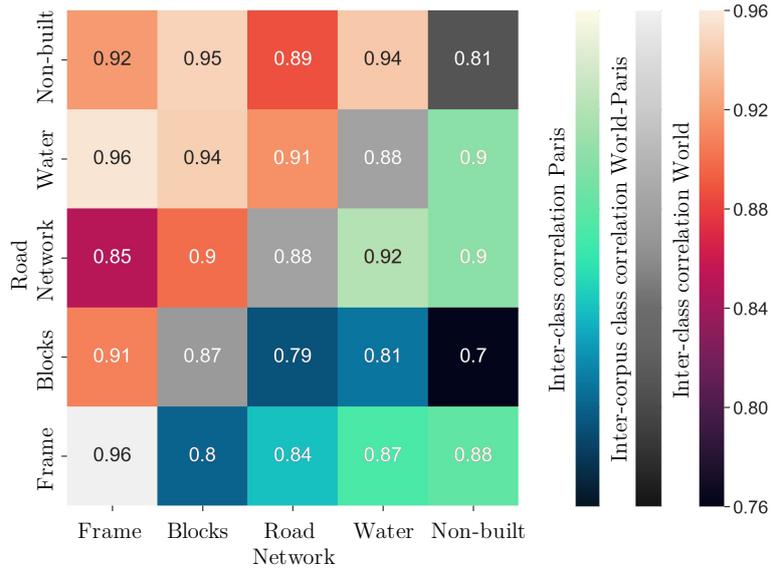


Figure 5: Aggregated correlation heatmap matrix of intra-corporal inter-class correlation (orange for the World, cyan for Paris) and inter-corporal class correlation (grey diagonal). ($r_{pearson} \in [-1, 1]$). All correlations are significant ($p_{value} < 0.05$).

Table 3

Mean intra-corporal inter-class correlations, per class

Corpus	World	Paris
Mean	0.917	0.842
Frame	0.908	0.846
Blocks	0.924	0.775
Road net.	0.888	0.862
Water	0.939	0.877
Non-built	0.924	0.847

3.2. Map segmentation

The results of the third experiment are summarized in Tab. 4 and Fig. 6. Some example prediction outputs can be observed in Fig. 7. The clearest finding is the consistent drop in performance, when increasing the number of classes from 2+1 to 4+1. The mean IoU (mIoU) on the 4+1 classes problems is not sufficient for reliable map segmentation. However, the mIoU is good on both World and Paris 2+1 corpora. The drop is more noticeable for precision than recall. The second clear difference occurs between the Paris and the World corpora, the first performing better. Again, the disparity is mostly due to a low recall. It is worth noticing that the top-50% of the World corpus performs very similarly to Paris average sample.

For the Parisian corpus, most confusion occurs as non-built is predicted as blocks, or to a lesser extent, as blocks are predicted as non-built, or as non-built is predicted as blocks. As it is also noticeable in Tab. 4, non-built is by far the class performing worse in the 4+1-classes problem. However, the water is the class suffering from the lowest precision score. In the 2+1-classes problem, the blocks are sometimes classified as road network, which heavily impacts the precision of the road network class.

Table 4

Performance achieved, per class and classes mean, on the two datasets, for 2+1 and 4+1 classes problems. For 2+1 classes problems, precision and recall were also computed on the top-50% of the dataset, selected by mIoU.

Metric	Class	Paris 2+1	Paris 2+1 Top 50%	World 2+1	World 2+1 Top 50%	Paris 4+1	World 4+1
IoU	Mean	0.8905	–	0.8055	–	0.6363	0.5595
	Frame	0.9953	–	0.9924	–	0.9810	0.9881
	Blocks	0.9181	–	0.9114	–	0.5657	0.3559
	Road Net.	0.7580	–	0.5147	–	0.7132	0.4682
	Water	–	–	–	–	0.4682	0.3318
	Non-built	–	–	–	–	0.4235	0.6538
Precision	Mean	0.9292	0.9679	0.8544	0.9205	0.7292	0.6986
	Frame	0.9959	0.9969	0.9935	0.9935	0.9838	0.9893
	Blocks	0.9689	0.9774	0.9730	0.9816	0.6872	0.4353
	Road Net.	0.8229	0.9295	0.5967	0.7863	0.7874	0.5348
	Water	–	–	–	–	0.4856	0.7098
	Non-built	–	–	–	–	0.7017	0.8240
Recall	Mean	0.9456	0.9698	0.9062	0.9445	0.8175	0.7187
	Frame	0.9996	0.9990	0.9989	0.9992	0.9971	0.9988
	Blocks	0.9448	0.9736	0.9350	0.9957	0.7618	0.6611
	Road Net.	0.8924	0.9368	0.7848	0.8686	0.8832	0.7898
	Water	–	–	–	–	0.9288	0.3839
	Non-built	–	–	–	–	0.5165	0.7599

