International Journal of Geo-Information

MDPI

Article

Building Block Extraction from Historical Maps Using Deep Object Attention Networks

Yao Zhao [1,2], Guangxia Wang [1,*], Jian Yang [1], Lantian Zhang [3] and Xiaofei Qi [4]

[1] School of Geospatial Information, PLA Strategic Support Force Information Engineering University, Zhengzhou 450001, China
[2] Speed China Technology Co., Ltd., Nanjing 210046, China
[3] Beijing Institute of Remote Sensing Information, Beijing 100011, China
[4] Xi'an Surveying & Mapping Institute, Xi'an 710054, China
* Correspondence: wangguangxia2011@163.com; Tel.: +86-(81)-639-662

Abstract: The geographical feature extraction of historical maps is an important foundation for realizing the transition from human map reading to machine map reading. The current methods for building block extraction from historical maps have many problems, such as low accuracy and poor scalability. Moreover, the high cost of annotating historical maps further limits its applications. In this study, a method for extracting building blocks from historical maps is proposed based on the deep object attention network. Based on the OCRNet framework, multiple attention mechanisms were used to improve the ability of the network to extract the contextual information of the target. Moreover, through the optimization of the feature extraction network structure, the impact of the down-sampling process on local information and boundary contours was reduced, in order to improve the network's ability to capture boundary information. Subsequently, the transfer learning method was used to jointly train the network model on both remote sensing datasets and few-shot historical map datasets to further improve the feature learning ability of the network, which overcomes the constraints of small sample sizes. The experimental results show that the proposed method can effectively improve the extraction accuracy of building blocks from historical maps.

Keywords: historical maps; building block; feature extraction; object attention; transfer learning

Citation: Zhao, Y.; Wang, G.; Yang, J.; Zhang, L.; Qi, X. Building Block Extraction from Historical Maps Using Deep Object Attention Networks. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 572. https://doi.org/10.3390/ijgi11110572

Academic Editors: Wolfgang Kainz and Maria Antonia Brovelli

Received: 26 September 2022
Accepted: 14 November 2022
Published: 16 November 2022

1. Introduction

Historical maps preserve the natural landscape and the traces of human activity on the Earth's surface for an extended period [1,2]. Such maps are valuable for historical and cultural heritage, and they provide an essential source for the description and communication of geographical features and their spatial relationships. Thus, historical maps are of great reference value for analyzing regional developments and changes. With the continuous development of digital technology, many countries have digitized paper historical maps and established different types of digital historical map archives to better preserve and utilize these historical documents (https://ngmdb.usgs.gov/topoview/viewer/ (accessed on 12 November 2022), https://www.oldmapsonline.org/ (accessed on 12 November 2022), http://www.map-cn.com/ (accessed on 12 November 2022), https://www.swisstopo.admin.ch/en/maps-geodata-online.html (accessed on 12 November 2022)). Although these digital historical maps are available to the public and can be easily read and understood by humans, an enormous amount of geographic information is locked in the images, which makes it challenging to perform quantitative calculations and spatial analysis through automatic machine reading [3]. In geographic feature extraction, each pixel is assigned with the accurate classification label (e.g., roads, waters, building blocks, and map annotations) such that computers can autonomously "read" the maps. As an important artificial facility, the building is an essential place for people's life and activities. The automatic extraction and identification of building block features from historical maps

has the following challenges due to the limitations of the era and mapping technology (Figure 1). (1) The sizes and styles of building blocks vary, but they can be simplified either as single buildings or as blank areas with boundaries. Thus, the multi-scale expression of features is a great challenge. (2) Building blocks can overlap and nest with each other, which increases the difficulty of filtering and screening boundaries. (3) Building blocks can be covered by other map contents, which interfere with the texture features of the building blocks. (4) Finally, the integrity of building blocks can be affected due to stains, creases, and map damage.

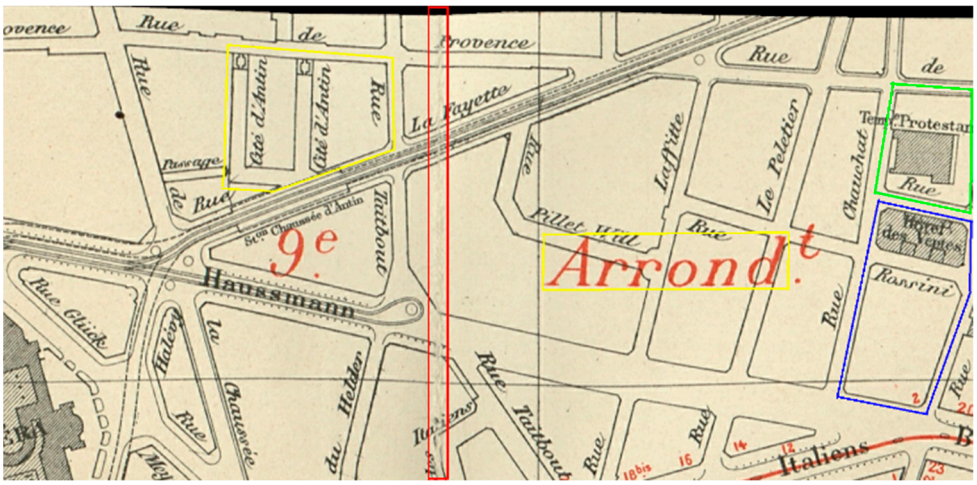

Figure 1. Exemplar challenges in a sample historical map for feature extraction: challenge 1: the shadows represent single buildings and the other has only building boundaries (blue box); challenge 2: building blocks overlap and nest with each other (green box); challenge 3: building blocks are covered by other map contents (yellow box); challenge 4: creases and damage (red box).

With the rapid development of deep learning techniques in the field of computer vision, Convolution Neural Networks (CNNs) have been widely applied in the field of geospatial information processing [4–6]. Compared with common images, building blocks account for only a small part of historical map images (Figure 2). Thus, local information may be missing, or the boundaries of the building blocks may be blurred due to down sampling in the training process [7], which affects the extraction results. The majority of the existing CNNs are composed of a series of sub-networks in high-resolution, low-resolution, and subsequent high-resolution sequences (e.g., ResNet [8], GoogLeNet [9], and VGGNet [10]). The network structure affects the ability to extract the features of building blocks to a certain extent. In addition, many public datasets are available for the semantic segmentation of natural images, such as ImageNet [11], COCO [12], and Cityscapes [13], which greatly increases the accuracy of semantic segmentation. In comparison, the benchmark datasets for the semantic segmentation of historical maps are relatively rare. Therefore, improving the accuracy of semantic segmentation based on limited label data is of great significance for feature extraction from historical map images [14].

