International Journal of Geo-Information

MDPI

Article

Automatic Road Extraction from Historical Maps Using Transformer-Based SegFormers

Elif Sertel [1,*], Can Michael Hucko [1] and Mustafa Erdem Kabadayı [2]

[1] Geomatics Engineering Department, Istanbul Technical University, Istanbul 34469, Turkey; huckoc21@itu.edu.tr
[2] Department of History, College of Social Sciences and Humanities, Koç University, Rumelifeneri Yolu, Istanbul 34450, Turkey; mkabadayi@ku.edu.tr
* Correspondence: sertele@itu.edu.tr

Abstract: Historical maps are valuable sources of geospatial data for various geography-related applications, providing insightful information about historical land use, transportation infrastructure, and settlements. While transformer-based segmentation methods have been widely applied to image segmentation tasks, they have mostly focused on satellite images. There is a growing need to explore transformer-based approaches for geospatial object extraction from historical maps, given their superior performance over traditional convolutional neural network (CNN)-based architectures. In this research, we aim to automatically extract five different road types from historical maps, using a road dataset digitized from the scanned Deutsche Heereskarte 1:200,000 Türkei (DHK 200 Turkey) maps. We applied the variants of the transformer-based SegFormer model and evaluated the effects of different encoders, batch sizes, loss functions, optimizers, and augmentation techniques on road extraction performance. Our best results, with an intersection over union (IoU) of 0.5411 and an F1 score of 0.7017, were achieved using the SegFormer-B2 model, the Adam optimizer, and the focal loss function. All SegFormer-based experiments outperformed previously reported CNN-based segmentation models on the same dataset. In general, increasing the batch size and using larger SegFormer variants (from B0 to B2) resulted in improved accuracy metrics. Additionally, the choice of augmentation techniques significantly influenced the outcomes. Our results demonstrate that SegFormer models substantially enhance true positive predictions and resulted in higher precision metric values. These findings suggest that the output weights could be directly applied to transfer learning for similar historical maps and the inference of additional DHK maps, while offering a promising architecture for future road extraction studies.

Keywords: transformers; multiclass road segmentation; road type detection; historical maps

check for updates

Citation: Sertel, E.; Hucko, C.M.; Kabadayı, M.E. Automatic Road Extraction from Historical Maps Using Transformer-Based SegFormers. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 464. https://doi.org/10.3390/ijgi13120464

Academic Editors: Florian Hruby and Wolfgang Kainz

Received: 27 September 2024
Revised: 12 December 2024
Accepted: 18 December 2024
Published: 21 December 2024

Copyright: © 2024 by the authors. Published by MDPI on behalf of the International Society for Photogrammetry and Remote Sensing. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Digital map archives house extensive collections of scanned raster maps from various time periods. Historical maps and imagery are important geospatial data sources for various geography related applications and offer valuable insights into past land use and land cover and provide valuable geospatial information on transportation infrastructures and urban areas, enabling quantitative and spatial analysis [1–3]. These extracted features not only aid in understanding the evolution of road networks but also contribute to reconstructing past landscapes and urban fabric [3,4].

Extracting roads, settlements, and other geospatial features from historical maps is a labor-intensive task that demands considerable time and effort, especially when performed manually through digitization, as highlighted by [5]. This manual process often introduces inconsistencies due to human error and variations in expert interpretations [6–8]. Moreover, the diverse quality of historical maps, along with differences in scale and geographical coverage, further complicates the extraction process. The variability in map styles, legends,

and conventions across different time periods and sources adds an additional layer of complexity [8,9]. As the demand for more accurate, efficient, and scalable solutions increases, the integration of advanced technologies for the automatic identification of geospatial features has become essential.

Traditional approaches to analyze historical road networks have primarily relied on manual digitization and analytical modeling, which are often limited in scalability. For instance, ref. [10] employed complex network techniques to reconstruct the road systems of Bourbon Spain, applying network metrics to assess connectivity and structural properties. Ref. [11] utilized a sequential cost–benefit analysis to spatiotemporally reconstruct ancient road networks. They systematically evaluated economic and logistical factors that influenced road development over time. Building on the foundational methods, recent advancements in artificial intelligence have enabled the automated extraction of geospatial features from historical maps.

With the developments on geospatial artificial intelligence (GeoAI) approaches, historical maps have been widely used to automatically extract geospatial features including human settlements, road lines, and road intersection points [12–14]. The automatic extraction of roads from historical maps is a crucial task for comprehending past transportation conditions and conducting spatiotemporal analysis to unveil insights into the expansion of transportation lines and nearby settlement patterns. Ekim et al. [2] was one of the first studies to implement the use of deep neural networks to study historical maps for the semantic segmentation of five different road classes. Their study achieved the highest accuracy with the U-Net++ architecture combined with the ResNeXt50_32×4d encoder, yielding an overall accuracy of 98.73%, an intersection over union (IoU) of 41.99%, precision of 37.94%, recall of 58.78%, and an F1 score of 46.61%. Following this, ref. [14] explored several segmentation architectures, including FPN-Net, Ma-Net, U-Net++, and Deeplabv3, to automatically extract different road types from the same dataset. They assessed the performance of different encoders and batch sizes for the segmentation with U-Net++ architecture. By increasing the batch size and integrating the Split-Attention Network (Timm-ResNest200e) encoder, they significantly improved precision by 13.47%, recall by 10.92%, and the F1 score by 11.11%, surpassing the results of Ekim et al. [2] on the same dataset.

Ref. [15] presented a benchmark for the automatic vectorization of historical maps, focusing on Paris atlases from the nineteenth and early twentieth centuries. They introduced a supervised pipeline that combined deep edge filtering and watershed transform to extract closed shapes. They also contributed with an improved training method, joint optimization strategies, the evaluation of deep edge filters, and topology-preserving loss functions. These contributions led to an improvement in the PQ (panoptic quality) score, with the proposed training configuration achieving a COCO PQ score of 46.7%, compared to the original configuration's 34.8%.

Ref. [9] automatically generated training data by converting the available geospatial vector data into symbolic representations. This training data, along with historical maps, was then used as an input for a U-Net model. The authors employed U-Net as the core of a fast and effective deep convolutional neural network (CNN) approach to extract road networks from the late nineteenth- and early twentieth-century Swiss Siegfried maps. They further enhanced this approach by incorporating the same maps into their model development [16]. Ref. [17] used a region-based CNN with an Inception-ResNet v2 encoder to automatically identify road intersections from scanned historical United States Geological Survey (USGS) maps of several cities in the United States. They also highlighted that certain characteristics of the map tiles—such as edge density (where brightness changes significantly), unique red-green-blue (RGB) combinations, and sharpness (related to blurriness)—significantly influence the accuracy of object detection from these maps. In their review, ref. [8] highlighted various methods for road feature extraction, including line extraction, image filtering, color image segmentation, single line tracing, machine learning, and CNNs. They stated that CNNs can distinguish between various map feature types or between the same type of feature represented by various symbols. The review concluded

that road feature extraction from raster maps is a complex task that requires a combination of various techniques.

