Detection of Pictorial Map Objects with Convolutional Neural Networks

Raimund Schnürer [a,*], René Sieber [a], Jost Schmid-Lanter [b], A. Cengiz Öztireli [c], Lorenz Hurni [a]

[a] *Institute of Cartography and Geoinformation, ETH Zurich, Zurich, Switzerland*
[b] *Abteilung Karten und Panoramen, Zentralbibliothek Zürich, Zurich, Switzerland*
[c] *Department of Computer Science and Technology, University of Cambridge, Cambridge, United Kingdom*

* Corresponding author: schnuerer@ethz.ch
Stefano-Franscini-Platz 5, 8093 Zurich, Switzerland

Abstract

In this work, realistically drawn objects are identified on digital maps by convolutional neural networks. For the first two experiments, 6200 images were retrieved from Pinterest. While alternating image input options, two binary classifiers based on Xception and InceptionResNetV2 were trained to separate maps and pictorial maps. Results showed that the accuracy is 95-97% to distinguish maps from other images, whereas maps with pictorial objects are correctly classified at rates of 87-92%. For a third experiment, bounding boxes of 3200 sailing ships were annotated in historic maps from different digital libraries. Faster R-CNN and RetinaNet were compared to determine the box coordinates, while adjusting anchor scales and examining configurations for small objects. As results, an average precision of 32% was obtained for Faster R-CNN and of 36% for RetinaNet. Research outcomes are relevant for crawling map images in the Internet and enhancing the advanced search of digital map catalogues.

Keywords

Artificial intelligence; convolutional neural networks; pictorial maps; map libraries; classification; object detection

Introduction

An increasing number of maps is available in digital form. On the one hand, contemporary maps are created almost exclusively using software applications and distributed via electronic devices. Keeping track of these newly published maps is a challenging endeavour: Maps need to be identified as such and distinguished from other types of digital images. Some curators, for instance, compile and feature new maps in books (e.g. Clarke, 2015) and on blogs (Agarwal, 2019). However, the selection of maps largely depends on personal preference and the amount of depicted maps is limited since the collection process is undertaken manually. Detecting maps automatically would revivify attempts of developing a global search engine for maps (Goel et al., 2011). Another challenge is to categorise maps according to thematic, stylistic, or other criteria. Pictorial maps, for example, can be found in specialised books (e.g. Antoniou and Kotmair, 2015) or by keyword search on social media websites (e.g. Instagram, n.d.). In the latter case, when being annotated by laymen, tags might be incorrect though. Recommendation systems for tags (Nguyen et al., 2017) could support users in creating or validating map metadata.

On the other hand, libraries digitise their stocks of historic printed maps to preserve these unique documents and to make them accessible over the Internet. Historic maps are not only acknowledged as a heritage, but also recognised as a rich data source of topographic and socio-geographic features. Map readers, such as historians, may be interested in these features, for example place names (Jordan et al., 2009), land cover (Fuchs et al., 2015), or illustrations like sea-monsters (Duzer, 2014). Given the 'Iconic Turn' (Alloa, 2016) in history research, illustrations within maps and on map borders get into the focus of investigations more often. In many library catalogues however, the presence of pictorial objects in maps can only be deduced from the title, the description, the map type, or keywords, if at all. An additional search filter option would enhance these catalogues and facilitate researchers answering semiotic and iconic questions (e.g. Baumgärtner et al., 2019). Considering the on-going trend of storytelling in cartography (Caquard and Cartwright, 2014), another use case would be to add the detected pictorial objects as protagonists to modern maps. Map objects, like persons or animals, could tell a personal story, give background information to a topic, or highlight interesting places on a map, for example for touristic purposes (Graça and Fiori, 2015).

Convolutional Neural Networks (CNNs) are a promising technology to tackle these challenges. CNNs are a type of artificial neural network, which is a computational model inspired by the biological neural network of the human brain (Dayhoff 1990 cited in Merwin et al., 2009). Artificial neural networks are more suited to fulfil complex perceptual tasks than other machine learning methods like support vector machines (Bengio and LeCun, 2007). The increase of parallel computing capabilities of the graphic processing units reduced training times of artificial neural networks, including CNNs, largely in the past years (Scherer et al., 2010). This enabled a growing number of researchers to experiment with different architectures, for instance to classify images (Krizhevsky et al., 2012). Other tasks include object detection and image segmentation in domains like autonomous driving (Siam et al., 2017), remote sensing (Audebert et al., 2016), or medicine (Ronneberger et al., 2015). CNNs take images, usually encoded as three-dimensional arrays (i.e. height, width, colour channels), as input. They output arrays of different dimensions and sizes, for example one-hot encoded categories (i.e. an array containing only zeros except a single one value), or arrays of the same dimension and size, as it is the case for autoencoders (Goodfellow et al., 2016). In between, CNNs perform mathematical operations in intermediate layers, such as convolutional, pooling, or fully-connected layers (O'Shea and Nash, 2015), to detect patterns in images. Layer parameters, for example values of convolution matrices (aka kernels), are gradually adapted to minimise differences between the actual and the desired output. Previous research focused mainly on the recognition of objects in natural images like photos (Girshick et al., 2013), and marginally on man-made images like artwork (Gonthier et al., 2018) or mangas (Yanagisawa et al., 2018), but only scarcely on maps (see Related work).

In this paper, we examine whether CNNs are able to detect objects in pictorial maps. Some of the oldest maps contain pictorial, that are realistically depicted, symbols and illustrations. The Bedolina map, for example, a rock engraving created in Val Camonica around 1500 BC, shows houses as well as fields with humans and animals as pictograms (Turconi, 1997). Pictorial maps flourished in the Middle Ages and Renaissance when painting and mapmaking were closely related (Rees 1980 cited in Kent, 2012). It was the Age of Discovery where monsters, for instance on Ortelius' (1585) Islandia map, symbolised the dangers of the sea and the fear of the unknown. In the following decades, pictorial maps declined due to different map materials and production techniques (Wallis and Robinson, 1987), and the growing sense of accurate geographic information (Child, 1956)**.** The next heyday of pictorial maps was in the 20th century, especially in the United States of America (Hornsby, 2017), which yielded Goodman and Neuhaus' (1930) map of Berkeley, for instance. Illustrative objects enlivened the map by portraying local customs and typical actions. On the other side of the coin, large menacing figures appeared on propaganda maps during the two world wars (Mason, 2016). Today, pictorial objects are often used in maps for tourism and leisure time. One of their purposes is to support underlying topographic features, like cars driving on roads. Graça and Fiori (2015) recommend the usage of pictorial symbols in tourist maps which shall encourage the reader to visit the represented locations. Sarjakoski et al. (2009) give an example of a project using comic-like icons on mobile maps for a national park to "possibly invoke positive emotions" (p. 113). Moreover, pictorial objects are used nowadays to teach map literacy to children. For example, different animals and other agricultural products are represented as pictograms in an agriculture map of a Bulgarian school atlas (Bandrova, 2003). Lastly, maps of fantasy books or in computer games, mimicking the style of the

Middle Ages and Renaissance, may also contain pictorial map objects (Lamb and Johnson, 2014).

Pictorial maps follow the typical design process of maps. Child (1956) lists, for example, the purpose of the map, the method of reproduction, colour, projection, and lettering as typical decisions which authors should consider when creating a pictorial map. Child further emphasises and exemplifies pictorial symbols for cultural or man-made features, water, relief, and vegetation. Those "should be clear and simple and if possible recognisable on sight" and the "size of the symbol should not be too large for the scale of the map" (Child, 1956 p. 68). Holmes (1991) states that an outline may not be enough for pictorial objects, therefore internal shading and patterns may be added complementary. Moreover, Holmes recommends the use of characteristic attributes, for example flying birds, to recognise pictorial symbols. According to Roman (2015), illustrative objects in maps shall be arranged based on the ABC-rule: (A) elements visible at first glance, (B) elements which support A, (C) elements which support the overall image. Beyond, Roman endorses the four I's as design functions: Identification (= what is the map about), Image (= visual relation to the map), Information (= additional literal facts), Incidentals (= engagement of the map-reader for the first three functions). While simplicity, distinct outlines, and the clear layout may facilitate the recognition of pictorial objects by CNNs, the different drawing styles may counteract it.

The overall goal of this work is to provide training datasets and first baselines to detect pictorial objects in maps with CNNs. As one dataset contains also depictions other than maps, maps are first separated from these non-maps with two state-of-the-art CNNs for image classification. To justify this division, we establish a modern definition of maps, which incorporates digital developments of recent years, thus contributing to the ICA Research Agenda (Virrantaus et al., 2009). Moreover, the trained CNNs may help to recognise maps when crawling images in the web. With the same two classifier CNNs, maps are next distinguished according to their level of abstraction, whereat pictorial maps are those with less abstract objects. This differentiation may be used to reveal inconsistencies in pictorial map tags on social media websites, amongst others. For the definition, we relate pictorial maps to decorative, illustrative, and figurative maps. Finally, as an example for a frequently encountered pictorial object on historic maps, sailing ships are identified with two CNNs targeted at object detection. These CNNs output bounding box coordinates of individual ancient ships. These detection results may further enhance the advanced search of digital map libraries when adding sailing ships to the filter options. The development of customised CNN architectures will be subject of future work to improve the accuracies of the individual tasks.

Related work

Artificial neural networks (ANNs) which have been often applied in cartography are self-organising maps and backpropagation neural networks. Sen et al. (2014), for instance, used a self-organising map for line generalisation, in particular for omitting rivers at certain scales. Merwin et al. (2009) interpolated values, such as population counts, of areas, where source and target zones are differently subdivided, with a backpropagation neural network. Other examples of ANNs in cartography comprise a particle swarm optimisation neural network (Wang et al., 2015) or a multilayer perceptron and radial basis function network combined with the Weighted Effective Area algorithm (Olszewski et al., 2018). CNNs though have been hardly used in cartography. Duan et al. (2018) extracted railroads and waterlines from historical topographic maps with a fully convolutional network, a CNN which discards fully-connected layers. The authors achieved promising results with accuracy rates at 85% to 93%. Feng et al. (2019) trained CNNs to generalise building footprints at different scales. The authors were successful in producing visually pleasing results and they plan to preserve also rectangularity and parallelisms of buildings in future. Kang et al. (2019) established a classifier for style-transferred maps which evaluates whether the design characteristics of the original map were preserved. Considering indoor mapping, CNNs helped to find walls (Dodge et al., 2017) and junctions (Liu et al., 2017), and to detect objects like doors (Ziran and Marinai, 2018; Dodge et al., 2017) in floor plans. Due to the scarcity of related work regarding maps, we give a brief overview of popular CNNs for classification of and object detection in natural images since we used some of them in our experiments.

