*remote sensing*

MDPI

*Review*

# Deep Learning Approaches Applied to Remote Sensing Datasets for Road Extraction: A State-Of-The-Art Review

**Abolfazl Abdollahi [1], Biswajeet Pradhan [1,2,*], Nagesh Shukla [1], Subrata Chakraborty [1] and Abdullah Alamri [3]**

1 The Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), School of Information, Systems and Modelling, University of Technology Sydney, Sydney 2007, Australia; abolfazl.abdollahi@student.uts.edu.au (A.A.); Nagesh.Shukla@uts.edu.au (N.S.); subrata.chakraborty@uts.edu.au (S.C.)

2 Department of Energy and Mineral Resources Engineering, Sejong University, Choongmu-gwan, 209 Neungdong-ro, Gwangjingu, Seoul 05006, Korea

3 Department of Geology & Geophysics, College of Science, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia; amsamri@ksu.edu.sa

* Correspondence: Biswajeet.Pradhan@uts.edu.au; Tel.: +61-2-9514-7937

Received: 24 March 2020; Accepted: 30 April 2020; Published: 2 May 2020


**Abstract:** One of the most challenging research subjects in remote sensing is feature extraction, such as road features, from remote sensing images. Such an extraction influences multiple scenes, including map updating, traffic management, emergency tasks, road monitoring, and others. Therefore, a systematic review of deep learning techniques applied to common remote sensing benchmarks for road extraction is conducted in this study. The research is conducted based on four main types of deep learning methods, namely, the GANs model, deconvolutional networks, FCNs, and patch-based CNNs models. We also compare these various deep learning models applied to remote sensing datasets to show which method performs well in extracting road parts from high-resolution remote sensing images. Moreover, we describe future research directions and research gaps. Results indicate that the largest reported performance record is related to the deconvolutional nets applied to remote sensing images, and the F1 score metric of the generative adversarial network model, DenseNet method, and FCN-32 applied to UAV and Google Earth images are high: 96.08%, 95.72%, and 94.59%, respectively.

**Keywords:** road extraction; common benchmarks; machine learning; deep learning; remote sensing

## 1. Introduction

Spaceborne, airborne, and drone-based sensors using advanced Earth observation and remote sensing technologies have obtained large amounts and different types of high-resolution images. Such images are extensively used in several applications, such as urban planning [1], disaster management [2], and emergency tasks [3]. Among topographic object classes, road objects are essential urban features. Therefore, the constant updating of road databases is necessary to achieve several geospatial information systems (GIS) goals, such as emergency functions, automated means of navigation, urban planning, and traffic control [4]. A road database can be created and updated using feature extraction from spatial high-resolution satellite imagery [5]. Consequently, generating automatic novel techniques for extracting road classes from high-resolution satellite images and keeping road networks up-to-date in GIS databases are useful for a variety of applications [6]. High-resolution remote sensing imagery can produce a massive amount of data and has become the main data source

for extracting road regions and updating geospatial databases in real time [7]. Although road extraction from remote sensing imagery recently gained considerable attention, this task remains challenging owing to irregular and complex road sections and structures [8]. Other features, such as building roofs, pedestrian areas, and car parking appear similar in satellite images, thereby resulting in insufficient road contexts in images [9]. Meanwhile, roadside buildings, tree shadows, and vehicles on roads can be identified from high-resolution remotely sensed imagery [10]. Given the aforementioned issues, road class extraction from high-resolution remotely sensed imagery is difficult. Manual and traditional approaches for road extraction from high-resolution remote sensing imagery are costly, time consuming, and fraught with errors owing to human operators [11]. Therefore, various road extraction approaches, such as supervised [12] and unsupervised [13] techniques, were suggested for extracting road regions from remotely sensed imagery. Such approaches use textural [14], geometric, and photometric [15] information to extract roads through classification [16]. Techniques for road extraction can be categorized into two categories: (1) automatic and semiautomatic approaches and (2) road area and centerline extraction methods. Automatic techniques are useful in real-time applications and do not require human collaboration, unlike semiautomatic approaches. Road area extraction techniques concentrate on road segmentation and classification, whereas road centerline extraction methods focus on road skeleton recognition [17]. Recently, artificial intelligence algorithms have shown considerable development in feature extraction and segmentation from remote sensing images, thereby persuading researchers to distinguish road sections from high-resolution remote sensing imagery owing to the considerable efficiency of deep learning approaches in different applications [18]. Deep learning is a rapidly growing area in machine learning and has become an effective tool for expediting image processing and object detection. Moreover, deep learning has been widely implemented in remote sensing images, especially in mapping urban land cover with highly accurate results [19].

In this study, we conduct a systematic review of road extraction in remote sensing images from a novel perspective by discussing the current deep learning techniques applied to remote sensing datasets for road extraction and semantic segmentation. Previously, [20] only summarized the research outcomes of road extraction based on heuristic approaches. The applied deep learning approaches are categorized in this study on the basis of the type of deep convolutional neural networks. In addition, their strengths and limitations are discussed, and further insights for future studies in this field are provided. The remainder of this review paper is organized as follows. Section 2 provides a short background on the application of deep learning models for remote sensing image classification and road extraction. Section 3 presents the methodology for gathering relevant works. Section 4 discusses the extensive literature by considering common and state-of-art methods, including methodology, data, proposed approaches, difficulties, opinions, and precision. Section 5 describes the advantages and disadvantages of the proposed deep learning models, and Section 6 concludes the study.

## 2. Background

This section provides a summary of traditional road extraction methods. In addition, it discusses the development of deep learning methods in processing remotely sensed images and computer vision, specifically, road semantic segmentation from high-resolution remote sensing imagery.

At present, road extraction and monitoring operations are performed manually, which is ineffective and costly. Therefore, the automatic extraction and detection of roads from high-resolution images would be efficient and cost effective. Previously, remotely sensed imagery, such as multispectral and hyperspectral images, with high-spectral bandwidths was used for traditional remote sensing-based road extraction [21]. The present application of extracting road sections from remote sensing imagery at the macrolevel can be used in urban planning given the huge volume of available high-spectral resolution satellite and low-spatial resolution remotely sensed images [22]. Road extraction methods principally use the depth of spectral information to extract road sections from hyperspectral and multispectral satellite images [23]. Within the last decade, extremely high-resolution remote sensing

imagery, such as orthophoto and unmanned aerial vehicle (UAV) images obtained by advanced remote sensing technologies, was increasingly utilized for shadow classification, road extraction, and vehicle detection. Such fields confirmed the potential of images with high spatial resolutions [24].

Various studies have extracted road parts from high-resolution remotely sensed images using two main techniques, namely, data-driven and heuristic methods. Data-driven methods generally use the information of large data to conduct road extraction from satellite images. Recently, several data-driven approaches were considered for extracting road classes from remote sensing imagery containing conditional random fields (CRFs) [25], clustering [26], and Markov random fields (MRFs) [27]. By contrast, heuristic methods involve texture progressive analysis [28] and mathematical morphology [29], and often use certain information about road sections. Thus, these approaches are useless in handling different types of roads compared with data-driven techniques. However, traditional segmentation approaches fail to achieve high accuracy in road extraction and cannot handle multiscale roads, particularly narrow road sections with high width variance. The reason for this inability is that compared with normal images, high-resolution remote sensing images gain more detail. Thus, narrow road regions become apparent in such images, thereby introducing novel difficulties for road segmentation from high-resolution satellite images. Also, most of the preliminary studies for road extraction are on the basis of unsupervised learning like global optimization and graph cut methods [30] that rely on the color features and they have one general constraint, which is color sensitivity. Therefore, if the colors of roads in remote sensing imagery consist of more than one color, these segmentation algorithms will not attain good results and not perform well in road extraction and classification. Therefore, new robust techniques, such as deep learning methods, are needed to extract road networks with various scales accurately from remote sensing imagery [30].

In different fields, such as image classification, scene recognition, object detection, and semantic segmentation, advanced cutting-edge convolutional neural networks (CNNs) presently exceed other methods [31]. Compared to the unsupervised approaches that rely on the color for segmentation, more than one feature other than color, such as texture, shape, and line can be extracted by deep learning methods, among others. One of the most well-known methods initially identified to generalize CNNs in computer vision is the AlexNet18 model, which won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) challenge in 2012 [32]. Recently, a CNN model called the fully convolutional network (FCN), which was suggested by [33], revealed promising results in dense semantic segmentation. In addition, remotely sensed image processing, such as object identification in high-resolution remote sensing images [34], semantic labeling of satellite images [35], and image classification [36], was conducted using modern CNN models. The FCN demonstrated satisfactory results in the semantic segmentation of high-resolution remote sensing imagery [37]. Specifically, CNNs and the FCN were also synthesized for road semantic segmentation from remotely sensed imagery to learn road features and extract road regions automatically [38,39]. One of the initial efforts of implementing deep learning methods for road extraction from remote sensing images was made by [40]. For detecting road parts from remote sensing data, they applied restricted Boltzmann machine (RBMs). Also, they used preprocessing and postprocessing steps for achieving better results. Saito, Yamashita, and Aoki [38] proposed a method for roads and buildings extraction from raw remotely sensed images that was different from [40]. This approach was applied on a Massachusetts road dataset that obtained better outcomes. In recent years, many studies proposed that a deeper neural framework showed better results [41]; however, training of such a model is challenging because of the gradient vanishing issue. To address this issue, a deep residual learning architecture is suggested by [42] to simplify training by using an identity mapping [43].

## 3. Methodology

A systematic review was implemented to identify and select related literature as well as to achieve our research purpose based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement regarding record objects chosen for meta-analyses and systematic reviews [44,45].

The Web of Science (WoS) database was used to search for relevant manuscripts [46]. We restricted the search results to peer-reviewed documents, such as journals and conference papers, to ensure the authenticity and quality of the outcomes. We applied main expressions such as "Road Extraction, Remote Sensing Images, and Deep Learning"; "Road Detection, Deep Learning, and High-Resolution Satellite Images"; and "Road Semantic Segmentation, Deep Learning, and Remote Sensing Images" over a 10-year period, from 2010 to 2019, to gather papers. The process flow of gathering manuscripts is depicted in Figure 1.