Ref. [18] proposed a semi-supervised road updating network (SRUNet) that integrates vector maps with satellite imagery, leading the model to understand the road features and facilitate road updating. SRUNet has three parts, which are the map encoding branch, the boundary enhancement module, and the residual refinement module. Their results showed that using historical vector maps improved the performance by 12.4%, superior to training with only satellite imagery.

Ref. [19] introduced using attention-based generative adversarial networks to generate realistic augmented versions of historical maps. Their study utilized a dataset containing seven road classes (cart roads, bridle paths, non-maintained roads, country roads, maintained roads, main roads, and footpaths) along with a background class. Then, they trained U-Net3+, enhanced by an attention mechanism using the augmented images. The training increased the F1 score from 45.70% to 61.68% and the IoU from 42.08% to 49.51%, showing an 18.26% and 7.62% increase, respectively, compared to the typical U-shaped network.

Recently, transformer-based architectures have become more applicable for accurate road extraction from remotely sensed images such as satellite or aerial images. Ref. [20] proposed the application of the dominant-transformer based global context modeling module in DT-RoadCDNet to bi-temporal remote sensing images for simultaneous road segmentation and change detection. Ref. [21] introduced a road segmentation network named DCS-TransUperNet that is built on the CSwin transformer architecture with dual resolution capabilities and applied this model to satellite images from Massachusetts and DeepGlobe datasets. Ref. [22] proposed the utilization of Swin transformers for enhancing contextual encoders in very-high-resolution (VHR) images, aiming to improve the accuracy and efficiency of the road extraction processes. Ref. [23] put forward a segmentation network to extract roads using transformer and CNN models with the DeepGlobe dataset.

While there has been considerable research on transformer-based road extraction from satellite images [20–26], there is limited work on applying transformer-based semantic segmentation architectures to historical maps. This study aims to address this gap in the literature by leveraging state-of-the-art transformer-based techniques to extract five different road types from historical maps and improve the accuracy metrics over those from traditional CNN approaches. We compare the performance of transformer-based segmentation methods with CNN-based methods for road extraction from historical maps. Specifically, by implementing the transformer-based SegFormer model variants with various encoders, batch sizes, and augmentation techniques, we aim to identify the most effective GeoAI solution for automatically segmenting different road types from historical maps. Our results demonstrate that the SegFormer model, when trained on historical maps, outperforms traditional CNN-based segmentation architectures, achieving superior performance compared to previous studies using the same dataset [2,14].

2. Dataset

UrbanOccupationsOETR_Deutsche_Heereskarte_200_Turkey, our publicly available dataset extracted from the DHK 200 Turkey map, covers a sizable area of around 150,000 square km in northwest Türkiye with diverse road types. This section of the map is scanned at 3000 dpi and covers approximately 65,000 × 18,000 pixels with a square grid size of 17 × 17 m. Detailed information on dataset preparation is available in [2]. This dataset includes 7076 historical map patches and their related digital ground truth masks for five distinct road types, with a size of 256 by 256 pixels. The DHK series has impressive coverage [27] including the entire of Europe, regions in Caucasia, and in North Africa using unified cartographical standards. The DHK was a military project prepared and used by the German Armed Forces (Wehrmacht) during World War II. The DHK 200 Turkey map was produced by the Main Survey Department (Hauptvermessungsabteilung) XIV in Vienna in 1942–1943. We obtained the full third special issue, version 6612, of the DHK Turkey from the Austrian Federal Office of Metrology and Surveying's (Bundesamt für

Eich- und Vermessungswesen, (BEV)) Cartography Department/Historical Map Archive in Vienna. This archive houses the records of Main Survey Department XIV. We are convinced that DHK is a good example of an early twentieth century historical pre-computer map production, due to its photo-mechanical, analogue, cartographic production methods compared to born digital maps emerging at the end of the twentieth century. The DHK is currently only partially available as there are sections of incomplete series in the original printed versions, or scanned copies thereof as jpeg or tiff files at public and university libraries, including collections acquired by the Allies of World War II during the war [28]. The reason the Austrian BEV has the complete DHK series in original printed version is due to an institutional continuity. The Main Survey Department XIV of the German Armed Forces, the producer of the DHK, was the official institutional predecessor of the Austrian BEV during World War II. The German Main Survey Department XIV also produced other DHK series for southeast Europe, such as the DHK Bulgaria. Furthermore, the DHK series for southeast Europe should be seen as a further development of an earlier Austro-Hungarian military cartography tradition and especially closely related to the 1:200,000 Map of Central Europe and the Balkans [29]. We are providing a section of an online available example sheet of DHK Turkey via McMaster University Digital Archive (http://digitalarchive.mcmaster.ca/islandora/object/macrepo:82386, accessed on 17 December 2024), used in this work in Figure 1 (sheet number D-V, edition 3). The notice on the left bottom corner on this sheet states that two other German maps of Turkey dated to 1928 and 1936 served as basemaps and the road network was improved from another German world map. To the best of our knowledge, none of the DHK sheets have yet been digitally or spatially transformed and therefore, they are not available in geo-rectified or semantically segmented formats.

The DHK was a cartographical work in progress throughout the tumultuous years of WWII. After June 1941, during the German Wehrmacht's invasion of the Soviet Union, the rapid advance of German forces led to the capture of materials from the USSR's geodetic and cartographic administrations. Among these were highly classified coordinate lists and new topographic maps. Upon examining the materials, the German Armed Forces were impressed by the exceptional quality of the Soviet work. This resulted in the direct integration of Soviet data into the German Military Survey (Deutsche Heeresvermessung) and the adaptation of war maps by aligning the Soviet system with the German unified system. Consequently, and exclusively for military purposes, Germany began using the Soviet system and designated it as the German Army Grid (Deutsches Heeresgitter). Since 1943, it served as the geometric foundation for all newly created DHK series [30]. This is why the German Army Grid used the Russian Gauss–Krüger with 6 degree zones and the geodetic datums of Rauenberg, RDN-1940, and DHDN77 [31]. In our previous work [2], we used Pulkovo datum for georeferencing the DHK Turkey, which could be revisited for georectification accuracy. However, we would like to stress that both in [2] and here we are primarily striving for higher efficiency for our feature extraction applications.