Classification is a task for CNNs aiming to tell what is depicted in an image. One of the first CNNs which solved this task adequately was *AlexNet* (Krizhevsky et al., 2012) consisting of five convolutional layers and three fully-connected layers. An improvement over AlexNet are the VGGNets (Simonyan and Zisserman, 2014) which use smaller kernels in the convolutional layers. The most popular versions of this CNN are *VGG16* and *VGG19*, whereat 16 and 19 correspond to the total number of convolutional and fully-connected layers. One of the next advancements were so-called *Inception* modules (Szegedy et al., 2015) which have convolutional layers with different kernel sizes and a pooling layer in parallel. In the following, an identity mapping in parallel to convolutional layers was introduced by *ResNet* (He et al., 2015), leading to better optimisations of changes (= residuals) from input to output. Both networks were combined to *InceptionResNetV2* (Szegedy et al., 2016) which further improved the classification accuracy. An alternative with fewer layers but about the same effectiveness is given by *Xception* (Chollet, 2016). Overall, the accuracy (top-1, single model, single crop) of distinguishing 1000 categories of the ImageNet (Stanford Vision Lab, 2016) dataset on natural images was improved from 62.5% in AlexNet to 79% in Xception and 80.1% in InceptionResNetV2, as reported in the cited articles.

Another task of CNNs is to detect locations of objects in images. At this, *R-CNN* (Girshick et al., 2013) pioneered by feeding resized bounding box proposals, obtained by the selective search algorithm (Uijlings et al., 2013), into AlexNet. The performance was increased by *Fast R-CNN* (Girshick, 2015) where the image on the whole next to the bounding box proposals are taken as inputs for a classification CNN, namely VGG16. In *Faster R-CNN* (Ren et al. 2015), another iteration, the region proposals are not pre-generated, but predicted by the CNN from anchor boxes. In parallel to the detectors above where the image is processed in multiple stages, CNNs have been developed which output labelled regions in one stage. Prominent examples for these one-stage detectors are *SSD* (Liu et al. 2016), *RetinaNet* (Lin et al., 2017), and *YOLO* (Redmon and Farhadi, 2018). RetinaNet, for instance, concatenates ResNet layers of different resolutions, whereat layers of lower resolutions are upsampled. Of the described architectures, YOLO is the fastest (< 50ms), but RetinaNet is most precise on average (37.8%).

Experiments

Classification of maps vs. non-maps

Definitions

Over the years, a multitude of map definitions have been established. In our work, we like to apply a modern definition which includes also trends like maps of fictional spaces, indoor maps, and 3D visualisations. Conventional definitions however do not take these current developments into account. For example, according to Cartwright (2014), “a map is a symbolised image of geographical reality, representing selected features or characteristics, resulting from the creative effort of its author’s execution of choices, and is designed for use when spatial relationships are of primary relevance” (p. 528). Based on this definition, maps are (static) images which clearly neglects the interactivity introduced by digital mapping. Therefore, Kraak and Fabrikant (2017) tried to establish a new definition by collecting responses of cartographers in a survey. They agreed on the least common denominator of their suggestions and proposed the definition: “A map is a visual representation of an environment” (p. 6). Clearly, this definition is not as restrictive as previous ones, however we would deduce that photos, paintings, circuit diagrams, and visualisations of non-spatial environments (e.g. social relationships) would be also counted as maps. For this reason, we would like to introduce a narrower definition for our work:

A map is a scaled-down 2D or 3D representation – optionally animated and interactive – of macroscopic spaces – possibly with additional temporal and thematic information – where features are symbolised and relationships between them are mainly preserved.

As the definition is formed of different aspects, we like to explain briefly our intentions in the following:

Scaled-down: Map scales shall be always smaller than the identical scale (1:1). A map of a model railroad set, for example, would have a very large scale (e.g. 1:5). Up-scaled representations, such as circuit diagrams of computers, shall be excluded.

2D or 3D: Maps shall cover 2D planes (e.g. printed map sheets) or 3D spaces (e.g. Augmented Reality maps). We would count 2.5D representations to 3D. 1D representations however, like stops of a certain bus line or a list of waypoints for route navigation, shall be excluded. We would refer the number of dimensions only to space, separately from time and theme (see additional temporal and thematic information) which are seen by some as 4D and nD representations. We do not distinguish between pseudo and true 3D and we would count cartographic 3D representations on 2D surfaces (e.g. on computer screens) as 3D maps.

Representations: Maps shall depict spatial entities in a certain manner (see also *Features are symbolised*), but they are not those entities themselves.

Animated and interactive: Maps shall include temporal (e.g. glacier motions) and non-temporal animations (e.g. adaptive generalisation when zooming in). Interactivity, which changes the map content by user inputs (e.g. dropdown menu selection), is especially relevant for digital maps.

Macroscopic: Maps shall depict spaces visible to the human eye. Maps of the outer space (i.e. celestial maps) and of indoor spaces (e.g. flats) would thus be included. Microscopic spaces on a cellular or atomic level would be excluded. Illustrations like the interior of a car or a wardrobe would be a border case.

Spaces: Maps shall depict the real world and fictional spaces (e.g. books, computer games, dreams).

Additional temporal and thematic information: Space-time cubes and thematic maps shall be included. Time lines and mind maps to a certain topic shall be excluded.

Features are symbolised: The creation process of maps from the data model to the visualisation shall follow certain rules and conventions (e.g. styling, generalisation, projection). This shall exclude paintings, where the painter has more freedom, and photos, which are not abstracted.

Relationships are mainly preserved: The topology of features shall be primarily maintained to allow orientation in space, however some distortions shall be possible (e.g. cartograms, small displacements). Depictions where features are arranged by other attributes than location, for example when sorting country shapes alphabetically, would rather be infographics.

Data

In total, 3100 maps and 3100 non-maps were collected from Pinterest (n.d.). On this social media website, people can share memorable images, whereat a preview of the image is shown and in many cases a link to the original image source is given. Among those images are a large number of maps varying in time, spatial extent, theme, and style. Since a method to query by text is not offered by the application programming interface (API) of Pinterest, we used Google's (2019) Custom Search API instead to retrieve about 8000 images with keyword 'illustrated map' and having the site restricted to Pinterest. Maps were then separated manually from non-maps to create training and validation data for the CNNs. We categorised an image as a map when all non-optional criteria of the above definition were fulfilled. In case one of the mandatory requirements was violated, we classified the image as a non-map. Mixtures between maps and non-maps, that are maps or map-related products appearing in the real world or real-world objects placed on maps, were excluded because they fit into both categories and their amount was about nine times less than the collected images of the other two categories. As the number of maps was higher than the number of non-maps, non-maps were enriched with 141 images having the keywords 'illustration', 'sketch', and 'painting'. Another 1569 random non-maps were added by the keyword 'pinimg' since this string is contained in all URLs of Pinterest images. Too closely zoomed maps, duplicate and very similar images were removed from the search results. The remaining images have a width of 566 pixels and height of 552 pixels on average. We split the images with a ratio of 60:40 into train and validation sets for the CNNs.

Procedure

We tested the CNNs Xception and InceptionResNetV2 to classify images as either maps or non-maps. These networks take RGB images with a size of 299x299px as input. As our images exceed the size in either height or width, we tested three methods for feeding images into the networks:

- Resized: Images are resized to the input size without maintaining the aspect ratio.
- Middle random crop: The smaller image side is downscaled to 299px while maintaining the aspect ratio. In case the smaller image side is already less than 299px, this image side is up-scaled to 299px while maintaining the aspect ratio. Afterwards, in both cases, a random crop is carried out along the larger image side to reduce the side to 299px.
- Random crop: A 299x299px random patch is cropped from the image. In case the smaller image side is less than 299px, this image side is first up-scaled to 299px while maintaining the aspect ratio and the other image side is reduced to 299px (which is identical to the second middle random crop case).

The Lanczos filter is used for resizing the images. We assume an equal performance of those methods since while the whole image is processed for the first option, undistorted details of images are taken into account in the third option. The second option is a mixture between the first and the third option. During training, crops are randomised for each image in each epoch.

Both CNNs are initialised with weights from models pre-trained on the ImageNet dataset and fed with images in batches of 16. The models are retrained for 40 epochs with a learning rate of 10^{-5}, binary crossentropy loss, and the Adam optimiser. Retraining one model took about 90 minutes with a NVIDIA GeForce GTX 1080 graphics board. We used the software library TensorFlow (n.d.) for Python with its high-level API Keras, where Xception and InceptionResNetV2 are pre-implemented.

Results

We averaged the validation results of three Xception and InceptionResNetV2 models, which have achieved the highest accuracy during training, while changing image options (Table 1). For our classification tasks, we define accuracy as the number of all correct predictions divided by the number of all predictions. A prediction is counted as correct when its class score is higher than 0.5. Overall, the accuracies are quite high and only marginally different between the different input options in our experiment. As the accuracies for random crop are lower than the other two input options, we calculated the accuracy additionally when splitting the image into cells of 299x299px along a regular grid and averaging results from these cells by applying the retrained model from the random crop. While this approach is more time-consuming, the accuracy is slightly higher than in the first two approaches. We also tried to train the models from scratch instead of initialising them with ImageNet weights, however this resulted in a significantly lower accuracy. When using a higher learning rate, the loss did not converge that smoothly. Using a 70:30 split between training and validation images led to an alternating loss.

The classification results for a threshold of 0.5 are nearly consistent to the areas under the Receiver Operating Characteristics (ROC) curves for the different image input options (Figure 1). A ROC curve shows the relationship between the true and false positive rate for varying classification thresholds, whereat the area under the curve (auc) is 1 in an ideal case. To distinguish between maps and non-maps, averaging the score of image grid cells leads to the largest auc (0.994) and the random crop to the smallest auc (0.991) for both classification models. Interestingly, Xception is slightly more performant than InceptionResNetV2 for the average over grid calculation considering the auc metric.

As the CNNs achieve a high categorisation accuracy between maps and non-maps, we only show the failures cases as qualitative results. An artistically styled world map, a street and store map of Los Angeles, and a perspective city map of Torun (Figure 2) were misclassified in all twelve runs of the two networks for resizing and averaging over grid. We restricted us to these two methods as they do not involve any randomisation during evaluation. According to our definition, the first example is considered as a map because it maintains shapes and spatial relationships between the continents, while the latter example annotates a 3D scene with enlarged buildings. The second example would be counted as a map even with a more conservative definition. Regarding non-maps, the CNNs wrongly categorised a graph showing relations between painters, a collage of letters and telegrams, and US states shaped as a heart (Figure 3) in eleven out of twelve runs. As thematic and not spatial relationships are depicted in the first case and as topology is not preserved in the latter case, we do not count them as maps based on our definition. The second case is clearly no map, even with a broader definition.

	Xception	InceptionResNetV2
Resized	96.47%	96.60%
Middle random crop	96.52%	96.41%
Random crop		
- random crop	95.50%	95.89%
- average over grid	**96.63%**	**96.76%**

Table 1: Correct classifications of maps and non-maps for the examined CNNs and image input options (as explained in Procedure). The values are averages of validation accuracies of three retrained models having achieved the highest accuracy during training.

