JOIV : Int. J. Inform. Visualization, 7(4) - December 2023 2321-2335

INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

Performance Improvement of Deep Convolutional Networks for Aerial Imagery Segmentation of Natural Disaster-Affected Areas

Deny Wiria Nugraha [a,c], Amil Ahmad Ilham [b], Andani Achmad [a,*], Ardiaty Arief [a]

[a] *Department of Electrical Engineering, Hasanuddin University, Bontomarannu, Gowa, 92171, Indonesia*
[b] *Department of Informatics, Hasanuddin University, Bontomarannu, Gowa, 92171, Indonesia*
[c] *Department of Information Technology, Tadulako University, Mantikulore, Palu, 94119, Indonesia*
*Corresponding author: *andani@unhas.ac.id*

***Abstract*— This study proposes a framework for improving performance and exploring the application of Deep Convolutional Networks (DCN) using the best parameters and criteria to accurately produce aerial imagery semantic segmentation of natural disaster-affected areas. This study utilizes two models: U-Net and Pyramid Scene Parsing Network (PSPNet). Extensive study results show that the Grid Search algorithm can improve the performance of the two models used, whereas previous research has not used the Grid Search algorithm to improve performance in aerial imagery segmentation of natural disaster-affected areas. The Grid Search algorithm performs parameter tuning on DCN, data augmentation criteria tuning, and dataset criteria tuning for pre-training. The most optimal DCN model is shown by PSPNet (152) (bpc), using the best parameters and criteria, with a mean Intersection over Union (mIoU) of 83.34%, a significant mIoU increase of 43.09% compared to using only the default parameters and criteria (baselines). The validation results using the k-fold cross-validation method on the most optimal DCN model produced an average accuracy of 99.04%. PSPNet(152) (bpc) can detect and identify various objects with irregular shapes and sizes, can detect and identify various important objects affected by natural disasters such as flooded buildings and roads, and can detect and identify objects with small shapes such as vehicles and pools, which are the most challenging task for semantic segmentation network models. This study also shows that increasing the network layers in the PSPNet-(18, 34, 50, 101, 152) model, which uses the best parameters and criteria, improves the model's performance. The results of this study indicate the need to utilize a special dataset from aerial imagery originating from the Unmanned Aerial Vehicle (UAV) during the pre-training stage for transfer learning to improve DCN performance for further research.**

***Keywords*— Semantic segmentation; aerial imagery; natural disaster; deep convolutional networks; the best parameters and criteria; grid search algorithm; k-fold cross-validation.**

Manuscript received 10 Nov. 2022; revised 18 Mar. 2023; accepted 25 Apr. 2023. Date of publication 31 Dec. 2023.
International Journal on Informatics Visualization is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.

I. INTRODUCTION

Image segmentation is an essential topic in image processing and computer vision. Classification, detection, localization, and segmentation are the four primary steps in identifying objects in an image through image processing. Image segmentation can be defined as a problem of classifying pixels with semantic labels (semantic segmentation) or partitioning individual objects. In contrast, image classification must identify what is in the image. Semantic segmentation performs pixel-level labeling with a set of object categories for all image pixels, so it is generally a more difficult task than image classification, predicting a single label for the entire image [1]. In computer vision, semantic segmentation is a fundamental task that assigns a label to each pixel, aka pixel-level classification [2]. Since the advent of deep neural networks, segmentation has made tremendous progress. We refer to [1]–[4] for a full description of deep learning techniques for semantic segmentation.

Numerous researchers are interested in image segmentation due to the numerous application domains that can be implemented. Conversely, increasing datasets are accessible over the internet and are becoming easier to acquire. Due to this convenience, it is necessary to automate image segmentation operations to handle various daily life challenges. These tasks can be in the form of urban remote sensing image segmentation to map land cover [5]–[7], river segmentation on remote sensing imagery [8], image segmentation for building extraction [9], [10], forest fire segmentation [11]–[14], segmentation of roads and buildings that diverse [15], and coconut tree segmentation [16].

In recent years, image segmentation in the context of natural disasters has attracted considerable interest. It is one of the essential research topics in artificial intelligence and image processing. The advent of advanced technology for capturing natural disaster events has increased. An Unmanned Aerial Vehicle (UAV) is one such device that captures aerial images of natural disaster damage and the affected area. The use of aerial images for monitoring and responding to natural disasters is gaining popularity. Even for difficult-to-explore areas on the ground, it is possible to create aerial imagery rapidly. This image can then identify the locations most in need of support. Such image analysis is typically performed manually (manual interpretation with ordinary eyesight). The procedure is time-consuming and frequently yields incorrect findings.

Until now, many methods have been proposed by researchers to produce accurate segmentation of aerial imagery. The current state-of-the-art methods are divided into two parts: models that rely on conventional handcrafted features, as done in studies [17]–[19], and deep neural networks. The benefit of using the latter is its ability to study end-to-end data. Driven by the highly developed success of Deep Convolutional Networks (DCN), several researchers used it for segmentation and natural disaster troubleshooting, including research [20] that used AlexNet to detect landslides and floods, detecting drifting buildings from aerial imagery before and after tsunamis used AlexNet and VGG [21], detecting objects and classifying damage after typhoons used Nazr-CNN [22], semantic segmentation of flooded areas with the integration of CNN and RNN networks [23], detecting flood areas used Generative Adversarial Networks (GAN) [24], extracting flooded areas from UAV imagery used the Fully Convolutional Network (FCN) based on Visual Geometry Group (VGG) [25], and identifying affected areas and access roads in post-disaster scenarios used multiple models for the binary semantic segmentation task and multi-class in aerial images, e.g., U-Net, LinkNet, and ENet [26].

Other studies analyzed and evaluated the performance of popular semantic segmentation methods named DeepLabv3+, Pyramid Scene Parsing Network (PSPNet), and ENet on problems related to natural disaster datasets [27], detecting and segmenting important objects in aerial footage of disaster locations used Mask-Region Based Convolutional Neural Networks (Mask-RCNN) and PSPNet [28], segmentation of damage to buildings after a natural disaster used MSNet [29], a self-attention-based semantic segmentation named ReDNet on a disaster UAV dataset and compared with three other advanced segmentation models: ENet, DeepLabv3+, and PSPNet [30], flood detection based on CNN AlexNet to extract flood-related features from disaster zone images [31], semantic segmentation of aerial images for post-flood landscape understanding by applying three advanced semantic segmentation networks namely ENet, PSPNet, and DeepLabv3+ [32], detecting buildings damaged after an earthquake used a network model Convolutional neural network VGG-16, VGG-19, and NASNet [33], semantic segmentation of natural disaster datasets used self-attention-based methods combined with Global Average Pooling and U-Net [34], semantic segmentation of post-flood datasets with U-Net, PSPNet, and DeepLabV3+ [35], detecting flooding used segmentation with three deep neural networks: PSPNet, DeepLabV3, and U-Net [36], and extracted residential buildings with a modified Mask R-CNN [37], semantic segmentation of volcanic ash eruptions used SegNet and U-Net convolutional neural networks for volcano monitoring in volcanic eruptions [38], landslide detection and identification used Lightweight Attention U-Net [39], finding buildings damaged by disasters used transfers-learning deep attention network (TDA-Net) [40], and semantic segmentation to detect landslides used U-Net [41], [42], and self-training method [43].