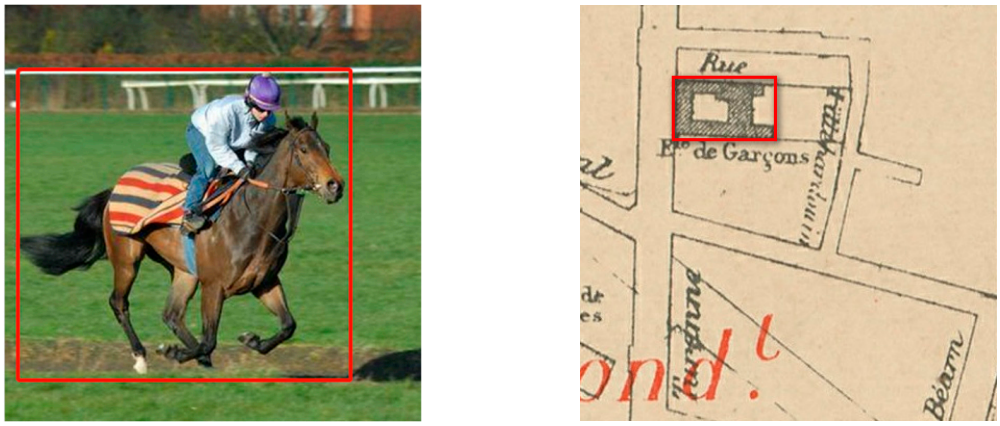

Figure 2. Comparison of object contexts. Rider on a horse in a natural image (**left**) and a building in a cropped historical map (**right**). The building block image represents only a small part of this theoretical map image compared to the natural image (area surrounded by red boxes as a proportion of the total image area).

To fully utilize the advantages of deep learning in feature extraction from historical maps, the object context features were incorporated with the attention mechanism in this study. Specifically, based on HRNet [15] and OCRNet [16], a Deep Object Attention Network (DOANet) was developed to extract fine-grained geographic features from historical maps under the condition of limited training samples. DOANet makes full use of the deep features in the limited training samples, and the attention mechanism accurately captures the global contextual information, thereby improving the ability of the model to learn the features of building blocks and suppress its responsiveness to other categories of features. Moreover, the transfer learning method was integrated into training to extract building blocks from the established few-shot historical map dataset. The results show that the proposed model can effectively reduce the cost of manual annotation under the condition of limited map samples and has good accuracy in historical map building block extraction.

The paper is organized as follows: in order to understand the value of this research, the paper describes the work related to this paper (Section 2), describes the idea of the algorithm (Section 3), and carries out the algorithm implementation and compares the experimental results with the analysis (Section 4). Finally, Section 5 concludes the paper with remarks on future work.

2. Related Works

As a critical task of digital map processing, information extraction from historical maps has received much attention from researchers in various fields [17–21]. Early studies mainly used color segmentation, template matching, shape descriptors, mathematical morphological operators, and other methods to extract geographic features, such as roads, contour lines, and buildings [22–28]. However, these methods often use customized processes and parameter configurations for specific map types or geographic elements, and thus, they suffer from low automation. As the scale of digital map data has increased, the cost of data processing has increased significantly. In recent years, CNN-based object detection, scene classification, and semantic segmentation methods have been applied to extract geospatial data from remote sensing images, and these approaches have shown better results than traditional methods [5,29,30]. As both remote sensing images and historical map images are pixel-based and both are objective representations of geographical elements, they have similarities at the semantic level. Hence, many studies have been carried out to apply deep learning methods to the extraction of information from historical maps [31–34].

To reduce the impact of scale on the feature extraction results of deep learning networks, Duan et al. [35] proposed a georeferencing method based on reinforcement learning. Through the automatic alignment of contemporary vector data and georeferenced historical maps, the precise locations of geographic features on scanned maps were annotated. In addition, Generative Adversarial Networks (GANs) have been used to generate data for historical maps. For example, Li [36] proposed an automatic method to generate a dataset from Open Street Map to train text detection systems to be able to work with historical maps. Andrade et al. [4] synthesized satellite-like urban images based on historical maps. Although these methods can increase the scale of datasets to a certain extent and improve the feature learning ability of deep learning networks, they still require a large amount of historical map data, and deviations between vector data and historical maps remain.

Saeedimoghaddam et al. [33] used Faster RCNN to extract the intersection points of single-lane and double-lane roads from the United States Geological Survey (USGS) historical map series; they also used the pre-trained Inception-Resnet-V2 on the Microsoft Common Objects in Context (COCO) dataset to improve the accuracy of the network. Because the target objects in the COCO dataset are quite different from geospatial elements in terms of the scale, direction, and shape of the data [37], the use of geospatial data (e.g., remote sensing images) for pre-training and transfer learning has been proposed.

Heitzler et al. [38] segmented single buildings from the Swiss Siegried map using U-Net and used methods based on contour tracing and orientation-based clustering to vectorize the segmentation results. Uhl et al. [3,7,39] studied the effects of different network structures to extract the footprint of human settlements from historical USGS topographic map series and used the weakly supervised CNN to solve the problem of the high costs related to manual annotation. They found that for the semantic segmentation of historical maps, the accuracy of the feature extraction network had a significant influence on the segmentation performance. It is worth noting that the above studies only focused on buildings in small-scale topographic maps that were represented by small rectangles with a regular shape and simple texture (most of them filled with a single color). However, in large-scale maps, building blocks have complex contours and different types of textures, even texture-free blanks, posing a great challenge to the algorithm's feature extraction capabilities.

3. Methods

3.1. Network Model

When training samples are limited, the number of target features (i.e., building blocks) in a map is small. Hence, improving the feature extraction ability of a network is necessary, and the problem of missing details during down sampling must be addressed. In this study, the encoding and decoding structures of HRNet were optimized to capture multi-scale deep features, and OCRNet was introduced to obtain contextual information in the samples. Then, the deep features were incorporated with the contextual information to increase the ability of the network to learn the building block features in few-shot datasets.