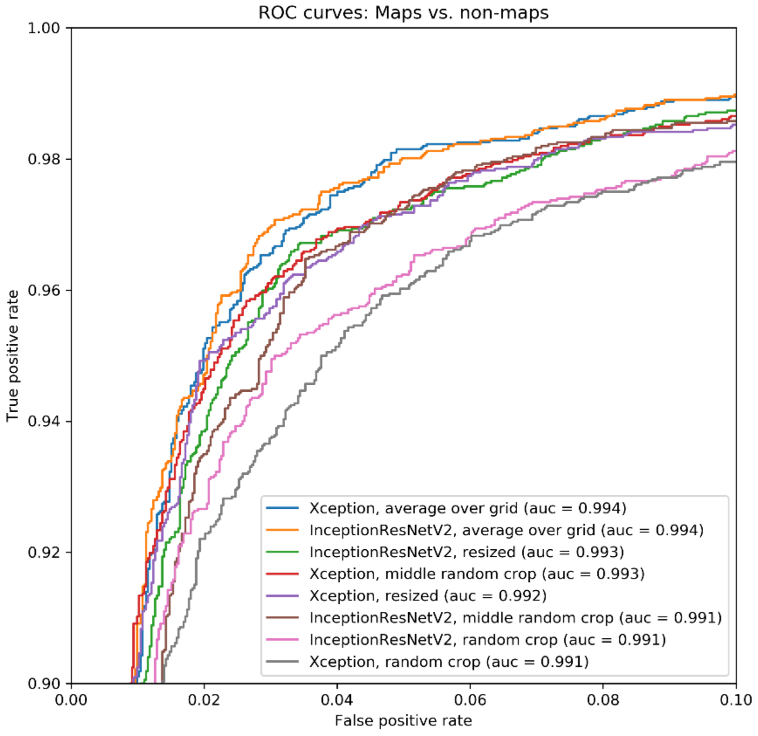

Figure 1: ROC curves (enlarged) and auc scores for the tested CNNs and image evaluation options to classify maps and non-maps

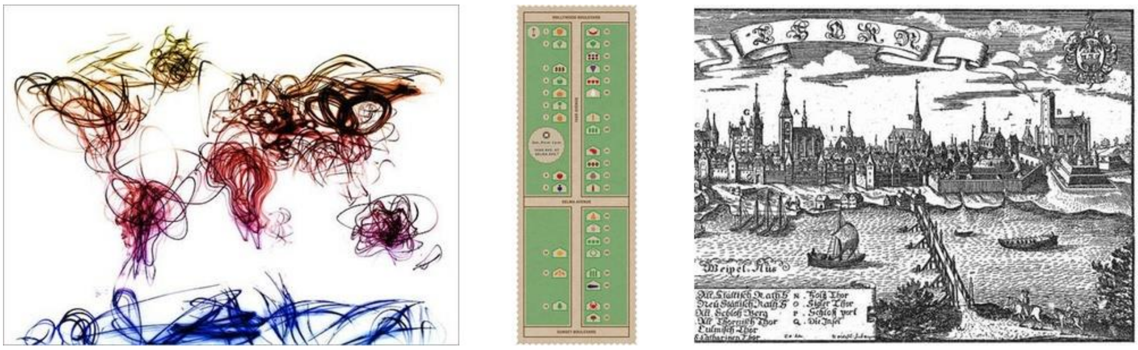

Figure 2: The three most frequently misclassified maps by both CNN models for resized and average over grid image evaluation options (image sources: Pinterest - https://i.pinimg.com/736x/9d/40/93/9d4093cebf375ef698c2022857b83de4--world-map-canvas-world-map-art.jpg, https://i.pinimg.com/736x/5e/94/a1/5e94a1054a88227364c82db48cbbd747--easter--food-design.jpg, https://i.pinimg.com/736x/f5/7d/38/f57d38e1acddef874dcb529fee617290--maps.jpg)

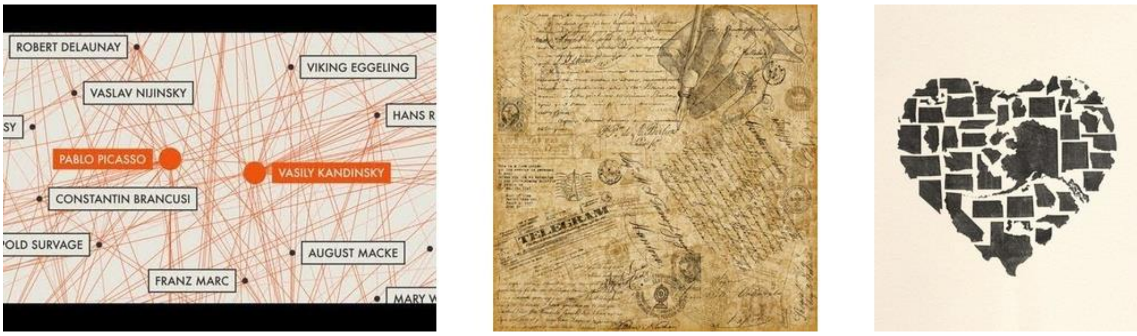

Figure 3: The three most frequently misclassified non-maps by both CNN models for resized and average over grid image evaluation options (image sources: Pinterest - https://i.pinimg.com/736x/bb/4c/52/bb4c5218917d2368937279be05deb528--moma-org-the-artist.jpg, https://i.pinimg.com/736x/bb/8d/1b/bb8d1b57149f189eaf3deed58e4a7482.jpg, https://i.pinimg.com/736x/c1/4c/0a/c14c0a9ec176a79addef054c1e134e95--heart-map-my-heart.jpg)

Classification of pictorial maps vs. non-pictorial maps

Definitions

In this section, we like to characterise pictorial maps and relate them to decorated/decorative, illustrated/illustrative, and figurative maps. "'Pictorial' as opposed to 'decorative' maps might be defined as those which are intended primarily for instruction conveyed by means of naturalistic or realistically drawn details" (Child, 1956, p. 6). Wallis and Robinson (1987) state that "topographical information is delineated by more or less realistic drawings, illustrations of features in elevation, and small bird's eye view sketches" (p. 43) in pictorial maps. Both definitions have in common choosing verisimilar (i.e. very close to reality) or indexed (i.e. parametric) representations (Bodum, 2005) for map features. These realistically drawn objects - or alternatively pictures (i.e. paintings or photos) in panels - could be placed inside the map, on the map border, or aside the map. Anthropomorphic maps, where figures are coalesced with the map content, would also count to pictorial maps. With instruction, Child seems to refer to educative maps for students or interested adults which explain certain topics (e.g. animals of the world, cuisine of a country, industries of a city) by pictorial symbols. These thematic maps are missing in Wallis and Robinson's definition as only topographic information is mentioned. In return, it is indicated that pictorial maps could be drawn from an oblique angle, thus panoramic/perspective maps would also be a subset of pictorial maps.

According to Child (1956), maps are decorative when "the composition, lettering and embellishments have all been considered as parts of the design" (p. 32). While we think that composition is important for all kind of maps, flourishes in lettering may be a distinct property of decorative maps. In addition, embellishments like cartouches and ornamentation seem to occur frequently in decorative maps. Other authors of books (e.g. Barron, 1990; Skelton, 1966) give many examples of decorative maps, but not a clear definition. Theoretically, people may declare maps as decorative when they beautify a place, for instance a wall map decorating a living room. Illustrated maps are also only vaguely defined. According to Roman (2015), illustrated maps "compress and distort the reality to fit the mental image of a place" (p. 6). A corresponding example would be a touristic map showing selected landmarks of a city. Moreover, Illustrative objects could support topographic features, such as parasols on a beach. Illustrated maps are rather artistic than technical, because they are created mainly by painters, architects, designers, geographers, historians, or reporters (Antoniou and Kotmair, 2015). Lastly, the term 'figurative map' is coined by the title of Minard's (1869) map of Napoleon's Russian Campaign. The map contains numbers (= figures) of troops but not any images. In other maps like 'The figurative map of Adriaen Block' (1614), a decorated scale bar and compass rose are present. Van Bleyswijck's (n.d.) 'Kaart Figuratief' depicts pictorial objects like houses and ships as well as images of places of interest in the city of Delft aside the map. As the term ‘figurative’ has different meanings, what is reflected in these titles, we can assume that some but not all figurative maps are pictorial maps.

Concluding, we would define pictorial maps as those with verisimilar and indexed representations, thus are rather individual than typified. Non-pictorial maps would be those with more abstract representations, which are icons/pictograms, geometric shapes, and labels referring to Bodum (2005). We would use the terms illustrated/illustrative maps synonymously to pictorial maps. We would refer to decorated/decorative maps when labels or elements like title, legend, north arrow, scale bar, map frame, etc. are embellished. When referring to symbols, all maps would be figurative as they convey a certain meaning and they are not meant to be interpreted literally. When referring to images, pictorial maps would be figurative. When referring to creatures, only maps with humans/humanoids, animals, or mythological creatures would be figurative.

Data

From the dataset used in the first experiment, we selected 1500 pictorial maps and 1500 non-pictorial maps. Pictorial maps contain realistically drawn objects (e.g. persons, cars, houses) and show space in 2D projection or 3D perspective. Non-pictorial maps are 2D representations which include abstract geometries (e.g. points, lines, polygons), icons, and labels. Maps of both types vary in creation dates, scales, locations, themes, and styles. We split the maps into train and validation sets with a ratio of 60:40. The maps have a width of 586 pixels and height of 553 pixels on average.

As we are mainly interested in finding pictorial objects, we excluded 100 of the maps from the first dataset for this experiment. Those are anthropomorphic maps (e.g. Eytzinger and Hogenberg’s (1583) Leo Belgicus), where pictorial objects cover a large area of the map, and maps showing 3D reliefs without any other pictorial objects (e.g. Berann’s (1989) Yosemite panorama). Maps depicting mountains in a molehill manner (e.g. Coronelli’s (1690) Abyssinia map) were also excluded as we see molehills as a mixture of an iconic and parametric representation.

Procedure

Similar to the first experiment, we retrained models for Xception and InceptionResNetV2 to categorise a map as either pictorial or non-pictorial. The CNNs use the same hyperparameters (i.e. weights, batch size, learning rate, loss function, optimiser) as in the first experiment. For feeding the images into the networks, two input options were compared:

- Resized: Map images are resized to 299x299px without maintaining the aspect ratio.
- Manual gridded crop: Map images are partitioned along a regular grid into cells of 299x299px, whereat cells may overlap and image sides smaller than 299px are

up-scaled to this size. Next, cells were manually identified where pictorial objects are present. Only those cells are taken into account as training data for pictorial maps, whereas all grids cells are available as candidates for non-pictorial maps. In every epoch, one cell is selected randomly for each of the pictorial and non-pictorial maps.