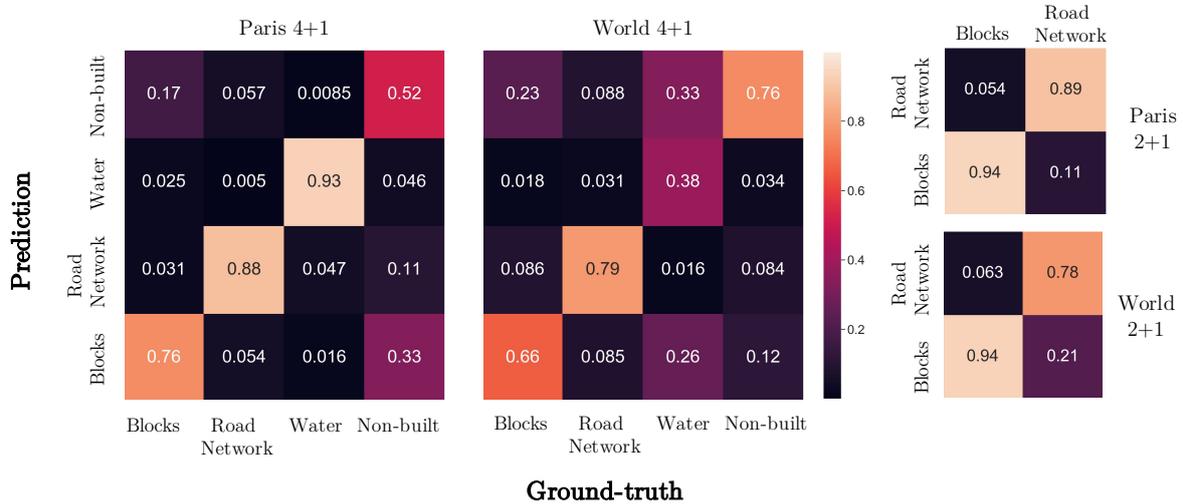


Figure 6: Confusion matrix, normalized according to ground-truth. The diagonal corresponds to recall.

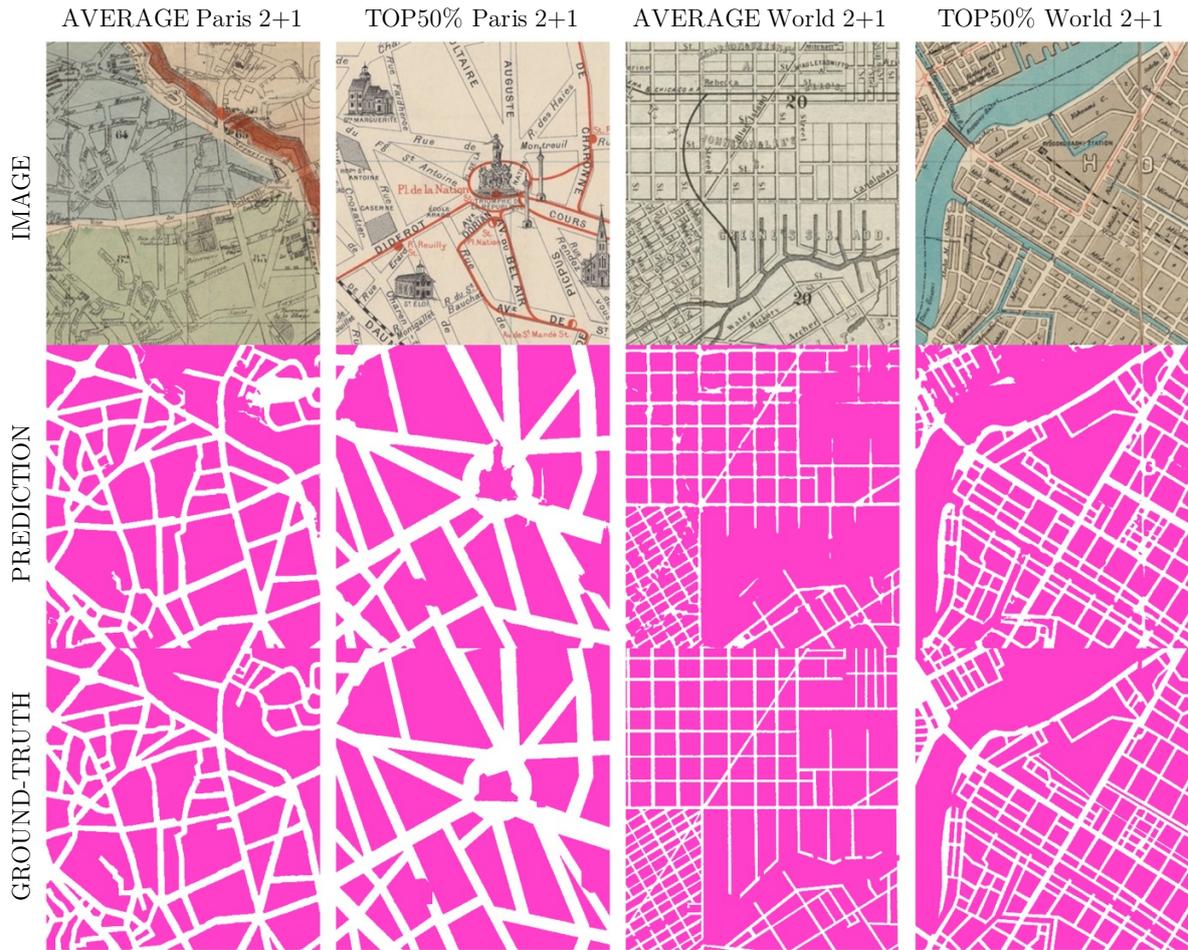


Figure 7: Examples of results. Each time, the first example is close to the median mIoU and the second is part of the top 50%. Values of the mIoU from left to right: 0.8965, 0.9506, 0.8079, and 0.8622.