3.1.1. Architecture of DOANet

The architecture of the proposed DOANet for building block extraction from historical maps based on the attention mechanism is shown in Figure 3. Based on OCRNet, DOANet is composed of a feature extraction module and an object attention module. In particular, the object attention module is further divided into the criss-cross attention module and the object context module. Specifically, the criss-cross attention module uses a large receptive field to obtain spatial distribution information and learns important features while ignoring irrelevant features. The object context module is designed to fuse receptive fields of different sizes to capture detailed contextual information. Then, the object attention module aggregates the spatial distribution information and object contextual information to enhance feature representation. The deep features extracted by the feature extraction module are further optimized by the criss-cross attention module, and the optimized deep features and

the coarse classification results of the intermediate layers are taken as the input of the object context module to obtain the object region features. Then, the optimized deep features and object region features are combined by the criss-cross attention module to obtain the object context features. Finally, the optimized deep features are spliced with the object context features to obtain the final feature representation with enhanced contextual information.

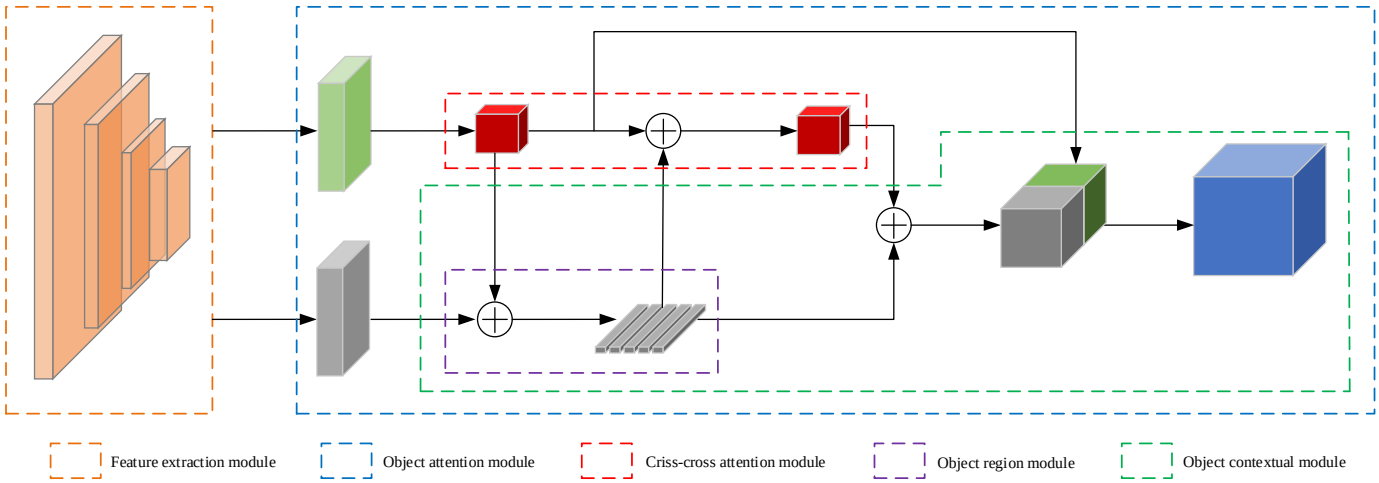

Figure 3. Architecture of DOANet.

3.1.2. Feature Extraction Module

Unlike most feature extraction networks, HRNet is composed of parallel high-resolution and low-resolution sub-networks and uses repeated multi-scale feature fusion. Low-resolution feature maps of the same depth and similar levels are used to improve high-resolution features, enabling the network to capture local information with a strong robustness. Based on HRNet, DOANet consists of six encoders and six decoders (Figure 4), which reduces the connections of the same-scale layers in the original network and strengthens the connections between layers of different scales. Moreover, due to the small number of encoders, fewer decoders are needed, thereby reducing the size of the network. Each encoder uses leaky-ReLU as the activation function, supplemented by batch normalization operations to improve the stability of the model parameters. Because the extraction of building information from historical maps is a binary classification problem, i.e., the labels only include the background and buildings, cross entropy is used as the loss function L:

$$L = \frac{1}{N}\sum_{i} -[y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i)] \quad (1)$$

where y_i is the label of sample i, which is one for positive classification and zero for negative classification, and p_i is the probability that sample i will be positively predicted.

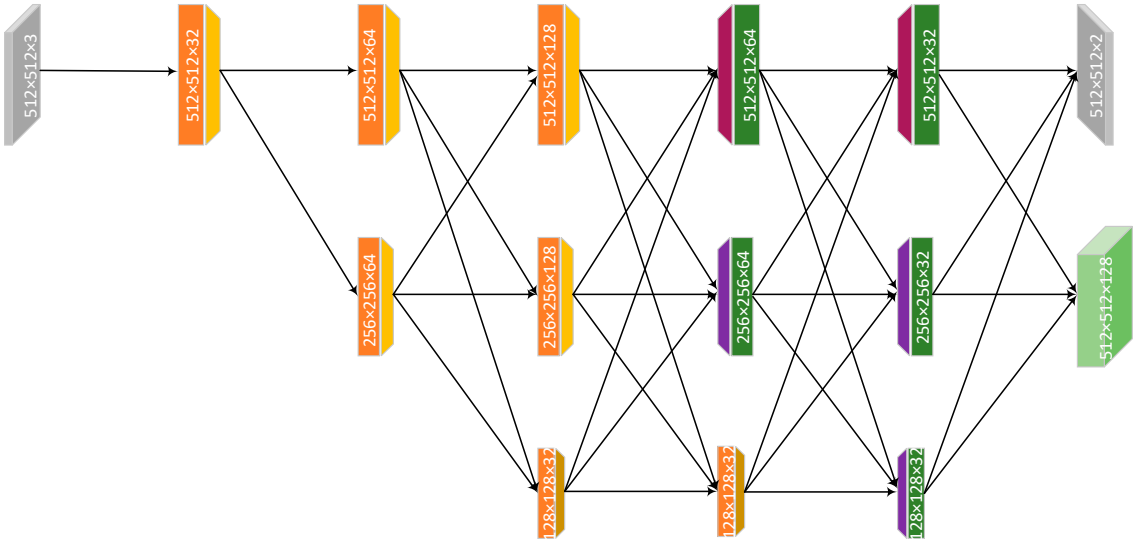

Figure 4. Structure of the feature extraction module.