Again, the Lanczos filter is used for resizing the images. Note that a middle and a complete random crop are not possible for this experiment since this may lead to regions which do not contain any pictorial objects.

Results

Again, we evaluated three Xception and InceptionResNetV2 models, which have reached the highest validation accuracy during training, and averaged their validation results while altering image options (Table 2). Overall, the number of correct categorisations between pictorial and non-pictorial maps is with 88-92% at a high level, though it is lower than in the first experiment. In all but one evaluation option (i.e. random crop), Xception is more accurate than InceptionResNetV2. Retraining the classification models with manually identified grid cells containing pictorial objects and applying the retrained models to images with randomly cropped cells led to a decreased accuracy compared to the resizing option. Considering a map as a pictorial one when at least one of the grid cell contains pictorial objects, then the accuracy is similar to the resizing option. When averaging the classification results of image grid cells, the accuracy improved about 2% in relation to those two options.

According to the ROC curves (Figure 4), the evaluation option to declare maps as pictorial when at least one of the image grid cells is predicted as pictorial seems to be more suited for higher classification thresholds. Higher thresholds reduce the number of positive outcomes, thus eventually lead to more true negatives but also more false positives. In contrast to a threshold of 0.5, the one pictorial cell within grid option has a similar performance to averaging the scores over the grid considering the auc scores of the two CNNs. The options to resize or to crop a random part of an image perform similarly to the previous metric. Overall, the auc scores of Xception are higher than those of InceptionResNetV2 for all matching image evaluation options.

As the CNNs categorised pictorial and non-pictorial maps mostly successfully, we selected only some failure cases, which occurred in all twelve runs. For evaluation, we selected the same options – resizing and averaging over grid - as in the first experiment. Examples for frequently misclassified pictorial maps (Figure 5) are a subway illustration on a Tokyo metro map, photos on a Beijing city map, and pictorial objects (e.g. lighthouse, rainbow, horse) on an Iceland map. In all three maps, pictorial objects are relatively small. Regarding non-pictorial maps (Figure 6), a fantasy indoor map, a Rome city map, and a papercraft world map were often wrongly classified. While the first example may be a border case of our definition, it is not clear which activation may have triggered the misclassification of the second and third example.

	Xception	InceptionResNetV2
Resized	89.64%	88.61%
Manual gridded crop		
- random crop	87.69%	88.00%
- one pictorial cell within grid	89.14%	88.67%
- average over grid	**91.89%**	**90.83%**

Table 2: Correct classifications of pictorial maps and non-pictorial maps for the examined CNNs and image input options (as explained in Procedure). The values are averages of validation accuracies of three retrained models having achieved the highest accuracy during training.

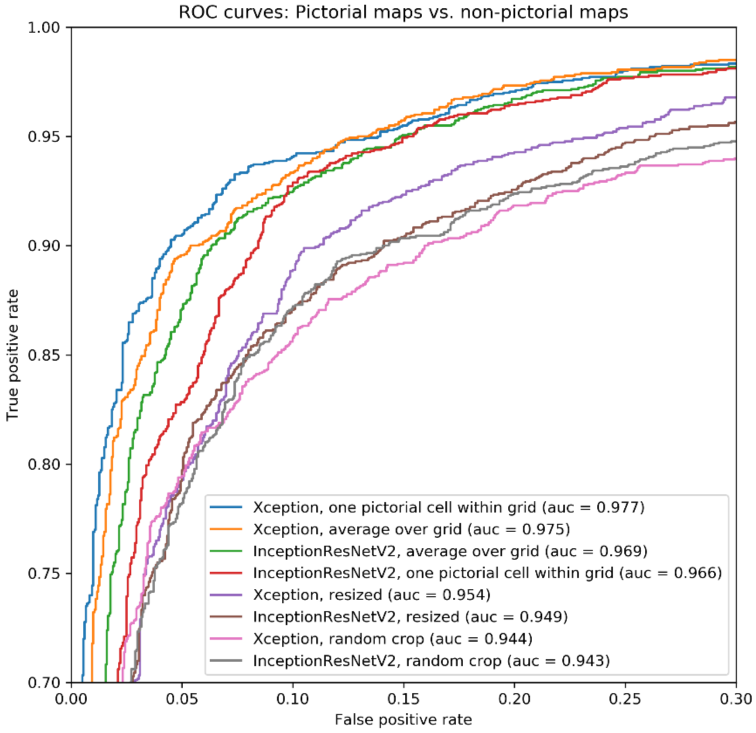

Figure 4: ROC curves (enlarged) and auc scores for the tested CNNs and image evaluation options to classify pictorial maps and non-pictorial maps

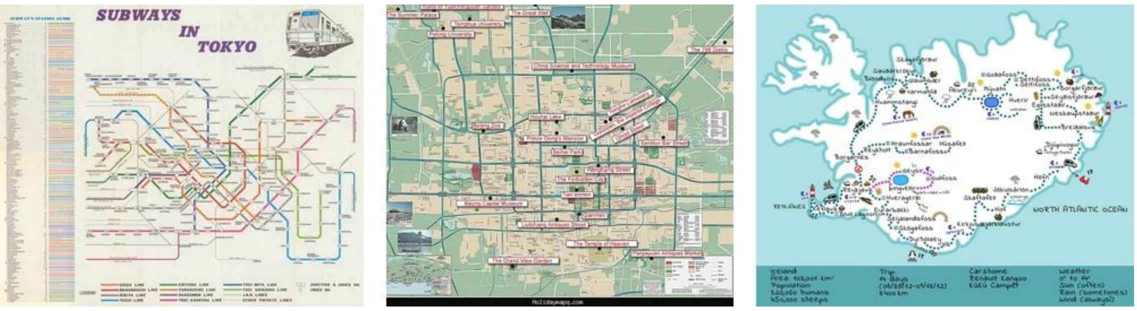

Figure 5: Selection of three frequently misclassified pictorial maps by both CNN models for resized and average over grid image evaluation options (image sources: Pinterest - https://i.pinimg.com/736x/b6/d2/d1/b6d2d1375ac9808cc2998c814862e5d8.jpg, https://i.pinimg.com/736x/51/d7/8b/51d78ba2f5a0f8217a20dbb7fe8ca883--nice-map-beijing.jpg, https://i.pinimg.com/736x/50/a8/5a/50a85a6ea7ee74f36b5141138d47a281.jpg)

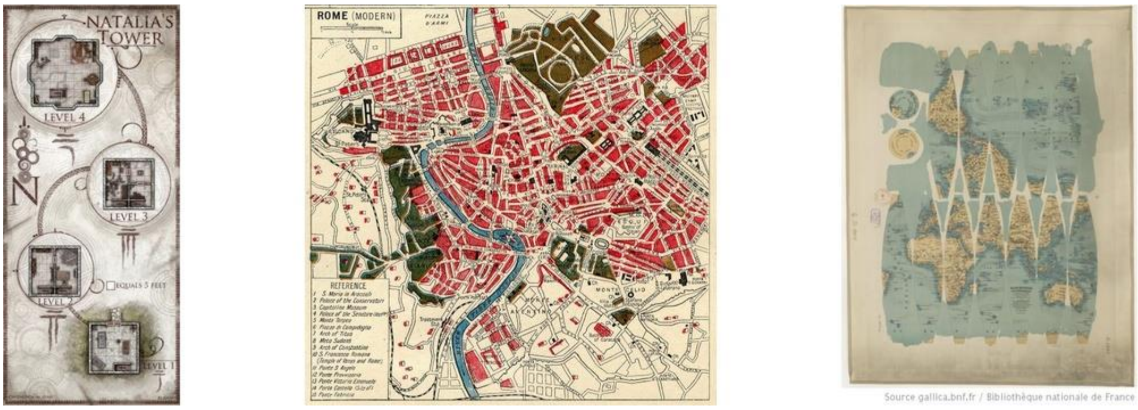

Figure 6: Selection of three frequently misclassified non-pictorial maps by both CNN models for resized and average over grid image evaluation options (image sources: Pinterest - https://i.pinimg.com/736x/5c/0a/a1/5c0aa1e7bfd262b50d2b6e43c237b833.jpg, https://i.pinimg.com/736x/6f/34/73/6f34737d2cc282e23c5668217ddf3544--printable-maps-vintage-printable.jpg, https://i.pinimg.com/736x/5f/f7/a2/5ff7a22d4cdf35580c12a654ac208ca5--map-mind-illustrated-maps.jpg)

Detection of sailing ships on maps

Definitions

Ships seem to be one of the most frequently appearing pictorial objects on maps from the late Middle Ages and Renaissance. At those times, ships were mainly used for exploration (e.g. Columbus discovering America), fishing (e.g. with harpoons and nets), trade (e.g. Italian merchants with the Far East), and battles (e.g. in the Anglo-Spanish War 1585–1604). Maps, such as portolan charts, were essential for navigation in all of these nautical endeavours. Map-makers seem to have placed ships on these maps as a symbol for the above uses, to attract attention, and to simply cover empty areas (Reinhartz, 2012). One of the earliest map which features ships was Cresques's (1375) Catalan Atlas. A junk and a boat with pearl fishers in the Indian Ocean, a galley near the Canaries, and another ship in the Caspian Sea are depicted on this world map (Unger, 2010). Ships mostly depict common types of the era and rarely represent specific instances – such as Magellan's Victoria on Ortelius's (1589) map 'Maris Pacifici'. As copyright laws were introduced not before the 18th century, cartographers often reused images of ships (Reinhartz, 2012), for example from Bruegel the Elder's (1565) template showing sixteen different ship types.

The word ship, which is a part of our everyday language, is defined as a "large sea-going vessel" and as "a vessel having a bowsprit and three masts" (OED 2019). For our purposes, we like to define ships as large sea-going vessels with at least one mast but not necessarily a bowsprit. Sails on the masts may be hoisted or lowered, and paddles and flags may be present. Barques, brigs, carracks, clippers, galleys, galleons, or junks would be exemplary ship types which we like to detect with CNNs. We like to differentiate ships from boats which are "small, typically open vessel[s] for travelling over water" (OED 2019). Commonly, it is distinguished that a ship can carry boats but a boat cannot carry ships. We also like to exclude submarines, which are able to travel underwater, and modern ships. To the latter would count for instance 19th century steam ships (e.g. paddle steamers) as well as 20th century passenger ships (e.g. cruise ships), cargo ships (e.g. container ships), fishing ships (e.g. trawlers), utility ships (e.g. icebreakers), and warships (e.g. aircraft carriers).

Data

We obtained 525 maps and illustrations with 3200 ships from 11 digital map libraries. Most of them were listed and described on the website 'Map History / History of Cartography' (Campbell, 2019). A complete list of libraries where we collected the maps is given in the Appendix (Table 7). As only a few libraries offered APIs to search and download maps programmatically, we retrieved the maps mainly by crawling the websites and parsing their HTML content with Python scripts. For smaller collections, we obtained maps manually via the graphical user interface of the websites. If possible, we restricted our search to maps from the 15th to 18th century.