Although DCN is highly dependent on architectural modifications, as shown in several studies above, tuning and selecting the appropriate parameters and criteria allows us to have enormous potential to improve further DCN performance for aerial imagery segmentation of natural disaster-affected areas. The main challenge is to improve the performance of DCN to accurately produce aerial imagery semantic segmentation of natural disaster-affected areas. Therefore, this study presents a framework and reveals practical knowledge through experimental studies for aerial imagery segmentation of natural disaster-affected areas. We summarize the experiments carried out and the knowledge gained in the trials, which are our main contributions to this paper as follows:

- Improve the performance of the network model for aerial imagery semantic segmentation of natural disaster-affected areas by integrating the Grid Search algorithm and DCN and validating the results with the k-fold cross-validation method.
- Conduct comprehensive testing using several parameters, data augmentation, evaluation dataset of aerial imagery of natural disaster-affected areas, large-scale datasets for pre-training, and DCN, which accurately produces the best parameters and criteria for aerial imagery semantic segmentation of natural disaster-affected areas.
- Based on the results of the best parameters, appropriate data augmentation criteria, and suitable pre-training dataset criteria, we conducted a comprehensive test and produced the most optimal DCN performance for aerial imagery semantic segmentation of natural disaster-affected areas. We utilize two semantic segmentation network models, U-Net [44] and PSPNet [45], two advanced semantic segmentation networks that have demonstrated promising performance on various segmentation benchmarks. We used the Residual Network (ResNet) architecture [46] as the backbone of an information encoder capable of extracting fine image patterns. Using the PSPNet model, we also determined the relationship between the number of layers and performance improvements, utilizing PSPNet-(18, 34, 50, 101, 152).
- The U-Net and PSPNet models with the best parameters and criteria, resulting from this study, were compared with the same network model, which only used default parameters and criteria (baselines). We also compared the results of our study with those in the literature review. We conducted these comparisons to prove that our proposed framework has significantly increased

DCN's performance for aerial imagery semantic segmentation of natural disaster-affected areas.

This study addresses recognizing aerial imagery of natural disaster-affected areas through semantic segmentation, resulting in improved DCN performance. Therefore, research that is closely related to the use of parameters, the use of data augmentation criteria, the use of dataset criteria in pre-training, and the use of natural disaster aerial imagery datasets, we describe representatively in the following.

A. Use of Parameters and Criteria on Deep Convolutional Networks for Aerial Imagery Segmentation

Previous researchers have used parameters such as learning rate, data split, optimizer, and data augmentation criteria on DCN for segmentation. Research [21] used Stochastic Gradient Descent (SGD) with a learning rate of 0.001 and data augmentation with vertical and horizontal flipping. CNN was trained using Adam optimization with a learning rate of 1e-5 and a batch size of 12, and RNN was trained using Adagrad with a learning rate of 0.01 and a batch size of 8 [23]. A learning rate of 0.0001 and a maximum epoch of 6 were used for all classes [25]. The model was trained using the Adam optimizer with a learning rate of 10^{-4} for 600 epochs and used horizontal and vertical flipping data augmentation [26]. [27], [32] used random shuffling, scaling, flipping, and random rotation data augmentation; the batch size was set to 2; for semantic segmentation implementing PSPNet used a learning rate of 0.0001; for ENet 0.0005; and DeepLabv3+ 0.01. The dataset was separated into training (60%), validation (20%), and testing (20%); the Mask-RCNN learning rate was set as 10^{-3}; and visual augmentation was done with zoom, blur, pixel dropout, adding clouds, and color equalization operation [28]. Research [29] trained the model using 80% of the dataset and tested it on the remaining 20% dataset, at 100 epochs, with an initial learning rate of 0.003, then reduced to 0.001 after 10 epochs, and SGD was used as an optimizer with a batch size equal to 8. [30] used a learning rate of 0.0001 and augmentation of random shuffling, scaling, flipping, and random rotation.

Divide the dataset into training, validation, and test sets, with 70% for training and 30% for validation and testing [32]. 70%, 15%, and 15% of the dataset were used for training, validation, and test sets, respectively, with 300 epochs, and the learning rate was 0.0001 [33]. The learning rate of 0.001 and used random shuffling, scaling, flipping, and random rotation augmentations [34]. For U-Net, the learning rate of 0.01; for PSPNet and DeepLabV3+, the learning rate of 0.001; image augmentation used shuffling, rotation, scaling, shifting, and brightness contrast; Adam optimizer; and a batch size of 24 was used for all models [35]. For PSPNet and U-Net training, a learning rate of 0.001 for 15 epochs; for DeepLabV3, a learning rate of 0.01 for 10 epochs; and used albumentation tools for image augmentation that implement various image transformation operations such as Resize, ShiftScaleRotate, RGBShift, RandomBrightnessContrast, and Normalize [36]. The optimizer used in Mask R-CNN during training was SGD, with a learning rate of 0.0025 and a batch size of 3 [37].

Research [38] used a learning rate of 0.0001, batch size equal to 4, the number of epochs 100, Adam optimizer, data augmentation with horizontal flips, zoom, random noise, and rotations, and the data set was divided into two sets: training and validation respectively in the proportion of 80% and 20%. Research [39] used a learning rate of 1×10^{-5}, the maximum number of epochs was 150, with a batch size of 16, the division of the dataset for training was 70%, and validation was 30%, and the Adam optimizer. Research [40] chose Adam as the optimizer, and the learning rate was 1×10^{-4}. Research [41] used augmentation consisting of random rotations and vertical and horizontal flips; the model was trained for 200 epochs with a dynamic learning rate of 0.001, Adam was used as optimization, the model was trained with four different batch sizes (16, 32, 64, 128), and 30% of each dataset was used as validation data. Research [42] used the Adam optimizer with 100 epoc; the learning rate was 0.01. For the overall training setting, research [43] used the SGD optimizer, batch size set to 16, and used data augmentation random flipping, random resizing, and cropping.

Research [20] resulted f-scores for landslide and flood detection in the range of 80%-90%, but only detected 1 class, [21] achieved a classification accuracy of 94%-96% in all conditions, but only detected 1 class, namely bulding, [23] obtained accuracy and mean Intersection over Union (mIoU) of semantic segmentation of 96% and 92%, but only used 1 class, namely flooded areas, [24] resulted accuracy for flood segmentation, in rural areas 89%-95.5% and in urban areas 80.5%-88%, [27] achieved the highest mIoU of 79.43% with the PSPNet method used 9 classes, [30] resulted a mIoU value of 80.27% for the PSPNet method with 9 classes, [31] had an accuracy of 91% for segmentation with only 1 class, namely flooding, [32] resulted a mIoU value for segmentation of 80.35% used the PSPNet method and used 9 classes, [33] achieved the highest accuracy of 70% for the VGG-19 model used 3 building classes, namely normal, less damaged, and damaged, [34] achieved a PSPNet mIoU value of 79.43% for 9 classes, and [35] resulted the best segmentation mIoU value of 52.23% used the DeepLabV3+ (pseudo-labels) method. Research [38] resulted in a mIoU of 90.13% obtained for the U-Net architecture, and for SegNet, a mIoU value of 88%, calculated using a validation dataset, to extract volcanic ash eruption forms automatically. Research [39] resulted in mIoU, and F1_score values of Lightweight Attention U-Net achieved 82.29% and 87.45%, which are the best performance for landslide segmentation. Research [41] achieved the highest mIoU value of 43%. Research [42] achieved an Area under the Precision-Recall curve (AUPRC) value exceeding 0.7.