3.1.3. Attention Module

The criss-cross attention module [40] in Figure 5 was used in this study. The module can capture the dependence of each pixel on the rest of the pixels in the image, thereby effectively improving the ability of the network to extract contextual information. First, the dimension of the feature map H of size $C \times W \times H$ is reduced by two 1×1 convolutions, and two feature maps, Q and K, are obtained. Then, for any pixel u on the feature map Q, a channel vector Q_u with a size of $1 \times 1 \times C'$ is obtained, and all pixels in the same row and column as pixel u are used to construct a feature vector Ω_u with a size of $(H + W - 1) \times C'$. Next, the affinity $d_{i,u}$ of each pixel u on the feature map Q to the feature vector Ω_u is calculated through the affinity operation:

$$d_{i,u} = Q_u \Omega_{i,u}^T \tag{2}$$

where $Q_{i,u}$ denotes the i-th channel vector of Ω_u. The attention map A of size $(H + W - 1) \times W \times H$ is obtained after the SoftMax layer. In addition, the feature map V of size $C \times W \times H$ is obtained through another 1×1 convolution of the feature map H. The feature vector $\Psi_{i,u}$ in the same row and column as each pixel u in V is dot multiplied with the feature vector $A_{i,u}$ in the corresponding position, and the dot products for all pixels are added to obtain the residual aggregation feature at the position, which is then added to the original feature vector H_u to obtain the feature vector H_u' with a stronger feature representation ability. The equation is outlined as follows:

$$H_u' = \sum_{i=0}^{H+W-1} A_{i,u} \cdot \Psi_{i,u} + H_u \tag{3}$$

Because a single criss-cross attention module only considers elements on the same row and column as a pixel, two criss-cross attention modules are connected to obtain the contextual information at all positions.

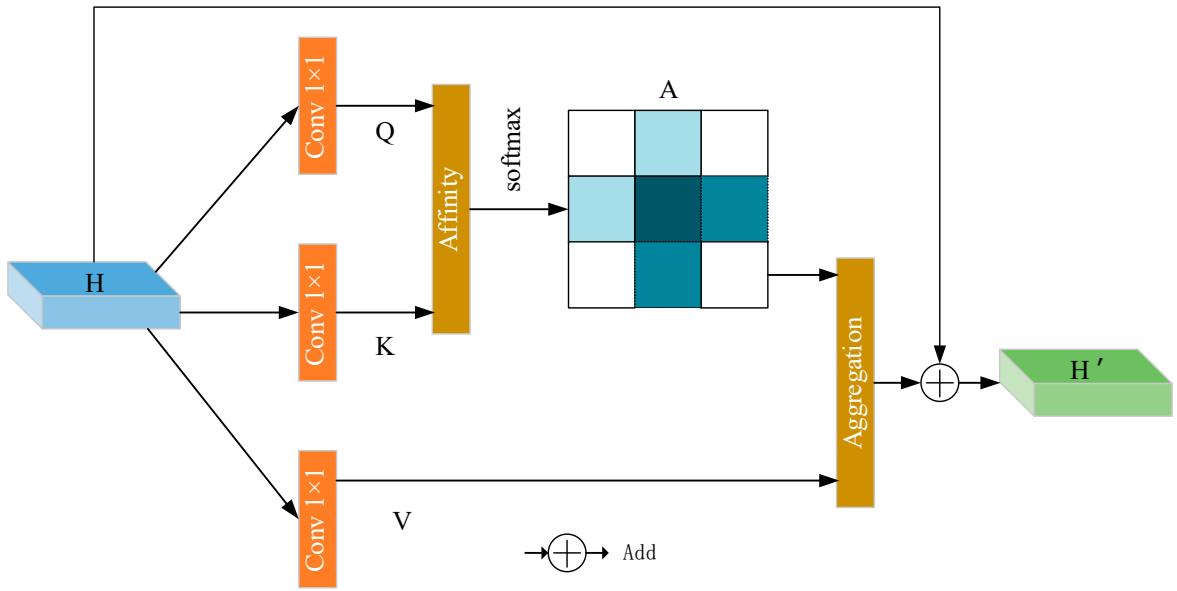

Figure 5. Structure of the attention module.

3.2. Transfer Learning

Transfer learning is a technique that converts the learning processes in the source domain, including the training data, model parameters, and tasks, into knowledge and then transfers them to the target domain to facilitate the learning of the prediction function in the target domain [41]. The transfer learning method used in this study is shown in Figure 6. A public dataset was used as the training dataset in the source domain; that is, the sample dataset in the source domain was imported into DOANet for learning, and the parameters and features of the source domain network were shared with the target domain through network replication. The target domain network was initialized with the network parameters in the source domain, while freezing the batch normalization layer in the target domain. The dataset in the target domain was used for training to fine-tune the network parameters in the target domain, thereby realizing the knowledge transfer.

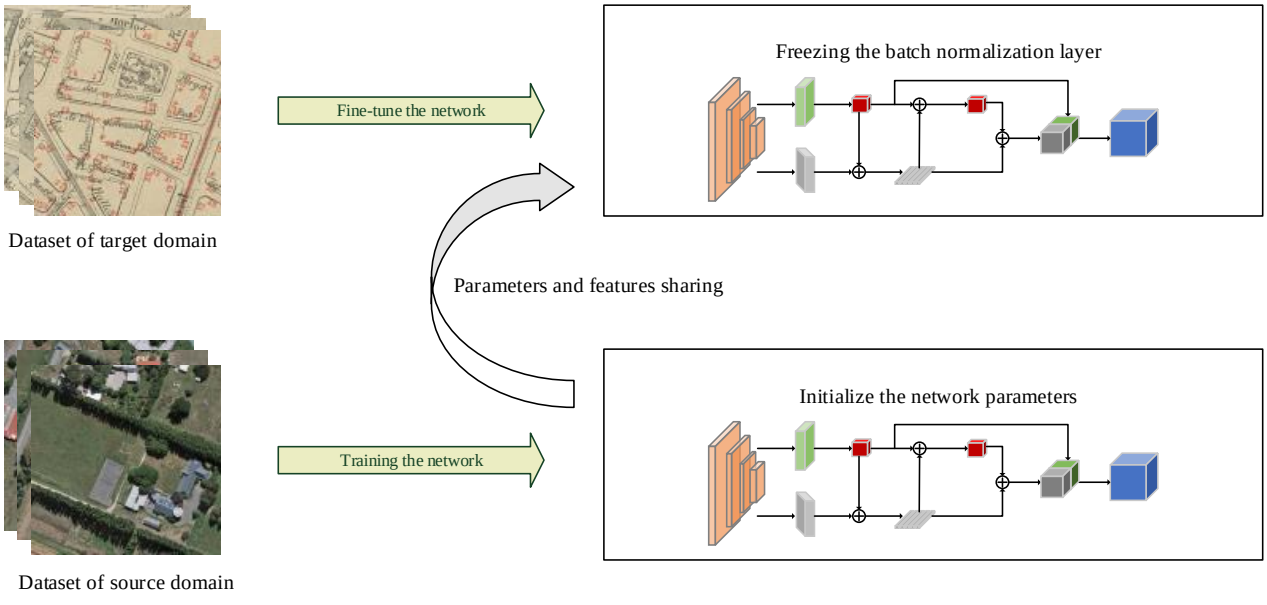

Figure 6. Integration of DOANet with transfer learning.