For the World corpus, water on the contrary benefits from a relatively high precision but a very low recall. It is heavily confused with non-built and blocks. Non-built is still sometimes predicted as blocks, but for this dataset, the contrary is more frequent. Both blocks and road network classes suffer from a low precision.

3.3. Semantic segmentation experiments

3.3.1. Impact of learning transfer on each class

As one can see in Fig. 8, the performance in class segmentation shows significant disparities between both datasets. If, in general, the Parisian corpus achieves much higher performances, the World corpus seems to be better at recognizing the non-built lands. The transfer learning is quite successful for the World corpus, when pretrained on Paris. The water ($pvalue = 0.0063 < 0.05$) in particular seems to be better recognized. The transfer learning from the World corpus to Paris is a bit less successful. However, the water class also demonstrates a significant improvement (0.0076). Overall, the improvement is trending for the World corpus, when pretrained on Paris (0.062).

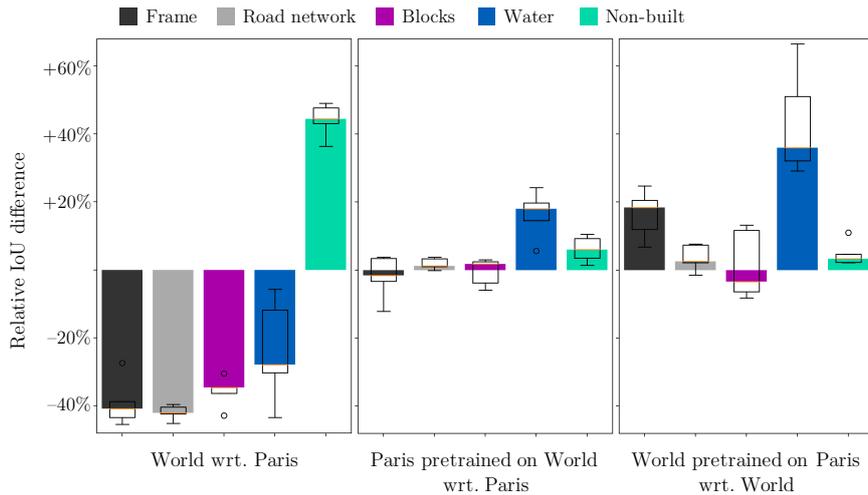


Figure 8: Per-class relative performance of regular training between World and Paris corpora (left), and per-class relative performance of transfer learning between World and Paris corpora (middle and right). The relative IoU is computed with regard to the median of the reference IoU. When the pretraining is not specified, the CNN is generically pretrained on ImageNet. Each experiment was repeated 5 times, the values are represented as boxplots.

3.3.2. Analysis of cross-cultural performances and dataset design

In total, the mIoU over the 8 experiments is 0.6112, which is noticeably better than the 0.5595 score obtained in the previous section for the same 4+1 World set. That means that the average performance is slightly better than the performance observed on the validation set. The mIoU can also be computed on each patch separately, which corresponds to an average of 0.5424. Fig. 9, shows a few examples of the high disparity between the top-50% and the bottom-50% of the World corpus, which was already noticed in the results of the previous experiment.

Maps that have been published by a Western country score 0.5608, while other maps score 0.4911. The region of the city represented is also impactful, with Subsaharian (0.6322) and North African (0.6280) cities scoring best, followed closely by Eastern Europe and Central Asia (0.5931), South America (0.5913), Western Europe (0.5891), and North America (0.5711). At the end of the line are the South Asian cities (0.3985). In the middle, one would find the Middle East (0.5048), East Asia (0.4896), Oceania (0.4685), and Central America (0.4635). The urban form also has a clear impact, as cities with a more regular (0.5453) or mixed (0.5535) urban form score better than cities with an irregular (0.5088) urban form. This performance drop is especially noticeable on the blocks class for regular (0.3548), mixed (0.3027), and irregular (0.2179) urban form.

3.3.3. Perspectives on confidence prediction

The average mIoU over the 10-folds is 0.6993. For this third setting, the correlation between the obtained confidence index and the reference is 0.571 ($p_{value} = 1.2 * 10^{-3}$) on the validation set, which represents an intermediate to high dependency.