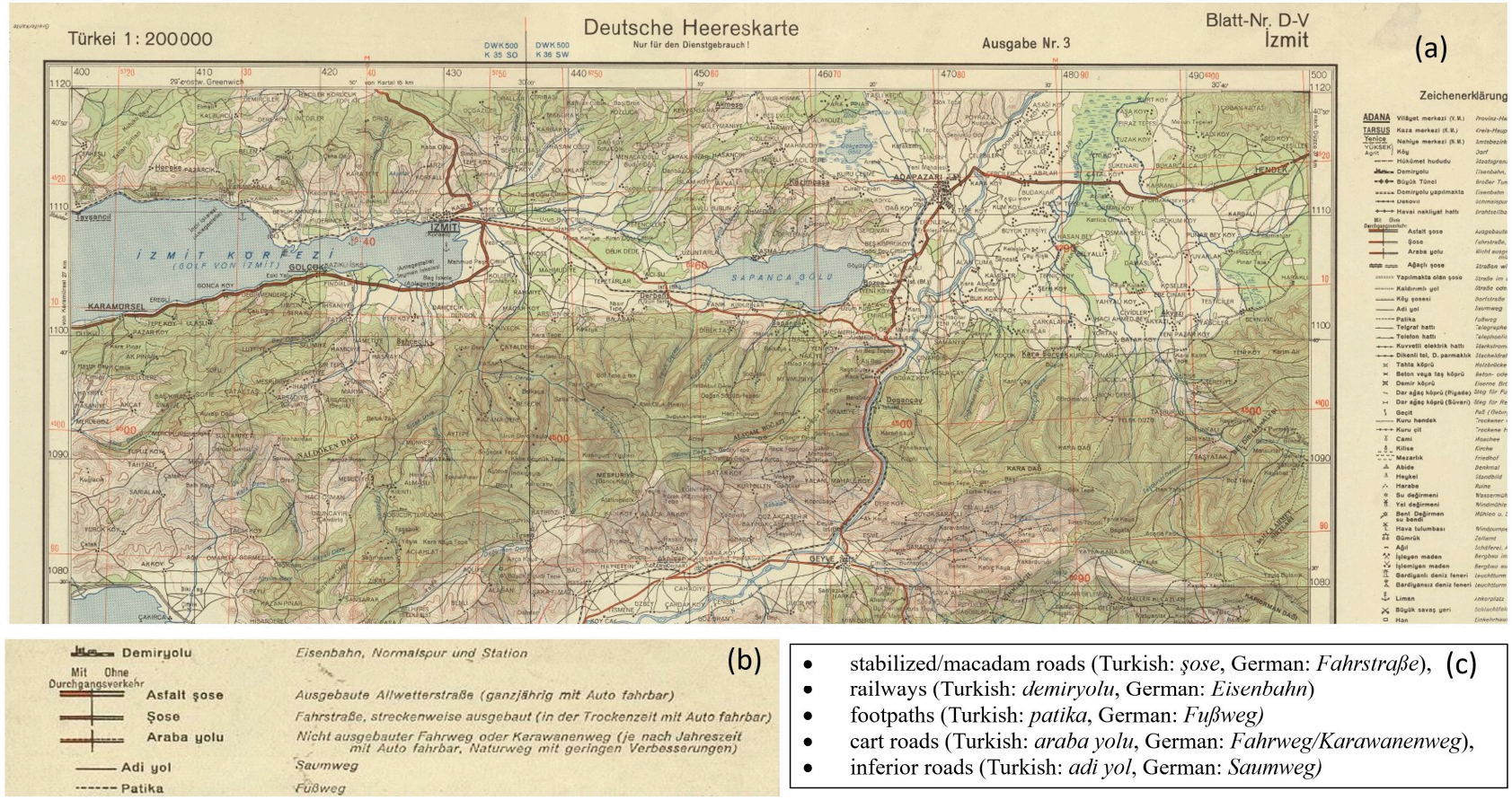


Figure 1. (**a**) An example of the DHK 200 Map Sheet D-V for İzmit, (**b**) part of the legend focusing on road types, (**c**) selected road type explanations with our English translations (Source: https://digitalarchive.mcmaster.ca/ accessed on 17 December 2024).

3. Methodology

We used SegFormer [32] to extract roads from the DHK 200 Turkey historical map. Transformer-based methods use a self-attention module to determine the associations between each pixel on a feature map.

3.1. SegFormer

SegFormer is a strong and effective framework for semantic segmentation that combines lightweight multi-layer perceptron (MLP) decoders with transformers. SegFormer, in contrast to conventional CNN-based methods, uses transformers to capture long-range dependencies between pixels, which enables it to effectively handle challenging segmentation problems.

The hierarchically structured transformer encoder within the framework produces multiscale features, improving its capacity to extract intricate details from images [32]. SegFormer includes two primary components:

1. A hierarchical transformer encoder designed to produce coarse features at high resolution and fine features at low resolution
2. A lightweight all-MLP decoder that integrates these multiscale features to generate the ultimate semantic segmentation mask.

SegFormer-B0 is the smallest variant in the SegFormer family, featuring fewer parameters, which makes it faster and less computationally demanding. The SegFormer-B1 model is slightly larger than B0, with more parameters and layers, leading to enhanced performance. SegFormer-B2, the largest variant, incorporates the most parameters and layers, offering the highest accuracy but requiring significantly more computational resources for both training and inference.

3.2. Implementation

The imbalance of the dataset forced us to implement various techniques such as sampling strategies or loss functions that prioritize underrepresented classes, enhancing the model's ability to learn from less frequent data samples and thus improving overall accuracy. We used a weighted random sampler and focal loss [33] $\alpha = 0.25$ and $\gamma = 2$ to overcome the class imbalance.

Another challenge to address was the difference in the shapes of the ground truth (GT) mask and the segmentation output mask. In order to input the output mask and GT into the loss function, both must be of the same size. Non-differentiable methods such as interpolating the output mask to the size of GT would not work since the loss function requires the ability to compute gradients for backpropagation. On the other hand, while downscaling the GT to match the output mask's size through interpolation is a viable approach, this method fails to preserve the labels on both smooth and sharp turns in the roads because of the narrow widths of road class labels. Finally, to downscale GT while preserving all the labels, we used 2D Max Pooling approach with stride.

After overcoming the challenges of the imbalance of data and the difference of shapes, we used AdamW (Adam with decoupled weight decay) [34] with $\gamma_{\text{learning rate}} = 10^{-4}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 10^{-8}$ as an optimizer. We used sixteen and eight batch sizes. We trained models until the validation loss no longer decreased for 10 epochs.