B. Exploration Study in Transfer Learning

Previous researchers have used several natural disaster aerial imagery datasets. Here, we show what datasets the researchers used. Research [20], [31] used datasets from Google Earth. The AIST Building Change Detection (ABCD) dataset was used in the study [21]. Typhoon disaster dataset [22], UAV dataset from flooded area [23], [25], OpenStreetMap (OSM) dataset [26], High Resolution UAV Dataset (HRUD) [27], [30], [34], Volan2019 dataset [28], and the Instance Segmentation in Building Damage Assessment (ISBDA) dataset [29], used in each of these studies. The FloodNet dataset was used in studies [32], [35], [36]. Datasets from open sources such as images.google.com and images.baidu.com were used in the study [33], and datasets taken from the Geospatial

Information Authority of Japan (GSI) were used in the study [37]. Research [38] used the Etna_NETVIS dataset. The Red Relief Image Map (RRIM) dataset was used in the study [39]. The xView2 dataset, WHU Building, and other data from Google Earth were used in the study [40]. Research [41] used three different datasets created by RapidEye, the Normalized Vegetation Index (NDVI), and the digital elevation model (DEM). Synthetic Aperture Radar (SAR) datacubes were used in the study [42]. Research [43] used Sentinel-2 and ALOS PALSAR data.

Several previous researchers have also used transfer learning with well-organized datasets; for example, research [22], [25], [26], [33], [36], [43] used ImageNet [47], research [28], [29] used Common Objects in Context (COCO) [48], and research [40] used a high-quality xBD dataset for the pre-training model.

The use of transfer learning in research [22] resulted in an overall accuracy for the best segmentation of 40.90% with the Nazr-CNN model for three damage classes, [25] resulted in the highest overall accuracy on FCN-8s of 95.520% for four classes: water, building, vegetation, and road, [26] achieved the highest mIoU for segmentation on the UNetUp (VGG16) model of 44.99% used only road, and building classes, [28] achieved the best performance mIoU of 32.17% and accuracy of 77.01% on the PSPNet model with class namely, flood area, debris, roads, and vegetation, [29] achieved an AP value (averaged over all IoU thresholds) for MSNet of 37.2%, [36] reached 56% mIoU on the PSPNet model, together with the Resnet-152 encoder, [37] resulted in the highest mAP value for segmentation used the Mask R-CNN model of 37.3% for four levels of damage, [40] resulted F1-scores (F1), precision (P), recall (R) on TDA-Net with respective values of 95.6%, 94.9%, and 96.4% for detected damaged buildings, and [43] resulted the highest F1-score of 73.50%.

The related studies above produced various accuracy values for aerial imagery segmentation according to the parameters and criteria used. The resulting accuracy value is quite high in several studies that only used one class or a small number of classes. However, the resulting accuracy is quite small in studies that used a large number of classes. High accuracy does not necessarily result in high mIoU values, so it is necessary to display mIoU values in each final test result in displaying segmentation results so that the accuracy of the DCN model used can be seen. Some of the mIoU values displayed in these related studies are still quite small, especially for segmentation tasks with many classes; this is due to the inaccuracy of the use of parameters and criteria for the DCN model. Our study used nine object classes and presented a complete performance evaluation consisting of accuracy, precision, recall, F1-score, and Intersection over Union (IoU).

The previous studies that have been described representatively above used parameters (such as learning rate, data split, and optimizer), data augmentation criteria, and pre-training dataset criteria that had been determined only based on their respective literature studies or only applied different settings in a trial-and-error manner, or only used the default parameters and criteria. None of these previous studies have improved DCN performance for aerial imagery semantic segmentation of natural disaster-affected areas, no one has used the Grid Search algorithm to tune DCN parameters (such as learning rate, data split, and optimizer), tuning data augmentation criteria, and tuning dataset criteria for pre-training comprehensively for aerial imagery semantic segmentation of natural disaster-affected areas, no one has yet searched for the best combination of parameters and criteria, and no one has validated using the k-fold cross-validation method on the most optimal DCN model.

These previous studies also have not conducted tests to verify the relationship between the number of layers and increased performance. No one has carried out transfer learning using a combination of general datasets (real-world images + urban images + road images) and a combination of special aerial imagery datasets originating from the UAV for aerial imagery semantic segmentation of natural disaster-affected areas. No one has tested and compared model performance with several scenarios, namely using default parameters and criteria, using the best parameters, and using the best parameters and criteria.

Our study proposes a framework for improving performance and exploring the application of DCN using the best parameters and criteria to accurately produce aerial imagery semantic segmentation of natural disaster-affected areas. Our study takes the initiative to perform aerial imagery semantic segmentation of natural disaster-affected areas by integrating the Grid Search algorithms and DCN. This study performs parameter and criteria tuning comprehensively using the Grid Search algorithm and validates the results using the k-fold cross-validation method, taking into account the parameters used in DCN, paying attention to various appropriate data augmentation methods, and paying attention to various datasets that are suitable for pre-training. Combinations of each parameter and criteria were tried to get the most optimal performance results in producing aerial imagery semantic segmentation of natural disaster-affected accurately. We provide the results of tuning the best combination of parameters and criteria and comparing performance with models using default parameters and criteria (baselines). We also try to optimize PSPNet with multiple layers using the best parameters and criteria. This effort is beneficial for revealing practical knowledge and fair comparison with several approaches/scenarios.

We believe that transfer learning considerations make the aerial imagery semantic segmentation of natural disaster-affected areas more reliable and knowledgeable. We validated the effect of general datasets (real-world, urban, or road images), a special dataset of aerial imagery derived from UAVs, and a combined dataset for transfer learning on the performance of DCNs for semantic segmentation. We also display the results of aerial imagery semantic segmentation of natural disaster-affected areas visually to see the accuracy of the DCN model.

We organize this paper as follows: Section II of Materials and Method describes Deep Convolutional Networks (DCN), the dataset used, the Grid Search algorithm, the k-fold cross-validation method, the implementation of semantic segmentation, and the proposed framework or method. The experimental results are presented and discussed in Results and Discussion in Section III. Finally, Section IV presents our conclusions and suggests further research in the future.

II. Materials and Method

A. Deep Convolutional Networks (DCN)

We mainly use U-Net [44] and PSPNet [45] as DCN models for semantic segmentation in this study and ResNet architecture [46] as the backbone. At the start of the test, we confirmed performance with the PSPNet(50) model and added layers to PSPNet(101). Next, we retest with various layers, such as PSPNet(18), PSPNet(34), and PSPNet(152). In addition, we compared the results with U-Net. All these network models use the best parameters, the appropriate data augmentation criteria, and the suitable pre-training dataset criteria and compare the results with the network models using the default parameters and criteria (baselines).