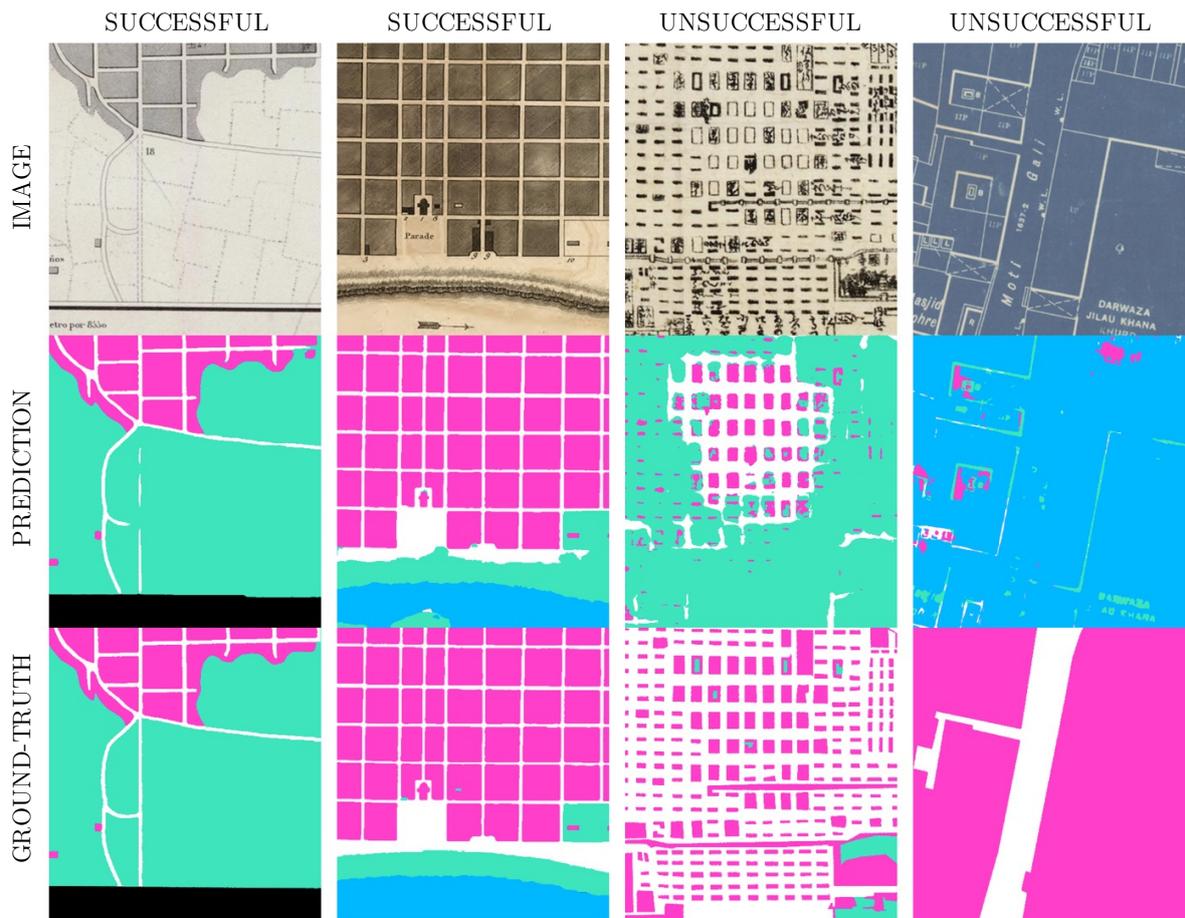


Figure 9: Examples of results. Each time, the first example is close to the median mIoU and the second is part of the top 50%. Values of the mean IoU from left to right: 0.8965, 0.9506, 0.8079, and 0.8622.

3.3.4. Importance of graphical cues for learning

The removal of color had noticeably almost no impact on performance. The median loss is only 1.3% with regard to the reference mIoU. The binarization of the values resulted in a drop of 7.2%. Finally, the disappearance of colors and textures led to a 10.4% decrease in performance. This experiment thus shows that even when most graphical cues are removed, most of the performance is conserved, and therefore that neural networks may also heavily rely on more abstract reasonings for image segmentation.

4. Discussion

As measured through the operationalization of the figuration, both corpora, Paris and the World, present a much greater figurative diversity than the other datasets used in the literature (Tab. 2). This is good news and validates the interest of the studied datasets. The USGS is massively less diverse than the other datasets, which is logical since it is a digital-born map. The representation of the different elements is therefore perfectly codified and reproduced. The ICDAR21 corpus also shows relatively little diversity. It is composed of plates published in

different years, but is still based on a single printed collection, and thus on a unified grammar. The Napoleonic cadaster for its part is famous for the high level of formalization of cartographic grammar. However, its manual execution explains a certain residual diversity. Finally, the two studied corpora show a similar level of figurative diversity, although Paris stems from a single cultural pool. This is consistent with the samples taken from this corpus, which show a great variety of grammars, a high density of information, and above all a high level of technicality, which also allows for an important figurative diversity.

Regarding the performances of semantic segmentation, while the Paris corpus shows excellent results which already indicate a potential for production (mIoU 0.8905, Tab. 4), the results of the World corpus seem globally still perfectible (mIoU 0.8055). However, the best half (top-50%) of the World corpus presents performances very similar to the Paris corpus and may thus also present an important automation potential (see also Fig. 9). The question therefore lies in identifying this outperforming half in the large map collections, to open automatic vectorization perspectives. As we demonstrated in the experiment on confidence prediction, the estimation of a confidence index is possible and could probably solve this problem by identifying promising maps beforehand. However, we consider that further research is still needed to maximize the reliability of such an index and that other paradigms should be explored.