3.3. Evaluation

We conducted both statistical and visual analysis to evaluate the performance of the implemented models. We employed the common evaluation metrics used in the semantic segmentation task, namely, precision, recall, F1 score, and intersection over union (IoU). Since we have multiple road classes in our work, we also generated a confusion matrix to analyze class-wise accuracy metrics.

Precision (P) measures the proportion of pixels correctly identified as belonging to a specific label out of all the pixels labeled as that class. It shows how many of the pixels

predicted by the model to be part of a certain class are actually correct. A high precision rate indicates that the model has a low rate of false positives.

On the other hand, recall (R) measures the proportion of actual pixels belonging to a specific class that were correctly identified by the model. It assesses the model's ability to find all the relevant instances of a class within an image. A high recall rate indicates that the model has a low rate of false negatives. Both precision and recall are crucial for understanding the model's performance, especially in applications where the balance between missing a relevant pixel (high recall) and ensuring identified pixels are correct (high precision) is essential.

F1 score is a single metric to assess the balance between precision and recall. It is calculated by the harmonic mean of precision and recall.

IoU shows the overlap between the predicted mask and the ground truth mask, and it is a common evaluation metric used in road extraction. IoU is calculated by dividing the area of overlap between the predicted and ground truth mask by the area covered by both the predicted and ground truth mask.

A confusion matrix is a detailed breakdown of the model's predictions across different classes. It shows how many pixels of each actual class were correctly or incorrectly classified into each predicted class. The diagonal of the matrix represents correctly classified pixels (true positives), and off-diagonal elements indicate errors or confusions between classes.

4. Results

We trained the SegFormer model with different sizes of the mix transformer (MiT) encoders, specifically MiT-B0, MiT-B1, and MiT-B2, using historical maps and corresponding road type masks as input. We generated model weights for our application. Initially, we trained the SegFormer with MiT-B0 (SegFormer-B0) using a batch size of 16 (experiment 1 in Table 1) and applied the same augmentations as in the previous work [14]: horizontal flip (HF), shift (SH), scale (SC), rotate (RO), perspective (PE), and Gaussian noise (GN). We referred to these augmentations collectively as "Old" and used "New" to refer to SH, SC, and GN only. Comparing our first experiment (experiment 1 in Table 2) with the previous best work (experiment 15 in Table 2), which used the Split-Attention Network (Timm-ResNest200e) encoder and U-Net++ architecture on the same dataset, we achieved better scores across all accuracy metrics, indicating the superiority of the SegFormer models.

Table 1. Experimental designs.

No	Model	Opt/Loss	Augmentations	Batch Size
1	SegFormer-B0	Adam/Dice	Old	16
2	SegFormer-B0	Adam/Focal	Old	16
3	SegFormer-B0	AdamW/Focal	Old	16
4	SegFormer-B0	AdamW/Focal	No	16
5	SegFormer-B0	Adam/Dice	New	16
6	SegFormer-B0	AdamW/Focal	New	16
7	SegFormer-B0	AdamW/Dice	New	16
8	SegFormer-B0	Adam/Focal	New	16
9	SegFormer-B0	AdamW/Focal	New	8
10	SegFormer-B1	AdamW/Focal	New	16
11	SegFormer-B1	Adam/Focal	New	16
12	SegFormer-B1	AdamW/Focal	New	8
13	SegFormer-B2	AdamW/Focal	New	8
14	SegFormer-B2	Adam/Focal	New	8
15	U-Net++-TIMM [14]	Adam/Dice	Old	16

We then used the same configuration without any augmentation techniques (experiment 4 in Tables 1 and 2), which yielded better results than its counterpart experiment with "Old" augmentations (experiment 3 in Tables 1 and 2). By trying different augmentation combinations, we achieved better accuracy metrics using SH, SC, and GN augmentation

techniques, as evidenced by the improvement in results between experiments 1 and 5 (Tables 1 and 2).

Table 2. Accuracy metrics for different experiments.

No	Precision	Recall	F1-Score	IoU
1	0.6031	0.7043	0.6490	0.4807
2	0.6876	0.5579	0.6158	0.4455
3	0.6605	0.5932	0.6242	0.4547
4	0.6751	0.6780	0.6756	0.5110
5	0.6293	0.7203	0.6697	0.5053
6	0.7057	0.6675	0.6853	0.5216
7	0.6668	0.6901	0.6780	0.5136
8	0.6966	0.6831	0.6889	0.5259
9	0.7015	0.6434	0.6717	0.5067
10	0.6928	0.6886	0.6905	0.5279
11	0.6822	0.6956	0.6878	0.5251
12	0.6842	0.6485	0.6632	0.4988
13	0.6893	0.6960	0.6920	0.5297
14	0.7061	0.6974	0.7017	0.5411
15	0.5141	0.6970	0.5772	0.4199

Our results indicate that all transformer model configurations outperform the best-performing CNN-based U-Net++ model (experiment 14) used in previous works. For instance, the least complex model, SegFormer-B0, achieved an F1 score of 0.6490 and an IoU of 0.48077 (experiment 1, Table 2) when using the same augmentations, optimizer, and loss function as the previous work, which obtained an F1 score of 0.5272 and an IoU of 0.4199 (Table 2).

We conducted a series of experiments using the SegFormer-B0 architecture (experiments 1 to 9) to investigate the effects of different optimizers, loss functions, and augmentation techniques. The highest F1 score and IoU values were achieved in experiments 8 and 6, respectively. Both experiments employed new augmentation techniques and focal loss, but experiment 8 used the Adam optimizer, while experiment 6 utilized AdamW.

Progressing to more advanced models, the SegFormer-B1 model achieved better performance, with an F1 score of 0.6905 and an IoU of 0.5279 (experiment 10, Table 2). In this case, the experiment with the AdamW optimizer produced slightly higher results than the experiment with Adam (experiment 11). The most advanced model tested, SegFormer-B2, achieved an F1 Score of 0.6920 and an IoU of 0.5297 for experiment 13 and an F1 Score of 0.7017 and an IoU of 0.5411 for experiment 14 despite a reduced batch size of eight due to hardware limitations. The best performing model was SegFormer-B2 with the Adam optimizer and the focal loss. To conduct a fair comparison among different encoders for SegFormer, we trained the SegFormer-B0 and SegFormer-B1 with a batch size of eight (experiments 5 and 9 in Tables 1 and 2). Increasing the batch size from eight to sixteen improved the results. For example, in the case of SegFormer-B0, experiments 9 and 6 are identical except for their batch sizes. The experiment with a batch size of eight (experiment 9) produced lower metric values, with an F1 score of 0.6717 and an IoU of 0.5067, compared to the experiment with a batch size of sixteen (experiment 6), which achieved an F1 score of 0.6853 and an IoU of 0.5216. This trend is also evident in the SegFormer-B1 experiments with different batch sizes, as seen in experiments 10 and 12 in Tables 1 and 2.

