applied sciences

MDPI

Article

Distorted Aerial Images Semantic Segmentation Method for Software-Based Analog Image Receivers Using Deep Combined Learning

Kalupahanage Dilusha Malintha De Silva and Hyo Jong Lee *

Division of Computer Science and Engineering, CAIIT, Jeonbuk National University, Jeonju 54896, Republic of Korea
* Correspondence: hlee@jbnu.ac.kr

Abstract: Aerial images are important for monitoring land cover and land resource management. An aerial imaging source which keeps its position at a higher altitude, and which has a considerable duration of airtime, employs wireless communications for sending images to relevant receivers. An aerial image must be transmitted from the image source to a ground station where it can be stored and analyzed. Due to transmission errors, aerial images which are received from an image transmitter contain distortions which can affect the quality of the images, causing noise, color shifts, and other issues that can impact the accuracy of semantic segmentation and the usefulness of the information contained in the images. Current semantic segmentation methods discard distorted images, which makes the available dataset small or treats them as normal images, which causes poor segmentation results. This paper proposes a deep-learning-based semantic segmentation method for distorted aerial images. For different receivers, distortions occur differently, and by considering the receiver specificness of the distortions, the proposed method was able to grasp the acceptability for a distorted image using semantic segmentation models trained with large aerial image datasets to build a combined model that can effectively segment a distorted aerial image which was received by an analog image receiver. Two combined deep learning models, an approximating model, and a segmentation model were trained combinedly to maximize the segmentation score for distorted images. The results showed that the combined learning method achieves higher intersection-over-union (IoU) scores than the results obtained by using only a segmentation model.

Keywords: semantic segmentation; deep learning; aerial images; image enhancement

Citation: De Silva, K.D.M.; Lee, H.J. Distorted Aerial Images Semantic Segmentation Method for Software-Based Analog Image Receivers Using Deep Combined Learning. *Appl. Sci.* **2023**, *13*, 6816. https://doi.org/10.3390/app13116816

Academic Editor: Zhengjun Liu

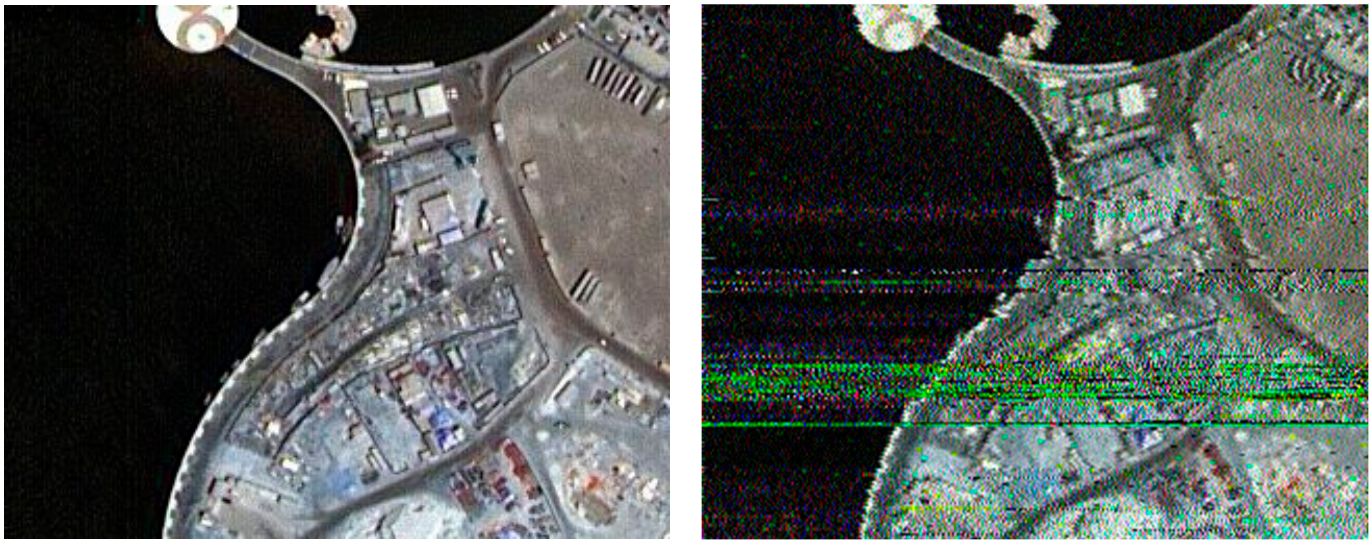
Received: 2 May 2023
Revised: 29 May 2023
Accepted: 31 May 2023
Published: 4 June 2023

1. Introduction

Earth-orbiting satellites and unmanned aerial vehicles (UAVs) are a crucial source of aerial images. Detailed and comprehensive studies on aerial images are important for useful land cover examination. For a specific area of land, the diversity of resources and the correct assessment of the availabilities and capabilities of each resource make a clear view for efficient land management. For a given duration of time, earth surfaces may go through many changes due to human activity, climate change, and natural disasters, hence frequent land cover monitoring is helpful for impact assessment. Undergoing changes such as natural processes and social and economic events made the change-detecting process an active research field [1]. Possibilities for good urban planning are largely dependent on aerial images [2], and vegetation planning [3] is another major activity. Pre- and post-disaster aerial images can be compared for detecting the impacts caused by the disaster [4]. For situations which need immediate attention, such as traffic monitoring [5] and the search and rescue of humans [6], aerial images provide crucial assistance.

Aerial photos, which are taken from earth-orbiting satellites or UAVs equipped with camera sensors, have a viewpoint from a higher altitude. Satellites and UAVs with longer airtime need to send taken images to image receivers. Many studies use aerial imagery

without considering technological details such as image transmission from source to ground stations. The quality of the images depends on the image-capturing mechanism of the source and the image transmission process. Sending an aerial image to a receiving ground station is practically achieved by establishing a wireless communication link between the satellite itself and the ground station. During transmission, radio communications can be disturbed in many ways, such as attenuation and interference. When this happens, inevitably, some information regarding the aerial image will be lost, causing receiver dissatisfaction. Image information which was originally modulated to the radio carrier will not be demodulated or recovered correctly. This will result in distorted images at the receiver's end. Figure 1 shows an example of a source (transmitted) image and the resulting received image.



(**a**) (**b**)

Figure 1. (**a**) Transmitted image and (**b**) received image.

Advancements in deep learning architectures for image classifications and big data analysis have demonstrated the use of such techniques in land cover classification and changed regions detection in aerial images [7]. Labor-intensive tasks, such as the mapping of road networks in aerial images, are becoming inexpensive with deep learning [8]. For damage assessment in timely disaster risk management, an adequate deep-learning-based framework has been proposed [9]. Many aerial image datasets, such as [10], are presented with fine annotations and readily available for semantic segmentation. Images and their corresponding annotations are applicable for popular deep learning segmentation models such as U-net [11] and Linknet [12]. However, in the case of certain transmission systems, consideration of the uncertainty of having some distorted aerial images to be semantically segmented and the results of such images with existing pretrained semantic segmentation models will not be better than images with no distortions. Another challenging fact is the lack of distorted aerial images to train a segmentation model. The manufacturing process of satellite image receivers is still unable to properly solve the problem of distortions in received images. Existing distortions have made it difficult to correctly examine aerial data.