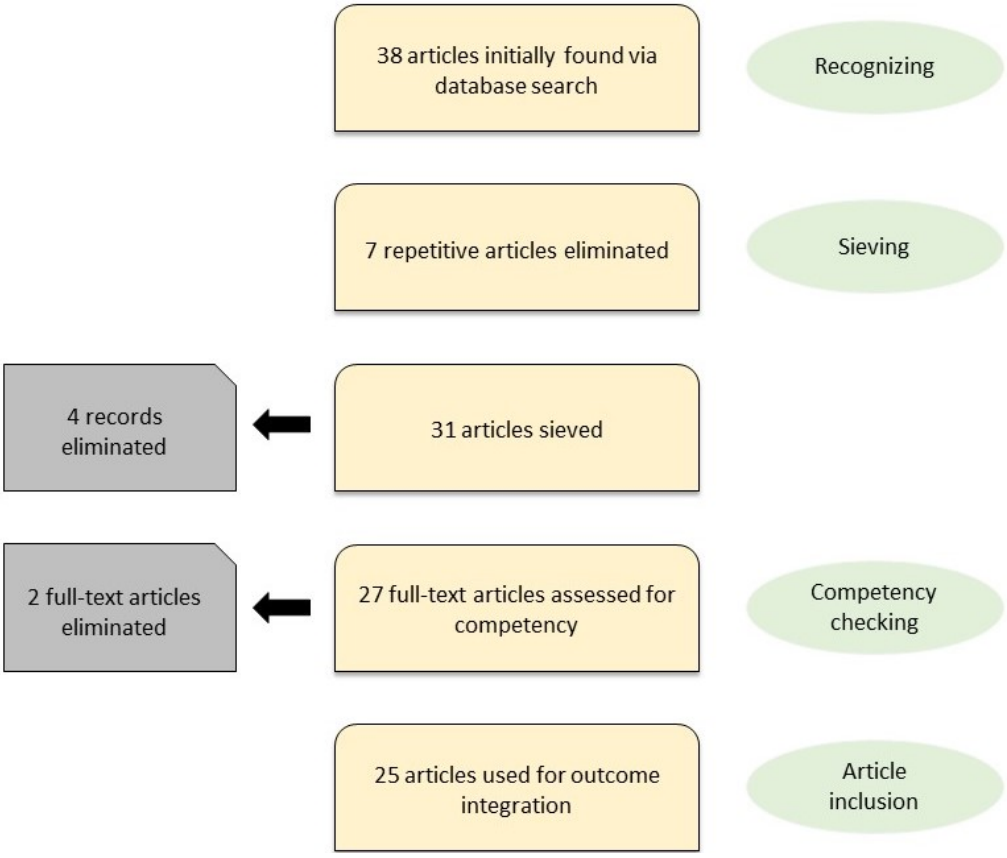


**Figure 1.** The process of extracting relevant papers based on the diverse combination of important search expressions.

A set of inclusion and exclusion criteria was ascertained as competency factors to identify previous studies and subjects based on the purpose of this work. The exclusion factors were as follows:

- The full text of the papers was not provided by publishers;
- Remote sensing images were not used in the papers.

The inclusion factors were as follows:

- Articles written in English;
- Peer-reviewed papers, such as conferences and journals;
- Published papers during the 10-year period (i.e., 2010–2019);
- Products that revealed a deep learning technique for road extraction from remote sensing images.

A total of 38 records were initially identified. Subsequently, we excluded redundant papers, and those that did not use remote sensing images for road extraction; thus, only 25 studies were accepted. Finally, we classified the documents selected based on purpose as an outcome integration process and revealed the consequences in detail in Section 4. In Section 4, we present major findings,

including the benefits and drawbacks of current products for road segmentation from remote sensing imagery via deep learning models, as well as evidence for each main outcome. We also discuss several recommendations for future research.

## 4. Results

This section elaborates on prior studies on deep learning methods that were applied to remote sensing images to extract road sections. We split the results into several subsections based on the type of deep learning methods used (Figure 2).

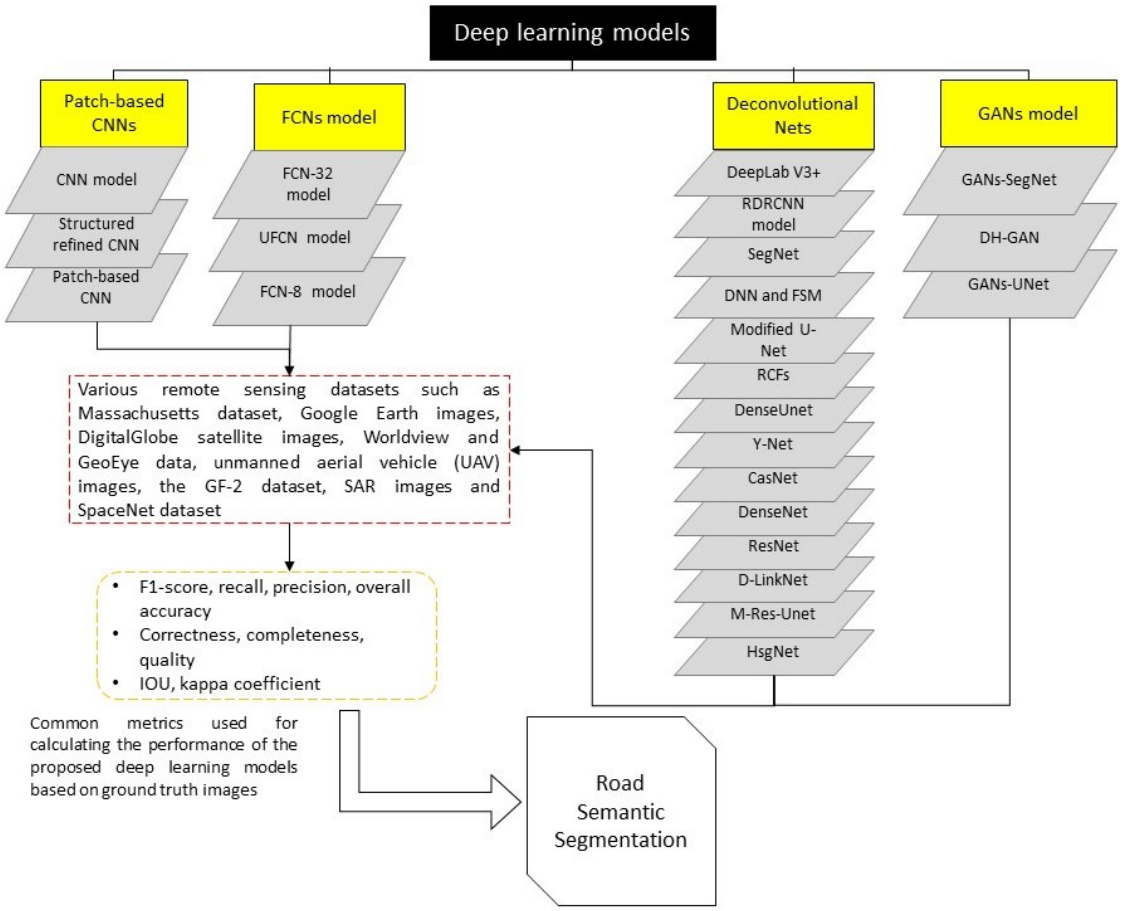


**Figure 2.** Road semantic segmentation using different deep learning models from remote sensing datasets.

### *4.1. Road Extraction Based on the Patch-Based CNN Model*

In the patch-based CNN model, the possibility of road dispensation is firstly predicted piece-by-piece with a particular stride and then the label map of the whole image is produced by assembling all of the label patches. Figure 3 illustrates a general architecture of the patch-level CNNs model. The initial section is convolutional and max pooling layers chased by fully connected layers acting as a linear discriminator. In this section, we describe the prior studies that used the CNN model for road extraction.

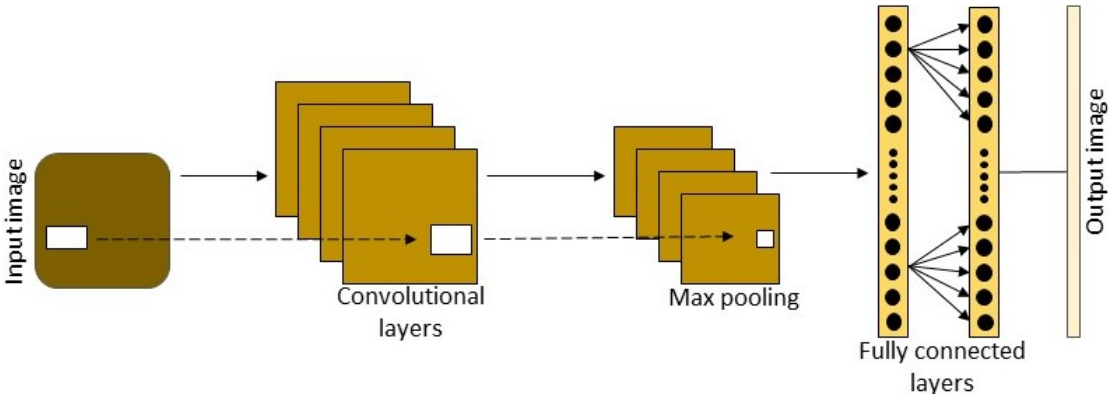


**Figure 3.** General architecture of the patch-level CNNs model.

Zhong et al. [39] provisionally implemented the newest CNN model to extract road and building objects from satellite imagery. The model fused low-level fine-grained features and high-level semantic meaning. In addition, further hyperparameters, such as the input image size, training epoch, and learning rate, were analyzed to specify the capability of the method in the context of high-resolution remote sensing images. The Massachusetts dataset, with a 1-m spatial resolution and 1500 × 1500 pixel size, containing 1711 images for the road and 151 images for the building datasets, was used for the evaluation. The Massachusetts dataset is related to the state of Massachusetts. The dataset covers over 2600 square kilometers with diverse rural, suburban, and urban areas [43]. With the integration of the pretrained FCN method with a novel four-stride pooling layer output to the last score layer, as well as fine-tuned with high-resolution spatial data, the extraction accuracy of the adjusted model was upgraded significantly to over 78%. Wei et al. [47] used a technique on aerial images for extracting road classes based on a road structure-refined CNN model, which provided road geometric information and spatial correlation. The proposed model was merged with fusion and deconvolutional layers to obtain structured output. Furthermore, a novel road structure-based loss function was applied to cross-entropy loss to yield a weight map by using the minimum Euclidean distance of every pixel to the road section and to model the road geometric structure. The Massachusetts road dataset, including 1172 images randomly divided into 49, 14, and 1108 images for testing, validation, and training, respectively, was used to calculate the proposed technique. Efficiency measures, namely, F1 score, recall, precision, and accuracy, were calculated for comparison, which were 66.2%, 72.9%, 60.6%, and 92.4%, respectively. The outcomes proved that the suggested model could extract roads effectively and achieve better accuracy compared with other existing road segmentation methods. However, postprocessing was needed to improve results. The link to download the public Massachusetts dataset and CNN code can be found in the online version, at https://www.cs.toronto.edu/~{}vmnih/data/, https://github.com/AhmedAhres/Satellite-Image-Classification.