The identification of the most promising maps can also be based on the results of cross-cultural validation. Indeed, we observed that maps with regular urban forms, like most colonial cities, or mixed forms, like most European capitals, are better segmented on average. This is consistent with the results on the importance of non-graphical cues for learning, which emphasize the crucial importance of non-visual features, such as morphology, topology, and semantic hierarchy, in CNN performance. In general, we notice that areas urbanized more recently, such as Africa, perform well. This is consistent with the findings on urban form. The poorer performance of South and East Asian cities can be explained on the one hand by atypical visual elements, from a Western point of view, by the irregularity of Indian and Islamic urban forms, as well as, in some cases, by the poor state of preservation of the documents. The little performance of the Oceanian maps could be explained by the relatively low representation of the region (around 3%) in the sample. The cross-cultural validation also highlights the greater ease of segmenting maps published in the West, although, as mentioned earlier, non-Western maps are on average more recent. These cultural biases tend to argue for a differential treatment of non-Western maps (e.g. Fig. 2 A.3-4, C.1-2).

The detailed analysis of the results of semantic segmentation brings many additional and interesting insights. First, we notice the particular case of the water class. This class was the least represented in the training corpora (Tab. 1), in particular for Paris, where it represents less than 3% of the surface. The proportion lies close to the road network for the World (12%). However, that class differs little from the other classes, figuratively ($r = 0.94$, Tab. 3). This explains its poor final performance ($IoU_{Paris} = 0.47$, $IoU_{World} = 0.33$, Tab. 4). The performance is higher for Paris. However, water seems to be mostly recognized by elimination of other classes, which is expressed in poor precision (0.49), compared to recall (0.88). Moreover, the water class is vulnerable to overfitting on the blue color, as demonstrated by the particular case of the blueprint (Fig. 9), while the results on the importance of graphical cues for learning show on the contrary that most of the performance relies on more abstract features. These elements point to an imbalance of the water class in the original sample. This imbalance is not specific to our corpus, since it is the result of geographical constraints. However, our research provides a first solution to this limitation. Indeed, the experiment on the impact of learning transfer on each class demonstrated that the water class could benefit from a significant transfer

(Fig. 8, +18% for Paris, +36% for the World). The constitution of large and diversified corpora intended for pretraining therefore seems to be justified by this example.

The second noteworthy result is that the non-built class scores much better for the World (0.6528), compared to Paris (0.4235), even though the World generally scores lower. For the World, however, the figuration of this class is not really outstanding (in average $r_{pearson} = 0.924$). In reality, this class benefits (and suffers) from a catch-all effect. Indeed, a third (0.33) of the water surfaces and nearly a quarter (0.23) of the blocks surfaces are wrongly classified as non-built, while the non-built class itself does not confuse these classes (0.03, respectively 0.12). This catch-all effect results in a relatively high recall, compared to Paris, while the impact on precision remains little, as this class represents a large area (0.359) of the dataset. As it can be seen in Fig. 6, this catch-all non-built class is the main reason for the underperformance of the World dataset. Solutions to this problem might include separating this class into two smaller and figuratively more specific classes.

For Paris, this same non-built class obtains lower results, though quite close to those obtained for water, for example. This class is less of a catch-all, as the balance between confusing and being confused is shifted. In particular, the non-built is abundantly confused with blocks (0.33), even though the figuration between those two is the least correlated (0.7). Conversely, blocks are also sometimes confused with the non-built (0.17). This confusion is by far the most important reason for this poor performance. Several explanations are possible. Notably, in Paris, this class mainly describes parks, as well as courtyards. Compared to the World, the topology of the non-built class is less outstanding. Some parks adopt forms that are quite similar to building blocks, especially in Paris. Moreover, the spatial juxtaposition of inner courtyards and buildings promotes confusion, especially in this direction, since an unrecognized inner courtyard will be classified as a non-built confused with blocks. The Parisian corpus also suffers from poor learning transfer from the World, which may be caused by a relatively large figuration distance between Paris and the World for this class ($r_{pearson} = 0.81$). This seems obvious, considering that the elements represented in one and the other corpus are also semantically quite different, as explained.

The arguments discussed above do not pretend to be exhaustive because neural networks have many as yet unknown springs. However, they do shed some light on a few points. We consider that the framework of the maps is ideal for this discussion on the ability of neural networks to combine figurative and topological cues and to open up avenues of understanding on their performances. Indeed, maps have a "textbook" figuration, with basic textures, such as hatching, some colors, and partly geometric morphology, squares, rectangles, trapezoids. They are therefore easier to characterize from this point of view. These conditions are met in very few fields of application of deep learning. In addition, for semantic segmentation, the ability to visualize the results and understand the errors is particularly useful for interpretation. This field therefore brings together all the elements that can help to better understand the performance of neural networks.

4.1. Conclusion

Developing a generic pipeline for processing historical maps will be an important milestone for massively extracting information from this rich family of cultural heritage documents. In this research, we made progress in understanding the figurative variance of cartographic corpora, and established a congruent metric to measure the figurative diversity of a collection based on the acuity of the multimodal distribution of graphical features. Through several

experiments, we have shown that neural networks are extremely robust in the face of figurative diversity, even if some map grammars seem more difficult to segment. This high performance is mostly due to the fact that neural networks can integrate highly abstract reasoning, such as morphology, topology, and semantic hierarchy, to supplement figurative features, without excluding the latter. This work, and its conclusions, paves the way for the generic segmentation of historical maps, highlighting the weaknesses of learning processes and outlining potential levers for action.