This paper proposes a semantic segmentation method for distorted aerial images which can easily be implemented in software-based radio receivers to provide it as an added functionality to the radio. The image approximating network included in the proposed method can improve the quality of a distorted image in such a way that it is more applicable to the segmentation model, hence the combined model can predict segmentations for distorted aerial images with a higher accuracy than the results produced by using only the pretrained segmentation model. This work skillfully combines several concepts, approaches, techniques, and components such as semantic segmentation, deep

learning, aerial images, and existing segmentation models if needed, such as U-net and Linknet. Further, according to the user's preference or a model's appropriateness in a certain problem domain, different segmentation model architectures can be used as a module in the proposed method. Contributions from the proposed method are listed as follows.

- We propose a combined deep learning model of an approximating network to be used with a segmentation network. A programmable interconnection between the approximating block and the segmentation block provides the possibility of changing, training, or adjusting participating deep learning models according to the relevance of the problem to be solved.
- A comprehensive loss function is proposed to train the combined model optimally.
- The proposed method provides a compatible implementation of the software-based image receiver and aerial image segmentation in a small-scale computer such as a single board computer (SBC).
- The developed segmentation model is compared with similar benchmark networks to demonstrate the robustness of the proposed method, verifying that the proposed method obtains relative improvements of up to 80% in terms of mean IoU.

2. Related Works

2.1. Classical Segmentation Methods

Classical segmentation includes edge-based, region-based, and threshold-based methods. Edge-based segmentation relies on detecting edges within an image by identifying local changes in image intensity. However, this method is not suitable for images with smooth edges or many edges. In contrast, region-based segmentation depends on a seed point to initiate the segmentation process, where the region grows by examining neighboring pixels' intensity to determine whether to include them, thus separating the regions. This technique is computationally expensive, and different seed points may result in varying segmentation outcomes, which undermines the segmentation accuracy. Combining edge-based and region-based techniques addresses their individual shortcomings and leads to a more robust segmentation technique [13,14]. Threshold-based segmentation is one of the simplest and most commonly used techniques [15,16]. It involves calculating an optimal threshold that distinguishes between two classes while minimizing intra-class variance and maximizing inter-class variance. This method performs well when the image histogram has a bimodal distribution but cannot process images with unimodal intensity distribution [17–19].

2.2. Deep-Learning-Based Semantic Segmentation

A pixel of an image can be in a part of a certain region in that image with its own characteristics to form a region type, and an image is a combination of one or more region types. Semantic segmentation correctly labels each pixel according to its relevance to a certain region type. In segmented output, spatial information must be retained until the end result. A software-based radio is likely to be a portable device, hence it can be implemented in a mini-computer or an SBC. Its computing capabilities must be considered before designing a proper deep leaning model to be implemented in such a device. Deep convolutional neural networks (CNNs) which are successful at image classifications such as VGG16 [20] have continuing convolutional layers for feature extraction. The work presented in [21] proves the capability of end-to-end training methods for pixel-to-pixel semantic segmentation with fully convolutional networks (FCNs). These networks take an arbitrary input size and produce a same-sized segmented output. In FCNs, pretrained deep CNNs such as AlexNet [22], VGG [20], and GoogLeNet [23] have been used for segmentation. AlexNet and GoogLeNet are much easier to test with a low power portable computer which can be used as a software-based image receiver. Internal mathematical operations other than tensor calculations of a deep learning model decide the additional processing power and the amount of memory needed by it. Deep deconvolutional neural

networks have eliminated the need of saving pooling indices [24,25]. Auto-encoders have been an inspiration for many segmentation techniques [12,26,27]. An input is encoded to a feature space which can be decoded into spatial categorization to achieve segmentation. SegNet [25] consists of an encoder which was followed by a decoder network, and it has topological similarities with VGG16. A better candidate to be implemented in an SBC is LinkNet [12], and it has a mechanism to bypass spatial information directly from its corresponding encoders to decoder blocks. At each encoder, there is a possibility of information loss, but LinkNet can preserve a considerably large amount of information without losing details. ParseNet is an end-to-end, effective CNN for semantic segmentation, which uses a technique to add global context to full CNNs [28].

Ronneburger et al. introduced U-net [11] for biomedical images semantic segmentation built upon FCNs [21]. Its internal concatenation operation can be performed in a system with low computational power. Its structure has a contracting path for context capturing and a symmetric expanding path for precise localization. Feature pyramid networks (FPNs), which are more tolerant of images with distortions because of their internal structure, showed significant improvement as a feature extractor and can be used for both object detection and semantic segmentation [29]. A robust segmentation method for noisy images was introduced by using an unsupervised denoising filter [30] for real-time images. A scalable subspace clustering method [31] was proposed, which includes a concise dictionary and robust subspace representation in a unified model.

2.3. Multiple Model Training Methods

The work of training and evaluating multiple neural networks within one training step is tested for adjusting their training parameters to achieve a collaborative result. For example, generative adversarial networks (GANs) brought a generator network and a discriminator network which can be connected and trained together [32]. A loss function for a GAN is formed by combining two different prediction error functions, one for the generator and one for the discriminator. SR-GAN [33] employees a deep residual network, a ResNet [34] structure, with skip connections as a generator combined with a discriminator. It also brought a more intuitive combined loss function. Pix2pix GAN was tested to use a segmentation model as its generator, which brought attention to the segmentation-oriented loss function to be in a combined loss definition [35]. AIDEDNet [36] and MATR [37] made use of multiple internal deep learning models for better results. Such multiple networks require the mandatory need of training each subnetwork simultaneously with the same dataset, preventing the option of training the subnetworks separately. For medical images segmentation, a dual adversarial attention mechanism was proposed [38] which has included two inputs for two sub networks.

3. Methodology

A segmentation model can be trained with an available aerial image dataset with thousands of images, but these datasets do not include aerial images with distortions. Pretrained segmentation models produce erroneous results for a distorted aerial image because they have not learnt about them. The expansion of the learning space of an existing segmentation model can be done so that already learned parameters are not changed. In this work, the main task is to fuse a limb of extra deep CNNs prior to a segmentation model while keeping the compatibility over the conjunction, so that the whole structure will be collaboratively acting as a segmentation model. Throughout the work, the possibility of implementing the proposed method without exceeding the capabilities of software-based radio hardware was a high concern. The inadequacy of distorted data prevents the proposed method from using the multiple network training strategies which were used in [32–35]. Because each participating network must be trained simultaneously, the accuracy will be a low value for unseen data as the training distorted aerial image set is insufficient.

The extra CNN to be fused prior to the pre-trained segmentation model is an approximator model which has the duty of producing a more applicable input to the segmentation model. Through the programmatically made conjunction, its output is presented to the segmentation model. In Section 5, the performance of the proposed approximator model will be tested. The following are details of the proposed combined method.

- The approximator model and the segmentation model must keep compatibility at their connecting point. The expected output of the approximator model is given to the segmentation model.
- Each model is composed of modules. Combinations of different approximators and different segmentation models are possible.
- The proposed method follows a modular approach, according to the user's preference, so different approximators and different segmentation models can be used.