1) U-Net: U-Net modifies and expands the FCN architecture so that the network uses fewer training images and generates more accurate segmentation. The objective and concept behind this strategy are to augment the conventional contract network with successive layers so that the upsampling operator replaces the pooling operator as this layer increases output resolution. One of the most significant changes to the U-Net architecture is upsampling. Many feature maps are included, allowing the network to propagate context information to higher-resolution layers. The architectural model is shaped like a U [44]. U-Net was initially designed for biomedical image segmentation tasks. In recent years, research has demonstrated that U-Net is also applicable and has significant potential for semantic segmentation of aerial imagery.

2) PSPNet: Scene decomposition is a fundamental concept in computer vision based on semantic segmentation. Scene parsing aims to comprehensively understand the scene by predicting object labels, locations, and shapes. Previously developed frameworks for advanced scene decoding relied heavily on Fully Connected Networks (FCN). The usage of CNN presents a number of challenges because it is difficult to examine the variety of scenes. To overcome these challenges, the Pyramid Scene Parsing Network (PSPNet) was released [45]. Pixel prediction is based on FCN in PSPNet. In addition, the pixel-level features have been expanded to a series of built global pyramids in which local and global values are merged to produce more accurate final predictions. In addition, optimization techniques with highly supervised losses have been integrated. For the previous global scene construction on the final layer feature map of the neural network, the Pyramid Pooling Module was implemented to reduce the loss of context information between distinct sub-regions. This module has operations under four different stages of the pyramid. PSPNet is a proven and effective pyramid scene parsing network for comprehending complex scenes. PSPNet achieves state-of-the-art performance on various datasets, including the 2016 ImageNet scene decoding, 2012 PASCAL VOC, and Cityscapes benchmarks. PSPNet utilizes ResNet as its backbone with an extended network to extract feature maps. Then a 4-level pyramid pooling is applied to the feature map to extract the previous global context. The final prediction map is produced by combining these global priorities with the original feature map, followed by a convolution layer.

3) ResNet: The PSPNet model utilizes a backbone capable of extracting fine patterns of images in the form of an information encoder. A Microsoft Research team developed deep Residual Learning for Image Recognition to solve the fundamental issues of VGG and AlexNet. The scalability of the network is a challenge for AlexNet and VGG. As increasingly deep networks begin to coalesce, the degradation problem becomes apparent. As the network depth increases, the accuracy saturates and then rapidly falls. ResNet is based on implementing a residual block of "identity shortcut connections" that traverse one or more layers. When the identity mapping reaches optimal, it pushes the residual to zero and matches the identity mapping. With these actions and modifications, ResNet outperforms current state-of-the-art convolution networks [46].

B. Datasets

The datasets used in this study are divided into pre-training and evaluation datasets for training, validation, and testing. The selection of the two kinds of datasets is based on the availability of datasets that include segmentation and annotations and are publicly available and easily accessible. The datasets used in this study and their characteristics are shown in Table I.

1) Pre-Training Dataset: For the pre-training dataset, we used two types of external datasets: general datasets (real-world images, urban images, or road images) and special datasets of aerial imagery derived from UAVs. Both types of datasets were tested to verify the effect of these datasets on DCN performance for semantic segmentation. In order to successfully optimize the DCN model for semantic segmentation, a large number of pre-trained datasets are required. We define a dataset that is larger in scale than the evaluation dataset, is easy to obtain, and has segmentation annotations. The transfer learning procedure consists of pre-training with a large-scale dataset and training by a relatively small evaluation dataset. However, due to the limited capabilities of personal computers and the availability of existing datasets, we limited the number of images in each dataset, as shown in Table I. We selected the COCO, VOC, Cityscapes, DSRS, and Mapillary Vistas datasets for the pre-training datasets containing real-world, urban, or road images and the USS and Semantic Drone datasets derived from UAV aerial imagery. The VOC and DSRS datasets have a single label on each image, while the COCO, Cityscapes, Mapillary Vistas, USS, and Semantic Drone datasets have multiple labels. We collected these datasets from the relevant sites (COCO, VOC, Cityscapes, Mapillary Vistas, USS, and Semantic Drone) and the data science community site Kaggle (DSRS). In transfer learning, a trained model is needed; this trained model is called a pre-trained model. Pre-trained models are usually already trained on larger, structured, and labeled datasets. Currently, many pre-trained models are provided for various needs, such as pre-trained models for image classification and object detection. Still, obtaining a pre-trained model for image segmentation that fits the overall DCN model we use in this study isn't easy. Therefore, in this study, we use all of the above datasets and their annotations in the pre-training process to create their pre-trained models to get a special pre-trained model for image segmentation and have a good quality pre-trained model.

TABLE I
DATASET DETAILS

Dataset name	Types of images	Objective	The number of images that match the segmentation	Resolution of images	Annotation type
Common Objects in Context (COCO) [48]	General (real-world images)	Pre-training	5000	Varies in size	Multiple labels
PASCAL Visual Object Classes (VOC) [49]	General (real-world images)	Pre-training	2913	Varies in size	Single label
Cityscapes [50]	General (urban images)	Pre-training	2975	2048 × 1024	Multiple labels
Dira-Simulator-Road-Segment (DSRS) [51]	General (road images)	Pre-training	5000	320 × 160	Single label
Mapillary Vistas [52]	General (urban images)	Pre-training	5000	Varies in size	Multiple labels
Combined general dataset (balanced)	General	Pre-training	6250	Varies in size	Multiple labels
Combined general dataset (unbalanced)	General	Pre-training	20888	Varies in size	Multiple labels
UAVid Semantic Segmentation (USS) [53]	Aerial imagery (UAV)	Pre-training	270	Average 3840 × 2160	Multiple labels
Semantic Drone [54]	Aerial imagery (UAV)	Pre-training	400	6000 × 4000	Multiple labels
Combined aerial imagery dataset	Aerial imagery (UAV)	Pre-training	670	Varies in size	Multiple labels
FloodNet [32]	Aerial imagery (UAV)	Evaluation (training, validation, and testing)	2343	Average 4000 × 3000	Multiple labels

2) Evaluation Dataset: We use FloodNet as an evaluation dataset for training, validation, and testing in image recognition for aerial imagery semantic segmentation of natural disaster-affected areas, which are aerial imagery datasets originating from UAVs. We obtained this dataset from research [32]using high-resolution aerial image data collection to understand post-disaster (flood) landscapes. FloodNet delivers high-resolution images taken from low altitudes, which have an advantage over satellite images captured from higher altitudes that clouds and smoke may obscure. The collection was acquired using a small UAV platform, DJI Mavic Pro quadcopters, at an altitude of 60 meters, resulting in images with a very high spatial resolution (about 1.5 centimeters) that distinguishes it from previous natural disaster datasets. Post-flood damage in the affected area is shown in all images. This dataset contains pixel-level semantic segmentation annotations. There are 2343 images and their respective annotations, categorized into 9 classes: building-flooded, building-non-flooded, road-flooded, road-flooded, road-non-flooded, water, tree, vehicle, pool, and grass.

C. Implementation Details

1) Optimization of Parameters and Criteria Using Grid Search Algorithm and K-Fold Cross-Validation Method: The Grid Search (GS) algorithm is a complete search method with a uniform grid in the search parameter space defined. The primary purpose of this method is to identify optimal model parameters so that model performance can be improved as much as possible [55]. The basic principle of the GS method is to divide the grid into a certain range and traverse all points in the network with the parameter values used. Finally, the parameter with the highest accuracy was determined as the best parameter [56]. GS was developed to match parameters and criteria and optimize the solution of complex problems, especially in this study in accurately producing aerial imagery semantic segmentation of natural disaster-affected areas.