Alshehhi et al. [48] implemented a patch-based CNN model for extracting road and building parts simultaneously from remote sensing imagery. Global average pooling was replaced with fully connected layers to consider a medium of feature maps from the final convolutional layer. Furthermore, the authors implemented a simple linear iterative clustering method during postprocessing to integrate CNN features with low-level features, such as the compactness and asymmetry of buildings and roads. This process integrated ungrouped areas of buildings and connected–disconnected road parts, as well as improved the performance of the proposed method. The Massachusetts dataset, including 10 images for testing, 137 images for training, and 4 images for validation, and the Abu Dhabi dataset with a 0.5 m spatial resolution per pixel, including 30 images for testing, 150 images for training, and 30 images for validation, were used for the evaluation. The authors used prevalent measure correctness to calculate the performance of the suggested approach, which was 91.7% for the Massachusetts dataset and 80.9% for the Abu Dhabi dataset. The results showed that the approach was effective in road and building extraction. However, further processing was needed to determine boundaries precisely.

Liu et al. [49] presented an approach for road centerline extraction from high-resolution remote sensing imagery that comprised four major stages. First, a CNN model was used to classify aerial images and learn features from raw images. Second, edge-preserving filtering was applied to the classified images with the original images to exploit road edges. Third, multidirectional morphological and shape feature filtering was used during postprocessing to obtain trustworthy roads. Finally, an integrated Gabor filter model and multiple directional nonmaximum suppression were applied to extract road centerlines. The suggested method was applied to two datasets, namely, the EPFL dataset and the Massachusetts road dataset. Three accuracy measures, namely, completeness, which was 95.40%; correctness, which was 89.97%; and quality, which was 86.21%, were used to quantify the performance, which indicated the advantage of the proposed method for road centerline extraction. However, certain centerlines were not single-pixel wide in the proposed method. Li et al. [50] employed a model based on a CNN to extract roads from high-resolution satellite imagery. First, a CNN model was applied to allocate labels to every pixel and anticipate the possibility of each pixel relating to road sections. Second, a line integral convolutional-based method was executed to maintain edge information, conjoin tiny gaps, and soften a rough map. Finally, several image-processing operations were implemented to acquire road centerlines. The authors used images from the Pleiades-1A satellite, with a spatial resolution of 0.5 m, and the Geoeye satellite to test their model. The completeness indicator was 80.57%, the correctness indicator was 96.57%, and the quality indicator was 78.27%, which showed that the proposed model achieved high precision for road extraction in terms of correctness. However, completeness and quality percentages were low, which was related to the complexity of the texture of various features in the images.

### *4.2. Road Extraction Based on the FCNs Model*

Compared to the CNN model that utilizes a dense layer to achieve a fixed-length feature vector and only accepts images with a fixed size, the FCNs model uses the interpolation layer after the final convolutional layer to upsample the feature map and restore the similar input size, as well as accepts input images of any size. In the FCNs, the final dense layers are replaced with convolutional layers and then output is a label map. A general architecture of FCNs model is presented in Figure 4. In the following, the previous research related to the FCNs model and road extraction are explained.

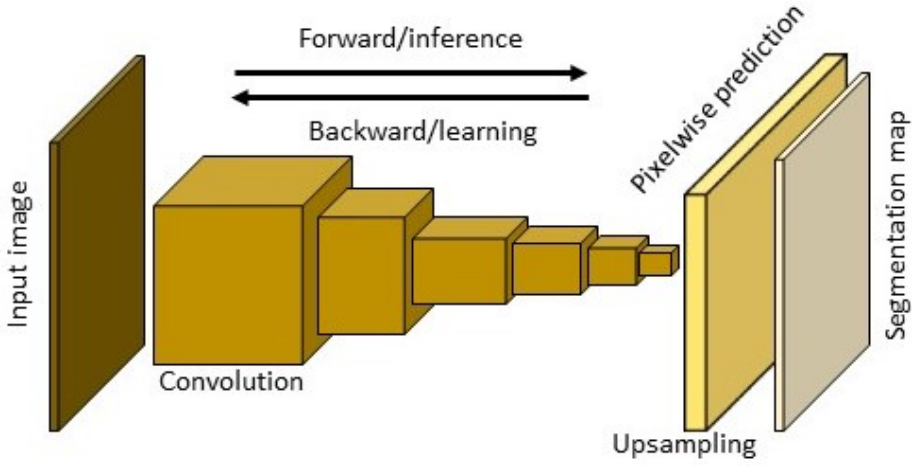


**Figure 4.** General architecture of FCNs model.

Varia et al. [51] applied a deep learning technique, namely, the FCN-32 for extracting road parts from extremely high-resolution UAV imagery. UAV-based imaging systems, which commonly use drones, can be used for the real-time assessment of several applications, monitoring tasks, and large-scale mapping, and are managed autonomously by onboard computers or remotely by human operators [52]. UAV-based remote sensing systems are used in various remote sensing applications, such as object

recognition [53] and digital elevation model (DEM) generation [54]. Compared with traditional remotely sensed systems, UAVs have multiple advantages, including improved security, high speed, low cost, and high flexibility. In addition, improved details can be provided by high-resolution images taken by drone systems for object extraction and detection. The suggested techniques were evaluated on a UAV image dataset with 189 training and 23 test images. The training time for the FCN-32 was approximately 370 s per image. The authors evaluated quality, correctness, and completeness assessment measures to show the models' efficiency for road extraction and found that the proposed models achieved satisfactory results. Moreover, they are effective for road extraction from UAV images. However, the models misclassified nonroad areas as road areas in certain areas with high complexity, thereby resulting in a large number of false negatives (FN) and reducing the percentage of completeness and quality in the final output. The suggested models were highly dependent on the number of images fed into them for training. Thus, they should be applied to many images with a large variety for better training and improved accuracy.

Kestur et al. [55] presented a novel architecture based on the FCN called the U-shaped FCN (UFCN) to extract roads from UAV images. The model was used on a UAV dataset with 109 images, approximately 70% of which were used for training and 30% for testing. The authors applied data augmentation during the training step to increase dataset size efficiently to improve training. The prediction took 1.95, 7.68, 43.87, and 1.09 s per image for UFCN, SVM, 1D-CNN, and 2D-CNN, respectively. The 1D-CNN model was slower than the UFCN model because of the computationally intensive architecture of the 1D-CNN network. Metric indicators, namely, F1 score, recall, precision, and overall accuracy, were calculated to assess classification performance, which were 89.6%, 86.8%, 92.5%, and 95.2%, respectively. The authors also compared their model with a two-dimensional CNN model, a one-dimensional CNN model, and an SVM model. They found that the model outperformed all the aforementioned methods in terms of accuracy and prediction time. Although the result achieved by the proposed model was promising, the dataset could be extended over a large area to use the suggested method for road extraction from extremely high-resolution remote sensing imagery. An FCN-8 network was proposed by [56] for road extraction from SAR images. The method was implemented on the TerraSAR-X dataset with 20% for testing and 80% for training. The experimental outcomes proved that the proposed model was able to extract the road part accurately. The access link to the open source code of FCN models for satellite image segmentaion can be found at https://github.com/Mattymar/satellite-image-segmentation.

### *4.3. Road Extraction Based on the Deconvolutional Neural Networks (Dense Net)*

Deconvolutional networks struggle to extract hierarchical features from images that closely pertain to a number of deep learning methods from the machine learning community. These models comprise an encoder and decoder part, which a bottom-up mapping from the input image to the latent feature space is provided by the encoder part while the latent features are mapped back to the input image using the decoder part. A general architecture of deconvolutional networks is shown in Figure 5. Following this, the previous works related to using deconvolutional models for road extraction from remote sensing datasets are highlighted.

Panboonyuen et al. [30] presented a technique based on a modified deep encoder–decoder neural network to extract road objects from remote sensing imagery. To improve the suggested model, the authors enhanced certain phases of the suggested approach containing the incorporation of the exponential linear unit (ELU) function against the rectified linear unit function. In addition, the authors increased the number of training datasets by rotating images to eight different angles incrementally and used a landscape metrics (LM) method to eliminate false road parts and improve the general accuracy of the output. The designed model was tested on the Massachusetts dataset containing 49, 14, and 1108 images for testing, validation, and training, respectively. The most common metrics, namely, F1 score, recall, and precision, were also used for the performance evaluation, which gained 85.7%, 86.1%, and 85.4%, respectively. The results proved that the suggested approach yields satisfactory

results and outperforms state-of-the-art approaches in road extraction from remote sensing imagery in terms of performance metrics. Wang et al. [57] introduced a semiautomatic technique based on the finite state machine (FSM) and DNN, including two main steps, namely, training and tracking, for road extraction from high-resolution remote sensing imagery. In the training step, the model was trained to recognize the pattern of an input image. To generate training samples, a vector-guided labeling approach that elicited huge image-direction mates from available vector road maps and images was defined. In the tracking step, a fusion strategy was used to detect the size of a detection window, and the trained DNN was used to recognize extracted image patches. In general, the DNN was applied to the proposed method to determine a pattern from complicated scenes, and the FSM was used to control the behavior of trackers and translate identified patterns into states. The model was applied to two datasets, namely, aerial and Google Earth images, which were divided into 60%, 20%, and 20% for training, testing, and validation, respectively. Completeness, correctness, and quality percentage indices were used for the performance assessment, which were 75%, 70%, and 74%, respectively, thereby proving that the suggested method could effectively exploit road classes from high-resolution remote sensing imagery in areas that were not highly complex. However, the proposed method could not operate properly in extremely complicated positions where road and other occlusions roughly contribute equal reflectance characteristics.