3.1. Dataset

Satellite images of the Dubai dataset [39] provide their volume with segmentation labels in six classes. The classes are building, land, road, vegetation, water, and unlabeled. The provided images and segmentation labels are cropped to a size of 320 × 256. The total number of images is 1404. Even though the dataset provides its own color scheme in segmentation labels, this work used more intuitive colors for each region type (buildings: blue, vegetation: green, water: light blue, roads: yellow, unlabeled: gray). Samples from the Dubai dataset are shown in Figure 2.

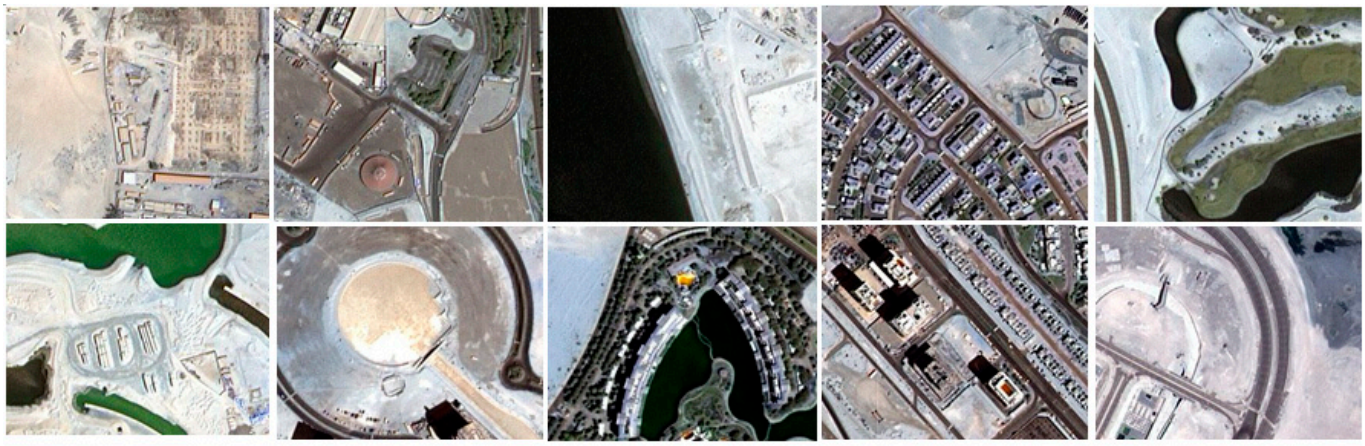


Figure 2. Sample images from the Dubai dataset [39]. All images were cropped to a size of 320 × 256.

3.2. Distorted Aerial Images

The workflow used in analog image transmitters and receivers is depicted in Figure 3. An aerial image from a proper source is converted to set of low-frequency tones between 0 Hz and 3 kHz. These tones are frequency-modulated with a radio frequency (RF) carrier. The characteristics of the RF carrier are determined according to the relevance of the situation, how the transmission process must be accomplished, and by considering the possibilities of acquiring a proper license to transmit. For the work conducted in this paper, a 433 MHz carrier frequency was used, which is a part of the non-licensed industrial, scientific, and medical (ISM) band. The modulated RF carrier is amplified and transmitted over the air by using a proper antenna. The image receiver obtains that signal from its antenna, demodulates it, then de-converts it to obtain the image which was sent.

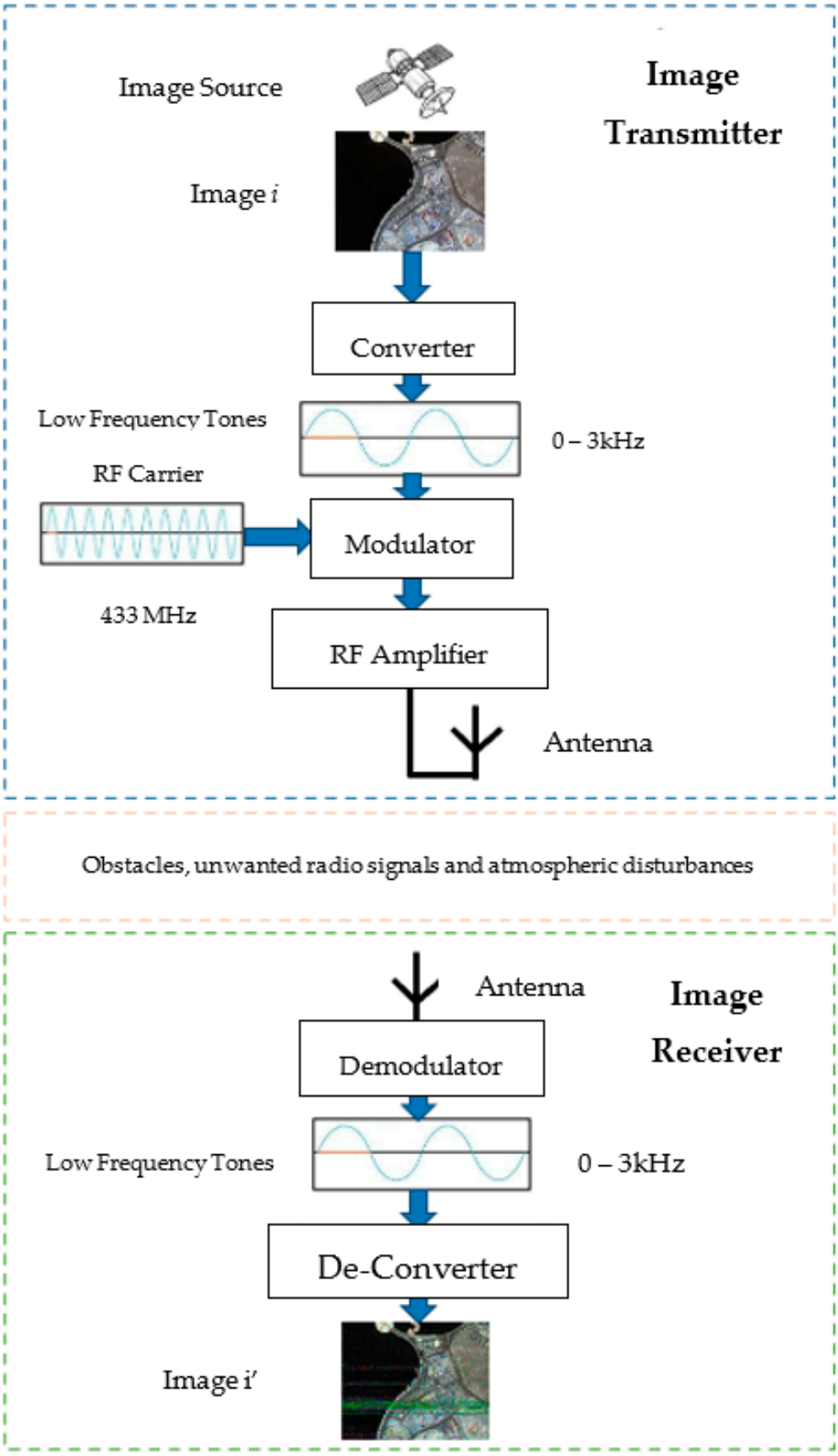


Figure 3. Basic workflow of the image transmission process from image source to image receiver.