The maps have an average width of 1116px and an average height of 907px. We split the maps with a ratio of 60:40 into train and validation sets for the CNNs. There are 294 maps with 1918 ships in the training set and 231 maps with 1283 ships in the validation set. Maps originating from the same digital map library are either in the training or validation set, but not in both. 41 maps of the validation set do not contain any ships.

Procedure

We compare two popular CNNs, Faster R-CNN and RetinaNet, to detect bounding boxes of ships in historic maps. Faster R-CNN proposes regions of interest in a first stage, and extracts features[1] and predicts bounding box coordinates in a second stage. In RetinaNet, these tasks are performed in one stage using a pyramid of feature maps[2] with multiple scales. In our experiments, we use TensorFlow implementations of Faster R-CNN (Huang et al., 2019) and RetinaNet (Gaiser, 2018). We chose ResNet50 as a sub-CNN for feature extraction since models pre-trained on natural images of ResNet architectures with more layers were not available for RetinaNet. When detecting objects in the COCO (2015) natural images dataset with ResNet50, Faster R-CNN achieves an average precision of 30% (Huang et al., 2019) and RetinaNet of 35% (Gaiser, 2018). In this metric, a detection is marked as correct when the intersection over union of a ground-truth bounding box and a predicted bounding box (i.e. area of overlap / area of union) lies above a certain threshold. The average precision, which is the primary COCO metric used for comparisons, is the arithmetic mean for ten different thresholds ranging from 50% to 95% in steps of 5%. Other COCO metrics consider only a certain threshold (i.e. 50% or 75%) or are applied only to bounding boxes of a certain size (small < 32^2px; large > 96^2px; medium in between). While COCO metrics were available for the Faster R-CNN implementation when training on custom datasets, we had to calculate them separately for RetinaNet with the Python COCO API. For both CNNs, we included bounding boxes with confidence scores > 0 in the calculations.

For our ship dataset, we trained Faster R-CNN with a learning rate of 10^{-4} for 50 epochs and RetinaNet with learning rate of 10^{-5} for 30 epochs. Both networks received images in batches of 1 (i.e. single images) and images were flipped randomly along their horizontal axes as the only augmentation technique. Objects within an image are linked to anchors, which are rectangles with different ratios and scales. The anchor centres are distributed in equal intervals (= strides) over the image. We did not modify the predefined anchors ratios (i.e. 2:1, 1:1, and 1:2), however we tested different anchor scales (see Tables 3-6). The existing Faster R-CNN model was trained on scales of 0.25, 0.5, 1.0, 2.0, whereas the pre-trained RetinaNet model was set to scales of 2^0, $2^{1/3}$, $2^{2/3}$ (≈ 1.0, 1.26, 1.59). As ships usually cover only a small area of the image, we optimised the CNNs accordingly:

- Faster R-CNN configuration for small objects: We set the first stage features stride as well as the height and width stride of anchors to 8 instead of 16. The modification of the first stage features stride increases the size of the output

[1] Features characterise objects in CNNs; they should not be confused with cartographic or geographic features.

[2] Feature maps are outputs of intermediate or final layers in CNNs.

feature map of ResNet50 so that more details of smaller objects will be preserved. The change of height and width stride results in smaller differences between anchors centres, thus leading to a finer virtual grid on the images where the anchors will be attached. In total, the number of trainable parameters stays the same.

- RetinaNet configuration for small objects: We used the first four out of five outputs of intermediate layers of ResNet50 instead of the last four. Anchor strides and anchor sizes are halved and feature pyramid levels 2 to 6 are used instead of 3 to 7. By this, images are less down-sampled so that details of smaller objects can be better preserved. As a positive side effect, the number of trainable parameters is halved.

On a NVIDIA GTX 1080, one epoch of training our ship dataset took both Faster R-CNN and RetinaNet about 1min30s for the normal configuration. For the configuration for small objects, the training time needed for one epoch increased to 2min20s for Faster R-CNN and to 2min10s for RetinaNet.

Results

We calculated the mean of the highest validation average precisions of three different training runs for Faster R-CNN (Table 3) and RetinaNet (Table 4) with different anchor scales. The predefined scales of Faster R-CNN reached the third highest average precision, while the predefined scales of RetinaNet were the second highest. For both CNNs, a reduction of predefined scale values led to the best result for our dataset. Scales with other values resulted in lower average precisions, whereat the predefined RetinaNet values scored astonishingly poor for Faster R-CNN. In general, the average precisions of RetinaNet were 8-10% higher than those of Faster R-CNN.

We observed an overall increase of average precisions for the Faster R-CNN configuration for small objects (Table 5) and diverging results for RetinaNet (Table 6). Besides to the RetinaNet’s preset scales, we note a 7% increase of average precisions for Faster R-CNN and a 1% increase for RetinaNet of the two scale combinations which achieved also the highest scores with the standard configuration. Three scale combinations for RetinaNet resulted in lower precisions while one remained on about the same level. Still, RetinaNet’s top two results are about 4% higher than the two highest average precisions of Faster R-CNN.

Qualitative results show that larger freestanding ships are recognised well (Figure 7). With smaller sizes and more occlusions between the ships, however, the detection accuracy drops (Figure 8).

Scales	AP	AP$_{50}$	AP$_{75}$	AP$_{\text{small}}$	AP$_{\text{medium}}$	AP$_{\text{large}}$
1.0, 1.26, 1.59	5.4%	15.35%	2.01%	0.83%	7.75%	12.95%
0.5, 1.0, 1.5	20.24%	48.15%	12.66%	9.34%	26.3%	34.36%
0.25, 0.5, 1.0, 2.0 *	23.4%	55.05%	14.99%	12.82%	30.17%	34.37%
0.25, 0.5, 1.0	23.6%	54.67%	14.91%	12.76%	30.52%	35.06%
0.125, 0.25, 0.5, 1.0	**24.93%**	**56.86%**	**16.91%**	**15.15%**	**31.46%**	**35.29%**
0.0625, 0.125, 0.25, 0.5, 1.0	23.55%	54.91%	15.85%	12.77%	30.35%	34.71%

Table 3: Average COCO metrics of the best Faster R-CNN models of three runs for different scales (* = preset)

Scales	AP	AP$_{50}$	AP$_{75}$	AP$_{\text{small}}$	AP$_{\text{medium}}$	AP$_{\text{large}}$
1.0, 1.26, 1.59 *	34.82%	58.56%	36.99%	20.72%	**44.84%**	45.25%
0.5, 1.0, 1.5	**35.37%**	**59.74%**	**37.69%**	**22.65%**	44.35%	**44.72%**
0.25, 0.5, 1.0, 2.0	33.10%	58.67%	33.44%	21.64%	41.36%	42.99%
0.25, 0.5, 1.0	32.18%	57.62%	32.94%	21.78%	39.34%	41.52%
0.125, 0.25, 0.5, 1.0	32.62%	59.20%	32.39%	21.82%	40.39%	41.04%
0.0625, 0.125, 0.25, 0.5, 1.0	32.95%	59.03%	33.80%	22.35%	40.47%	40.96%

Table 4: Average COCO metrics of the best RetinaNet models of three runs for different scales (* = preset)

Scales	AP	AP$_{50}$	AP$_{75}$	AP$_{\text{small}}$	AP$_{\text{medium}}$	AP$_{\text{large}}$
1.0, 1.26, 1.59	6.7%	15.66%	4.4%	1.6%	9.18%	16.44%
0.5, 1.0, 1.5	27.61%	53.12%	26.2%	14.75%	36.28%	38.38%
0.25, 0.5, 1.0, 2.0 *	31.48%	61.04%	28.91%	18.75%	40.04%	42.68%
0.25, 0.5, 1.0	30.58%	60.57%	26.58%	17.78%	38.76%	42.54%
0.125, 0.25, 0.5, 1.0	**32.26%**	**62.92%**	28.97%	**19.39%**	**40.66%**	**43.44%**
0.0625, 0.125, 0.25, 0.5, 1.0	31.77%	61.06%	**29.78%**	18.39%	40.64%	42.64%

Table 5: Average COCO metrics of the best Faster R-CNN models for small objects of three runs configuration and different scales (* = preset)

Scales	AP	AP_{50}	AP_{75}	AP_{small}	AP_{medium}	AP_{large}
1.0, 1.26, 1.59 *	35.99%	63.02%	37.04%	25.91%	**43.91%**	41.10%
0.5, 1.0, 1.5	**36.24%**	**63.35%**	**38.57%**	**26.25%**	43.65%	**42.10%**
0.25, 0.5, 1.0, 2.0	28.45%	56.63%	24.45%	19.12%	35.79%	33.48%
0.25, 0.5, 1.0	32.37%	60.19%	31.73%	23.06%	39.87%	35.76%
0.125, 0.25, 0.5, 1.0	30.04%	59.04%	27.65%	20.20%	37.70%	34.11%
0.0625, 0.125, 0.25, 0.5, 1.0	29.68%	58.80%	26.73%	20.70%	36.92%	32.80%

Table 6: Average COCO metrics of the best RetinaNet models for small objects of three runs configuration and different scales (* = preset)

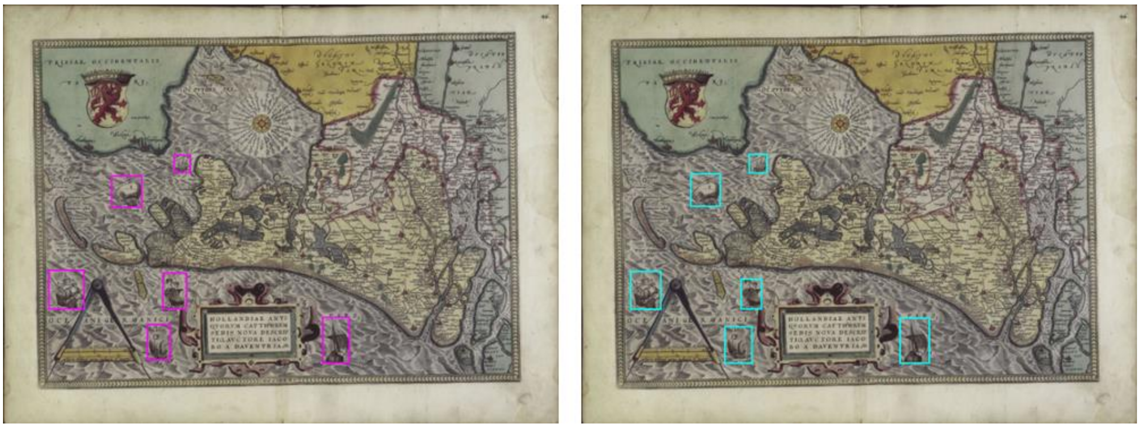

Figure 7: Ground truth (left) and detected bounding boxes (right) with the best trained Faster R-CNN model (AP: 32.8%) for large, freestanding ships (original image source: Sammlung Ryhiner - https://biblio.unibe.ch/web-apps/maps/zoomify.php?col=ryh&pic=Ryh_3106_1)