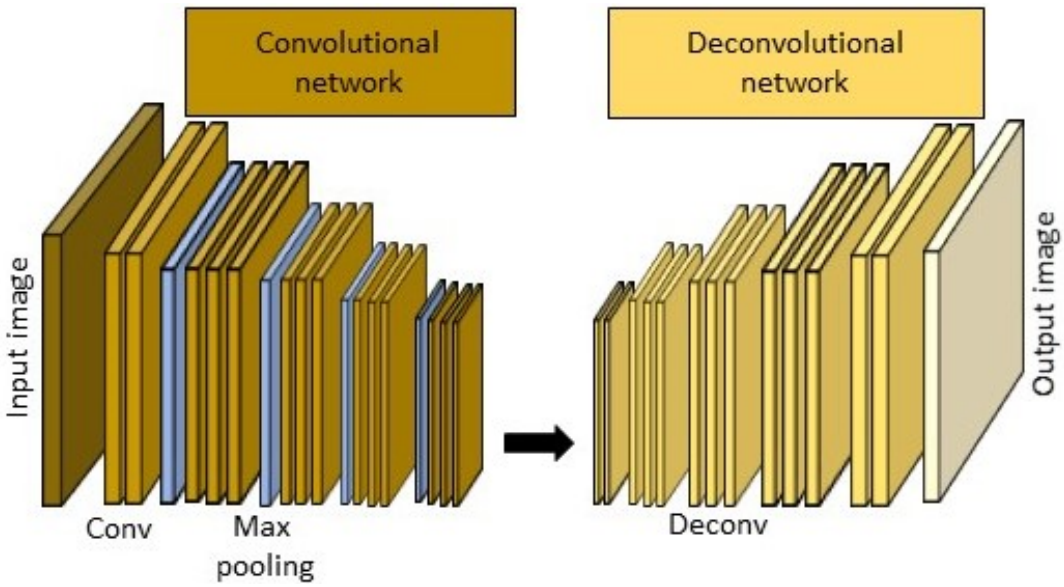


**Figure 5.** General architecture of deconvolutional networks.

Panboonyuen et al. [58] developed a new enhanced deep convolutional encoder–decoder model based on SegNet to segment road classes from high-resolution remote sensing imagery. A new activation function, namely, the ELU, was incorporated into the model to improve accuracy. The LM method was applied to remove falsely categorized road classes and identify road patterns. In the final step, the authors used CRFs to sharpen extracted roads. The proposed model was applied to two aerial and satellite datasets: (1) the Massachusetts dataset, including 1171 images divided into 1108, 14, and 49 images for training, validation, and testing, respectively, and (2) the Thailand Earth Observation System (THEOS) dataset containing 855 satellite images. The authors used F1 score, recall, and precision performance measures, which achieved 87.6%, 89.4%, and 85.8%, respectively, for the Massachusetts dataset and 64.9%, 58.4%, and 75.1%, respectively, for the THEOS dataset. The results indicated that the suggested approach outperforms other existing road segmentation techniques. However, this framework only works on extremely high-resolution remote sensing images, and distinguishing road sections from low- and medium-resolution remote sensing imagery is challenging. Constantin et al. [59] introduced a modified U-net CNN for extracting road classes

from high-resolution remote sensing imagery. The authors applied a novel binary cross entropy loss function and Jaccard distance fusion to train the model to decrease the number of false positives (FPs) and enhance the accuracy of binary classification. The proposed method was tested on the Massachusetts dataset, including 49 aerial test images, 14 validation data, and 1108 training data, with extra data augmentation to extend the dataset. For the accuracy assessment, overall accuracy, F1 score, recall, and precision were calculated, which were 97.14%, 74.54%, 75.48%, and 74.15%, respectively. Although the proposed model achieved a high accuracy of over 97%, its accuracy for other parameters was low. Therefore, additional pre- and postprocessing operations are necessary to improve the classification efficiency of the proposed approach for road extraction.

Zhang et al. [60] developed a deep residual U-net model similar to a U-net architecture for road semantic segmentation from high-resolution remote sensing imagery. The proposed network was designed based on residual units, which simplify network training. Rich skip connections were also used inside the model, which allowed few parameters and facilitated information propagation while achieving improved performance. The authors used their model on the Massachusetts road dataset, including 1171 images divided into 49, 14, and 1108 images as the test, validation, and training data, respectively. The authors compared the suggested model with the U-net model and two other deep networks for road extraction and found that the suggested technique was more efficient in extracting roads from high-resolution remote sensing imagery in terms of precision and recall. However, the introduced approach could not identify road sections in parking lots and under trees. Hong et al. [61] employed a method based on richer convolutional features (RCFs) for road segmentation from high-resolution remote sensing imagery. The proposed model contains four principal phases. (1) Training and testing samples were generated based on dataset preprocessing on the main image. (2) The RCF network was trained on the training samples and implemented on the testing images to generate strict road feature maps. (3) Autothreshold segmentation was applied to remove nonroad information and produce a road binary map. (4) Finally, road sections were extracted and vectorized. The authors applied their method on the Massachusetts road dataset, including 865 images. Four metrics, namely, precision, recall, F1 score, and overall accuracy, were used to determine the capability of the proposed method for road extraction, which were 85.8%, 98.5%, 91.5%, and 96.3%, respectively. Although the suggested approach achieved high accuracy for road class extraction from high-resolution remote sensing imagery, it could not gain precise road width information owing to combined pixel and model structure issues.

Xin et al. [62] applied the DenseUNet model for road extraction from remote sensing images. The DenseUNet model included skip connection and dense connection units that facilitated the merging of various scales by joints at different network layers. Two main datasets, namely, the Massachusetts and Conghua datasets, were used to calculate model efficiency. The image resolution of the Conghua dataset was 0.2 m and consisted of three red, blue, and green bands (RGB). A total of 47 aerial images were used in this dataset, with each image consisting of $3 \times 6000 \times 6000$ pixels. In this dataset, 80% of the data were used for training and the remaining 20% were used for model validation. The Massachusetts dataset was separated into 49 images, 14 data items, and 1108 data items for testing, validation, and training, respectively. The authors used precision, recall, F1 score, Intersection Over Union (IOU), and the Kappa coefficient to calculate the efficiency of the proposed method for road extraction. The respective values were 78.25%, 70.41%, 74.07%, 74.47%, and 70.32% for the Massachusetts dataset and 85.55%, 78.51%, 76.25%, 80.89%, and 80.11% for the Conghua dataset. The outcomes showed that the suggested technique has the advantage of low noise and high precision.

Li et al. [63] suggested a new convolutional neural network called the Y-Net, which includes two main fusion and feature extraction modules for extracting road parts from high-resolution remote sensing imagery. A feature extraction module consisting of a deep downsampling-to-upsampling subnetwork was introduced for semantic feature extraction, and a convolutional subnetwork without downsampling was introduced for detail feature extraction. The authors applied a fusion module to combine features for segmenting road classes. Moreover, the proposed technique was tested on the

public Massachusetts dataset and a private dataset from the Jlin 1 business satellite. Both datasets were split into a training dataset with 12,376 images, a validation dataset with 474 images, and a testing dataset with 531 images. The authors calculated mean region IOU (mean IOU), the Dice coefficient, mean accuracy, the Matthew correlation coefficient, and pixel accuracy for the accuracy assessment of the proposed model, which were 77.09%, 85.58%, 82.53%, 71.56%, and 97.36%, respectively. The experiment results showed the superiority and potential of the model for road semantic segmentation from remote sensing imagery. However, the proposed approach possesses several road extraction limitations. A small portion of the remote sensing imagery is occupied by a number of road pixels; thus, class imbalance is a considerable dilemma in road segmentation, particularly in narrow road sections. Thus, the method does not perform well in such areas. In addition, the proposed method requires additional time for training, which could be reduced by introducing transfer learning and generative adversarial network (GAN) fusion in the model, thereby improving accuracy. In general, deep learning models can achieve high accuracy in road extraction from remote sensing imagery compared with other machine learning approaches.

Cheng et al. [64] presented a new deep learning technique called the cascaded end-to-end (CasNet) deep learning model for detecting road classes and extracting road centerlines from extremely high-resolution remote sensing imagery. The suggested model includes two networks. The first is for detecting road regions, and the second is for extracting road centerlines, which are cascaded to the previous one and take full advantage of feature maps provided previously. The authors used a thinning method to achieve a single-pixel width and smooth road centerline. The model was evaluated on Google Earth images with 224 images. The Earth images obtained using Google Earth were in the form of aerial or satellite images with RGB color and different spatial resolutions based on the data source [47]. The dataset was randomly divided into 180, 14, and 30 images for training, validation, and testing, respectively. Several regularization methods and data augmentation approaches were applied to reduce overfitting and increase the size of the dataset. Classification metrics, namely, quality, correctness, and completeness, were introduced to evaluate the road extraction performance of the proposed model, which were 88%, 92%, and 94%, respectively. The results showed that the method is effective for road centerline extraction and road detection. However, the proposed method does not perform well in areas where roads are covered by tree occlusions. Therefore, additional high-level semantic information is needed to improve the performance of the method and to extract obstructions effectively. Xu et al. [65] used a new technique based on a densely connected convolutional network (DenseNet) by introducing local and global road information to segment roads from high-resolution remote sensing images. The method was applied to Google Earth data with a 1.2-m spatial resolution containing 224 images. The authors calculated F1 score, accuracy, precision, and recall measurement indicators for the accuracy evaluation, which were 95.72%, 96.3%, 96.30%, and 95.15%, respectively. The results proved that the introduced technique is efficient for road extraction. The experiment results were compared with other semantic segmentation methods, such as the DeepLab V3+, FCN, and U-net models, and showed that the proposed method outperformed the others.

Buslaev et al. [66] developed a deep learning technique based on the U-net family to extract roads from remote sensing imagery. The authors used an encoder similar to the RezNet-34 network, and a decoder was used based on the vanilla U-Net decoder. The authors also produced a loss function that considers binary cross-entropy and IOU simultaneously. In addition, data augmentation was used to improve the performance of the method. The model was evaluated on a dataset collected by the DigitalGlobe satellite, with a 50 cm pixel resolution and 6226 images. Furthermore, 1243 validation images were provided to calculate the performance of the model. IOU was used as a metric for the accuracy assessment of the suggested method, which was 64%, thereby indicating satisfactory results for road extraction. However, the model can be further improved by preparing high-quality labeled masks and amending data augmentation. Zhou et al. [67] introduced the D-LinkNet model for road semantic segmentation from remote sensing imagery. The proposed model contains an encoder–decoder structure, dilated convolution, and a pretrained encoder for extracting road sections. A dilated

convolution is a beneficial alternative to pooling layers, which is a valuable kernel for expanding and modifying receptive feature point fields and keeping detailed information, such as narrowness, connectivity, and complexity, without reducing the resolution of feature maps. The proposed technique was tested on the DigitalGlobe road dataset with 6226, 1243, and 1101 data items for training, validation, and testing, respectively. The IOU metric was evaluated and showed that the method has road extraction capabilities but retains several issues concerning road connectivity and recognition.