Acknowledgments

The authors declare to have no conflicts of interest. We would like to thank our former collaborators, Raphaël Barman and Nils Hamel, for their support on this project.

References

- [1] M. G. Arteaga. “Historical map polygon and feature extractor”. In: *Proceedings of the 1st ACM SIGSPATIAL International Workshop on MapInteraction*. MapInteract '13. Orlando, Florida: Association for Computing Machinery, 2013, pp. 66–71. DOI: 10.1145/2534931.2534932.
- [2] S. Banda, A. Agarwal, C. R. Rao, and R. Wankar. “Contour layer extraction from colour topographic map by feature selection approach”. In: *2011 IEEE Symposium on Computers Informatics*. 2011, pp. 425–430. DOI: 10.1109/isci.2011.5958953.
- [3] R. Barvir and V. Vozenilek. “Developing Versatile Graphic Map Load Metrics”. In: *ISPRS International Journal of Geo-Information* 9.12 (2020), p. 705. DOI: 10.3390/ijgi9120705.
- [4] R. Brügelmann. “Recognition of hatched cartographic patterns”. In: *International Archives of Photogrammetry and Remote Sensing* 31.B3 (1996), pp. 82–87.
- [5] Bysyk. *Github bigjpg*. 2019. URL: <https://github.com/by-syk/bigjpg-app>.
- [6] Y. S. Can, P. J. Gerrits, and M. E. Kabadayi. “Automatic Detection of Road Types From the Third Military Mapping Survey of Austria-Hungary Historical Map Series With Deep Convolutional Neural Networks”. In: *IEEE Access* 9 (2021), pp. 62847–62856. DOI: 10.1109/access.2021.3074897.
- [7] J. Chazalon, E. Carlinet, Y. Chen, J. Perret, B. Duménieu, C. Mallet, T. Géraud, V. Nguyen, N. Nguyen, J. Baloun, L. Lenc, and P. Král. *ICDAR 2021 Competition on Historical Map Segmentation*. 2021. URL: <https://arxiv.org/abs/2105.13265>.
- [8] Y.-Y. Chiang, W. Duan, S. Leyk, J. H. Uhl, and C. A. Knoblock. *Using Historical Maps in Scientific Studies: Applications, Challenges, and Best Practices*. SpringerBriefs in Geography. Cham: Springer International Publishing, 2020. DOI: 10.1007/978-3-319-66908-3.
- [9] Y.-Y. Chiang and C. A. Knoblock. “A general approach for extracting road vector data from raster maps”. In: *Ijdar* 16.1 (2013), pp. 55–81. DOI: 10.1007/s10032-011-0177-1.
- [10] Y.-Y. Chiang, S. Leyk, and C. A. Knoblock. “A Survey of Digital Map Processing Techniques”. In: *ACM Comput. Surv.* 47.1 (2014), 1:1–1:44. DOI: 10.1145/2557423.

- [11] Y.-Y. Chiang, S. Leyk, and C. A. Knoblock. “Efficient and Robust Graphics Recognition from Historical Maps”. In: *Graphics Recognition. New Trends and Challenges*. Ed. by Y.-B. Kwon and J.-M. Ogier. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2013, pp. 25–35. DOI: 10.1007/978-3-642-36824-0_3.
- [12] A. Cordeiro and P. Pina. “Colour map object separation”. In: *Remote Sensing: From Pixels to Processes*. 2006, pp. 243–247.
- [13] N. Dalal and B. Triggs. “Histograms of Oriented Gradients for Human Detection”. In: vol. 1. IEEE Computer Society, 2005, pp. 886–893. DOI: 10.1109/cvpr.2005.177.
- [14] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. “ImageNet: A large-scale hierarchical image database”. In: *IEEE Conference on Computer Vision and Pattern Recognition*. 2009, pp. 248–255. DOI: 10.1109/cvpr.2009.5206848.
- [15] D. B. Dhar and B. Chanda. “Extraction and recognition of geographical features from paper maps”. In: *Ijdar* 8.4 (2006), pp. 232–245. DOI: 10.1007/s10032-005-0010-9.
- [16] H. Ernstson, S. E. van der Leeuw, C. L. Redman, D. J. Meffert, G. Davis, C. Alfsen, and T. Elmqvist. “Urban Transitions: On Urban Resilience and Human-Dominated Ecosystems”. In: *Ambio* 39.8 (2010), pp. 531–545. DOI: 10.1007/s13280-010-0081-9.
- [17] A. Garcia-Molsosa, H. A. Orengo, D. Lawrence, G. Philip, K. Hopper, and C. A. Petrie. “Potential of deep learning segmentation for the extraction of archaeological features from historical map series”. In: *Archaeological Prospection* 28.2 (2021), pp. 187–199. DOI: 10.1002/arp.1807.
- [18] X. Glorot and Y. Bengio. “Understanding the difficulty of training deep feedforward neural networks”. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. Ed. by Y. W. Teh and M. Titterton. Vol. 9. Proceedings of Machine Learning Research. Chia Laguna Resort, Sardinia, Italy: Pmlr, 2010, pp. 249–256. URL: <http://proceedings.mlr.press/v9/glorot10a.html>.
- [19] B. Graeff, R. Carosio, B. Graeff, and R. Carosio. “Automatic Interpretation of Raster-Based Topographic Maps by Means of Queries”. In: *FIG XXII International Congress Washington, D. C., published on CD-ROM* (2002).
- [20] K. He, X. Zhang, S. Ren, and J. Sun. *Deep Residual Learning for Image Recognition*. 2015. URL: <https://arxiv.org/abs/1512.03385>.
- [21] M. Heitzler and L. Hurni. “Cartographic reconstruction of building footprints from historical maps: A study on the Swiss Siegfried map”. In: *Transactions in GIS* 24.2 (2020), pp. 442–461. DOI: 10.1111/tgis.12610.
- [22] K. Hosseini, K. McDonough, D. van Strien, O. Vane, and D. C. S. Wilson. “Maps of a Nation? The Digitized Ordnance Survey for New Historical Research”. In: *Journal of Victorian Culture* 26.2 (2021), pp. 284–299. DOI: 10.1093/jvcult/vcab009.
- [23] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. “Densely Connected Convolutional Networks”. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017, pp. 2261–2269. DOI: 10.1109/cvpr.2017.243.
- [24] A. Karmiloff-Smith. “Constraints on representational change: evidence from children’s drawing”. In: *Cognition* 34.1 (1990), pp. 57–83. DOI: 10.1016/0010-0277(90)90031-e.