To denote a situation where distortions are inevitable, an analog image transmitter was emulated to capture noisy aerial images. The transmitter and receiver were kept at a distance with many obstacles between them. The details of an image are converted to many low-frequency tones to be frequency-modulated with a radio carrier wave at the transmitter. The narrowband frequency modulation (NBFM) type was used, and the carrier frequency was 433 MHz. Image i was transmitted through the air, received as image i'. With this process, 120 images were collected. Figure 4 shows dissimilarities between the transmitted image and the received image.

(**a**) (**b**)

Figure 4. The source (transmitted) image (**a**) and the received image (**b**).

3.3. Approximator Model

For the approximator model, which is a deep CNN, we propose a network which is a combination of convolutional blocks and residual blocks. The expectation is to find better pixel values which are near appropriate for the distorted regions and to keep the original information for a given input. The proposed network has a symmetrical structure which is capable of learning with end-to-end mapping from input distorted images to non-distorted targets. Unlike the generator of SR-GAN, the proposed structure does not have residual connections from start to end. The proposed approximator network is shown in Figure 5.

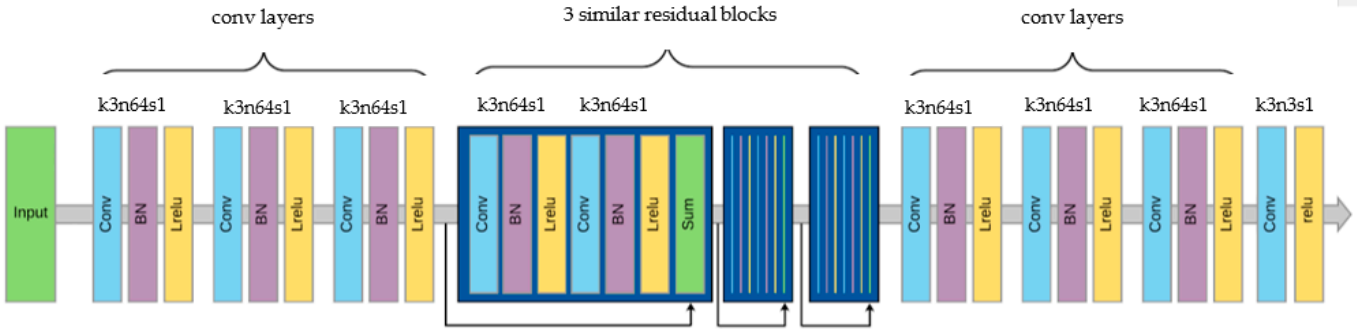


Figure 5. The proposed approximator network.

A single convolutional layer has a 3 × 3 kernel with 64 feature maps that use the same padding and (1, 1) strides. It is followed by a batch normalization operation (momentum = 0.5) and leaky rectified linear unit (Lrelu) activation function (α = 0.2). The architecture has residual blocks in the middle with skip connections. A single residual block contains two convolutional layers which increase the depth of the structure and involve

symmetric skip connections to improve efficiency in training. The skip connections keep forwarding information to the beginnings of the next blocks to retain spatial information which helps faster convergence. The final convolutional layer has three filters, followed by a rectified linear unit (relu) activation function.

3.4. Segmentation Model

From the chosen dataset, 1404 available aerial images and their corresponding segmentation labels were used to train the segmentation model. The dataset provides six different classes so that the output layer of the segmentation model is made for six categories. In other words, for each input pixel, the output will be a list with a length of six for each pixel, which provides the probabilities as to which class that pixel belongs to. As an accuracy metric, the basic definition of intersection-over-union (IoU) is used. A is the prediction set and B is the label set.

$$IoU = (A \cap B)/(A \cup B) \tag{1}$$

Since there are six region types in the chosen dataset, the total number of classes $C = 6$. For each class c, the IoU_c is defines as:

$$IoU_c = \frac{TP_c}{TP_c + FP_c + FN_c} \tag{2}$$

TP_c is the number of true positive pixels for a class $c \in C$, and it is divided by the total number of pixels in the union in the prediction set and label set. From (2), the mean IoU is defined as:

$$mIoU = \frac{1}{6}\sum_{c=1}^{6} IoU_c \tag{3}$$

The Jaccard loss was utilized as a loss function for training. The following is its definition for prediction set A and label set B.

$$L(A, B) = 1 - (A \cap B)/(A \cup B) \tag{4}$$

3.5. Combined Loss

Basically, two models are being used for accomplishing better approximation and segmentation. To decipher the pixel-wise difference per approximation, the mean squared error (MSE) loss function is used in the approximator model training. For an input distorted image I^{Inp} with a size of $W \times H$, the referencing image is I^{Ref}, and its approximation $A(I^{Inp})$ and loss of the approximator L_{AE} is defined as follows.

$$L_{AE} = \left(\frac{1}{WH} \sum_{x=1}^{W} \sum_{y=1}^{H} \left(I_{x,y}^{Ref} - A\left(I^{Inp}\right)_{x,y} \right)^2 \right) \tag{5}$$

The proposed method is not fixating a specific loss function for the approximating model or the segmentation model. The method itself is using models as modules. If the loss of the approximator model is L_{AM} and the loss of the segmentation model is L_{SM}, then the combined loss L_C is defined as:

$$L_C = \lambda a \, L_{AM} + \lambda s \, L_{SM} \tag{6}$$

where λa and λs are constants. Based on training observations with Jaccard loss and MSE, $\lambda a = 1$ and $\lambda s = 10^{-2}$ are the best for training the combined model properly.

4. Experiments

First, the proposed approximator network is trained end-to-end with distorted aerial images targeting referencing images to make a proper approximation for the segmentation model. Mainly it is conducted to see the efficiency of the derived combined loss function. For that task, we considered only the loss of the approximator model L_{AM}. Distorted and approximated aerial images can be segmented using a pre-trained U-net to compare the effects of approximation. Finally, by using the proposed combined learning method, the proposed approximator network is trained again, and the results are segmented. Figure 6 depicts the combined learning method. The loss of the approximator model is L_{AM}, and the loss of the segmentation model is L_{SM}. Learning is conducted according to the combined loss L_C. However, the weights of the segmentation model are not updated during the training process.