Doshi [68] applied an integrated model based on the ResNet and an inception-style encoder called the residual inception skip net to extract roads from satellite images. The introduced model was implemented on a dataset with a 0.5-m pixel resolution and 6226 images. The dataset was gathered by DigitalGlobe satellites. The dataset was randomly divided into 85% and 15% for training and testing, respectively. The IOU metric was calculated to assess the accuracy of the model, which was 61.3%, thereby showing that the suggested united method can generally exceed the two other baseline approaches (i.e., U-Net and DeepLab). However, various postprocessing strategies, such as the use of CRFs, can definitely promote and optimize the performance of the suggested method. Xu et al. [69] applied a deep CNN based on deep residual networks to extract roads from WorldView-2 satellite images. A Gaussian filter was first applied as a preprocessing operation to eliminate noise. Next, the M-Res-U-Net model was introduced for road semantic segmentation. The authors calculated precision, recall, and F1 score to assess the classification performance, which were 90.04%, 95.17%, and 92.77%, respectively. The proposed method could extract road classes efficiently and achieve improvements for the assessment factors. However, the approach did not perform well in certain areas wherein objects such as cars and building roofs had similar colors and spatial distributions. The authors generated ground truths using vector maps by setting a buffer in which all road areas with similar widths affected the accuracy of the model. Therefore, generating trustworthy labels and considering topological relationships could improve accuracy. Henry, Azimi, and Merkle [56] used DeepLabV3+ and Deep Residual U-Net to extract road sections from SAR images. The authors also used a control variable and mean squared error in the training process over the spatial tolerance of the network to promote the capability of the method. Each road was manually labeled, from major apparent highways to minor detectable roads. The authors applied the proposed approaches on a TerraSAR-X dataset with 80% for training and 20% for testing. For the accuracy evaluation, IOU, precision, and recall indices were calculated, which were 45.46%, 71.69%, and 75.17%, respectively. The results showed that though the FCNN models obtained satisfactory quantitative outcomes, the models missed multiple road sections and predicted unanticipated features, such as forests and hills.

He et al. [70] implemented a transfer learning technique for road segmentation from high-resolution remote sensing imagery. First, the authors applied a deep network based on an improved U-net model for training. Second, cross-modal data were used to fine tune the first two layers of a pretrained network to adjust the local features of the cross-modal data. An autoencoder was used to convert the data into three bands and extract local features for the cross-modal data of various bands. For the evaluation, the proposed method was tested on 6626 WorldView-3 images with a 0.5-m spatial resolution per pixel. The images were split into 6035 and 591 images for training and testing, respectively. F1 score, precision, recall, and IOU indicators were used to evaluate performance, which were 58.03%, 59.23%, 59%, and 42.03%, respectively. According to the results, the suggested model could extract road sections efficiently but could not achieve high accuracy in complex environments where other objects exhibited reflectances similar to road classes. Xia et al. [71] applied a DeepLab architecture for road extraction from high-resolution satellite images. The authors first implemented a semiautomatic approach to produce labeled data. A road benchmark was generated automatically then revised manually based on the construction characteristics and road patterns built by the transportation industry. The authors studied data influenced by color distortion as a type of road. Subsequently, they trained a DCNN model with deep layers to learn different road attributes. The designed method was tested on a GF-2 dataset, with spatial resolutions of 1 and 4 m for the panchromatic and multispectral scanners, respectively. The experiment results illustrated that the suggested approach can recognize road

classes from complicated positions with an accuracy of more than 80% in indistinguishable regions. However, smoothness estimation for curved lines is not successfully achieved by the proposed approach. Gao et al. [72] introduced a new framework called the refined deep residual CNN to extract roads from high-resolution satellite imagery. The proposed method comprises two main units, namely, residual connected and dilated perception units. The authors applied a postprocessing step based on a tensor-voting technique and math morphology to incorporate split roads and promote the performance of the proposed model. The suggested method was implemented on two datasets: (1) Massachusetts road images with a 1-m spatial resolution per pixel, including 60, 6, and 10 images for training, validation, and testing, respectively, and (2) GF-2 road images with a 0.8-m spatial resolution consisting of 60, 16, and 10 images for training, validation, and testing, respectively. The authors calculated IOU, accuracy, recall, precision, and F1 score indicators to assess the quantitative performance of the suggested approach, which were 65.91%, 98.10%, 77.94%, 83.88%, and 80.58%, respectively. The experimental results confirmed the efficiency advantage of the proposed technique for road extraction from remote sensing imagery. However, further processing is needed to achieve high accuracy in outline boundaries and complex urban areas. Xie et al. [73] applied a new road extraction method using a high-order spatial information global perception framework (HsgNet), which uses LinkNet as its basic network and embeds a middle block between encoder and decoder. The middle block learns to maintain various feature dependencies and channels' information, long-distance spatial relationship and information, and global-context semantic information. They implemented the proposed model on the DeepGlobe dataset that consists of 622 test images, 622 validation images and 4971 training images with spatial resolution of 0.5 m and image resolution of 1024 × 1024, as well as the SpaceNet dataset that includes 567 test images and 2213 training images with image resolution of 512 × 512. For evaluating the performance of the proposed method for road extraction, they calculated measurement metrics such as precision, recall, F1 score and IOU that obtained 83%, 82%, 71.1%, and 71.1%, respectively, for the DeepGlobe dataset and 81.6%, 84.5%, 83%, and 71%, respectively, for the SpaceNet dataset. The experimental results showed that the suggested model performed well for road extraction from high-resolution remote sensing imagery. The links to download the public datasets and official code repositories of aforementioned deep learning models can be found in the online version at https://github.com/robmarkcole/satellite-image-deep-learning; https://github.com/jeradhoy/DeepSatelliteData, https://github.com/divamgupta/image-segmentation-keras.

### 4.4. Road Extraction Based on the GANs Model

GANs comprises two main generative and discriminator models, in which the generative term tries to obtain the data dispensation and the discriminator part tries to determine the likelihood that a representation refers to training data instead of being created by a generative model [74]. The generic architecture of GANs model is presented in Figure 6. In this section, previous work related to applying the GANs model for road segmentation is highlighted.

Costea et al. [75] presented a new method named dual-hot generative adversarial networks (DH-GAN) to detect intersections and roads from UAV images at the higher semantic level of road graphs during the first step. Then, they applied a smoothing-based graph optimization method for pixelwise road segmenting and finding the road graph. They used F1 score, precision, and recall for evaluating the performance of the model, which were 86%, 89.84%, and 82.48% that proved the efficiency of the proposed model for road extraction, and also was able to minimize the memory costs. Varia, Dokania, and Senthilnath [51] applied the GANs model for road extraction from UAV images. They used the U-Net model for the generator part and the model was trained on 189 UAV images and evaluated on 23 test images. The training took 300 s per image for GANs-UNet. They achieved an accuracy of 96.08 for the F1 score, which shows that the proposed model was more efficient for road extraction from UAV images. Shi, Liu, and Li [74] implemented the GANs model for attaining a smooth road segmentation map from Google Earth images with 550 images: 320 images were used for training, 100 images for validation, and 130 images for testing. They also used data augmentation procedures to

increase the size of the dataset. An encoder–decoder SegNet model was used for generative part to generate a high-resolution segmentation map. The accuracy that they achieved for recall, precision, and F1 score was 91.01%, 88.31%, and 89.63%, respectively, that shows the superiority of the proposed model for road extraction. The access link to the GANs model code for image segmentation can be found at https://github.com/eriklindernoren/Keras-GAN/tree/master/pix2pix.

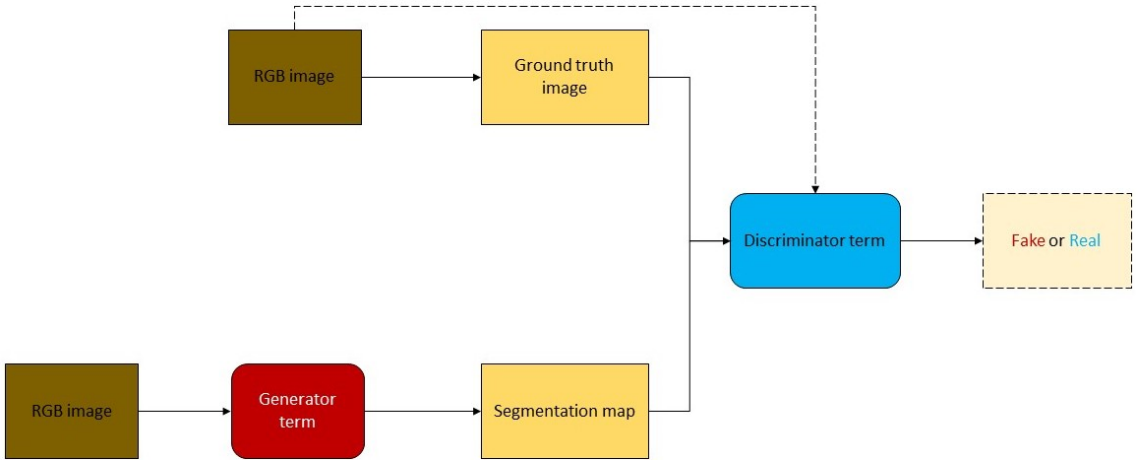


**Figure 6.** Generic architecture of GANs model.

## 5. Discussion

Several deep learning techniques have been suggested for extracting road classes from high-resolution remote sensing imagery. However, demands to obtain improved precision for segmented road outcome sets remain. Compared with other machine learning methods, deep learning techniques have shown notable development in object segmentation from images. However, their efficiency in road extraction can be scaled based on the processing power, model complexity, and the size of the training data. This review of existing research proves that compared with other machine learning and traditional techniques, deep learning methods have obtained higher precision in extracting road parts from high-resolution remote sensing imagery.