While factors like a larger batch size, new augmentation methods, and focal loss played a role in improving accuracy metrics, the comparison between the Adam and AdamW optimizers revealed that the Adam optimizer generally resulted in better accuracy across most cases.

Figure 2 shows the visual results of the seven best-performing experiments in terms of F1 score and IoU (experiments 1, 6, 8, 10, 11, 13, and 14), along with the previous best-performing model (experiment 15). These are presented alongside the original input image

and the GT for comparison. For the first patch, there is a short, stabilized road crossing the railway, as can be seen in the original image but not in the GT; however, experiments 6, 11, and 15 can successfully extract both the railway and the stabilized road. The results of experiment 15 illustrate that there is a mixture of stabilized roads with the coordinate grids, as can be seen in patches 6 and 8. However, for SegFormer-based models, this issue does not exist. In most cases, the best visual results were produced with experiment 14, such as patches 2, 4, 6, and 7. In some cases, all models generated good predictions, such as patches 3, 8, and 9.

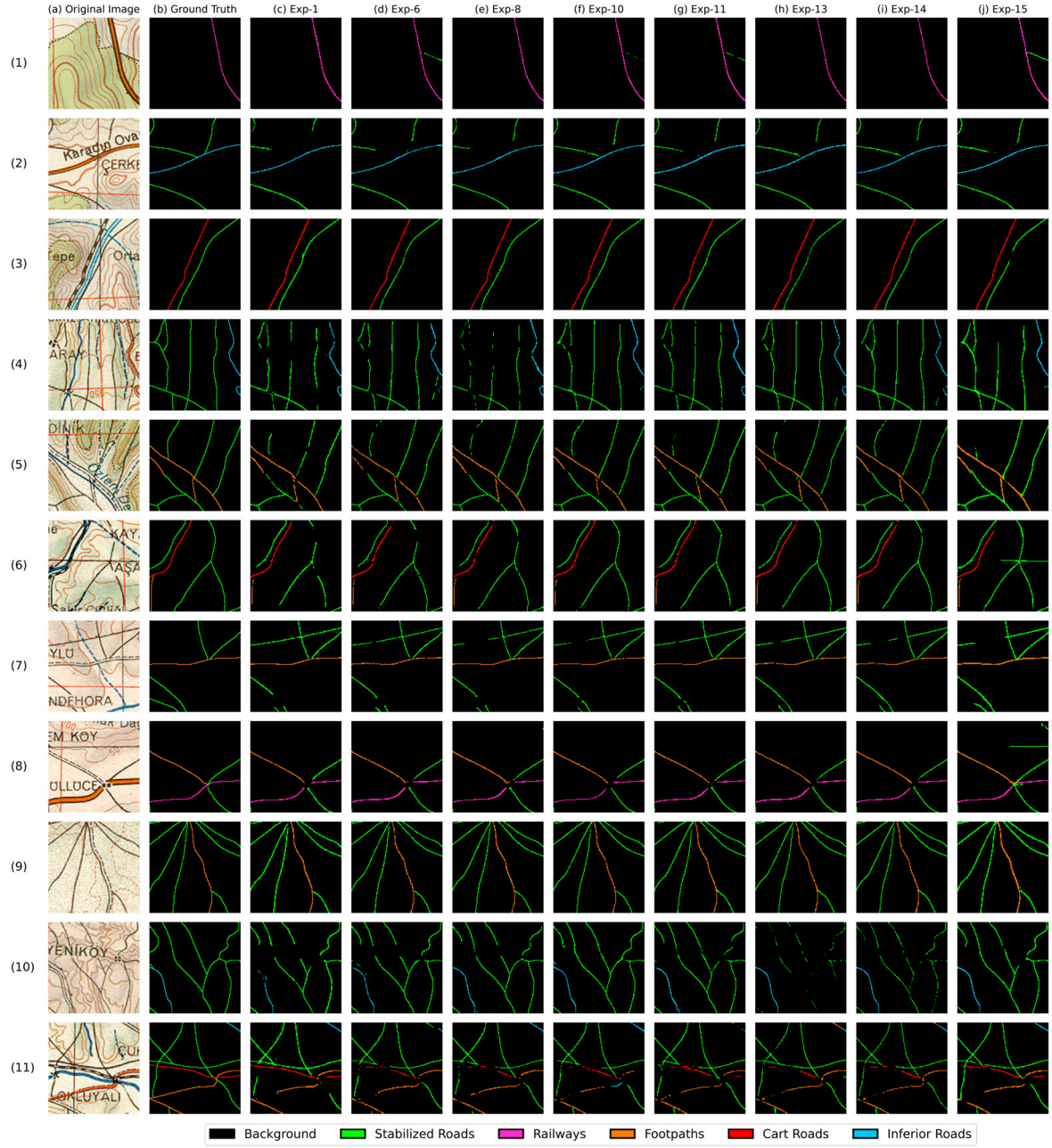


Figure 2. Predictions on the test set of DHK 200 Turkey map. (**a**) Input image. (**b**) Ground truth. (**c**) Experiment 1 (B0-Adam-Dice-Old). (**d**) Experiment 6 (B0-AdamW-Focal-New). (**e**) Experiment 8 (B0-Adam-Focal-New). (**f**) Experiment 10 (B1-AdamW-Focal-New). (**g**) Experiment 11 (B1-Adam-Focal-New). (**h**) Experiment 13 (B1-AdamW-Focal-New). (**i**) Experiment 14 (B1-Adam-Focal-New). (**j**) Experiment 15 (U-Net++-TIMM-Adam-Dice-Old) [11].

The class-wise results for the best performing model (experiment 14) provide insightful information regarding its performance on different road types. Analyzing the metrics, we notice a disparity in performance across classes, suggesting that the model has varying degrees of success for different categories of roads.

The stabilized roads class has the highest F1 score of 0.7341 and an IoU of 0.5799. Similarly, the cart roads class has an F1 score of 0.7155 and an IoU of 0.5570, indicating a reliable detection rate. In contrast, the railways, footpaths, and inferior roads classes show lower performance metrics. The inferior roads class has the lowest F1 score of 0.6598 and an IoU of 0.4923 among all classes.

The class-wise evaluation (Table 3 and Figure 3) across the five best experiments highlights that the stabilized roads are most accurately detected, with F1 scores ranging from 0.724 to 0.734. Railways and footpaths show similar scores, with most of the models falling into the 0.67–0.71 range. Cart roads have the biggest variability among the models, with scores between 0.671 (B0-Adam-Focal) and 0.715 (B2-Adam-Focal), suggesting that the experiment configuration has significant effects for the detection of this road type. On the other hand, inferior roads presented the lowest F1 scores, varying between 0.633 and 0.662, indicating the identification of this road type is the most challenging one for this dataset.