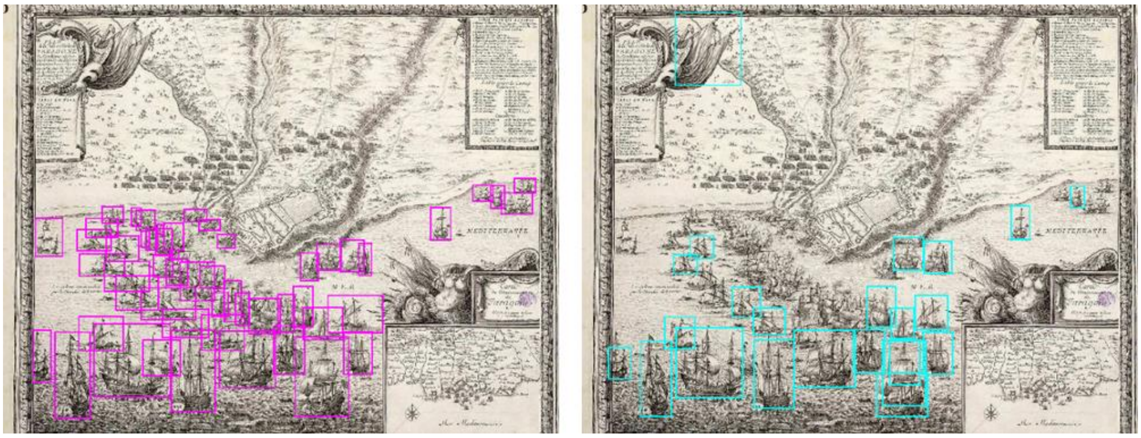

Figure 8: Ground truth (left) and detected bounding boxes (right) with the best trained RetinaNet model (AP: 36.8%) for occluded and small ships (original image source: Biblioteca Digital Hispánica - http://bdh-rd.bne.es/viewer.vm?id=0000022147)

Discussion

For our experiments, we prepared datasets for training and evaluating CNNs to detect pictorial map objects. Maps and images in the first dataset originate from Pinterest. We tried to limit similarly styled maps and similar motifs of images to two to three examples since some user pinned map series and a large amount of similar object types (e.g. cars, stitchery). To increase the difficulty, we added other artificial images than maps, such as sketches or invitation cards. The proportion of maps covering Europe and North America is higher than of other continents; similar seems to be the case for the mapmakers, who are largely unknown in Pinterest. Thus, applying the retrained CNNS to maps of other cultural background and of other data sources would probably lead to decreased accuracies. The geographic coverage of the second dataset with its ancient sailing ships is similar to the first dataset, but here maps originate from different digital map libraries. The amount of training data is not as high as in other training datasets, however the categorisation of maps and pictorial maps as well as the annotation of bounding boxes should be consistent as it has been curated by one person. Used definitions are not meant to be carved in stone; they can still be altered and networks fed with other training data accordingly. Although our map definition mentions interactive or animated maps, only static maps were used in our experiments, but theoretically screenshots from 2D or 3D map applications, web map services, or single frames from videos could be also taken as inputs for the CNNs. Concerning map applications and services, we would suggest that rather machine-readable interfaces should be provided to verify their identify and access their source data, what is already partly specified by OGC and ISO standards, than having to parse the map content with CNNs, what may also cause copyright issues.

Overall, the accuracy of correctly distinguishing maps from non-maps as well as pictorial maps from non-pictorial maps is with more than 91% quite high. The CNNs show a better performance than other machine learning methods like support vector machines or k-nearest neighbour, which achieved a F1 score (i.e. another metric for accuracy) of 74% for a similar map classification task (Goel et al., 2011). With image augmentation techniques - like translation, rotation, and scaling - or model ensembles - where different CNNs models are combined, our accuracy may be increased even more. We did not apply these techniques as we would primarily investigate differences between Xception and InceptionResNetV2. Although newer CNN models exist with better accuracies, we chose those models as they share the same input size of images and they are available in the same library and programming language. Similar applies to comparing Faster R-CNN and RetinaNet as both have the same backbone (i.e. ResNet50) and backend (i.e. Tensorflow) but different APIs. Faster R-CNN is harder to modify because it uses lower level API, that is why only configuration parameters were changed and not the network architecture as done for RetinaNet for small objects. Quantitative results from the object detection task cannot be directly compared to those of the classification tasks as we reported the values in the common COCO format. Qualitative results are convincing when ships are not too crowded, too small, or too blurry. How these difficulties may be improved is addressed in the future work section.

Our classification experiments to identify maps and pictorial maps resulted in a speed-accuracy trade-off: While feeding resized images into the CNNs is faster than splitting the images into grid cells and inputting those for validation, averaging the predicted classification scores of the cells is more accurate. The input size of 299 pixel images for classification tasks is justified by the CNN architectures. Enlarging the input size would increase the number of learnable parameters, what may lead to memory shortages on the graphic board. Current attempts like EfficientNet (Tan and Le, 2019) optimise the number of parameters, but they target rather at improving the performance than feeding in high resolution images. We would argue that it is not necessary to input maps at high resolutions because humans can also recognise maps at some distance without inspecting all details. Only if details were important, higher resolutions would be advantageous, for instance to identify small pictorial objects. This may explain why the accuracy of cropping images was higher than the resizing strategy in our second experiment, whereas input strategies had a similar accuracy in the first experiment.

Only the resizing option was available in the object detection libraries, however other hyperparameters like anchor scales and strides could be tuned. It cannot be excluded that other anchor scale values yield to better results since ours were manually determined, partly with help of a debugging tool for RetinaNet. Reducing the anchor strides increased the detection accuracy of Faster R-CNN, however, it was not possible to reduce the strides further due to constraints in the network configuration. The adaptation to detect smaller objects with RetinaNet also increased the accuracy and reduced the number of trainable parameters. Reducing the number of ResNet50 levels to three would have caused a reduction of levels in the feature pyramid network which we assume is not desirable since the factor between the smallest resized (2x2px) and largest resized ship (298x345px) is about 2^7 to 2^8. The size of smallest resized ship also demonstrates that it is only barely detectable for the CNNs. It is not clear at this point if both configurations for small objects can be combined in one or both CNNs to increase the accuracy even more.

Summary and future work

In this paper, we examined identifying pictorial objects in historic and contemporary maps with CNNs. We reached an accuracy of about 97% to classify maps and non-maps with Xception and InceptionResNetV2. With about 92%, the accuracy was lower to distinguish between pictorial and non-pictorial maps. For the first task, the accuracy of Xception and InceptionResNetV2 was about the same, for the second task Xception was slightly more accurate than InceptionResNetV2. From the examined input options, calculating the average over regular image grid cells achieved the highest accuracy, however this method is more computationally intensive than resizing the images. An average precision of about 32% could be obtained with Faster R-CNN and of about 36% with RetinaNet, both having ResNet50 as a backbone, to recognise sailing ships in maps. Configuring the networks to detect small objects increased the accuracy for both CNNs, in case of Faster R-CNN more than of RetinaNet. Reducing the anchor scale values from the original setup let also to higher accuracies. With our modifications, the average precision is slightly higher than the baselines for detecting objects in natural images with these networks.

Future work may extend our datasets with additional training data, for example by harvesting images from other websites or by synthetically creating special cases with maps in real-world images or zoomed-in maps. Also other map types than pictorial maps could be classified, for instance based on the visualisation type (e.g. chart maps), the dimensionality (2D/3D), or the level of representation (abstract-realistic). Map types requiring a semantic understanding of the content, like usage (e.g. hiking) or theme (e.g. weather), would go beyond the visual recognition capabilities of CNNs though. For object detection, datasets with other types of pictorial objects could be prepared, such as persons, animals, or sea-monsters. Eventually these objects could be detected class-agnostically with weakly supervised methods (Gonthier et al., 2018). Another ability of CNNs is to detect visually salient objects (Borji et al., 2015), what could help cartographers to quantify A-level pictorial objects according to Roman's (2015) ABC-rule.

Our detection accuracy may be improved by special CNN architectures for small objects (Eggert et al., 2017) as well as for crowded and occluded objects (Wang et al. 2018). Also new trends like dilated (Hamaguchi et al., 2018) and deformable convolutions (Dai et al., 2017) may further increase the accuracy. Even hyperparameters may be optimised and CNNs architectures may be created automatically (Zoph and Le, 2016). Novel CNN architectures like Mask R-CNN (He et al., 2017) and DeepLab (Chen et al., 2016) would be able to extract not only bounding boxes but also silhouettes of objects. Similarity metrics (Krizhevsky et al., 2012) could enable finding map series with a certain style of an author (e.g. artist, map agency) or maps produced with the same software. Calculating similarity metrics for single map objects (e.g. ships) would facilitate detecting duplicates, which could reveal hidden relationships between ancient cartographers. In combination with a metric on map readability (i.e. how accurately can map features be extracted), map producers could develop a style which is well-readable, yet distinguishable from others.

The overarching goals for cartographic research on CNNs would be identify maps in a first step and to vectorise, georeference and semantically attribute them, and extract metadata in a second step. Similarity metrics to other maps or map objects could be derived from the CNN outputs in a third step. This would allow creating a global search engine, which indexes maps in the Internet. Next to a simple text-based search, more sophisticated search filters could be provided, for example for map features (e.g. rivers, place names) or metadata (e.g. coordinate reference system, map style). Tools like an inverse map search or recommendations of similar maps are also thinkable due to the similarity metrics. With our three experiments, we contributed to finding maps in the Internet as well as extracting data (i.e. certain map objects) and metadata (i.e. a certain map type).

Supplemental online material

Datasets, code, and models are published on: http://narrat3d.ethz.ch.

Acknowledgements

The authors express their gratitude to Magnus Heitzler, Diego Gonzalez, and Benjamin Kellenberger for their hints on CNNs. Thanks go also to our colleagues from the Institute of Cartography and Geoinformation, ETH Zurich for helping to annotate the training data.

Funding

This work was supported by ETH Research Grant ETH-11 17-1.

Disclosure statement

The authors declare no conflict of interest.

Appendix

Library	Website	Maps used
Beinecke Rare Book & Manuscript Library	https://brbl-dl.library.yale.edu/	159
Biblioteca Digital Hispánica	http://www.bne.es/	35
Bibliotheque Nationale de France	https://gallica.bnf.fr/	63
Bodleian Library	https://digital.bodleian.ox.ac.uk/	16
Norman B. Leventhal Map & Education Center	https://collections.leventhalmap.org/	17
David Rumsey Map Collection	https://www.davidrumsey.com/	33
John Carter Brown Library	https://jcb.lunaimaging.com/	33
Library of Congress	https://www.loc.gov/	6
New York Public Library	https://www.nypl.org/	7
Royal Museum Greenwich	https://pro.europeana.eu/	7
Sammlung Ryhiner	https://www.unibe.ch/universitaet/dienstleistungen/universitaetsbibliothek/recherche/sondersammlungen/kartensammlungen/index_ger.html	149

Table 7: Digital libraries from which historic maps with sailing ships were retrieved for training Faster R-CNN and RetinaNet

References

Agarwal, A. (2019) *Top 40 Cartography Blogs & Websites for Cartographers To Follow in 2019* [Online]. Available at http://blog.feedspot.com/cartography_blogs/ (Accessed 6 May 2019).