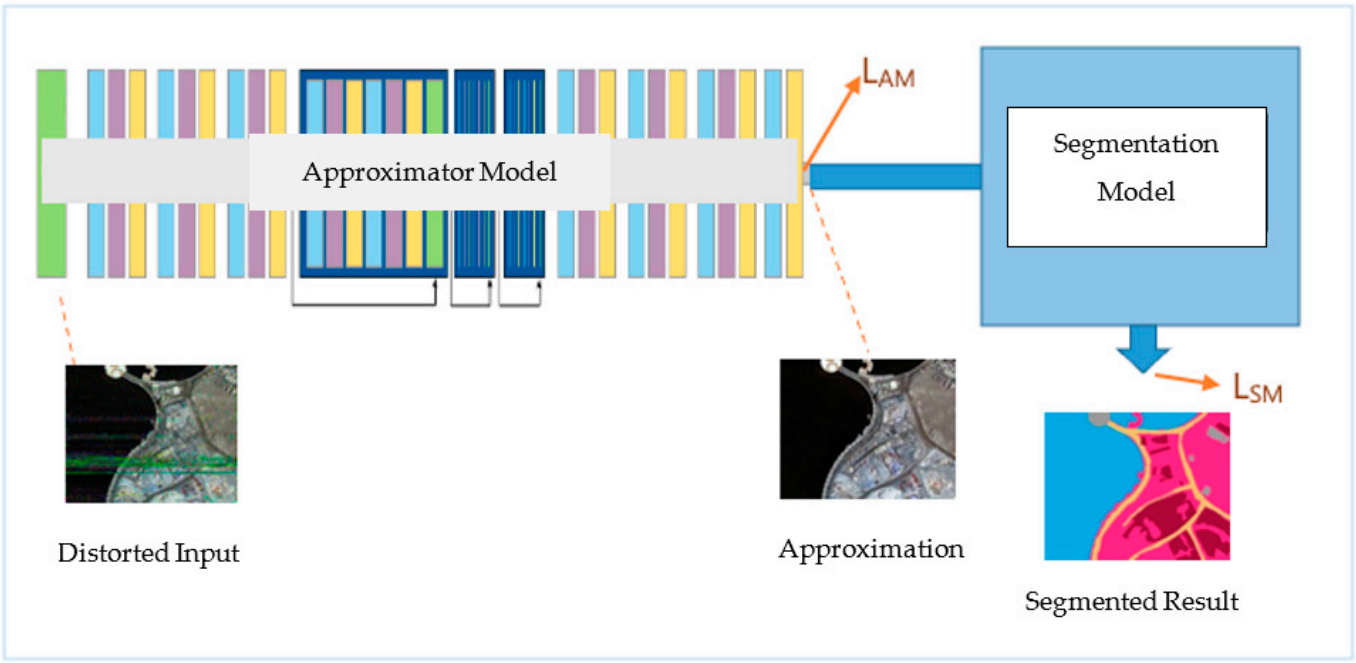


Figure 6. The proposed combined learning method. Approximator and segmentation model are fused together to behave as a single integrated model.

In the training processes, the inputs to the combined model are distorted images, for the approximator model targeting source images, and for the segmentation model targeting segmentation labels. We wanted the combined model to predict segmentations for both distorted and non-distorted images, so from the collected 120 images, 100 of them and their source images were used in the training of the approximator. The remaining 20 images were kept for testing. However, for the segmentation model, all 1404 images were used. Figure 7 shows 5 source images from the remaining 20, which were used to transmit, their corresponding segmentation labels and received instances.

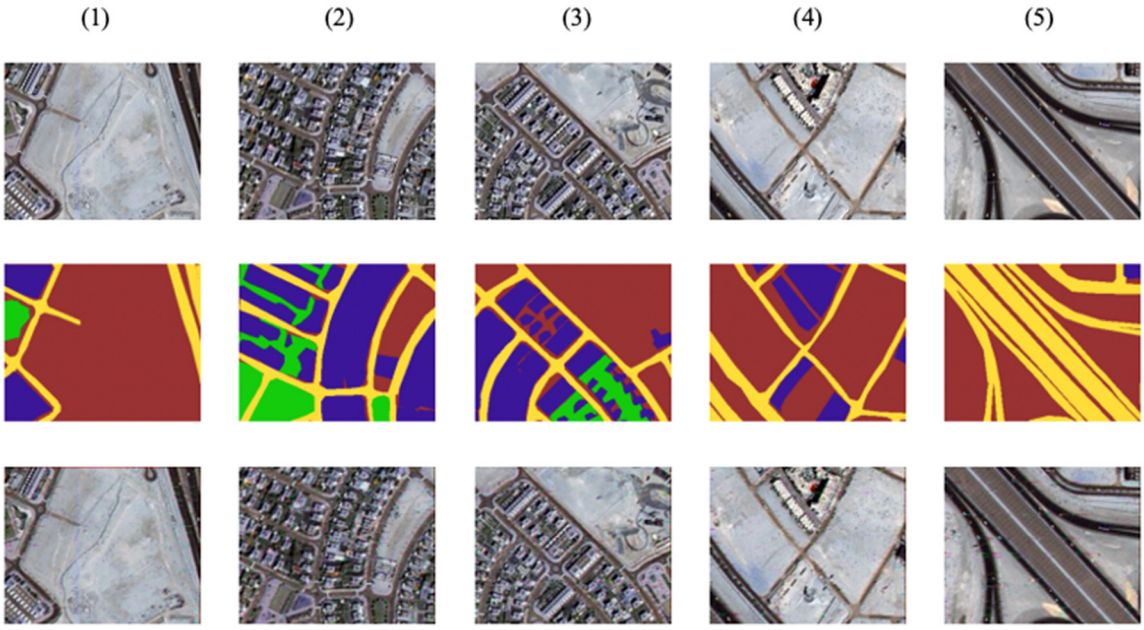


Figure 7. Samples of testing images (**top**), their corresponding segmentation labels (**middle**), and the received images (**bottom**) for results comparison. Images are named 1 to 5 for cross referencing them respectively between other figures.

5. Results

In each training scenario described in this paper, the segmentation models were trained with 500 epochs, and approximator models were trained for 1000 epochs with the Adam optimizer [40], whose parameters were set as follows: the learning rate was 0.001, β_1 was 0.9, and β_2 was 0.999. Training was completed on an NVIDIA Quadro 6000 RTX GPU (by NVIDIA corporation in Santa Clara, CA, USA) with the tensorflow framework. For a pretrained segmentation model, U-net was trained with the 1404 available images. The mean IoU was the evaluation metric used for the test segmentation results. Usually, the mean IoU is calculated for a set of images, but the same function can be utilized for a single image by setting the number of images per set to one. We tested three inference methods. (1) Segmentation model only—the received distorted image is directly passed to the segmentation model, (2) approximator trained separately—approximator network is trained only according to L_{AM}, and (3) the proposed method—approximator was trained according to both L_{AM} and L_{SM} (Figure 6). The obtained IoU values are shown below for each result.

All approximations were segmented by using pretrained U-net, and the IoU scores were obtained and are shown below for each segmentation result for the proposed approximator model. Based on the observations, the proposed approximator model which was trained by using the proposed combined learning method presented a low error in approximation and a higher IoU score for semantic segmentation.

Since this research field is narrow with limited related works, popular additional segmentation models, Linknet [12] and FPN [29], were tested to prove the modularity of the proposed combined learning method. These methods can be incoporated with many backbones such as Resnet18 [34], SeResnet18 [41], DenseNet121 [42], InceptionV3 [23], MobileNetV2 [43], and EfficientNetB0 [44]. The proposed approximator model was trained separately and combined with Linknet and FPN. The results are shown in Figures 8–10 for five testing images and their corresponding segmentation labels (the obtained IoU score is shown below each result). Table 1 summarizes the effects of the proposed combined learning method in terms of the mean IoU score. The obvious fact is that the segmentation models themselves could not have a higher segmentation score for distorted images. We have reached a point where we are closer to the related literature, which is worked out as

the approximator trained separately, and it presented an improvement of 5.52% for Unet, 10.0% for Linknet, and a 0.6% reduction in FPN. However, by using the proposed combined method, we have achieved 65.83% improvement for Unet, 75.36% for Linknet, and 53.6% for FPN. Table 2 shows a summary of the results for different backbones. Even without the proposed method, some backbones presented a relatively higher IoU score than that which was obtained with the separately trained approximator as they are much more tolerant of images with distortions. In this case, backbones such as DenseNet121 produced a slightly lower IoU score with a separately trained approximator. However, with the proposed method, the same backbone presented a higher IoU score.