4. Experiments

To evaluate the efficiency of DOANet and compare the results with those of the existing semantic segmentation algorithms, the dataset for task 1 (building block detection) in the ICDAR2021 Competition on Historical Map Segmentation [42] was used to train, validate, and test the algorithms. The dataset consists of large-scale urban maps of Paris dating from 1860 to 1940 collected by the French National Library, which includes one training image, one validation image, and three test images. The resolution of each image is at least 8000 × 6000. The building blocks in the training and validation images were manually annotated. A 512 × 512 sliding window with a step size of 200 was used to crop the training image and the validation image (including the corresponding annotated images), and the cropped image blocks were then used as the dataset in the experiment. The dataset was then divided into a training set and a validation set at a ratio of 8:2. In total, 2237 training samples and 559 validation samples were obtained. Prior to training, each training sample is flipped up and down and left and right in a mirror image, and then randomly rotated once at 45°. ICDAR2021 provides a standard indicator to evaluate the test results, which is calculated as follows:

$$PQ = SQ \times RQ = \frac{\sum_{(p,g)\in TP} IoU(p,g)}{TP} \times \frac{TP}{TP + \frac{1}{2}FP + \frac{1}{2}FN} \quad (4)$$

where PQ is the aggregated score, SQ is the mean Intersection Over Union ($mIoU$), RQ is the F-score, and TP, FP, and FN represent true positive, false positive, and false negative, respectively.

The test platform was a 64-bit Ubuntu 18.04 operating system equipped with eight GeForce RTX™ 2080 Ti GPUs (11 GB VRAM). The PaddlePaddle v2.1 framework was used to build the algorithm. The batch size was set to 8 according to the characteristics of the GPUs, the initial learning rate of the network was 0.0125, and the Stochastic Gradient Descent (SGD) was used as the optimizer. The momentum was 0.9, the weight decay rate was 4×10^{-5}, and the training epoch was set to 10,000 times. The open source WHU building dataset [43] was used as the source domain dataset, and the divided dataset was used as the target domain dataset to train DOANet. Building block extraction was carried out on the three test images after training, and the results were evaluated using the above indicators in Equation (4).

4.1. Results

The scores of DOANet were compared with the official results of ICDAR2021 (Table 1). The PQ of DOANet was higher than that of the other methods for all three test images. The visualization result of the method is shown in Figure 7. Compared with that of the other methods, the PQ of DOANet was increased by at least 7.2% suggesting that the feature extraction method based on the attention mechanism effectively aggregated the local features that did not pass through the receptive field and that the impact of blurred boundaries was reduced due to the scale change. Thus, the proposed algorithm showed a high level of detection accuracy. However, the method's ability to solve the problem of overlapping nesting within building blocks leaves much to be desired, as shown by the second result in Figure 7, where DOANet incorrectly judges several single buildings (shaded sections) as a whole.

Table 1. Comparison of the evaluation results.

Team	MapI *PQ* (%)	MapII *PQ* (%)	MapIII *PQ* (%)	Mean *PQ* (%)
L3IRIS	74.4	69.8	78.2	74.1
CMM (1)	59.8	61.4	66.7	62.6
CMM (2)	52.6	47.9	58.1	44.0
WUU (1)	7.7	5.9	5.7	6.4
WUU (2)	4.7	4.0	3.9	4.2
DOANet	83.5	79.2	81.2	81.3