Alloa, E. (2016) 'Iconic Turn: A Plea for Three Turns of the Screw', *Culture, Theory and Critique*, vol. 57, no. 2, pp. 228–250 [Online]. DOI: 10.1080/14735784.2015.1068127.

Antoniou, A. and Kotmair, A. A. (2015) *Mind the map: Illustrated Maps and Cartography*, Berlin, Gestalten.

Audebert, N., Boulch, A., Lagrange, A., Saux, B. L. and Lefèvre, S. (2016) 'Deep Learning for Remote Sensing', *Communications of the 16th ONERA-DLR Aerospace Symposium*, Oberpfaffenhofen.

Bandrova, T. (2003) 'Atlas Rodinoznanie', *International Research in Geographical and Environmental Education*, vol. 12, no. 4, pp. 354–358.

Barron, R. (1990) *Decorative Maps*, Poster Art Series, London, Crescent Books.

Baumgärtner, I., Debby, N. B.-A. and Kogman-Appel, K. (2019) *Maps and Travel in the Middle Ages and the Early Modern Period: Knowledge, Imagination, and Visual Culture*, Das Mittelalter. Beihefte, Berlin, De Gruyter.

Bengio, Y. and LeCun, Y. (2007) 'Scaling learning algorithms towards AI', *Large-scale kernel machines*, vol. 34, no. 5, pp. 1–41.

Berann, H. (1989) 'Panoramic drawing of the Yosemite National Park', [Online]. Available at https://commons.wikimedia.org/wiki/File:Heinrich_Berann_NPS_Yosemite.jpg (Accessed 6 May 2019).

Block, A. (1614) 'The figurative map of Adriaen Block', [Online]. Available at http://digitalcollections.nypl.org/items/510d47d9-7bf7-a3d9-e040-e00a18064a99 (Accessed 6 May 2019).

Bodum, L. (2005) 'Modelling Virtual Environments for Geovisualization: A Focus on Representation', in Dykes, J., MacEachren, A. M., and Kraak, M.-J. (eds), *Exploring Geovisualization*, International Cartographic Association, Oxford, Elsevier, pp. 389–402 [Online]. (Accessed 21 February 2017).

Borji, A., Cheng, M.-M., Jiang, H. and Li, J. (2015) 'Salient object detection: A benchmark', *IEEE transactions on image processing*, vol. 24, no. 12, pp. 5706–5722.

Brueghel the Elder, P. (1565) 'Sixteen Boats of Different Structure', [Online]. Available at https://commons.wikimedia.org/wiki/File:Pieter_Brueghel_I_-_Sixteen_Boats_of_Different_Structure,_c._1565_RP-P-1997-159.jpg (Accessed 6 May 2019).

Campbell, T. (2019) *Map History / History of Cartography* [Online]. Available at https://www.maphistory.info/ (Accessed 22 February 2019).

Caquard, S. and Cartwright, W. (2014) 'Narrative Cartography: From Mapping Stories to the Narrative of Maps and Mapping', *The Cartographic Journal*, vol. 51, no. 2, pp. 101–106.

Cartwright, W. (2014) 'Rethinking the definition of the word "map": an evaluation of Beck's representation of the London Underground through a qualitative expert survey', *International Journal of Digital Earth*, vol. 8, no. 7, pp. 522–537 [Online]. DOI: 10.1080/17538947.2014.923942.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K. and Yuille, A. L. (2016) 'DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs', *arXiv:1606.00915 [cs]* [Online]. Available at http://arxiv.org/abs/1606.00915 (Accessed 22 February 2019).

Child, H. (1956) *Decorative Maps*, The 'HOW TO DO IT' Series, First edition., London & New York, Studio Publications.

Chollet, F. (2016) 'Xception: Deep Learning with Depthwise Separable Convolutions', *arXiv:1610.02357 [cs]* [Online]. Available at http://arxiv.org/abs/1610.02357 (Accessed 6 December 2018).

Clarke, V. (2015) *Map: Exploring the World*, London, Phaidon Press Limited [Online]. Available at https://www.goodreads.com/book/show/25208271-map (Accessed 7 December 2018).

COCO Consortium (2015) *Common Objects in Context (COCO)* [Online]. Available at http://cocodataset.org (Accessed 6 May 2019).

Coronelli, V. M. (1690) 'Map of Ethiopia, Abyssinia, and the Source of the Blue Nile', [Online]. Available at https://commons.wikimedia.org/wiki/File:1690_Coronelli_Map_of_Ethiopia,_Abyssinia,_and_the_Source_of_the_Blue_Nile_-_Geographicus_-_Abissinia-coronelli-1690.jpg (Accessed 6 May 2019).

Cresques, A. (1375) 'Catalan Atlas', [Online]. Available at https://commons.wikimedia.org/wiki/File:1375_Atlas_Catalan_Abraham_Cresques.jpg (Accessed 6 May 2019).

Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H. and Wei, Y. (2017) 'Deformable Convolutional Networks', *arXiv:1703.06211 [cs]* [Online]. Available at http://arxiv.org/abs/1703.06211 (Accessed 21 February 2019).

Dodge, S., Xu, J. and Stenger, B. (2017) 'Parsing floor plan images', *2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA)*, pp. 358–361 [Online]. DOI: 10.23919/MVA.2017.7986875.

Duan, W., Chiang, Y.-Y., Knoblock, C. A., Uhl, J. H. and Leyk, S. (2018) 'Automatic Generation of Precisely Delineated Geographic Features from Georeferenced Historical Maps Using Deep Learning.', *UCGIS/AutoCarto*.

Duzer, C. V. (2014) *Sea Monsters on Medieval and Renaissance Maps*, Reprint edition., London, British Library.

Eggert, C., Zecha, D., Brehm, S. and Lienhart, R. (2017) 'Improving Small Object Proposals for Company Logo Detection', *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval*, ICMR '17, New York, NY, USA, ACM, pp. 167–174 [Online]. DOI: 10.1145/3078971.3078990 (Accessed 21 February 2019).

Eytzinger, M. and Hogenberg, F. (1583) 'Leo Belgicus', [Online]. Available at https://commons.wikimedia.org/wiki/File:1583_Leo_Belgicus_Hogenberg.jpg (Accessed 6 May 2019).

Feng, Y., Thiemann, F. and Sester, M. (2019) 'Learning Cartographic Building Generalization with Deep Convolutional Neural Networks', *ISPRS International Journal of Geo-Information*, vol. 8, no. 6, p. 258 [Online]. DOI: 10.3390/ijgi8060258.

Fuchs, R., Verburg, P. H., Clevers, J. G. P. W. and Herold, M. (2015) 'The potential of old maps and encyclopaedias for reconstructing historic European land cover/use change', *Applied Geography*, vol. 59, pp. 43–55 [Online]. DOI: https://doi.org/10.1016/j.apgeog.2015.02.013.

Gaiser, H. (2018) *Keras RetinaNet*, Python [Online]. Available at https://github.com/fizyr/keras-retinanet (Accessed 6 May 2019).

Girshick, R. (2015) 'Fast R-CNN', *arXiv:1504.08083 [cs]* [Online]. Available at http://arxiv.org/abs/1504.08083 (Accessed 7 December 2018).

Girshick, R., Donahue, J., Darrell, T. and Malik, J. (2013) 'Rich feature hierarchies for accurate object detection and semantic segmentation', *arXiv:1311.2524 [cs]* [Online]. Available at http://arxiv.org/abs/1311.2524 (Accessed 7 December 2018).

Goel, A., Michelson, M. and Knoblock, C. A. (2011) 'Harvesting maps on the web', *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 14, no. 4, pp. 349–372 [Online]. DOI: 10.1007/s10032-010-0136-2.

Gonthier, N., Gousseau, Y., Ladjal, S. and Bonfait, O. (2018) 'Weakly Supervised Object Detection in Artworks', *arXiv:1810.02569 [cs]* [Online]. Available at http://arxiv.org/abs/1810.02569 (Accessed 18 December 2018).

Goodfellow, I., Bengio, Y. and Courville, A. (2016) 'Autoencoders', in *Deep Learning*, MIT Press.

Goodman, M. B. and Neumaus, E. (1930) 'A map of Berkeley, Oakland & Alameda', San Francisco & Oakland, Sather Gate Book Shop [Online]. Available at https://www.davidrumsey.com/luna/servlet/detail/RUMSEY~8~1~268595~90042837:A-map-of-Berkeley,-Oakland-&-Alamed (Accessed 6 May 2019).

Google Developers (2019) *Custom Search JSON API*, Google [Online]. Available at https://developers.google.com/custom-search/v1/overview (Accessed 6 May 2019).

Graça, A. J. S. and Fiori, S. R. (2015) ‘Proposal for a Tourist Web Map of the South Area of Rio: Cartographic Communication and the Act of Representing the Landscape in Different Scales and Levels of Abstraction’, *Revista Brasileira de Cartografia*, vol. 67, no. 5, pp. 1079–1090.

Hamaguchi, R., Fujita, A., Nemoto, K., Imaizumi, T. and Hikosaka, S. (2018) ‘Effective Use of Dilated Convolutions for Segmenting Small Object Instances in Remote Sensing Imagery’, *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Lake Tahoe, NV, IEEE, pp. 1442–1450 [Online]. DOI: 10.1109/WACV.2018.00162 (Accessed 21 February 2019).

He, K., Gkioxari, G., Dollár, P. and Girshick, R. (2017) ‘Mask R-CNN’, *arXiv:1703.06870 [cs]* [Online]. (Accessed 6 November 2018).

He, K., Zhang, X., Ren, S. and Sun, J. (2015) ‘Deep Residual Learning for Image Recognition’, *arXiv:1512.03385 [cs]* [Online]. Available at http://arxiv.org/abs/1512.03385 (Accessed 6 December 2018).

Holmes, N. (1991) *Pictorial maps: [history, design, ideas, sources]*, New York, NY, Watson-Guptill.

Hornsby, S. J. (2017) *Picturing America: The Golden Age of Pictorial Maps*, 1 edition., Chicago ; London, University of Chicago Press.

Huang, J., Rathod, V., Votel, R., Chow, D., Sun, C., Zhu, M., Fathi, A. and Lu, Z. (2019) *Tensorflow Object Detection API*, Python [Online]. Available at https://github.com/tensorflow/models (Accessed 6 May 2019).