The Grid Search algorithm is used in this study to be tuning and completely identify the parameters and criteria that lead to the highest accuracy. The parameters consist of the learning rate, data split, and optimizer. The criteria used consisted of data augmentation criteria and pre-training dataset criteria. We use several augmentation methods on the data augmentation criteria, namely photometric distortion, geometric distortion, cutout, and a combination of all data augmentation methods. The pre-training dataset criteria consist of general and special aerial imagery datasets using the datasets described in Table I. The highest accuracy of the DCN model for semantic segmentation with all parameters and criteria is compared to determine the best combination of parameters and criteria to produce the most optimal model performance in accurately producing aerial imagery semantic segmentation of natural disaster-affected areas.

The results of the most optimal model using the best parameters and criteria were validated based on the cross-validation method using the k-fold cross-validation method.

The performance of the DCN model can be improved by using a combination of GS and k-fold cross-validation, and the model's performance can be evaluated based on the cross-validation method. In k-fold cross-validation, the training set is first divided into k subsets of equal size. The model will be trained and tested k times. In each training process, one data set will be used as a test, while the rest will be used as train data. Sequentially, each subset is tested by a model trained on another k−1 subset. Therefore, each sample in the training set is tested once. As a result, the cross-validation accuracy will be the percentage of data tested correctly. The estimated k-fold cross-validation of all model accuracy is calculated by the average of each k-model accuracy measurement (Equation 1), where A is the accuracy of the model and k is the number of subsets or groups used.

$$\text{Cross Validation Accuracy (CVA)} = \frac{1}{k}\sum_{i=1}^{k} A_i \qquad (1)$$

The flowchart of the Grid Search algorithm with k-fold cross-validation proposed in this study is shown in Fig. 1, and the complete parameters and criteria used in the Grid Search algorithm are shown in Table II.

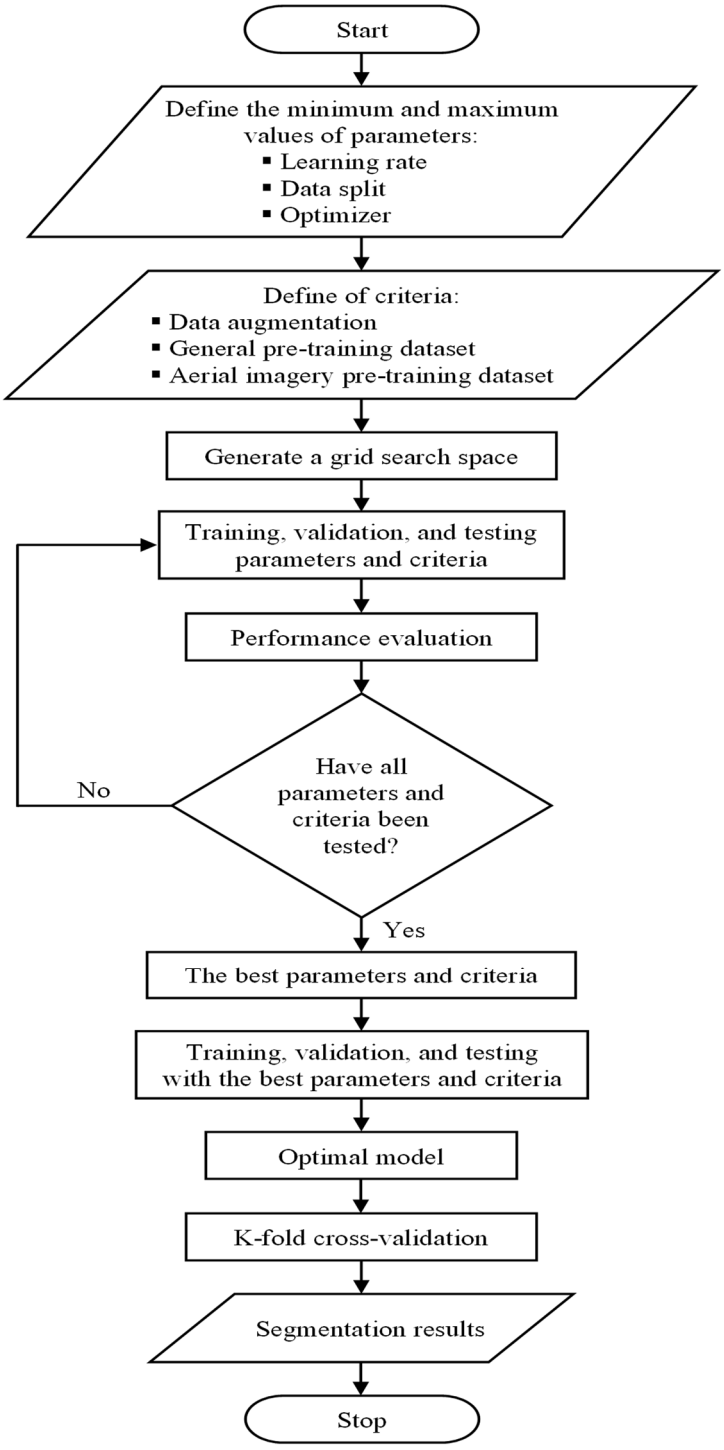


Fig. 1 Flowchart of the grid search algorithm with k-fold cross-validation

TABLE II
PARAMETERS AND CRITERIA USED IN THE GRID SEARCH ALGORITHM

Parameters/criteria	Range
Learning rate	0.0000001 … 0.01
Data split	50% for training and validation, and 50% for testing 60% for training and validation and 40% for testing 70% for training and validation, and 30% for testing 80% for training and validation, and 20% for testing 90% for training and validation, and 10% for testing
Optimizer	Adaptive Moment Estimation (Adam), Stochastic Gradient Descent (SGD), and Root Mean Square Propagation (RMSProp)
Data augmentation	Photometric distortion: brightness, contrast, saturation, dan noise; Geometric distortion: flipping (horizontal dan vertical), rotating (30°, 60°, 90°), random scaling (X scale, Y scale); Cutout; and Combination of all data augmentation methods
Pre-training dataset	General pre-training dataset: COCO, VOC, Cityscapes, DSRS, Mapillary Vistas, combined general dataset (balanced), combined general dataset (unbalanced); and Aerial imagery pre-training dataset: USS, Semantic Drone, combined aerial imagery dataset

2) Implementation of Semantic Segmentation: This section explains how to implement DCN for aerial imagery semantic segmentation of natural disaster-affected areas. The Keras framework was used to build the model and implement the segmentation network with the TensorFlow backend. All semantic segmentation experiments were run using a personal computer (PC) with a 12th Gen Intel® Core™ i7 processor, with turbo frequency up to 4.90 GHz 12-core, 32 GB of RAM, and a 10 GB NVIDIA GeForce RTX 3080 GPU.

This study uses a categorical cross-entropy loss function (Equation 2), where y_i is the true label (the ground truth label for each image labeled), $\hat{y}_i$ is the predicted label (the predicted result of an image classified), N represents the total number of samples used for each epoch, and Loss is the average cross-entropy between the desired distribution $\hat{y}_i$ and the ground truth y_i.