Based on previous studies, we categorize all the CNNs into four main models: the patched-based CNN model [40]; the FCN-based model [76,77]; deconvolutional net-based models, such as U-Net [78], SegNet [79], and DeepLab [80]; and the GAN-based model [81]. GANs contain two sections called the generator and discriminator parts, which have recently gained considerable attention [82]. The generator part struggles to make fake images from realistic ones, whereas the discriminator part strives to identify feigned images from actual images. Finally, dynamic balance can be achieved by the two parts, and an image can be segmented by the generator portion. In FCN models, each pixel can be inferred end-to-end by examining the patch-to-pixel anticipation. In these models, convolutional layers are replaced by final dense layers in which the output of the label map is the last convolutional layer. Deconvolutional net-based models are identified by deconvolutional layers, which are called decoder sections. Finally, the image block around a pixel can be used to train and anticipate input in the patch-based CNN model. The throughput outcomes of the aforementioned studies have shown that the deconvolutional networks are the most popular models that most of the researchers apply for the purpose of road semantic segmentation from high-resolution remote sensing imagery. We elaborate on the advantages and disadvantages of the discussed approaches to develop a general comparison (Table 1).

Table 1 shows that each model has its own limitations and strengths. For example, simple interpolation is utilized in the upsampling of the FCN models, thereby causing the models to achieve low precision. However, pixel-to-pixel reasoning can be obtained as well as end-to-end can be learned by FCNs inspired by CNN-based models that need expansive samples, ignore the

correlation among neighboring pixels, and require a high processing step to recognize precise road boundaries. While FCNs models encounter problems with road connectivity and cannot make smoothness predictions for curved lines as well as the segmentation map encounters with low spatial constancy, the DeconvNet model can obtain higher spatial precision and contains high adaptability compared with FCNs, as the former uses low-level information in deconvolutional layers. However, a large amount of storage and memory as well as a high computing process is required for applying this model. By contrast, the GANs model is more efficient because this model can achieve a constant segmentation map with road boundary information. However, the model encounters problems with a lack of convergence, gradient destruction, and complex training.

**Table 1.** Strengths and limitations of various deep learning methods for road extraction.

| Approaches | Complexity | Output | Smoothness |
| --- | --- | --- | --- |
| **Models based on GANs** | • Model breakdown and lack of convergence for complex and large data<br>• Complex training | • Efficient and robust<br>• Provide constant output | • Capable of achieving boundary information and smooth segmentation map |
| **Models based on CNNs** | • Require few parameters<br>• Require extensive samples<br>• Low computing process | • Not highly efficient in providing constant output<br>• Do not perform well in highly complex positions<br>• Ignore thecorrelation among neighboring pixels<br>• Attain pixel-to-pixel reasoning | • Require high processing to identify boundaries and create a smooth segmentation map |
| **Models based on FCNs** | • Low adaptability with complex data and depend on images and masks for training | • Issues with road connectivity<br>• Low position accuracy, lack of spatial consistency | • Cannot successfully achieve smoothness estimation for curved lines |
| **Models based on deconvolutional nets** | • Require large amounts of memory and storage<br>• Require additional time for training and high computing process | • High spatial accuracy<br>• Efficient and robust for achieving consistent output | • Able to obtain a smooth segmentation map |

In addition, we attempt to compare the accuracy of different deep learning models applied to remote sensing datasets based on the common metrics [83] used to evaluate the efficiency of the proposed approaches for road extraction. Popular evaluation measures are calculated based on a confusion matrix comprising four main factors, namely, false negative (FN), true negative, true positive, and FP [83,84]. A general comparison of all the methods used in all datasets is provided to elaborate on the most efficient technique for road extraction (Figure 7). All the aforementioned works and corresponding values are plotted using an x-axis and y-axis, respectively. Only the methods that include a dataset and research performance reports are compared.

We consider the F1 score metric, which is a trade-off measure between recall and precision, to compare the results achieved by different deep learning models for road extraction, except for the models such as U-Net, D-linkNet, and RISN applied to the DigitalGlobe satellite images, as the authors utilized only the IOU indicator for the performance evaluation. However, this indicator is only approximately 65% and does not demonstrate high precision. Figure 8 shows that the F1 score percentage is high for the GANs-UNet model, DenseNet method, and FCN-32 applied to UAV and Google Earth images, with accuracies of 96.08%, 95.72%, and 94.59%, respectively. One of the elegant fully convolutional neural networks named U-Net model was used for a generative model in the GANs framework to create a high-resolution segmentation map with more accuracy. Also, the model was applied on UAV images that consist of very high spatial resolution with a variety for the angle of capture, color, shapes, and orientation, which lead to achieving a highly precise road segmentation

map compared to the other deep learning models. Figure 8 illustrates the results achieved for road segmentation from UAV images (Figure 8a,b) with image dimension of 128 × 128, Google Earth images (Figure 8c) with a spatial resolution of 1.2 m and image dimension of 256 × 256, and the Massachusetts dataset (Figure 8d) with a spatial resolution of 1 m and image dimension of 375 × 375, by using the FCN-32, GANs-UNet, DenseNet, DeepLab V3+, CNN, and RSRCNN methods. The first and second columns are original and ground truth images, while the third and fourth columns depict the results achieved by the state-of-the-art methods. As it can be seen from Figure 8, the GANs model applied on UAV images performed better and predicted less FP and FN pixels when compared to other methods. Also, a smooth segmentation map with more details of boundary information is attained by the proposed model. In contrast, the CNN model applied on the Massachusetts dataset was unable to achieve high accuracy in road extraction compared to the RSRCNN method that was applied on the same dataset. The extracted road parts by CNN has a significant issue of fuzzy boundaries and "salt and pepper" phenomena because the CNN model only counts on texture and spectral features; the mixed pixels in road borders lead to misclassification while the other methods improve the classification performance by restraining the effect of mixed pixels by the segmentation process. In the models such as DenseNet and GANs, road features are extracted from every convolutional layer and then integrated on multiscales. Multiscale merging of road features not only uses high-level semantic information to avoid influence of width changes, curvatures, and shadows to achieve precise road boundaries, but also utilizes low-level information to preserve detailed information of road features. As a result, the CNN model predicted more nonroad pixels that lead to extract larger road parts compared to the reference map with low accuracy.