Instagram (n.d.) *#pictorialmaps* [Online]. Available at https://www.instagram.com/explore/tags/pictorialmaps/ (Accessed 6 May 2019).

Jordan, P., Bergmann, H., Cheetham, C. and Hausner, I. (2009) *Geographical Names as a Part of the Cultural Heritage*, Wiener Schriften zur Geographie und Kartographie, Institut für Geographie und Regionalforschung der Universität Wien, vol. 18.

Kang, Y., Gao, S. and Roth, R. E. (2019) ‘Transferring multiscale map styles using generative adversarial networks’, *International Journal of Cartography*, vol. 5, no. 2–3, pp. 115–141 [Online]. DOI: 10.1080/23729333.2019.1615729.

Kent, A. J. (2012) ‘From a Dry Statement of Facts to a Thing of Beauty: Understanding Aesthetics in the Mapping and Counter-Mapping of Place’, *Cartographic Perspectives*, no. 73, pp. 39–60.

Kraak, M.-J. and Fabrikant, S. I. (2017) ‘Of maps, cartography and the geography of the International Cartographic Association’, *International Journal of Cartography*, vol. 3, no. sup1, pp. 9–31 [Online]. DOI: 10.1080/23729333.2017.1288535.

Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012) ‘ImageNet classification with deep convolutional neural networks’, *Communications of the ACM*, vol. 60, no. 6, pp. 84–90 [Online]. DOI: 10.1145/3065386.

Lamb, A. and Johnson, L. (2014) 'Middle Earth to Panem: Maps of Imaginary Places as Invitations to Reading', *Teacher Librarian*, vol. 42, no. 1, pp. 60–63.

Lin, T.-Y., Goyal, P., Girshick, R., He, K. and Dollár, P. (2017) 'Focal Loss for Dense Object Detection', *arXiv:1708.02002 [cs]* [Online]. Available at http://arxiv.org/abs/1708.02002 (Accessed 7 December 2018).

Liu, C., Wu, J., Kohli, P. and Furukawa, Y. (2017) 'Raster-to-Vector: Revisiting Floorplan Transformation', *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2214–2222 [Online]. DOI: 10.1109/ICCV.2017.241.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y. and Berg, A. C. (2016) 'SSD: Single Shot MultiBox Detector', *arXiv:1512.02325 [cs]*, vol. 9905, pp. 21–37 [Online]. DOI: 10.1007/978-3-319-46448-0_2.

Mason, B. (2016) *These Colorful Propaganda Maps Fueled 20th-Century Wars* [Online]. Available at https://news.nationalgeographic.com/2016/10/propaganda-war-maps-gallery/ (Accessed 6 May 2019).

Merwin, D., Cromley, R. and Civco, D. (2009) 'A Neural Network-based Method for Solving "Nested Hierarchy" Areal Interpolation Problems', *Cartography and Geographic Information Science*, vol. 36, no. 4, pp. 347–365.

Minard, C. (1869) 'Carte Figurative des pertes successives en hommes de l'armée française dans la campagne de Russie 1812-1813', [Online]. Available at https://commons.wikimedia.org/wiki/File:Minard.png (Accessed 6 May 2019).

Nguyen, H. T. H., Wistuba, M. and Schmidt-Thieme, L. (2017) 'Personalized Tag Recommendation for Images Using Deep Transfer Learning', Ceci, M., Hollmén, J., Todorovski, L., Vens, C., and Džeroski, S. (eds), *Machine Learning and Knowledge Discovery in Databases*, Cham, Springer International Publishing, pp. 705–720.

Olszewski, R., Gnat, M. and Fiedukowicz, A. (2018) 'Artificial neural networks and fuzzy inference systems for line simplification with extended WEA metric', *Geodesy and Cartography*, vol. 67, no. 2, pp. 255–269 [Online]. DOI: 10.24425/118708.

Ortelius, A. (1585) 'Islandia', [Online]. Available at https://commons.wikimedia.org/wiki/File:Islandia_(Abraham_Ortelius).jpg (Accessed 6 May 2019).

Ortelius, A. (1589) 'Maris Pacifici', [Online]. Available at https://commons.wikimedia.org/wiki/File:Ortelius_-_Maris_Pacifici_1589.jpg (Accessed 6 May 2019).

O'Shea, K. and Nash, R. (2015) 'An Introduction to Convolutional Neural Networks', *arXiv:1511.08458 [cs]* [Online]. Available at http://arxiv.org/abs/1511.08458 (Accessed 22 February 2019).

Pinterest (n.d.) *Pinterest* [Online]. Available at https://www.pinterest.com/.

Redmon, J. and Farhadi, A. (2018) 'YOLOv3: An Incremental Improvement', p. 6.

Reinhartz, D. (2012) *The art of the map: an illustrated history of map elements and embellishments*, New York, Sterling.

Ren, S., He, K., Girshick, R. and Sun, J. (2015) ‘Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks’, *arXiv:1506.01497 [cs]* [Online]. Available at http://arxiv.org/abs/1506.01497 (Accessed 22 February 2019).

Roman, J. (2015) *The Art of Illustrated Maps: A Complete Guide to Creative Mapmaking’s History, Process and Inspiration*, First edition., Cincinnati, OH, HOW Books.

Ronneberger, O., Fischer, P. and Brox, T. (2015) ‘U-Net: Convolutional Networks for Biomedical Image Segmentation’, Navab, N., Hornegger, J., Wells, W. M., and Frangi, A. F. (eds), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Cham, Springer International Publishing, pp. 234–241.

Sarjakoski, L. T., Sarjakoski, T., Koskinen, I. and Ylirisku, S. (2009) ‘The Role of Augmented Elements to Support Aesthetic and Entertaining Aspects of Interactive Maps on the Web and Mobile Phones’, in *Cartography and Art*, Lecture Notes in Geoinformation and Cartography, Springer Berlin Heidelberg, pp. 107–122 [Online]. (Accessed 7 February 2017).

Scherer, D., Schulz, H. and Behnke, S. (2010) ‘Accelerating Large-Scale Convolutional Neural Networks with Parallel Graphics Multiprocessors’, *Artificial Neural Networks – ICANN 2010 Proceedings*, Springer, Berlin, Heidelberg, pp. 82–91 [Online]. (Accessed 17 February 2017).

Sen, A., Gokgoz, T. and Sester, M. (2014) ‘Model generalization of two different drainage patterns by self-organizing maps’, *Cartography and Geographic Information Science*, vol. 41, no. 2, pp. 151–165.

Siam, M., Elkerdawy, S., Jagersand, M. and Yogamani, S. (2017) ‘Deep semantic segmentation for automated driving: Taxonomy, roadmap and challenges’, *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1–8 [Online]. DOI: 10.1109/ITSC.2017.8317714.

Simonyan, K. and Zisserman, A. (2014) ‘Very Deep Convolutional Networks for Large-Scale Image Recognition’, *arXiv:1409.1556 [cs]* [Online]. Available at http://arxiv.org/abs/1409.1556 (Accessed 6 December 2018).

Skelton, R. A. (1966) *Decorative Printed Maps of the 15th to 18th Centuries*, Second impression., London, Spring Books.

Stanford Vision Lab (2016) *ImageNet* [Online]. Available at http://www.image-net.org/ (Accessed 6 May 2019).

Szegedy, C., Ioffe, S., Vanhoucke, V. and Alemi, A. (2016) ‘Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning’, *arXiv:1602.07261* *[cs]* [Online]. Available at http://arxiv.org/abs/1602.07261 (Accessed 6 December 2018).

Szegedy, C., Wei Liu, Yangqing Jia, Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A. (2015) 'Going deeper with convolutions', *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, IEEE, pp. 1–9 [Online]. DOI: 10.1109/CVPR.2015.7298594 (Accessed 6 December 2018).

Tan, M. and Le, Q. V. (2019) 'EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks', *arXiv preprint arXiv:1905.11946*.

TensorFlow Developers (n.d.) *Keras*, TensorFlow Core, Google [Online]. Available at https://www.tensorflow.org/guide/keras (Accessed 6 May 2019).

Turconi, C. (1997) 'The map of Bedolina, Valcamonica Rock Art', *TRACCE no. 9*, DARFO BOARIO TERME [Online]. Available at http://www.rupestre.net/tracce/?p=2422 (Accessed 22 February 2019).

Uijlings, J. R. R., van de Sande, K. E. A., Gevers, T. and Smeulders, A. W. M. (2013) 'Selective Search for Object Recognition', *International Journal of Computer Vision*, vol. 104, no. 2, pp. 154–171.

Unger, R. W. (2010) *Ships on Maps: Pictures of Power in Renaissance Europe*, Early modern history, Basingstoke, Palgrave Macmillan.

Van Bleyswyck, D. (n.d.) 'Kaart Figuratief', [Online]. Available at http://www.essentialvermeer.com/maps/delft/kaart.html#.XNAEA1VMRaR (Accessed 6 May 2019).

Virrantaus, K., Fairbairn, D. and Kraak, M.-J. (2009) 'ICA Research Agenda on Cartography and GIScience', *Cartography and Geographic Information Science*, vol. 36, no. 2, pp. 209–222 [Online]. DOI: 10.1559/152304009788188772.

Wallis, H. and Robinson, A. H. (eds.) (1987) *Cartographical Innovations: An International Handbook of Mapping Terms to 1900*, New edition edition., Tring, Herts, Map Collector Publications Ltd.

Wang, X., Xiao, T., Jiang, Y., Shao, S., Sun, J. and Shen, C. (2018) 'Repulsion Loss: Detecting Pedestrians in a Crowd', *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, IEEE, pp. 7774–7783 [Online]. DOI: 10.1109/CVPR.2018.00811 (Accessed 21 February 2019).

Wang, Y., Lv, H., Chen, X. and Du, Q. (2015) 'A PSO-Neural Network-Based Feature Matching Approach in Data Integration', in Sluter, C. R., Cruz, C. B. M., and Menezes, P. M. L. de (eds), *Cartography - Maps Connecting the World*, Lecture Notes in Geoinformation and Cartography, Springer International Publishing, pp. 189–219 [Online]. (Accessed 7 February 2017).

Yanagisawa, H., Yamashita, T. and Watanabe, H. (2018) 'A study on object detection method from manga images using CNN', *2018 International Workshop on*

Advanced Image Technology (IWAIT), pp. 1–4 [Online]. DOI: 10.1109/IWAIT.2018.8369633.

Ziran, Z. and Marinai, S. (2018) ‘Object Detection in Floor Plan Images’, Pancioni, L., Schwenker, F., and Trentin, E. (eds), *Artificial Neural Networks in Pattern Recognition*, Lecture Notes in Computer Science, Springer International Publishing, pp. 383–394.

Zoph, B. and Le, Q. V. (2016) ‘Neural Architecture Search with Reinforcement Learning’, *arXiv:1611.01578 [cs]* [Online]. Available at http://arxiv.org/abs/1611.01578 (Accessed 6 December 2018).