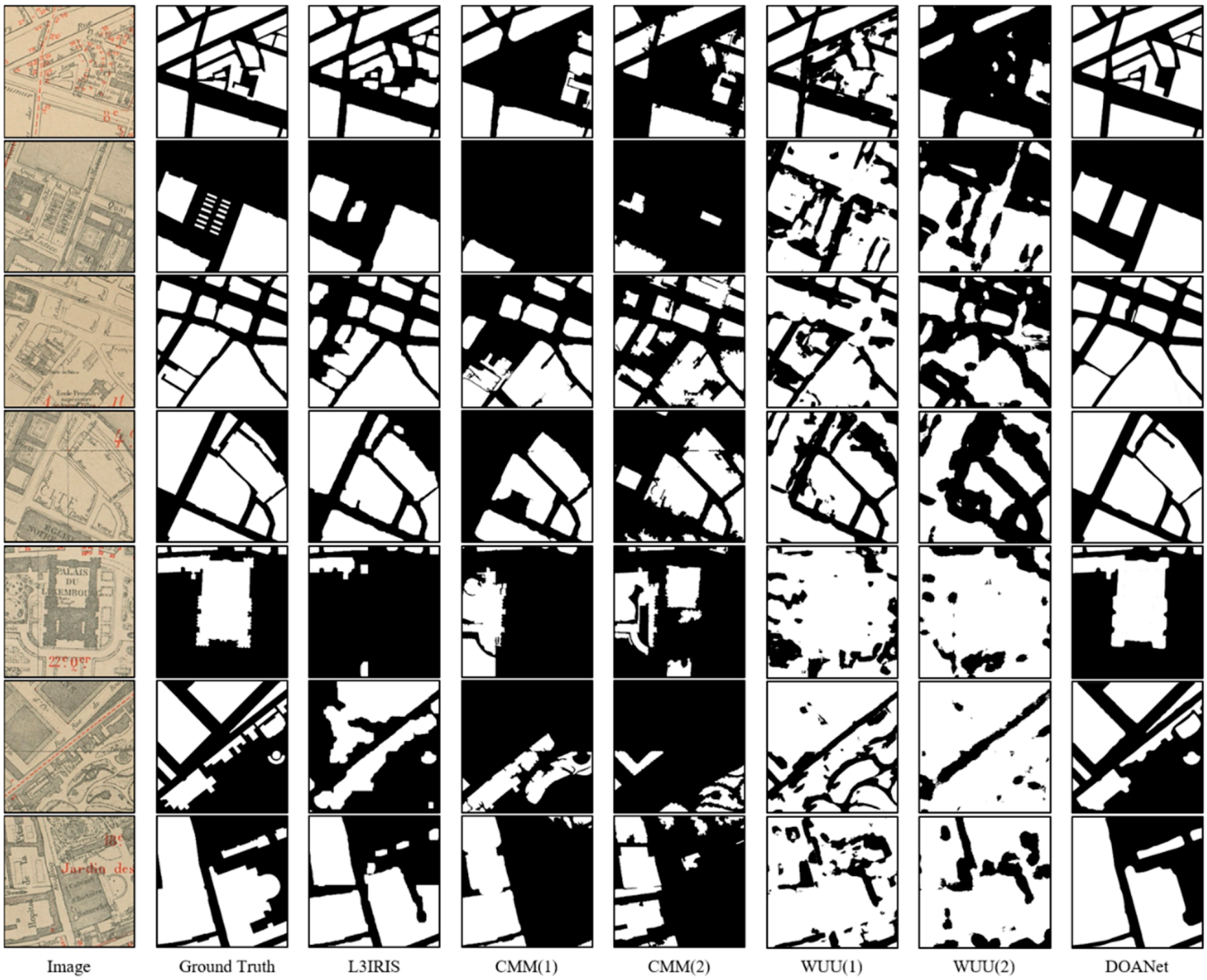

Figure 7. Sample results of the building block (white) extraction.

4.2. Ablation Analysis

To further analyze the role of the object attention module and transfer learning in the extraction of geographic features from historical maps, five algorithms were designed to investigate the modular performance of DOANet:

ANet: the feature extraction module was unchanged, and only the criss-cross attention module was retained.

ONet: the feature extraction module was unchanged, and only the object context module was retained.

OOANet: the object attention module was unchanged, the number of encoders and decoders in the feature extraction module was changed, and the original HRNet structure (Figure 8, left) was used [15].

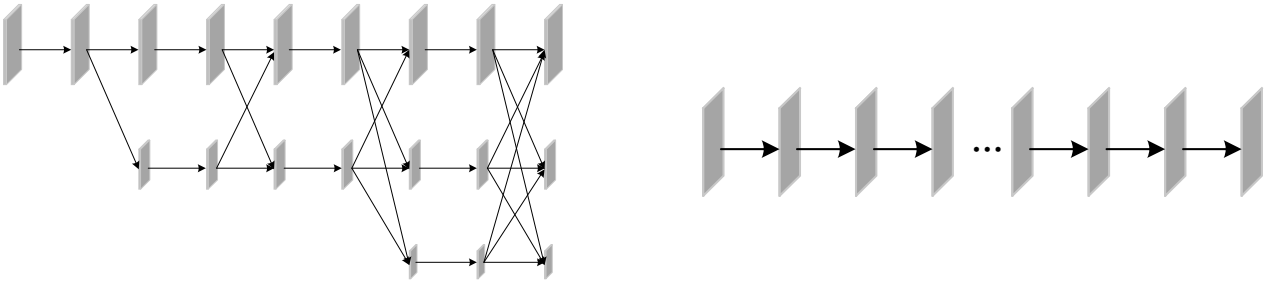

Figure 8. (**Left**): schematic diagram of the HRNet structure; (**Right**): schematic diagram of the ResNet structure.

ROANet: the object attention module was unchanged, and the feature extraction network ResNet (Figure 8, right), which has a series structure, was used [8].

DOANet-: the network structure was unchanged, and the transfer learning module was removed.

The quantitative evaluation indicators of each experiment are shown in Table 2. Due to the absence of part of the core modules, both ANet and ONet yielded poor building block extraction results. OOANet has a deeper network structure and should theoretically yield better extraction results, yet the results showed that OOANet failed to enhance the features. Moreover, the results of ROANet were also inferior to those of DOANet due to the loss of local information, which indirectly demonstrates the advantages of parallel networks in the extraction of geographical features from historical maps. In addition, in the absence of external knowledge transfer, the performance of DOANet- was affected to a certain extent due to the limited number of samples. As shown in Figure 9, when the number of training samples was small, the scores of DOANet- without transfer learning decreased. As the training dataset increased, the difference between DOANet and DOANet- gradually decreased. To achieve a score of >75%, DOANet- required 1200 training samples, whereas DOANet only needed 600 image samples. Hence, the method proposed in this study can effectively deal with the problem of limited training samples and thereby reduce the cost of annotation.

Table 2. Evaluation results of DOANet with varying network configurations.

Team	MapI *PQ* (%)	MapII *PQ* (%)	MapIII *PQ* (%)	Mean *PQ* (%)
ANet	69.4	72.1	73.0	71.5
ONet	75.5	72.8	72.8	73.7
OOANet	80.1	76.8	80.7	79.2
ROANet	68.5	71.2	70.9	70.2
DOANet-	79.4	78.1	81.3	79.6
DOANet	83.5	79.2	81.2	81.3

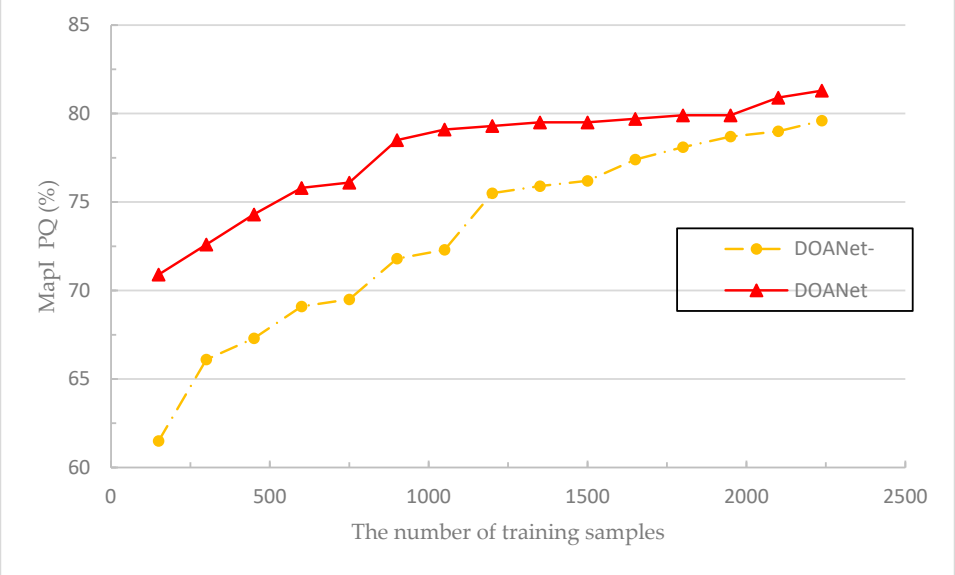

Figure 9. The number of training samples compared to scores.