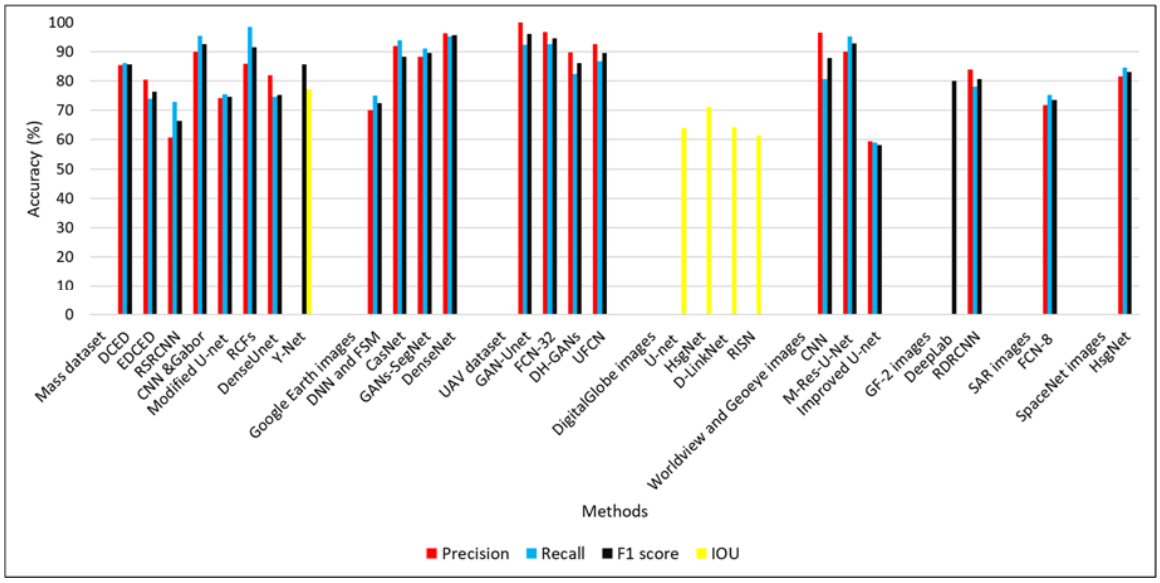


**Figure 7.** General comparison of deep learning models applied to different road datasets.

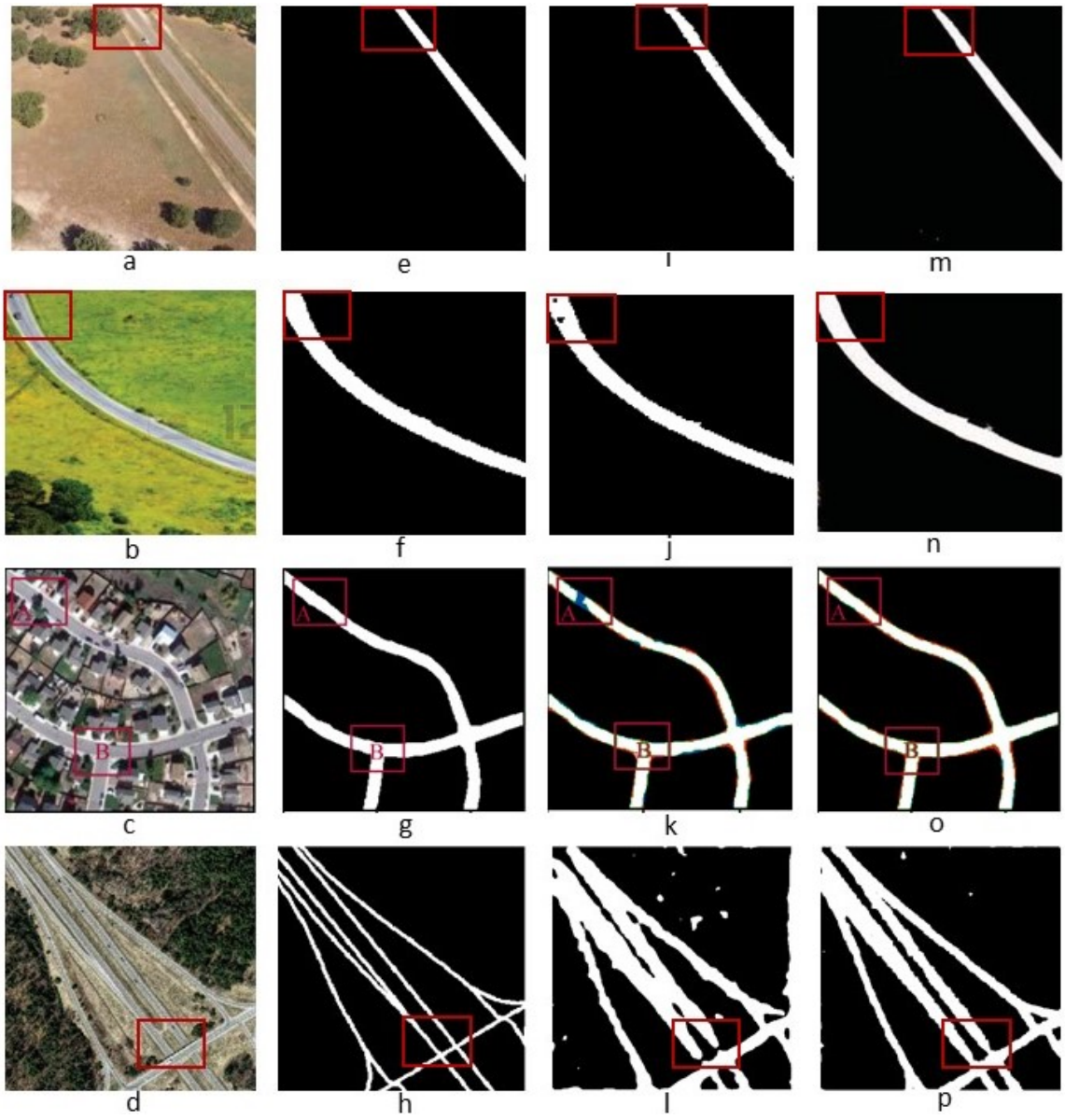


**Figure 8.** Extracted road parts using deep learning methods from high-resolution remote sensing images: (**a**–**d**) original images; (**e**–**h**) corresponding reference maps; (**i**,**j**) results of FCN-32 and (**k**) result of DeepLab V3+; (**m**,**n**) results of GANs-UNet and (**o**) result of DenseNet model; and (**l**,**p**) results of CNN and RSRCNN methods, respectively.

## 6. Conclusions

Spatial data, especially for road networks, should be updated regularly owing to rapid changes in artificial and natural features. Providing road data using traditional methods is ineffective, as such approaches are costly and time consuming. By contrast, extracting road types using advanced remote sensing technologies can be economically and practically efficient. Numerous proposed methods for road extraction and road data updates using remote sensing images are described in this review. In this research, we discover that most studies concentrate on using powerful methods to overcome constraints. Therefore, the development of advanced machine learning methods, such as deep CNNs, for feature segmentation and extraction from remote sensing images has encouraged researchers to apply such models to extract road networks from high spatial resolution remote sensing imagery, owing to the considerable efficiency of deep convolutional approaches in different applications.

Although the methods utilized for road extraction used different data, this study can provide the following important outcomes.

1. The capabilities of deep learning methods for road extraction are more effective than those of regular approaches.
2. When the complexity of images is high and various road types are present, the accuracy of the models is low. Therefore, mixing robust pre- and postprocessing techniques is recommended and useful to achieve satisfactory results.
3. The appropriateness of deep learning approaches for road extraction pertaining to different variables, such as architecture, data, and hyperparameters, is determined.
4. The low efficiency of the proposed methods in terms of data quality, training dataset, and model hyperparameters is presented.
5. Occlusions, such as shadows, cars, and buildings, are similar to road features, such as colors, reflectance, and patterns. Road extraction remains challenging owing to such issues.
6. Further research is required to build detailed techniques with high precision. CNNs trained by one dataset may be inconsistent with other scenes. Nonetheless, if training datasets are adequate and a deep learning model can be created effectively, then the model can be implemented properly on most prevalent datasets.

In this review, state-of-the-art deep convolutional models that represent common and newly advanced methodologies are described. In conclusion, introducing several new methods related to road semantic segmentation is important, and research on different proposed techniques with cutting-edge technology is increasing.

**Author Contributions:** A.A. (Abolfazl Abdollahi) wrote the manuscript; B.P. supervised, edited, restructured; B.P., N.S., S.C., and A.A.A. (Abdullah AlAmri) professionally optimized the manuscript; B.P. acquired the funding. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research is supported by the Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), Faculty of Engineering and IT, the University of Technology Sydney (UTS). This research was also supported by Researchers Supporting Project number RSP-2019/14, King Saud University, Riyadh, Saudi Arabia.

**Acknowledgments:** Authors would like to thank three anonymous reviewers and Editor for their valuable comments which helped us to improve the manuscript. Also, thanks to the funding and support of University of Technology Sydney which we gratefully appreciate.

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

## References

1. Abdullahi, S.; Pradhan, B.; Jebur, M.N. GIS-based sustainable city compactness assessment using integration of MCDM, Bayes theorem and RADAR technology. *Geocarto Int.* **2015**, *30*, 365–387. [CrossRef]
2. Youssef, A.M.; Sefry, S.A.; Pradhan, B.; Alfadail, E.A.; Risk. Analysis on causes of flash flood in Jeddah city (Kingdom of Saudi Arabia) of 2009 and 2011 using multi-sensor remote sensing data and GIS. *Geomat. Nat. Hazards* **2016**, *7*, 1018–1042. [CrossRef]
3. Weng, Q. Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends. *Remote Sens. Environ.* **2012**, *117*, 34–49. [CrossRef]
4. Wijesingha, J.S.J. Automatic road feature extraction from high resolution satellite images using LVQ neural networks. *Asian J. Geoinform.* **2013**, *13*, 30–36.
5. Kahraman, I.; Turan, M.K.; Karas, I.R. Road detection from high satellite images using neural networks. *Int. J. Modeling Optim.* **2015**, *5*, 304–307. [CrossRef]
6. Shi, W.; Miao, Z.; Debayle, J.; Sensing, R. An integrated method for urban main-road centerline extraction from optical remotely sensed imagery. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 3359–3372. [CrossRef]
7. Zhang, J.; Chen, L.; Wang, C.; Zhuo, L.; Tian, Q.; Liang, X. Road recognition from remote sensing imagery using incremental learning. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 2993–3005. [CrossRef]
8. Hormese, J.; Saravanan, C. Automated road extraction from high resolution satellite images. *Procedia Technol.* **2016**, *24*, 1460–1467. [CrossRef]

9. Abdollahi, A.; Bakhtiari, H.R.R.; Nejad, M.P. Investigation of SVM and level set interactive methods for road extraction from google earth images. *J. Indian Soc. Remote Sens.* **2018**, *46*, 423–430. [CrossRef]
10. Bakhtiari, H.R.R.; Abdollahi, A.; Rezaeian, H. Semi automatic road extraction from digital images. *Egypt. J. Remote Sens. Space Sci.* **2017**, *20*, 117–123. [CrossRef]
11. Liu, B.; Wu, H.; Wang, Y.; Liu, W. Main road extraction from zy-3 grayscale imagery based on directional mathematical morphology and vgi prior knowledge in urban areas. *PLoS ONE* **2015**, *10*, e0138017. [CrossRef] [PubMed]
12. Miao, Z.; Shi, W.; Gamba, P.; Li, Z. An object-based method for road network extraction in VHR satellite images. *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.* **2015**, *8*, 4853–4862. [CrossRef]
13. Grinias, I.; Panagiotakis, C.; Tziritas, G.; Sensing, R. MRF-based segmentation and unsupervised classification for building and road detection in peri-urban areas of high-resolution satellite images. *ISPRS J. Photogramm.* **2016**, *122*, 145–166. [CrossRef]
14. Sghaier, M.O.; Lepage, R. Road extraction from very high resolution remote sensing optical images based on texture analysis and beamlet transform. *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.* **2016**, *9*, 1946–1958. [CrossRef]
15. He, C.; Liao, Z.-X.; Yang, F.; Deng, X.-P.; Liao, M.-S. Road extraction from SAR imagery based on multiscale geometric analysis of detector responses. *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.* **2012**, *5*, 1373–1382.
16. Cheng, J.; Ding, W.; Ku, X.; Sun, J. Road Extraction from High-Resolution SAR Images via Automatic Local Detecting and Human-Guided Global Tracking. *Int. J. Antennas Propag.* **2012**, *2012*, 1–10. [CrossRef]
17. Alshehhi, R.; Marpu, P.R. Hierarchical graph-based segmentation for extracting road networks from high-resolution satellite images. *ISPRS J. Photogramm. Remote Sens.* **2017**, *126*, 245–260. [CrossRef]
18. Xu, Y.; Chen, Z.; Xie, Z.; Wu, L. Quality assessment of building footprint data using a deep autoencoder network. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 1929–1951. [CrossRef]
19. Audebert, N.; Le Saux, B.; Lefèvre, S. Segment-before-detect: Vehicle detection and classification through semantic segmentation of aerial images. *Remote Sens.* **2017**, *9*, 368. [CrossRef]
20. Wang, W.; Yang, N.; Zhang, Y.; Wang, F.; Cao, T.; Eklund, P. A review of road extraction from remote sensing images. *J. Traffic Transp. Eng.* **2016**, *3*, 271–282. [CrossRef]
21. Wang, J.; Qin, Q.; Gao, Z.; Zhao, J.; Ye, X. A new approach to urban road extraction using high-resolution aerial image. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 114. [CrossRef]
22. Hu, F.; Xia, G.-S.; Hu, J.; Zhang, L. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sens.* **2015**, *7*, 14680–14707. [CrossRef]
23. Li, W.; Wu, G.; Zhang, F.; Du, Q. Hyperspectral image classification using deep pixel-pair features. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 844–853. [CrossRef]
24. Senthilnath, J.; Dokania, A.; Kandukuri, M.; Ramesh, K.; Anand, G.; Omkar, S. Detection of tomatoes using spectral-spatial methods in remotely sensed RGB images captured by UAV. *Biosyst. Eng.* **2016**, *146*, 16–32. [CrossRef]
25. Wegner, J.D.; Montoya-Zegarra, J.A.; Schindler, K. A higher-order CRF model for road network extraction. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 1698–1705.
26. Maurya, R.; Gupta, P.; Shukla, A.S. Road extraction using k-means clustering and morphological operations. In Proceedings of the 2011 International Conference on Image Information Processing, Shimla, India, 3–5 November 2011; pp. 1–6.
27. Mattyus, G.; Wang, S.; Fidler, S.; Urtasun, R. Enhancing road maps by parsing aerial images around the world. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1689–1697.
28. Mena, J.B.; Malpica, J.A. An automatic method for road extraction in rural and semi-urban areas starting from high resolution satellite imagery. *Pattern Recognit. Lett.* **2005**, *26*, 1201–1220. [CrossRef]
29. Zhu, C.; Shi, W.; Pesaresi, M.; Liu, L.; Chen, X.; King, B. The recognition of road network from high-resolution satellite remotely sensed data using image morphological characteristics. *Int. J. Remote Sens.* **2005**, *26*, 5493–5508. [CrossRef]
30. Panboonyuen, T.; Vateekul, P.; Jitkajornwanich, K.; Lawawirojwong, S. An enhanced deep convolutional encoder-decoder network for road segmentation on aerial imagery. In *International Conference on Computing and Information Technology*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 191–201. [CrossRef]