- [25] E. Katona and G. Hudra. “An interpretation system for cadastral maps”. In: *Proceedings 10th International Conference on Image Analysis and Processing*. 1999, pp. 792–797. DOI: 10.1109/iciap.1999.797692.
- [26] N. W. Kim, J. Lee, H. Lee, and J. Seo. “Accurate segmentation of land regions in historical cadastral maps”. In: *Journal of Visual Communication and Image Representation* 25.5 (2014), pp. 1262–1274. DOI: 10.1016/j.jvcir.2014.01.001.
- [27] D. H. Laboratory. *dhSegment-torch*. 2021. URL: <https://github.com/dhlab-epfl/dhSegment-torch>.
- [28] R. G. Laycock, D. Drinkwater, and A. M. Day. “Exploring cultural heritage sites through space and time”. In: *J. Comput. Cult. Herit.* 1.2 (2008), 11:1–11:15. DOI: 10.1145/1434763.1434768.
- [29] S. D. Laycock, P. G. Brown, R. G. Laycock, and A. M. Day. “Aligning archive maps and extracting footprints for analysis of historic urban environments”. In: *Computers & Graphics. Virtual Reality in Brazil* 35.2 (2011), pp. 242–249. DOI: 10.1016/j.cag.2011.01.002.
- [30] S. Leyk. “Segmentation of Colour Layers in Historical Maps Based on Hierarchical Colour Sampling”. In: *Graphics Recognition. Achievements, Challenges, and Evolution*. Ed. by J.-M. Ogier, W. Liu, and J. Lladós. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2010, pp. 231–241. DOI: 10.1007/978-3-642-13728-0_21.
- [31] S. Leyk and R. Boesch. “Colors of the past: color image segmentation in historical topographic maps based on homogeneity”. In: *Geoinformatica* 14.1 (2009), p. 1. DOI: 10.1007/s10707-008-0074-z.
- [32] J. Long, E. Shelhamer, and T. Darrell. “Fully Convolutional Networks for Semantic Segmentation”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015, pp. 3431–3440. URL: <https://arxiv.org/abs/1411.4038>.
- [33] C. Mello, D. Costa, and T. d. Santos. “Automatic image segmentation of old topographic maps and floor plans”. In: *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 2012, pp. 132–137. DOI: 10.1109/icsmc.2012.6377689.
- [34] Q. Miao, P. Xu, T. Liu, Y. Yang, J. Zhang, and W. Li. “Linear Feature Separation From Topographic Maps Using Energy Density and the Shear Transform”. In: *IEEE Transactions on Image Processing* 22.4 (2013), pp. 1548–1558. DOI: 10.1109/tip.2012.2233487.
- [35] T. Miyoshi, W. Li, K. Kaneda, H. Yamashita, and E. Nakamae. “Automatic extraction of buildings utilizing geometric features of a scanned topographic map”. In: *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004*. Vol. 3. 2004, 626–629 Vol.3. DOI: 10.1109/icpr.2004.1334607.
- [36] S. Muhs. “Computational Delineation of Built-up Area at Urban Block Level from Topographic Maps: A Contribution to Retrospective Monitoring of Urban Dynamics”. PhD thesis. Dresden, Germany: Technische Universität Dresden, 2019. URL: <https://nbn-resolving.org/urn:nbn:de:bsz:14-qucosa2-340364>.