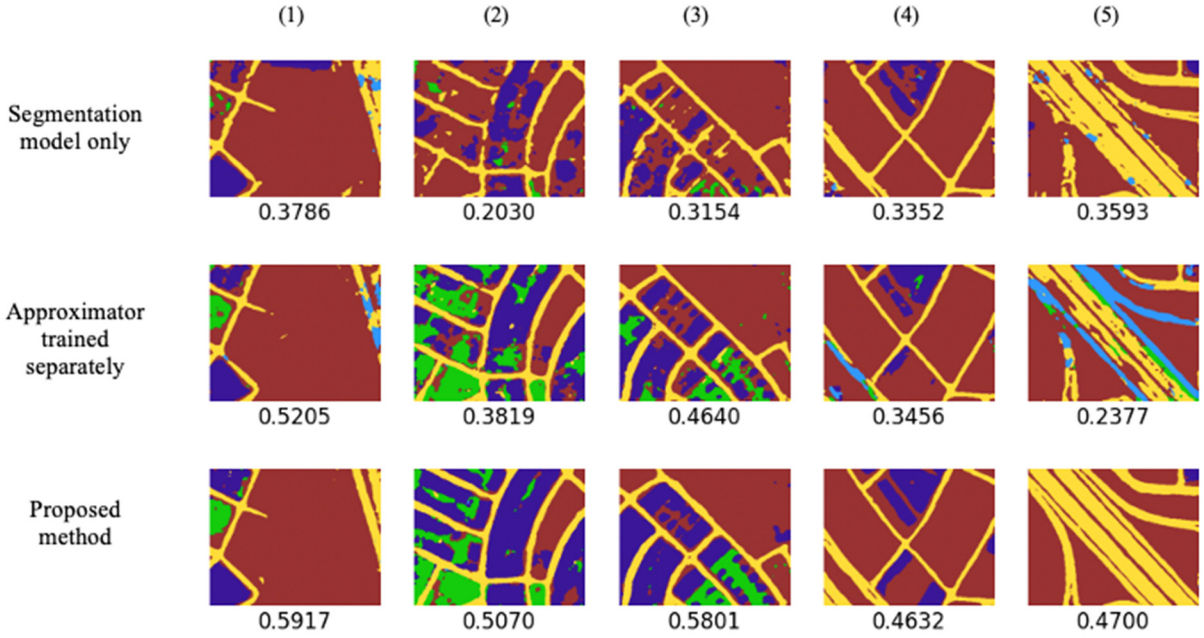


Figure 8. Results (IoU score) obtained from the proposed approximator network and U-net. Images are named 1 to 5 for cross referencing them respectively between other figures.

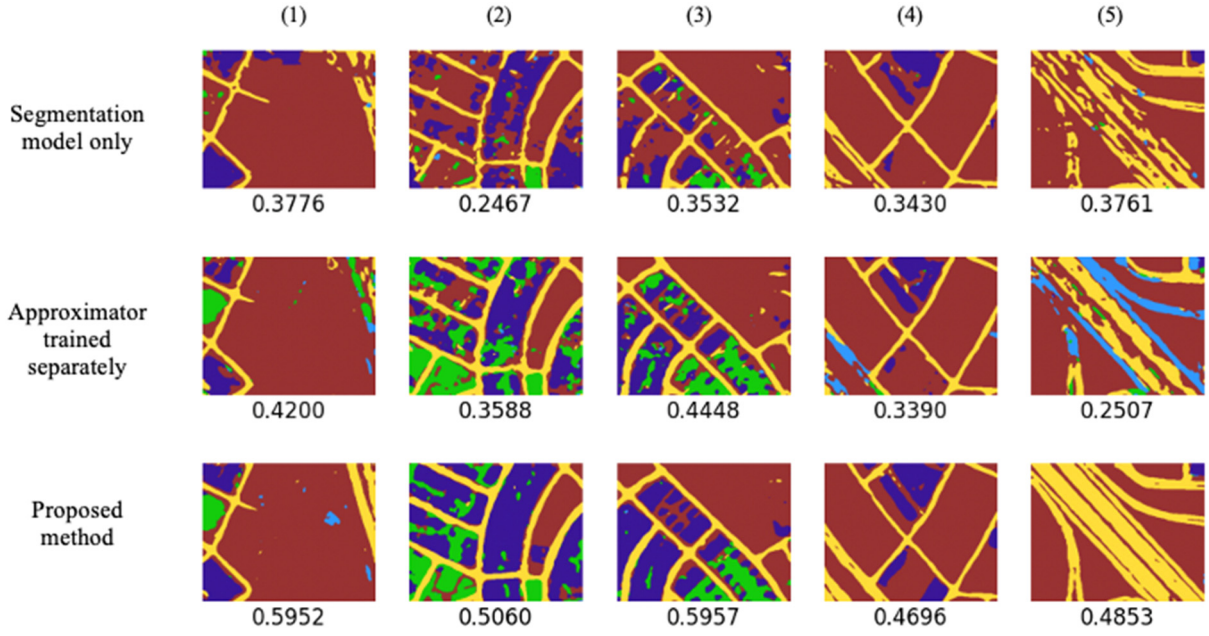


Figure 9. Results (IoU score) obtained from the proposed approximator network and Linknet. Images are named 1 to 5 for cross referencing them respectively between other figures.