31. Tang, S.; Yuan, Y. *Object Detection Based on Convolutional Neural Network*; Stanford University: Stanford, CA, USA, 2015; Available online: http://cs231n.stanford.edu/reports/2015/pdfs/CS231n_final_writeup_sjtang.pdf (accessed on 5 January 2020).
32. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, Nevada, 3–6 December 2012; pp. 1097–1105.
33. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
34. Ševo, I.; Avramović, A. Convolutional Neural Network Based Automatic Object Detection on Aerial Images. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 740–744. [CrossRef]
35. Volpi, M.; Tuia, D. Dense Semantic Labeling of Subdecimeter Resolution Images With Convolutional Neural Networks. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 881–893. [CrossRef]
36. Maggiori, E.; Tarabalka, Y.; Charpiat, G.; Alliez, P. Convolutional neural networks for large-scale remote-sensing image classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 645–657. [CrossRef]
37. Maggiori, E.; Tarabalka, Y.; Charpiat, G.; Alliez, P. Fully convolutional neural networks for remote sensing image classification. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 June 2016; pp. 5071–5074.
38. Saito, S.; Yamashita, T.; Aoki, Y. Multiple object extraction from aerial imagery with convolutional neural networks. *Electron. Imaging* **2016**, *2016*, 1–9. [CrossRef]
39. Zhong, Z.; Li, J.; Cui, W.; Jiang, H. Fully convolutional networks for building and road extraction: Preliminary results. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 June 2016; pp. 1591–1594. [CrossRef]
40. Mnih, V.; Hinton, G.E. *Learning to Detect Roads in High-Resolution Aerial Images*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 210–223. [CrossRef]
41. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
42. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
43. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity mappings in deep residual networks. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 630–645.
44. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G.; The, P.G. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med.* **2009**, *6*, e1000097. [CrossRef] [PubMed]
45. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Ann. Intern. Med.* **2009**, *151*, 264. [CrossRef]
46. Min, H.-S.J.; Beyeler, W.; Brown, T.; Son, Y.J.; Jones, A.T. Toward modeling and simulation of critical national infrastructure interdependencies. *IIE Trans.* **2007**, *39*, 57–71. [CrossRef]
47. Wei, Y.; Wang, Z.; Xu, M. Road Structure Refined CNN for Road Extraction in Aerial Image. *IEEE Geosci. Remote Sensing Lett.* **2017**, *14*, 709–713. [CrossRef]
48. Alshehhi, R.; Marpu, P.R.; Woon, W.L.; Dalla Mura, M. Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks. *ISPRS J. Photogramm. Remote Sens.* **2017**, *130*, 139–149. [CrossRef]
49. Liu, R.; Miao, Q.; Song, J.; Quan, Y.; Li, Y.; Xu, P.; Dai, J. Multiscale road centerlines extraction from high-resolution aerial imagery. *Neurocomputing* **2019**, *329*, 384–396. [CrossRef]
50. Li, P.; Zang, Y.; Wang, C.; Li, J.; Cheng, M.; Luo, L.; Yu, Y. Road network extraction via deep learning and line integral convolution. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 June 2016; pp. 1599–1602. [CrossRef]
51. Varia, N.; Dokania, A.; Senthilnath, J. DeepExt: A Convolution Neural Network for Road Extraction using RGB images captured by UAV. In Proceedings of the 2018 IEEE Symposium Series on Computational Intelligence (SSCI), Bangalore, India, 18–21 November 2018; pp. 1890–1895. [CrossRef]
52. Abdollahi, A.; Pradhan, B.; Shukla, N. Extraction of road features from UAV images using a novel level set segmentation approach. *Int. J. Urban Sci.* **2019**. [CrossRef]

53. Moranduzzo, T.; Melgani, F. Detecting cars in UAV images with a catalog-based approach. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 6356–6367. [CrossRef]
54. Yang, B.; Chen, C. Automatic registration of UAV-borne sequent images and LiDAR data. *ISPRS J. Photogramm. Remote Sens.* **2015**, *101*, 262–274. [CrossRef]
55. Kestur, R.; Farooq, S.; Abdal, R.; Mehraj, E.; Narasipura, O.; Mudigere, M. UFCN: A fully convolutional neural network for road extraction in RGB imagery acquired by remote sensing from an unmanned aerial vehicle. *J. Appl. Remote Sens.* **2018**, *12*, 016020. [CrossRef]
56. Henry, C.; Azimi, S.M.; Merkle, N. Road Segmentation in SAR Satellite Images With Deep Fully Convolutional Neural Networks. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 1867–1871. [CrossRef]
57. Wang, J.; Song, J.; Chen, M.; Yang, Z. Road network extraction: A neural-dynamic framework based on deep learning and a finite state machine. *Int. J. Remote Sens.* **2015**, *36*, 3144–3169. [CrossRef]
58. Panboonyuen, T.; Jitkajornwanich, K.; Lawawirojwong, S.; Srestasathiern, P.; Vateekul, P. Road Segmentation of Remotely-Sensed Images Using Deep Convolutional Neural Networks with Landscape Metrics and Conditional Random Fields. *J. Remote Sens.* **2017**, *9*, 680. [CrossRef]
59. Constantin, A.; Ding, J.-J.; Lee, Y.-C. Accurate Road Detection from Satellite Images Using Modified U-net. In Proceedings of the 2018 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS), Chengdu, China, 26–30 October 2018; pp. 423–426. [CrossRef]
60. Zhang, Z.; Liu, Q.; Wang, Y. Road extraction by deep residual u-net. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 749–753. [CrossRef]
61. Hong, Z.; Ming, D.; Zhou, K.; Guo, Y.; Lu, T. Road Extraction From a High Spatial Resolution Remote Sensing Image Based on Richer Convolutional Features. *IEEE Access* **2018**, *6*, 46988–47000. [CrossRef]
62. Xin, J.; Zhang, X.; Zhang, Z.; Fang, W. Road Extraction of High-Resolution Remote Sensing Images Derived from DenseUNet. *Remote Sens.* **2019**, *11*, 2499. [CrossRef]
63. Li, Y.; Xu, L.; Rao, J.; Guo, L.; Yan, Z.; Jin, S. A Y-Net deep learning method for road segmentation using high-resolution visible remote sensing images. *Remote Sens. Lett.* **2019**, *10*, 381–390. [CrossRef]
64. Cheng, G.; Wang, Y.; Xu, S.; Wang, H.; Xiang, S.; Pan, C. Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3322–3337. [CrossRef]
65. Xu, Y.; Xie, Z.; Feng, Y.; Chen, Z. Road Extraction from High-Resolution Remote Sensing Imagery Using Deep Learning. *Remote Sens.* **2018**, *10*, 1461. [CrossRef]
66. Buslaev, A.; Seferbekov, S.; Iglovikov, V.; Shvets, A. Fully convolutional network for automatic road extraction from satellite imagery. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018; pp. 207–210. [CrossRef]
67. Zhou, L.; Zhang, C.; Wu, M. D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 182–186. [CrossRef]
68. Doshi, J. Residual inception skip network for binary segmentation. In Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 216–219. [CrossRef]
69. Xu, Y.; Feng, Y.; Xie, Z.; Hu, A.; Zhang, X. A Research on Extracting Road Network from High Resolution Remote Sensing Imagery. In Proceedings of the 2018 26th International Conference on Geoinformatics, Kunming, China, 28–30 June 2018; pp. 1–4. [CrossRef]
70. He, H.; Yang, D.; Wang, S.; Wang, S.; Liu, X. Road segmentation of cross-modal remote sensing images using deep segmentation network and transfer learning. *Ind. Robot Int. J.* **2018**. [CrossRef]
71. Xia, W.; Zhang, Y.-Z.; Liu, J.; Luo, L.; Yang, K. Road Extraction from High Resolution Image with Deep Convolution Network—A Case Study of GF-2 Image. *Proceedings* **2018**, *2*, 325. [CrossRef]
72. Gao, L.; Song, W.; Dai, J.; Chen, Y. Road Extraction from High-Resolution Remote Sensing Imagery Using Refined Deep Residual Convolutional Neural Network. *Remote Sens.* **2019**, *11*, 552. [CrossRef]
73. Xie, Y.; Miao, F.; Zhou, K.; Peng, J. HsgNet: A Road Extraction Network Based on Global Perception of High-Order Spatial Information. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 571. [CrossRef]
74. Shi, Q.; Liu, X.; Li, X. Road detection from remote sensing images by generative adversarial networks. *IEEE Access* **2018**, *6*, 25486–25494. [CrossRef]

75. Costea, D.; Marcu, A.; Slusanschi, E.; Leordeanu, M. Creating roadmaps in aerial images with generative adversarial networks and smoothing-based optimization. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Venice, Italy, 22–29 October 2017; pp. 2100–2109.
76. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
77. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Venice, Italy, 22–29 October 2017; pp. 3431–3440. [CrossRef]
78. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
79. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef]
80. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv* **2014**, arXiv:1412.7062.
81. Luc, P.; Couprie, C.; Chintala, S.; Verbeek, J. Semantic segmentation using adversarial networks. *arXiv* **2016**, arXiv:1611.08408.
82. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; pp. 2672–2680.
83. Dai, J.; Zhu, T.; Wang, Y.; Ma, R.; Fang, X. Road extraction from high-resolution satellite images based on multiple descriptors. *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.* **2020**, *13*, 227–240. [CrossRef]
84. Ghasemkhani, N.; Vayghan, S.S.; Abdollahi, A.; Pradhan, B.; Alamri, A. Urban Development Modeling Using Integrated Fuzzy Systems, Ordered Weighted Averaging (OWA), and Geospatial Techniques. *Sustainability* **2020**, *12*, 809. [CrossRef]