All models were trained for 50 epochs for a fair comparison between different models. During the training and validation process, we resized all original images to 473 × 473 pixels, the batch size was equal to 2, and the number of steps (no_of_step) was equal to the number of datasets used divided by the number of batches. We also use a checkpoint callback operation to save the best model for the duration of the epoch. In addition, we apply the early stopping method to prevent overfitting by stopping the training process when the loss train does not decrease.

To assess the performance of the DCN model for semantic segmentation, this study presents a performance evaluation consisting of accuracy, precision, recall, F1-score, and Intersection over Union (IoU), which is used based on a confusion matrix with four main factors, such as false negative (FN), false positive (FP), true negative (TN), and true positive (TP).

TP is a pixel correctly predicted according to its class, which includes 9 object classes: building-flooded, building-non-flooded, road-flooded, road-non-flooded, water, tree, vehicle, pool, and grass. FP is a pixel incorrectly identified as belonging to a class but actually does not belong to that class. FP represents the number of false positives that occur when a pixel is not of class, incorrectly identified as a certain object class. FN is a pixel incorrectly identified as not belonging to a class when in fact, it does. FN represents the number of false negatives that occur when the actual class of an object is incorrectly identified as a pixel instead of its class. TN is a correctly predicted pixel that does not belong to all classes.

Accuracy is the ratio between the number of correctly predicted pixels and the total number of pixels. Accuracy is calculated as the number of TP and TN pixels for each class divided by the total number of pixels (Equation 3). Precision counts how many positive predictions belong to the positive class (Equation 4). Recall represents the number of positive predictions from all positive samples (Equation 5). The F1-score provides a numerical value to balance precision and recall problems (Equation 6). For each class, the IoU pixels are calculated by dividing all the TP pixels corresponding to that class by the number of TP, FP, and FN cases (Equation 7). The average pixel IoU (mIoU) across all classes reflects the overall performance of the DCN model.

$$Loss = -\frac{1}{N}\sum_{i=1}^{N} y_i.log\hat{y}_i + (1-y_i).log(1-\hat{y}_i) \quad (2)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

$$F1\text{-}score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (6)$$

$$IoU = \frac{TP}{TP + FP + FN} \quad (7)$$

The proposed overall framework or methodology for improving the performance of deep convolutional networks for aerial imagery segmentation of natural disaster-affected areas is shown in Fig. 2. Segmentation results display performance evaluation, segmented objects, object class labels, object class probability, and the number of each object.

III. RESULTS AND DISCUSSION

A. Parameters and Criteria Testing Results

This section presents the test results of comprehensively tuning the parameters and criteria using the Grid Search algorithms and DCN for aerial imagery semantic segmentation of natural disaster-affected areas. The detailed settings for tuning parameters and criteria according to the parameters and criteria are shown in Table II. The test results for tuning parameters on DCN using the Grid Search algorithm are shown in Fig. 3, which produces 90 combinations of parameters. The best parameters of the DCN model for aerial imagery semantic segmentation of natural disaster-affected areas with the highest accuracy of 98.48% on a combination of parameters, namely: learning rate of 0.0001, data split with 90% for training and validation (70% training and 20% validation), and 10% for testing, and the optimizer used is RMSProp.

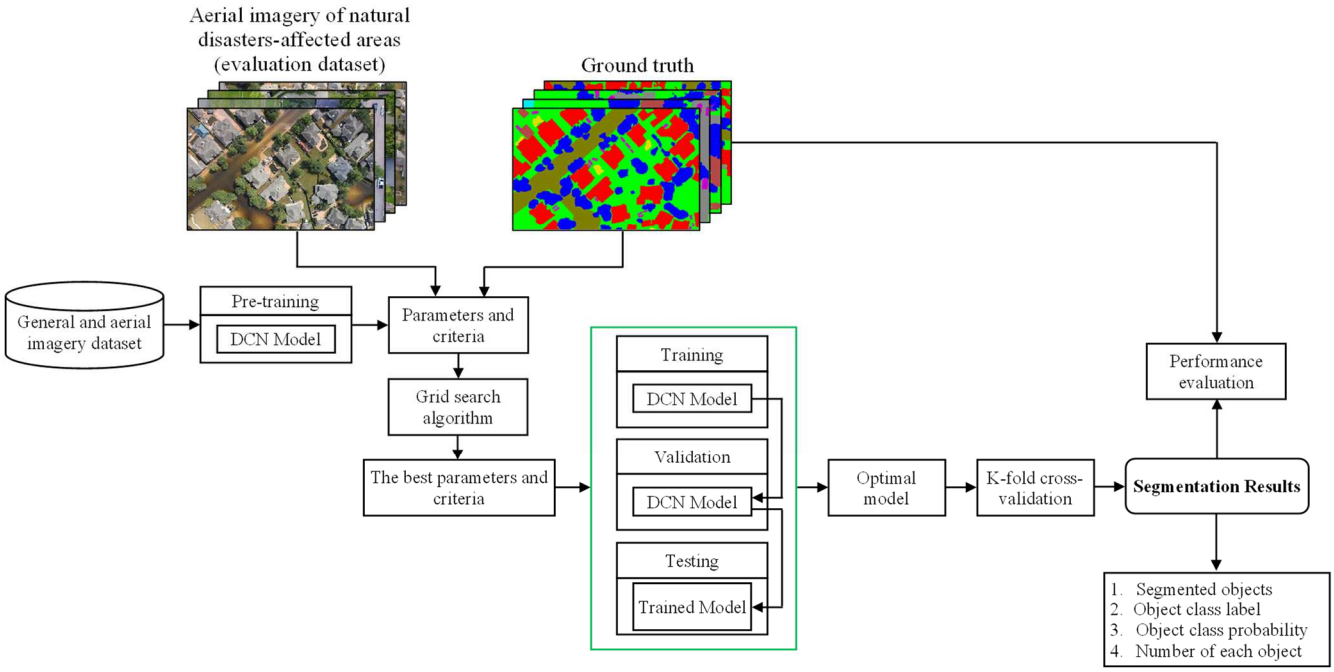


Fig. 2 Overview of the proposed framework for improving the performance of deep convolutional networks for aerial imagery semantic segmentation of natural disaster-affected areas