5. Conclusions

To address the problems associated with information extraction from historical maps, a building block extraction network for historical maps, DOANet, was developed based on the attention mechanism. Built upon the HRNet and OCRNet structures, DOANet uses parallel subnetworks to fuse multi-scale features and includes a criss-cross attention module. Moreover, the transfer learning method is integrated with DOANet to improve the detection accuracy of the model in the case of limited training samples. The experimental results show that the *PQ* of the proposed method increased by at least 7.2% compared with that of existing algorithms. The proposed method effectively solves the problem of poor network performance caused by insufficient training samples in the task of building block extraction from historical maps and provides a reference for extracting other features from historical maps. However, the ability of the method to solve the problem of overlapping nesting within building blocks leaves much to be desired, and we will focus on solving this problem in our next work. In addition, applying the method to different styles of maps (e.g., different languages, different time periods) will also be part of our future work in order to improve the generalizability of the method in this paper.

Author Contributions: Conceptualization, Yao Zhao; Methodology, Yao Zhao; Software, Yao Zhao; Validation, Yao Zhao; Formal Analysis, Yao Zhao; Data Curation, Yao Zhao; Writing—Original Draft Preparation, Yao Zhao; Writing—Review and Editing, Guangxia Wang, Jian Yang and Lantian Zhang; Visualization, Yao Zhao; Supervision, Guangxia Wang, Jian Yang and Xiaofei Qi. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by China's National Key R&D Program (No. 2017YFB0503500) and National Natural Science Foundation of China (No. 41901335).

Data Availability Statement: The data presented in this study are openly available in [Zenodo] at [10.5281/zenodo.4817662], reference number [42].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dominik, K.; Jacek, K.; Natalia, K.; Elżbieta, Z.; Krzysztof, O.; Katarzyna, O.; Urs, G.; Catalina, M.; Volker, C.R. Broad Scale Forest Cover Reconstruction from Historical Topographic Maps. *Appl. Geogr.* **2016**, *67*, 39–48. [CrossRef]
2. Shbita, B.; Knoblock, C.A.; Duan, W.W.; Chiang, Y.Y.; Uhl, J.H.; Leyk, S. Building Linked Spatio-Temporal Data from Vectorized Historical Maps. In Proceedings of the Extended Semantic Web Conference 2020, Heraklion, Greece, 1–4 June 2020; pp. 409–426.
3. Uhl, J.H.; Leyk, S.; Chiang, Y.Y.; Duan, W.W.; Knoblock, C.A. Automated Extraction of Human Settlement Patterns from Historical Topographic Map Series Using Weakly Supervised Convolutional Neural Networks. *IEEE Access* **2020**, *8*, 6978–6996. [CrossRef]
4. Andrade, H.J.; Fernandes, B.J. Synthesis of Satellite-Like Urban Images From Historical Maps Using Conditional GAN. *IEEE Geosci. Remote Sens. Lett.* **2020**, *19*, 3000504. [CrossRef]
5. Yuan, X.H.; Shi, J.F.; Gu, L.C. A Review of Deep Learning Methods for Semantic Segmentation of Remote Sensing Imagery. *Expert Syst. Appl.* **2021**, *169*, 114–147. [CrossRef]
6. Can, Y.S.; Gerrits, P.J.; Kabadayi, M.E. Automatic Detection of Road Types from the Third Military Mapping Survey of Austria-Hungary Historical Map Series with Deep Convolutional Neural Networks. *IEEE Access* **2021**, *9*, 62847–62856. [CrossRef]
7. Uhl, J.H.; Leyk, S.; Chiang, Y.Y.; Duan, W.W.; Knoblock, C.A. Extracting Human Settlement Footprint from Historical Topographic Map Series Using Context-Based Machine Learning. In Proceedings of the 8th International Conference of Pattern Recognition Systems (ICPRS 2017), Madrid, Spain, 28–31 October 2017; pp. 1–6.
8. He, K.M.; Zhang, X.Y.; Ren, S.Q.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016), Las Vegas, NV, USA, 26–30 June 2016; pp. 770–778.
9. Szegedy, C.; Liu, W.; Jia, Y.Q.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015), Boston, MA, USA, 7–9 June 2015; pp. 1–9.
10. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015), Boston, MA, USA, 7–9 June 2015; pp. 16–24.
11. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.H.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]
12. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In Proceedings of the European Conference on Computer Vision (ECCV 2014), Zürich, Switzerland, 7–9 September 2014; pp. 740–755.
13. Cordts, M.; Omran, M.; Ramos, S.; Rehfeld, T.; Enzweiler, M.; Benenson, R.; Franke, U.; Roth, S.; Schiele, B. The Cityscapes Dataset for Semantic Urban Scene Understanding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016), Las Vegas, NV, USA, 26–30 June 2016; pp. 3212–3223.
14. Chiang, Y.Y.; Duan, W.W.; Leyk, S.; Uhl, J.H.; Knoblock, C.A. *Using Historical Maps in Scientific Studies: Applications, Challenges, and Best Practices*; Springer: Berlin, Germany, 2020; pp. 10–25. ISBN 978-3-319-66907-6.
15. Sun, K.; Xiao, B.; Liu, D.; Wang, J.D. Deep High-Resolution Representation Learning for Human Pose Estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2019), Long Beach, CA, USA, 16–20 June 2019; pp. 5693–5703.
16. Yuan, Y.H.; Chen, X.K.; Chen, X.L.; Wang, J.D. Segmentation Transformer: Object-Contextual Representations for Semantic Segmentation. In Proceedings of the European Conference on Computer Vision (ECCV 2020), Glasgow, UK, 23–28 August 2020; pp. 1417–1438.
17. Howe, N.R.; Weinman, J.; Gouwar, J.; Shamji, A. Deformable Part Models for Automatically Georeferencing Historical Map Images. In Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ICAGIS 2019), Chicago, IL, USA, 5–8 November 2019; pp. 540–543.
18. Heitzler, M.; Hurni, L. Unlocking the Geospatial Past with Deep Learning–Establishing a Hub for Historical Map Data in Switzerland. In Proceedings of the 29th International Cartographic Conference (ICC 2019), Tokyo, Japan, 15–20 July 2019; pp. 1–10.
19. Asokan, A.; Anitha, J.; Ciobanu, M.; Gabor, A.; Naaji, A.; Hemanth, J. Image Processing Techniques for Analysis of Satellite Images for Historical Maps Classification—An Overview. *Appl. Sci.* **2020**, *10*, 4207. [CrossRef]
20. Garcia-Molsosa, A.; Orengo, H.A.; Lawrence, D.; Philip, G.; Hopper, K.; Petrie, C.A. Potential of Deep Learning Segmentation for the Extraction of Archaeological Features from Historical Map Series. *Archaeol. Prospect.* **2021**, *28*, 187–199. [CrossRef]
21. Chen, Y.Z.; Carlinet, E.; Chazalon, J.; Mallet, C.; Duménieu, B.; Perret, J. Combining Deep Learning and Mathematical Morphology for Historical Map Segmentation. In Proceedings of the International Conference on Discrete Geometry and Mathematical Morphology (DGMM 2021), Uppsale, Sweden, 24–27 May 2021; pp. 79–92.
22. Chiang, Y.Y.; Knoblock, C.A. Extracting Road Vector Data from Raster Maps. In Proceedings of the 8th International Workshop on Graphics Recognition (GREC 2009), La Rochelle, France, 11–13 July 2009; pp. 93–105.
23. Chiang, Y.Y.; Leyk, S.; Knoblock, C.A. Efficient and Robust Graphics Recognition from Historical Maps. In Proceedings of the 8th International Workshop on Graphics Recognition (GREC 2011), Seoul, Korea, 21–22 September 2011; pp. 25–35.
24. Chen, Y.; Wang, R.S.; Qian, J. Extracting Contour Lines from Common-conditioned Topographic Maps. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 1048–1057. [CrossRef]