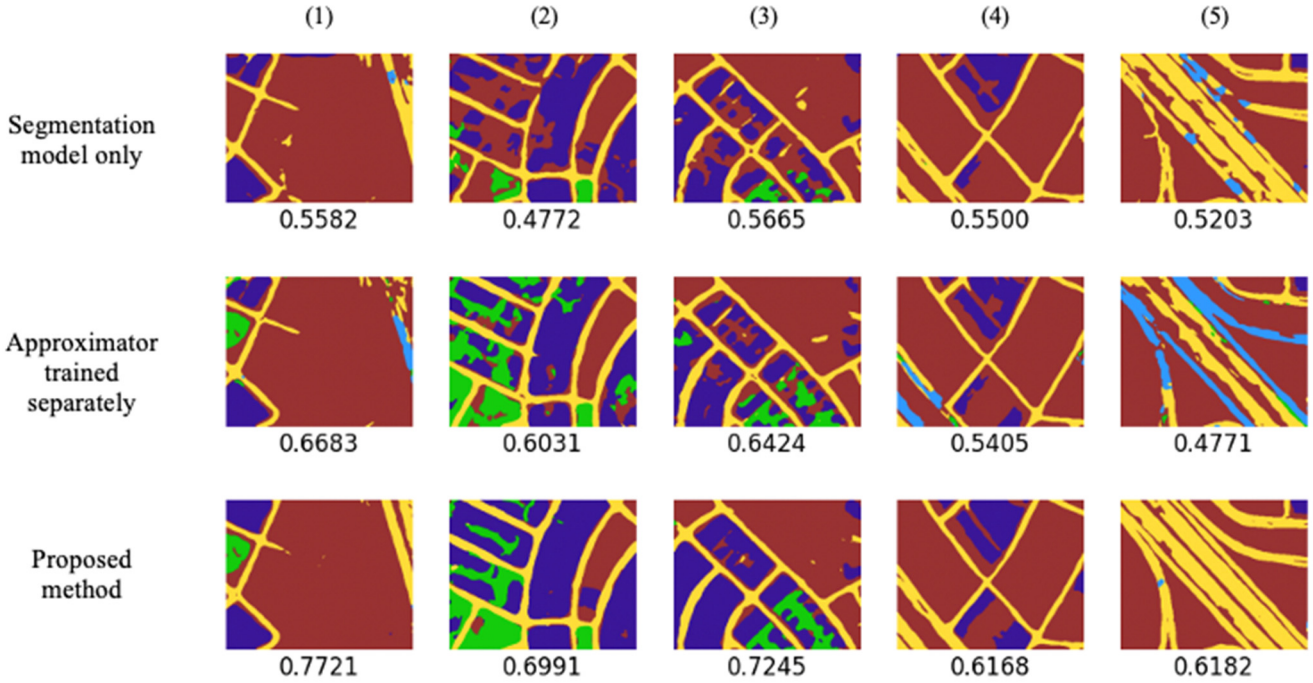


Figure 10. Results (IoU score) obtained from the proposed approximator network and FPN. Images are named 1 to 5 for cross referencing them respectively between other figures.

Table 1. Effects of the proposed method (mean IoU with standard dev.).

Inference Method	U-Net	Linknet	FPN
Segmentation model only	0.3870 (0.008)	0.3970 (0.012)	0.4615 (0.006)
Approximator trained separately	0.4084 (0.004)	0.4367 (0.005)	0.4587 (0.007)
Proposed method	0.6418 (0.026)	0.6962 (0.034)	0.7089 (0.027)

Table 2. Summary of results for different backbones (mean IoU with standard dev.).

Segmentation Model	Backbone	Segmentation Model Only	Approximator Trained Separately	Proposed Method
Unet	Resnet18 [34]	0.7022 (0.015)	0.6238 (0.002)	0.7749 (0.0004)
	SeResnet18 [41]	0.7109 (0.017)	0.7229 (0.020)	0.7873 (0.0011)
	DenseNet121 [42]	0.6433 (0.004)	0.5423 (0.001)	0.7139 (0.0015)
	InceptionV3 [23]	0.3541 (0.050)	0.4881 (0.008)	0.7065 (0.0022)
	MobileNetV2 [43]	0.4024 (0.031)	0.5283 (0.002)	0.7330 (0.0004)
	EfficientNetB0 [44]	0.4836 (0.009)	0.5773 (2×10^{-6})	0.7689 (0.0002)
Linknet	Resnet18 [34]	0.7133 (0.017)	0.5866 (6×10^{-5})	0.7706 (0.0002)
	SeResnet18 [41]	0.6759 (0.009)	0.6426 (0.004)	0.7752 (0.0004)
	DenseNet121 [42]	0.6946 (0.013)	0.6109 (0.001)	0.7569 (9×10^{-6})
	InceptionV3 [23]	0.4850 (0.008)	0.5034 (0.005)	0.7252 (0.0008)
	MobileNetV2 [43]	0.4009 (0.031)	0.4426 (0.018)	0.6289 (0.0156)
	EfficientNetB0 [44]	0.6847 (0.011)	0.6132 (0.001)	0.7586 (2×10^{-5})
FPN	InceptionV3 [23]	0.7327 (0.023)	0.7086 (0.016)	0.9189 (0.0272)
	MobileNetV2 [43]	0.4293 (0.022)	0.5128 (0.004)	0.7357 (0.0003)

To compare different backbones quantitatively, Table 3 summarizes parameter counts and computational complexity. The number of parameters is calculated for different combinations of the approximator network and segmentation models with different backbones. The evaluation time per a batch size of 20, which was taken by each combined model

entity, was measured and included. Larger combined models with backbones, such as InceptionV3, took a longer time to evaluate. Some combined models with backbones such as DenseNet121 took a longer amount of time because of the relatively higher internal mathematical operations which are needed to perform the evaluation.

Table 3. Parameters and computational complexity of combined networks.

Segmentation Model	Backbone	Parameters in Combined Form (M)	Evaluation Time (ms/batch)
Unet	Resnet18	14.75	288
	SeResnet18	14.84	286
	DenseNet121	12.55	306
	InceptionV3	30.34	316
	MobileNetV2	8.46	288
	EfficientNetB0	10.52	283
Linknet	Resnet18	11.93	266
	SeResnet18	12.02	275
	DenseNet121	8.76	294
	InceptionV3	26.68	311
	MobileNetV2	4.55	290
	EfficientNetB0	6.50	305
FPN	InceptionV3	25.44	367
	MobileNetV2	5.62	342

6. Conclusions

In this paper, we proposed an image segmentation method for distorted aerial images, which can be used in software-based aerial images reception and analyzing processes. Depending on an individual or an organization, methods of communication differ according to the scale of materials to be transmitted and received. In image communications, a variety of image receivers have suffered distortions in received contents. With a few hundred available distorted aerial images, the proposed method gained the advantage of using a segmentation model which was trained with thousands of aerial images and segmentation labels. Without a need for training a segmentation model, the added approximator model has taken the obligation of presenting a proper input to a pre-trained segmentation model. The proposed method employed a modular approach so that different combinations of approximator models and segmentation models can be used. Deep learning models which are yet to be introduced can be used as modules, hence perfectly futureproofing the software and hardware compatibility for receivers to use the proposed method. For controller software with a GUI, users can choose their preferred models from a list. The results obtained by using the proposed method showed a significant improvement in the semantic segmentation of distorted aerial images. For future works, we will increase the distorted aerial image dataset by performing more image transmitting activities and expand the concept into the general image communications over radio-waves-related applications and for satellite images from other planets. It is a transformation to receive images from much larger standards of image transmitting methods and to test different software-based image receiver hardware for seeking more implementations for the proposed combined learning method.

Author Contributions: K.D.M.D.S. designed and developed the proposed method, conducted the experiments, and wrote the manuscript. H.J.L. designed the new concept, provided the conceptual idea and insightful suggestions to refine it further, and reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by project for Joint Demand Technology R&D of Regional SMEs funded by Korea Ministry of SMEs and Startups in 2023 (RS-2023-00207672).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Unavailable due to further research.