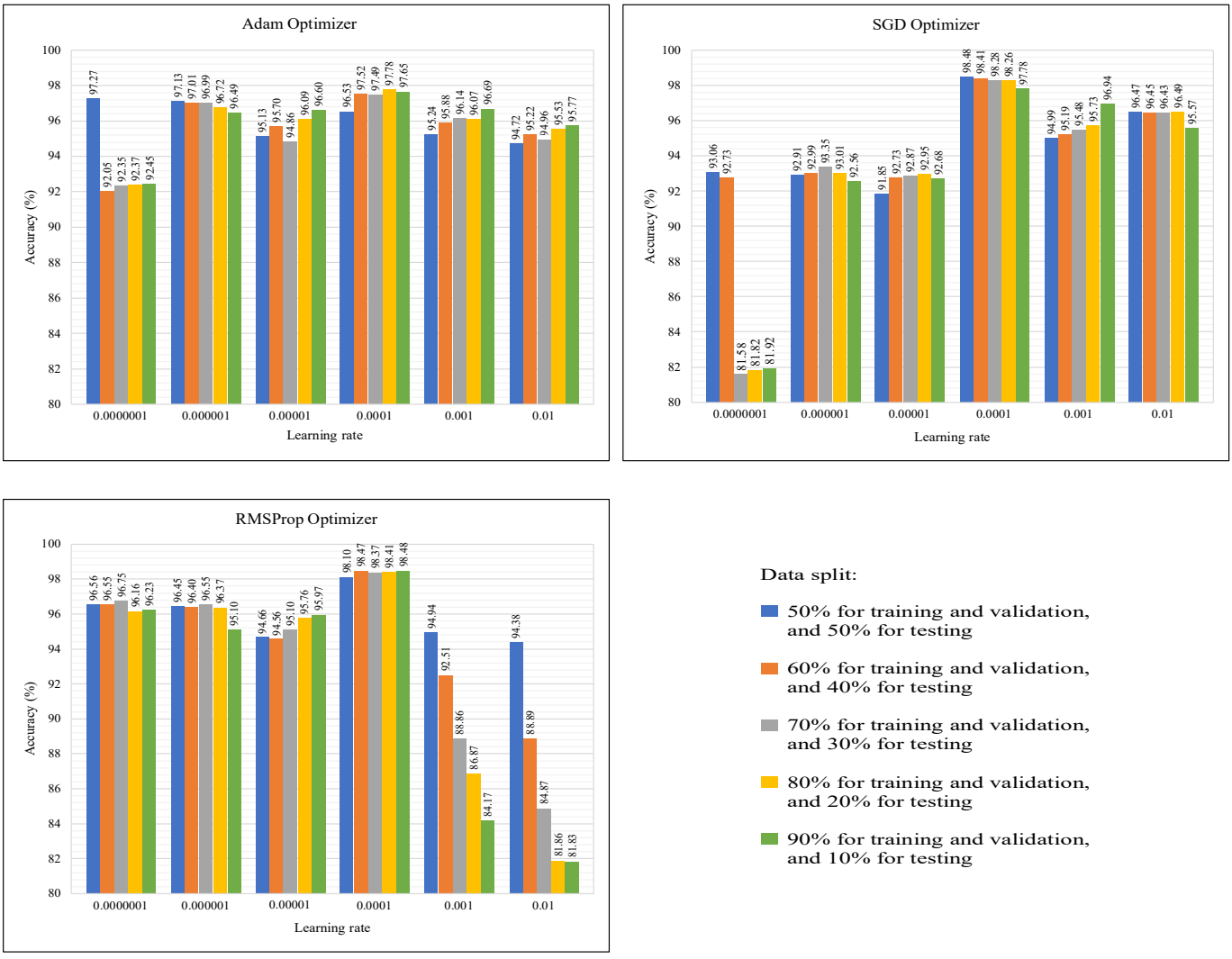


Fig. 3 The test results of tuning parameters on deep convolutional networks using a grid search algorithm for aerial imagery semantic segmentation of natural disaster-affected areas

Fig. 4 shows the test results of tuning the data augmentation criteria. We confirm that the appropriate data augmentation method is used to improve DCN performance and help prevent overfitting with the highest accuracy of 91.10% by using a geometric distortion data augmentation method consisting of flipping (horizontal and vertical), rotating (30°, 60°, 90°), and random scaling (X scale, Y scale).

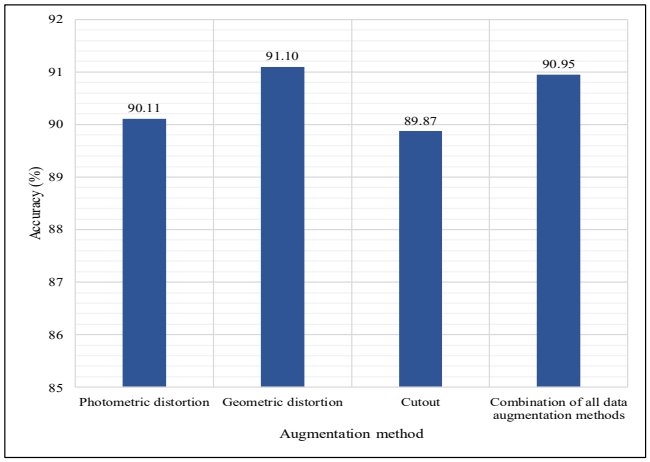


Fig. 4 The test results of tuning data augmentation criteria using a grid search algorithm for aerial imagery semantic segmentation of natural disaster-affected areas

What kind of dataset is suitable for transfer learning in aerial imagery segmentation of natural disaster-affected areas is shown in the test results of tuning dataset criteria for pre-training using a Grid Search algorithm, which leads to the highest accuracy to produce the most optimal performance of the DCN model. We compared several datasets in line with transfer learning. To confirm the effect of pre-training in a dataset, we performed transfer learning using a general dataset (real-world images, urban images, or road images) and a special dataset of aerial imagery derived from UAVs against an evaluation dataset (FloodNet). We used five general datasets, two special datasets of aerial imagery, and three combined datasets, which are in line with transfer learning and according to the object classes used in aerial imagery segmentation to improve DCN performance. Fig. 5 shows the effect of pre-training with all datasets.

As shown in Fig. 5, the Cityscapes pre-training model achieved the best level of performance for transfer learning with a single pre-training dataset, which had an accuracy of 92.396%. This is because the Cityscapes dataset is a general dataset that contains images of urban landscapes and multiple labels and has a fairly high image resolution, so it is still suitable for transfer learning, specifically for a single dataset, to aerial imagery of natural disaster-affected areas datasets which contain images of more complex urban and natural landscapes.

We want to improve accuracy by using combined datasets; we are trying to organize larger datasets. As shown in Fig. 5, the combined general dataset (unbalanced) pre-training model achieved the best performance level for transfer learning with the combined pre-training dataset, which had an accuracy of 92.4%. The combined dataset combines all general pre-training datasets (COCO + VOC + Cityscapes + DSRS + Mapillary Vistas), multiple labels, good segmentation annotations, and complex images (real-world images + urban images + road images).