Table 3. Class-wise metrics of the best model.

Class	F1 Score	Precision	Recall	IoU
Stabilized Roads	0.7341	0.7413	0.7269	0.5799
Railways	0.6848	0.6862	0.6833	0.5207
Footpaths	0.7144	0.7121	0.7167	0.5557
Cart Roads	0.7155	0.7246	0.7065	0.5570
Inferior Roads	0.6598	0.6663	0.6534	0.4923
Overall	0.7017	0.7061	0.6974	0.5411

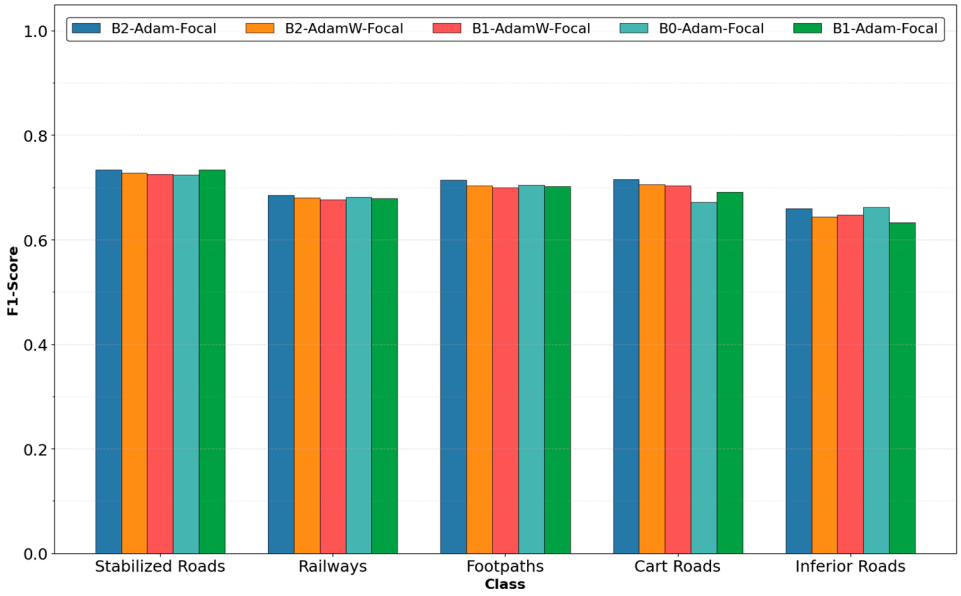


Figure 3. Class-wise F1-scores of best five experiments.

The class-wise IoU scores (Figure 4) follow a similar but with a lower pattern than the F1 scores, which is expected because of the nature of the metric itself. Stabilized roads again performed best, ranging between 0.567 and 0.579 with the best model (B2-Adam-Focal). Railways and footpaths show IoUs mostly in the 0.510–0.555 range, remaining consistent with the F1 score results. Cart roads again show more variability than other classes, showing results between 0.505 (B0-Adam-Focal) and 0.556 (B2-Adam-Focal). Inferior roads have the

lowest IoU scores, suggesting that the category is the hardest one for models to learn in this dataset.

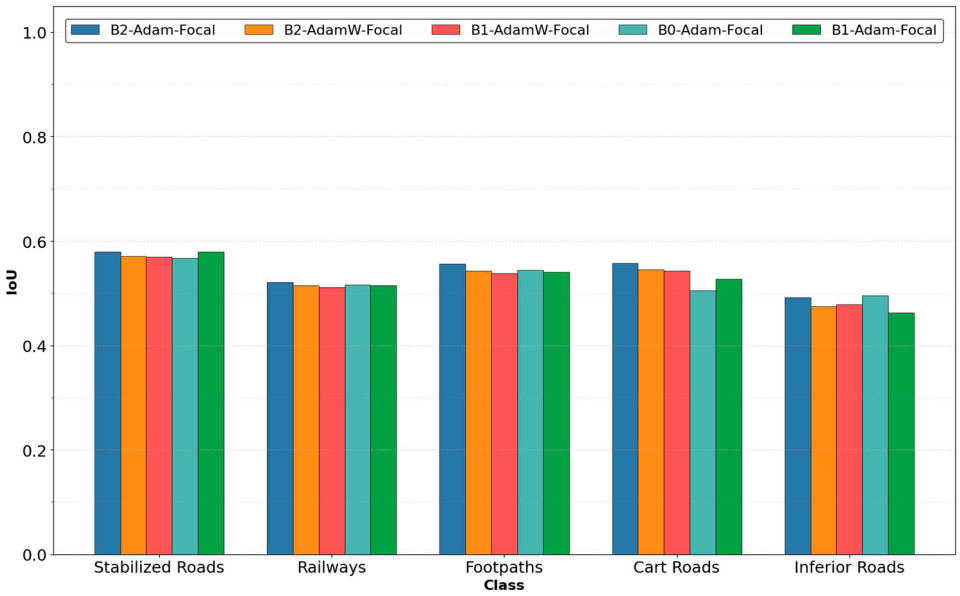


Figure 4. Class-wise IoUs of best five experiments.

5. Discussion

Road extraction from historical maps is crucial for understanding the development and evolution of transportation networks and urban landscapes over time. High-quality, well-annotated datasets are essential for achieving superior segmentation performance. Techniques like data augmentation, which increase training diversity and model robustness, have shown significant improvements in this area [35]. Recent advancements have integrated deep learning approaches to further enhance road extraction from historical maps. For example, ref. [36] introduced a probabilistic road classification framework, using deep learning models trained on synthetic data to accurately differentiate five road classes: walking path, dirt road or mule track, driveway without reinforcement, reinforced road (3–5 m wide), and reinforced road (wider than 5 m). This framework improved the accuracy of road network reconstructions.

Our dataset includes five standard road classes—stabilized roads, inferior roads, cart roads, footpaths, and railways—that provide insights into the historical evolution of transportation, especially when compared to current conditions. The high-resolution dataset used in this study is key to advancing automatic road extraction. Its unique characteristics, such as the diverse road classes and digitalization from the original DHK 200 Turkey maps, allow models to learn from a variety of geospatial features. Training models with this high-quality dataset enhances their ability to generalize across different map styles and types, supporting transfer learning for future research.