- [37] S. Muhs, H. Herold, G. Meinel, D. Burghardt, and O. Kretschmer. “Automatic delineation of built-up area at urban block level from topographic maps”. In: *Computers, Environment and Urban Systems* 58 (2016), pp. 71–84. DOI: 10.1016/j.compenvurbsys.2016.04.001.
- [38] S. Muhs, G. Meinel, D. Burghardt, and H. Herold. “Automatisierte Baublockabgrenzung in Topographischen Karten”. In: *Flächennutzungsmonitoring. ... Flächennutzungsmonitoring V: Methodik, Analyseergebnisse, Flächenmanagement*. IÖR Schriften 61. Berlin: Rhombos-Verl, 2013, pp. 211–219.
- [39] J.-M. Ogier, R. Mullot, J. Labiche, and Y. Lecourtier. “Technical Map Interpretation: A Distributed Approach”. In: *Pattern Analysis & Applications* 3.2 (2000), pp. 88–103. DOI: 10.1007/pl00010983.
- [40] T. Ojala, M. Pietikainen, and T. Maenpaa. “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24.7 (2002), pp. 971–987. DOI: 10.1109/tpami.2002.1017623.
- [41] S. Oliveira, I. di Lenardo, B. Tourenc, and F. Kaplan. “A deep learning approach to Cadastral Computing”. In: Utrecht, Netherlands, 2019. URL: <https://dev.clariah.nl/files/dh2019/boa/0691.html>.
- [42] S. A. Oliveira, B. Seguin, and F. Kaplan. “dhSegment: A generic deep-learning approach for document segmentation”. In: *16th International Conference on Frontiers in Handwriting Recognition (ICFHR)* (2018), pp. 7–12. DOI: 10.1109/icfhr-2018.2018.00011.
- [43] N. Otsu. “A Threshold Selection Method from Gray-Level Histograms”. In: *IEEE Transactions on Systems, Man, and Cybernetics* 9.1 (1979), pp. 62–66. DOI: 10.1109/tsmc.1979.4310076.
- [44] R. Petitpierre. *Generic Semantic Segmentation of Historical Maps - Github repository*. 2021. URL: <https://github.com/RPetitpierre/Generic%5C%5FSemantic%5C%5FSegmentation%5C%5Fof%5C%5FHistorical%5C%5FMaps>.
- [45] R. Petitpierre. *Historical City Maps Semantic Segmentation Dataset*. 2021. DOI: 10.5281/zenodo.5513639.
- [46] R. Petitpierre. “Neural networks for semantic segmentation of historical city maps: Cross-cultural performance and the impact of figurative diversity”. In: *CoRR* abs/2101.12478 (2021). URL: <https://arxiv.org/abs/2101.12478>.
- [47] “Procédé d’impression des cartes géographiques en couleur”. In: *L’Echo du monde savant, journal analytique des nouvelles et des cours scientifiques*. 2. Paris, 1840, pp. 401–402. URL: <https://go.epfl.ch/procede%5C%5Fdimpression%5C%5F1840>.
- [48] O. Ronneberger, P. Fischer, and T. Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: *arXiv:1505.04597 [cs]* (2015). URL: <http://arxiv.org/abs/1505.04597>.
- [49] S. Salvatore and P. Guitton. “Contour line recognition from scanned topographic maps”. In: (2004). URL: <http://dSPACE5.zcu.cz/handle/11025/1744>.
- [50] A. Savitzky and M. J. E. Golay. “Smoothing and Differentiation of Data by Simplified Least Squares Procedures.” In: *Anal. Chem.* 36.8 (1964), pp. 1627–1639. DOI: 10.1021/ac60214a047.

- [51] D. Schemala. “Semantische Segmentierung historischer topographischer Karten”. PhD thesis. Dresden, Germany: Technische Universität Dresden, 2016.
- [52] G. Touya, B. Decherf, M. Lalanne, and M. Dumont. “Comparing image-based methods for assessing visual clutter in generalized maps”. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* II-3/w5 (2015), pp. 227–233. DOI: 10.5194/isprsannals-II-3-W5-227-2015.
- [53] J. H. Uhl, S. Leyk, Y.-Y. Chiang, W. Duan, and C. A. Knoblock. “Automated Extraction of Human Settlement Patterns From Historical Topographic Map Series Using Weakly Supervised Convolutional Neural Networks”. In: *IEEE Access* 8 (2020), pp. 6978–6996. DOI: 10.1109/access.2019.2963213.
- [54] J. H. Uhl, S. Leyk, Y.-Y. Chiang, W. Duan, and C. A. Knoblock. “Map Archive Mining: Visual-Analytical Approaches to Explore Large Historical Map Collections”. In: *ISPRS International Journal of Geo-Information* 7.4 (2018), p. 148. DOI: 10.3390/ijgi7040148.
- [55] J.-M. Viglino and M. Pierrot-Deseilligny. “A vector approach for automatic interpretation of the French cadastral map”. In: *7th International Conference on Document Analysis and Recognition. Proceedings*. 2003, 304–308 vol.1. DOI: 10.1109/icdar.2003.1227678.
- [56] J. Wu, P. Wei, X. Yuan, Z. Shu, Y.-Y. Chiang, Z. Fu, and M. Deng. “A New Gabor Filter-Based Method for Automatic Recognition of Hatched Residential Areas”. In: *IEEE Access* 7 (2019), pp. 40649–40662. DOI: 10.1109/access.2019.2907114.
- [57] H. Yamada, K. Yamamoto, and K. Hosokawa. “Directional mathematical morphology and reformalized Hough transformation for the analysis of topographic maps”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15.4 (1993), pp. 380–387. DOI: 10.1109/34.206957.