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

References

1. Gerard, F.; Petit, S.; Smith, G.; Thomson, A.; Brown, N.; Manchester, S.; Wadsworth, R.; Bugár, G.; Halada, L.; Bezák, P.; et al. Land cover change in Europe between 1950 and 2000 determined employing aerial photography. *Prog. Phys. Geogr. Earth Environ.* **2010**, *34*, 183–205. [CrossRef]
2. Zhou, W.; Huang, G.; Cadenasso, M.L. Does spatial configuration matter? Understanding the effects of land cover pattern on land surface temperature in urban landscapes. *Landsc. Urban Plan.* **2011**, *102*, 54–63. [CrossRef]
3. Ahmed, O.; Shemrock, A.; Chabot, D.; Dillon, C.; Wasson, R.; Franklin, S. Hier-archical land cover and vegetation classification using multispectral data acquired from an unmanned aerial vehicle. *Int. J. Remote Sens.* **2017**, *38*, 2037–2052. [CrossRef]
4. Gupta, A.; Watson, S.; Yin, H. Deep learning-based aerial image segmentation with open data for disaster impact assessment. *Neurocomputing* **2021**, *439*, 22–33. [CrossRef]
5. Kyrkou, C.; Timotheou, S.; Kolios, P.; Theocharides, T.; Panayiotou, C.G. Optimized vision-directed deployment of UAVs for rapid traffic monitoring. In Proceedings of the 2018 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 12–14 January 2018; pp. 1–6. [CrossRef]
6. Petrides, P.; Kyrkou, C.; Kolios, P.; Theocharides, T.; Panayiotou, C. Towards a holistic performance evaluation framework for drone-based object detection. In Proceedings of the 2017 International Conference on Unmanned Aircraft Systems (ICUAS), Miami, FL, USA, 13–16 June 2017; pp. 1785–1793. [CrossRef]
7. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.-S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [CrossRef]
8. Bastani, F.; He, S.; Abbar, S.; Alizadeh, M.; Balakrishnan, H.; Chawla, S.; Madden, S.; DeWitt, D. RoadTracer: Automatic Extraction of Road Networks from Aerial Images. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4720–4728.
9. Gupta, A.; Welburn, E.; Watson, S.; Yin, H. CNN-Based Semantic Change Detection in Satellite Imagery. In Proceedings of the Artificial Neural Networks and Machine Learning–ICANN 2019: Workshop and Special Sessions: 28th International Conference on Artificial Neural Networks, Munich, Germany, 17–19 September 2019; pp. 669–684. [CrossRef]
10. Boguszewski, A.; Batorski, D.; Ziemba-Jankowska, N.; Dziedzic, T.; Zambrzycka, A. LandCover.ai: Dataset for Automatic Mapping of Buildings, Woodlands, Water and Roads from Aerial Imagery. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Nashville, TN, USA, 20–25 June 2021; pp. 1102–1110.
11. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Munich, Germany, 5–9 October 2015; pp. 234–241.
12. Chaurasia, A.; Culurciello, E. LinkNet: Exploiting encoder representations for efficient semantic segmentation. In Proceedings of the 2017 IEEE Visual Communications and Image Processing (VCIP), St. Petersburg, FL, USA, 10–13 December 2018; pp. 1–4. [CrossRef]
13. Sethi, G.; Saini, B.; Singh, D. Segmentation of cancerous regions in liver using an edge-based and phase congruent region enhancement method. *Comput. Electr. Eng.* **2016**, *53*, 244–262. [CrossRef]
14. Wu, K.; Zhang, D. Robust tongue segmentation by fusing region-based and edge-based approaches. *Expert Syst. Appl.* **2015**, *42*, 8027–8038. [CrossRef]
15. Priyanka, V.P.; Patil, N.C. Gray Scale Image Segmentation using OTSU Thresholding Optimal Approach. *J. Res.* **2016**, *2*, 20–24.
16. Aja-Fernández, S.; Curiale, A.H.; Vegas-Sánchez-Ferrero, G. A local fuzzy thresholding methodology for multiregion image segmentation. *Knowl.-Based Syst.* **2015**, *83*, 1–12. [CrossRef]
17. Zaitoun, N.M.; Aqel, M.J. Survey on Image Segmentation Techniques. *Procedia Comput. Sci.* **2015**, *65*, 797–806. [CrossRef]
18. Niu, S.; Chen, Q.; de Sisternes, L.; Ji, Z.; Zhou, Z.; Rubin, D.L. Robust noise region-based active contour model via local similarity factor for image segmentation. *Pattern Recognit.* **2017**, *61*, 104–119. [CrossRef]
19. Er, A.; Kaur, E.R. Review of Image Segmentation Technique. *Int. J.* **2017**, *8*, 36–39.
20. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
21. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. *arXiv* **2015**, arXiv:1411.4038.
22. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
23. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
24. Noh, H.; Hong, S.; Han, B. Learning deconvolution network for semantic segmentation. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1520–1528.

25. 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]
26. Ranzato, M.; Huang, F.; Boureau, Y.; LeCun, Y. Unsupervised learning of invariant feature hierarchies with applications to object recognition. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; pp. 1–8.
27. Ngiam, J.; Khosla, A.; Kim, M.; Nam, J.; Lee, H.; Ng, A. Multimodal deep learning. In Proceedings of the 28th International Conference on Machine Learning (ICML-11), Washington, DC, USA, 28 June–2 July 2011; pp. 689–696.
28. Liu, W.; Rabinovich, A.; Berg, A.C. ParseNet: Looking Wider to See Better, Computer Vision and Pattern Recognition. *arXiv* **2016**, arXiv:1506.04579.
29. TsLin, Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
30. Zhang, L.; Liu, J.; Shang, F.; Li, G.; Zhao, J.; Zhang, Y. Robust segmentation method for noisy images based on an unsupervised denosing filter. *Tsinghua Sci. Technol.* **2021**, *26*, 736–748. [CrossRef]
31. Huang, S.; Zhang, H.; Pizurica, A. Subspace Clustering for Hyperspectral Images via Dictionary Learning With Adaptive Regularization. *IEEE Trans. Geosci. Remote. Sens.* **2021**, *60*, 1–17. [CrossRef]
32. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks. *arXiv* **2014**, arXiv:1406.2661. [CrossRef]
33. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.P.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
34. 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 (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
35. Isola, P.; Zhu, J.-Y.; Zhou, T.; Efros, A.A. Image-to-Image Translation with Conditional Adversarial Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.
36. Zhang, J.; He, F.; Duan, Y.; Yang, S. AIDEDNet: Anti-interference and detail enhancement dehazing network for real-world scenes. *Front. Comput. Sci.* **2022**, *17*, 172703. [CrossRef]
37. Tang, W.; He, F.; Liu, Y.; Duan, Y. MATR: Multimodal Medical Image Fusion via Multiscale Adaptive Transformer. *IEEE Trans. Image Process.* **2022**, *31*, 5134–5149. [CrossRef] [PubMed]
38. Chen, X.; Kuang, T.; Deng, H.; Fung, S.H.; Gateno, J.; Xia, J.J.; Yap, P.-T. Dual Adversarial Attention Mechanism for Unsupervised Domain Adaptive Medical Image Segmentation. *IEEE Trans. Med. Imaging* **2022**, *41*, 3445–3453. [CrossRef] [PubMed]
39. Satellite Images of Dubai Dataset. Available online: https://www.kaggle.com/datasets/humansintheloop/semantic-segmentation-of-aerial-imagery (accessed on 5 March 2021).
40. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference for Learning Representations, San Diego, CA, USA, 7–9 May 2015. [CrossRef]
41. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. *arXiv* **2017**, arXiv:1709.01507.
42. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks, CVPR 2017. *arXiv* **2017**, arXiv:1608.06993.
43. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861.
44. Tan, M.; Le, Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv* **2019**, arXiv:1905.11946.