The dataset used in this research was previously applied to CNN-based segmentation architectures, including DeepLabV3, U-Net+, FPN-Net, and Ma-Net, as reported by [2,14]. The best IoU value achieved in those studies was 0.4199. In contrast, by implementing various encoder combinations of the SegFormer model, we achieved an IoU of 0.5411, improving the metric by 0.1212. As shown in Table 2, all SegFormer-based experiments, with different encoder, optimizer, and/or loss function combinations, outperformed the CNN-based approaches presented by [2,14] in terms of accuracy. While the recall values across SegFormer-based models were similar to those reported by [2,14], there was a significant improvement in precision, which increased from 0.5141 to 0.7061.

Tao et al. [23] proposed a transformer-based semantic segmentation model, Seg-Road, for road extraction from satellite images. They evaluated their model on the DeepGlobe and Massachusetts remote sensing datasets and compared its performance against CNN-based

models such as D-LinkNet, PSPNet, and CoANet. Similar to our work, their approach demonstrated superior performance using a transformer-based architecture. Specifically, they achieved a 3.84% increase in IoU on the DeepGlobe dataset and a 3.42% increase on the Massachusetts dataset, compared to CNN-based models. In a related study, Liu et al. [24] adopted another popular transformer approach, the Swin transformer, as the backbone for road extraction on the same datasets. Their proposed model, RoadFormer, achieved a 4.97% IoU improvement using the Swin-B backbone over ResNet-50. Both of these studies highlight the effectiveness of transformer-based architectures for road extraction from satellite imagery. In our research, we apply a similar transformer-based semantic segmentation approach to identify multi-class roads from historical maps, a valuable source of geospatial information, though one that presents unique challenges due to issues with map quality, symbolic representation, and edge definitions.

6. Conclusions

While transformer-based segmentation methods have been predominantly applied to satellite images, there is a growing need to leverage these approaches for geospatial object extraction from historical maps due to their superior performance over traditional convolutional neural network-based architectures. In this study, we used the transformer-based SegFormer model variants to address the historical road extraction problem. By experimenting with different encoders, batch sizes, and augmentation techniques, the research aimed to optimize the SegFormer model's performance for segmenting road types from historical maps.

In general, advancing the SegFormer model, increasing the batch size, and using focal loss and the proposed "new" augmentation configuration produced better results. However, for the optimizer, in some cases, Adam, but in some cases, AdamW, achieved higher accuracy. Therefore, both of these optimizers should be considered for future experiments on similar segmentation tasks. The results demonstrated that the SegFormer models, particularly Segformer-B2, outperformed previous CNN-based architectures in terms of accuracy metrics, achieving an F1 score of 0.7017 and an IoU of 0.5411. Based on the trend that increased model complexity leads to higher accuracy, further experiments with more advanced models like SegFormer B3, B4, and B5 may result in even better accuracy.

This study highlights the significant contribution of transformer-based architectures in enhancing the performance of road extraction from historical maps. The superior performance of the SegFormer model over CNN-based methods shows the potential for understanding complex patterns and long-range dependencies found in historical maps. Transformers especially excel at capturing global context and modeling spatial relationships, which are needed when dealing with historical maps, because of the cartographic symbolizations and possible degradations in map image quality. With these strengths, our approach showed that the transformer-based models can overcome the limitations of CNN architectures in this context.

In conclusion, our research findings underscored the potential of transformer-based approaches for accurate road extraction from historical maps and provided valuable insights for future studies in this domain. By extending the application of transformer architectures from satellite imagery to historical cartography, this approach offers a solution to accurate and reliable road extraction from historical maps. The results both validate the applicability of transformer-based deep learning models in this context and provide a foundation for future research in geospatial data processing and analysis

Author Contributions: Conceptualization, Elif Sertel, Can Michael Hucko, and Mustafa Erdem Kabadayı; methodology, Elif Sertel and Can Michael Hucko; software, Can Michael Hucko; validation, Elif Sertel and Can Michael Hucko; formal analysis, Can Michael Hucko; investigation, Mustafa Erdem Kabadayı; resources, Elif Sertel and Mustafa Erdem Kabadayı; data curation, Can Michael Hucko; writing—original draft, Elif Sertel and Can Michael Hucko; writing—review and editing, Elif Sertel and Mustafa Erdem Kabadayı; visualization, Can Michael Hucko; Supervision, Elif Sertel;

project administration, Mustafa Erdem Kabadayı; funding acquisition, Mustafa Erdem Kabadayı. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the European Research Council (ERC) projects: "A GeoAI-based Land Use Land Cover Segmentation Process to Analyse and Predict Rural Depopulation, Agricultural Land Abandonment, and Deforestation in Bulgaria and Turkey, 1940–2040" Grant Agreement No. 101100837 and "Industrialisation and Urban Growth from the mid-nineteenth century Ottoman Empire to Contemporary Turkey in a Comparative Perspective, 1850–2000" Grant Agreement No. 679097. M. Erdem Kabadayı was the principal investigator of both projects.

Data Availability Statement: The data that support the findings of this study are openly available at https://urbanoccupations.ku.edu.tr/historical-road-types-for-turkey-1940s and https://github.com/RSandAI/HistRoadSegformer (accessed on 17 December 2024).