25. Miao, Q.G.; Xu, P.F.; Liu, T.G.; Yang, Y.; Zhang, J.Y. Linear Feature Separation from Topographic Maps Using Energy Density and the Shear Transform. *IEEE Trans. Image Process.* **2013**, *22*, 1548–1558. [CrossRef]
26. Liu, Y. An Automation System: Generation of Digital Map Data from Pictorial Map Resources. *Pattern Recognit.* **2002**, *35*, 1973–1987. [CrossRef]
27. Miyoshi, T.; Li, W.Q.; Kaneda, K.; Yamashita, H.; Nakamae, E. Automatic Extraction of Buildings Utilizing Geometric Features of a Scanned Topographic Map. In Proceedings of the 17th International Conference on Pattern Recognition (ICPR 2004), Stockholm, Sweden, 26 August 2004; pp. 626–629.
28. Leyk, S.; Boesch, R. Colors of the Past: Color Image Segmentation in Historical Topographic Maps Based on Homogeneity. *Geoinformatica* **2010**, *14*, 1–21. [CrossRef]
29. Alganci, U.; Soydas, M.; Sertel, E. Comparative Research on Deep Learning Approaches for Airplane Detection from Very High-Resolution Satellite Images. *Remote Sens.* **2020**, *12*, 458. [CrossRef]
30. Cheng, G.; Xie, X.X.; Han, J.W.; Guo, L.; Xia, G.S. Remote Sensing Image Scene Classification Meets Deep Learning: Challenges, Methods, Benchmarks, and Opportunities. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 3735–3756. [CrossRef]
31. Wu, S.D.; Heitzler, M.; Hurni, L. Leveraging Uncertainty Estimation and Spatial Pyramid Pooling for Extracting Hydrological Features from Scanned Historical Topographic Maps. *Gisci. Remote Sens.* **2022**, *59*, 200–214. [CrossRef]
32. Ekim, B.; Sertel, E.; Kabaday, M.E. Automatic Road Extraction from Historical Maps Using Deep Learning Techniques: A Regional Case Study of Turkey in a German World War II Map. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 492. [CrossRef]
33. Saeedimoghaddam, M.; Stepinski, T.F. Automatic Extraction of Road Intersection Points from USGS Historical Map Series using Deep Convolutional Neural Networks. *Int. J. Geogr. Inf. Sci.* **2020**, *34*, 947–968. [CrossRef]
34. Schlegel, I. Automated Extraction of Labels from Large-Scale Historical Maps. *Agile Giss.* **2021**, *2*, 1–14. [CrossRef]
35. Duan, W.W.; Chiang, Y.Y.; Leyk, S.; Uhl, J.H.; Knoblock, C.A. Automatic Alignment of Contemporary Vector Data and Georeferenced Historical Maps Using Reinforcement Learning. *Int. J. Geogr. Inf. Sci.* **2019**, *34*, 824–849. [CrossRef]
36. Li, Z. Generating Historical Maps from Online Maps. In Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Chicago, IL, USA, 5–8 November 2019; pp. 610–611.
37. Xia, G.S.; Bai, X.; Ding, J.; Zhu, Z.; Belongie, S.; Luo, J.B.; Datcu, M.; Pelillo, M.; Zhang, L.P. DOTA: A Large-scale Dataset for Object Detection in Aerial Images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2018), Salt Lake City, GA, USA, 18–22 June 2018; pp. 3974–3983.
38. Heitzler, M.; Hurnim, L. Cartographic Reconstruction of Building Footprints from Historical Maps: A Study on the Swiss Siegfried map. *Trans. GIS* **2020**, *24*, 442–461. [CrossRef]
39. Uhl, J.H.; Leyk, S.; Chiang, Y.Y.; Duan, W.W.; Knoblock, C.A. Spatialising Uncertainty in Image Segmentation using Weakly Supervised Convolutional Neural Networks: A Case Study from Historical Map Processing. *IET Image Process.* **2018**, *12*, 2084–2091. [CrossRef]
40. Huang, Z.L.; Wang, X.G.; Huang, L.C.; Huang, C.; Wei, Y.C.; Liu, W.Y. CCNet: Criss-Cross Attention for Semantic Segmentation. In Proceedings of the IEEE International Conference on Computer Vision (ICCV 2019), Seoul, Korea, 27–31 October 2019; pp. 603–612.
41. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [CrossRef]
42. Chazalon, J.; Carlinet, E.; Chen, Y.Z.; Perret, J.; Duménieu, B.; Mallet, C.; Géraud, T.; Nguyen, V.; Nguyen, N.; Baloun, J.; et al. ICDAR 2021 Competition on Historical Map Segmentation. In Proceedings of the 16th International Conference on Document Analysis and Recognition (ICDAR 2021), Lausanne, Switzerland, 5–10 September 2021; pp. 693–707.
43. Ji, S.P.; Wei, S.Q. Building extraction via convolutional neural networks from an open remote sensing building dataset. *Acta Geod. Cartogr. Sin.* **2019**, *48*, 448–459. [CrossRef]