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

References

1. Chiang, Y.Y.; Duan, W.; Leyk, S.; Uhl, J.H.; Knoblock, C.A. Training deep learning models for geographic feature recognition from historical maps. In *Using Historical Maps in Scientific Studies*; SpringerBriefs in Geography; Springer International Publishing: Cham, Switzerland, 2020; pp. 65–98. Available online: http://link.springer.com/10.1007/978-3-319-66908-3_4 (accessed on 22 September 2024).
2. Ekim, B.; Sertel, E.; Kabadayı, 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]
3. Sertel, E.; Kabadayi, M.E.; Sengul, G.S.; Tumer, I.N. HexaLCSeg: A historical benchmark dataset from Hexagon satellite images for land cover segmentation [Software and Data Sets]. *IEEE Geosci. Remote Sens. Mag.* **2024**, *12*, 197–206. [CrossRef]
4. Uhl, J.H.; Leyk, S.; Chiang, Y.Y.; Knoblock, C.A. Towards the automated large-scale reconstruction of past road networks from historical maps. *arXiv* **2022**, arXiv:220204883. [CrossRef]
5. Turner, H.; Lahoorpoor, B.; Levinson, D.M. Creating a dataset of historic roads in Sydney from scanned maps. *Sci. Data* **2023**, *10*, 683. [CrossRef]
6. Laycock, S.D.; Brown, P.G.; Laycock, R.G.; Day, A.M. Aligning archive maps and extracting footprints for analysis of historic urban environments. *Comput. Graph.* **2011**, *35*, 242–249. [CrossRef]
7. 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]
8. Jiao, C.; Heitzler, M.; Hurni, L. A survey of road feature extraction methods from raster maps. *Trans. GIS* **2021**, *25*, 2734–2763. [CrossRef]
9. Jiao, C.; Heitzler, M.; Hurni, L. A fast and effective deep learning approach for road extraction from historical maps by automatically generating training data with symbol reconstruction. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *113*, 102980. [CrossRef]
10. Pablo-Martí, F.; Alañón-Pardo, Á.; Sánchez, A. Complex networks to understand the past: The case of roads in Bourbon Spain. *Cliometrica* **2021**, *15*, 477–534. [CrossRef]
11. Stahlberg, M.J.; Sagnol, G.; Ducke, B.; Klimm, M. Spatiotemporal reconstruction of ancient road networks through sequential cost–benefit analysis. *PNAS Nexus.* **2023**, *2*, pgac313. [CrossRef]
12. Chiang, Y.Y.; Chen, M.; Duan, W.; Kim, J.; Knoblock, C.A.; Leyk, S.; Li, Z.; Lin, Y.; Namgung, M.; Shbita, B.; et al. GeoAI for the digitization of historical maps. In *Handbook of Geospatial Artificial Intelligence*, 1st ed.; CRC Press: Boca Raton, FL, USA, 2023; pp. 217–247. Available online: https://www.taylorfrancis.com/books/9781003308423/chapters/10.1201/9781003308423-11 (accessed on 22 September 2024).
13. Uhl, J.H.; Leyk, S.; Chiang, Y.Y.; Duan, 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]
14. Avci, C.; Sertel, E.; Kabadayi, M.E. Deep learning-based road extraction from historical maps. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
15. Chen, Y.; Chazalon, J.; Carlinet, E.; Ôn Vũ Ngoc, M.; Mallet, C.; Perret, J. Automatic vectorization of historical maps: A benchmark. *PLoS ONE* **2024**, *19*, e0298217. [CrossRef] [PubMed]
16. Jiao, C.; Heitzler, M.; Hurni, L. A novel framework for road vectorization and classification from historical maps based on deep learning and symbol painting. *Comput. Environ. Urban Syst.* **2024**, *108*, 102060. [CrossRef]
17. 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]
18. Chen, X.; Yu, A.; Sun, Q.; Guo, W.; Xu, Q.; Wen, B. Updating road maps at city scale with remote sensed images and existing vector maps. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–21. [CrossRef]
19. Zhao, Y.; Wang, G.; Yang, J.; Li, T.; Li, Z. AU3-GAN: A method for extracting roads from historical maps based on an attention generative adversarial network. *J. Geovisualization Spat. Anal.* **2024**, *8*, 26. [CrossRef]

20. Zhou, M.; Sui, H.; Chen, S.; Chen, X.; Wang, W.; Wang, J.; Liu, J. UGRoadUpd: An unchanged-guided historical road database updating framework based on bi-temporal remote sensing images. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 21465–21477. [CrossRef]
21. Zhang, Z.; Miao, C.; Liu, C.; Tian, Q. DCS-TransUperNet: Road segmentation network based on CSwin transformer with dual resolution. *Appl. Sci.* **2022**, *12*, 3511. [CrossRef]
22. Chen, T.; Jiang, D.; Li, R. Swin transformers make strong contextual encoders for VHR image road extraction. In Proceedings of the IGARSS 2022–2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; pp. 3019–3022. Available online: https://ieeexplore.ieee.org/document/9883628/ (accessed on 22 September 2024).
23. Tao, J.; Chen, Z.; Sun, Z.; Guo, H.; Leng, B.; Yu, Z.; Wang, Y.; He, Z.; Lei, X.; Yang, J. Seg-Road: A segmentation network for road extraction based on transformer and CNN with connectivity structures. *Remote Sens.* **2023**, *15*, 1602. [CrossRef]
24. Liu, X.; Wang, Z.; Wan, J.; Zhang, J.; Xi, Y.; Liu, R.; Miao, Q. RoadFormer: Road extraction using a swin transformer combined with a spatial and channel separable convolution. *Remote Sens.* **2023**, *15*, 1049. [CrossRef]
25. Gao, L.; Zhou, Y.; Tian, J.; Cai, W. DDCTNet: A deformable and dynamic cross-transformer network for road extraction from high-resolution remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–19. [CrossRef]
26. Chen, T.; Liu, Y.; Jiang, H.; Li, R. Swin transformer coupling CNNs makes strong contextual encoders for VHR image road extraction. *Int. J. Remote. Sens.* **2023**, *44*, 5788–5813.
27. Scharfe, W. German Army Map of Spain 1:50.000: 1940–1944. In Proceedings of the 21st International Cartographic Conference Cartographic Renaissance, Durban, South Africa, 10–16 August 2003; pp. 2475–2495.
28. Powell, S.; Mühr, H. Capturing the complex histories of German World War II captured maps. *J. Map Geogr. Libr.* **2021**, *16*, 166–193. [CrossRef]
29. Arnaud, J.L. The long life of a 1:200,000 map of Central Europe and the Balkans. In *History of Cartography: International Symposium of the ICA, 2012 [Internet]*; Liebenberg, E., Collier, P., Török, Z.G., Eds.; Springer: Berlin/Heidelberg, Germany, 2014; pp. 111–128. [CrossRef]
30. Buchroithner, M.F.; Pfahlbusch, R. Geodetic grids in authoritative maps–new findings about the origin of the UTM Grid. *Cartogr. Geogr. Inf. Sci.* **2017**, *44*, 186–200. [CrossRef]
31. Timár, G.; Lévai, P.; Molnár, G.; Varga, J. A második világháború német katonai térképeinek koordinátarendszere [GIS Integration of the German Army Grid (DHG) and Its Geodetic Datums]. *Geodézia És Kartográfia* **2004**, *56*, 28–35.
32. Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Alvarez, J.M.; Luo, P. SegFormer: Simple and efficient design for semantic segmentation with transformers [Internet]. *arXiv* **2021**, arXiv:2105.15203.
33. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal loss for dense object detection. *arXiv* **2018**, arXiv:1708.02002.
34. Loshchilov, I.; Hutter, F. Decoupled weight decay regularization [Internet]. *arXiv* **2019**, arXiv:1711.05101.
35. Jiao, C.; Heitzler, M.; Hurni, L. A novel data augmentation method to enhance the training dataset for road extraction from Swiss historical maps. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2022**, *2*, 423–429. [CrossRef]
36. Mühlematter, D.J.; Schweizer, S.; Jiao, C.; Xia, X.; Heitzler, M.; Hurni, L. Probabilistic road classification in historical maps using synthetic data and deep learning [Internet]. *arXiv* **2024**, arXiv:2410.02250.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.