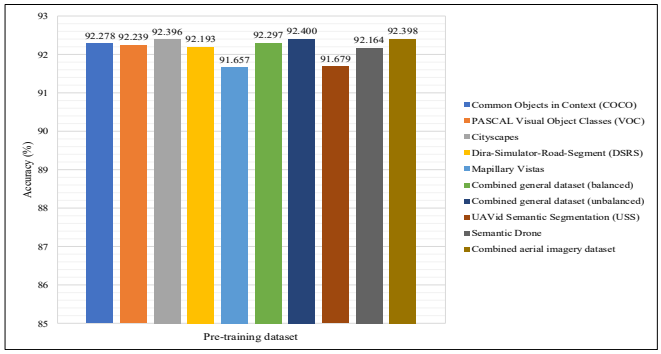


Fig. 5 Transfer learning on the evaluation dataset (FloodNet) with all pre-training datasets

Overall, the combined general dataset (unbalanced) pre-training model achieved the best performance across all transferred tasks, the second best was the combined aerial imagery dataset pre-training model, and the third best was the Cityscapes pre-training model. The difference between the combined general dataset (unbalanced) and the second best dataset, the combined aerial imagery dataset, is 0.002% (92.4% − 92.398%), and 0.004% (92.4% − 92.396%) in the third best dataset, Cityscapes. It can be seen that the difference in the results of the combined general dataset (unbalanced) and combined aerial imagery dataset is very small (0.002%), even though the difference in the number of images owned is very large; namely, the combined general dataset (unbalanced) has 20,888 images, while the combined aerial imagery dataset only has 670 images. We gained knowledge that the combined aerial imagery pre-training dataset derived from the UAV, which has high image resolution, good segmentation annotations, and multiple labels, is very precise and has a great opportunity for transfer learning to the aerial imagery of natural disaster-affected areas dataset, provided that it has a large enough number of images. The combined aerial imagery dataset as a pre-training dataset has similar characteristics to the evaluation dataset (FloodNet), namely the similarity of aerial imagery from UAVs, which contains images of more complex urban and natural landscapes. However, this cannot be done because the special dataset of aerial imagery from UAVs for pre-training has not been available with a large enough size to date.

Another piece of knowledge we gained was to practically increase the amount of data simply by combining multiple datasets and proving that the combined datasets increase the level of performance compared to a pre-training of only one dataset. Combining multiple datasets containing complex images (real-world images + urban images + road images) and multiple labels can improve accuracy. We also reveal that the carefully annotated composite of pre-training datasets effectively trains aerial imagery datasets for semantic segmentation tasks.

B. Results of Deep Convolutional Networks (DCN) Performance Testing

After testing the parameter tuning on the DCN, testing the data augmentation criteria tuning, and testing the dataset

criteria tuning for pre-training comprehensively using the Grid Search algorithm for aerial imagery semantic segmentation of natural disaster-affected areas, which produces the best parameters and criteria with the highest accuracy results, we apply it to two semantic segmentation network models, namely: U-Net and PSPNet, to produce the most optimal DCN performance. We also carried out comprehensive testing of the PSPNet model with multiple layers to verify the relationship between the number of layers and performance improvements. We used PSPNet-(18, 34, 50, 101, 152).

Tests were carried out with several scenarios, and performance comparisons were made with models using default parameters and criteria (baselines). Each PSPNet model in all layers and the U-Net model were tested with several scenarios: using the default parameters and criteria, the best parameters, and the best parameters and criteria. The default parameters and criteria for the U-Net model were obtained from the study [44], and the default parameters and criteria for the PSPNet model were obtained from the study [45]. To distinguish all these scenarios, we added a letter abbreviation behind the model that uses the best parameters and criteria. The abbreviation with the letter "bp" means that the model uses the best parameters, while the abbreviation with the letter "bpc" means that the model uses the best parameters and criteria. A model that does not have an additional letter abbreviation behind it is a model that uses default parameters and criteria (baselines). The model, which uses additional abbreviations for the letters “bp” and “bpc,” is the result of our proposed framework (this study).

The comparison of the results of the overall performance testing of the U-Net and PSPNet models is presented in Table III, and the results of the network model testing with intersection over union values for each object class are shown in Table IV.

Based on the test results shown in Table III and Table IV, the most optimal DCN performance is achieved by the PSPNet(152) (bpc) model that uses the ResNet-152 architecture as the backbone. The network model fully uses the best parameters and criteria.

TABLE III
PERFORMANCE TESTING RESULTS OF DEEP CONVOLUTIONAL NETWORKS MODELS FOR AERIAL IMAGERY SEMANTIC SEGMENTATION OF NATURAL DISASTER-AFFECTED AREAS

Model	Parameters and criteria	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
U-Net	Default	95.44	71.20	61.83	64.37
U-Net (bp)	Best parameters	96.71	76.28	63.01	64.69
U-Net (bpc)	Best parameters and criteria	96.51	76.73	67.29	69.41
PSPNet(18)	Default	95.57	76.98	48.97	54.30
PSPNet(18) (bp)	Best parameters	94.62	71.47	57.07	60.44
PSPNet(18) (bpc)	Best parameters and criteria	96.65	81.93	60.68	65.74
PSPNet(34)	Default	94.27	65.83	55.02	56.66
PSPNet(34) (bp)	Best parameters	96.89	84.12	60.94	67.21
PSPNet(34) (bpc)	Best parameters and criteria	96.60	80.97	63.34	70.17
PSPNet(50)	Default	96.16	75.65	52.92	57.69
PSPNet(50) (bp)	Best parameters	96.75	75.35	72.02	73.44
PSPNet(50) (bpc)	Best parameters and criteria	97.23	77.47	73.97	75.44
PSPNet(101)	Default	93.46	53.81	43.13	44.40
PSPNet(101) (bp)	Best parameters	97.44	89.44	71.58	77.52
PSPNet(101) (bpc)	Best parameters and criteria	98.88	90.39	89.07	89.64
PSPNet(152)	Default	95.02	71.12	50.33	54.17
PSPNet(152) (bp)	Best parameters	97.79	82.34	81.39	81.73
PSPNet(152) (bpc)	Best parameters and criteria	**98.99**	**90.84**	**90.47**	**90.65**

TABLE IV
PERFORMANCE TESTING RESULTS OF DEEP CONVOLUTIONAL NETWORKS MODELS WITH INTERSECTION OVER UNION VALUE (IN %) FOR EACH OBJECT CLASS

Model	Building-flooded	Building-non-flooded	Road-flooded	Road-non-flooded	Water	Tree	Vehicle	Pool	Grass	mIoU
U-Net	49.28	54.45	30.28	52.67	53.25	64.79	22.44	39.80	74.81	49.09
U-Net (bp)	0.10	51.40	28.12	71.86	62.43	75.54	41.53	58.53	83.68	52.58
U-Net (bpc)	65.93	60.99	30.39	62.73	59.42	69.92	30.57	35.26	81.73	55.22
PSPNet(18)	0.77	61.10	6.79	67.70	49.51	72.82	18.96	34.04	73.44	42.79
PSPNet(18) (bp)	57.65	49.88	27.04	55.07	50.25	46.04	9.32	43.23	71.38	45.54
PSPNet(18) (bpc)	70.27	52.91	35.34	69.15	57.70	72.54	27.23	8.47	82.17	52.86
PSPNet(34)	55.63	50.55	28.81	46.26	52.25	35.10	2.19	38.37	70.93	42.23
PSPNet(34) (bp)	67.83	53.73	32.70	68.53	66.05	73.80	9.19	34.87	82.74	54.38
PSPNet(34) (bpc)	49.01	66.97	29.93	66.65	59.75	73.80	28.64	49.00	80.91	56.07
PSPNet(50)	45.57	51.56	35.74	53.53	59.04	68.47	8.83	2.80	80.13	45.07
PSPNet(50) (bp)	66.05	67.77	49.03	60.11	63.67	72.18	26.82	49.02	81.87	59.61
PSPNet(50) (bpc)	67.24	66.36	53.70	66.48	67.09	73.44	32.04	47.04	84.33	61.97
PSPNet(101)	4.70	38.11	26.62	28.13	47.32	48.73	11.79	11.62	63.72	31.19
PSPNet(101) (bp)	74.54	76.13	61.68	76.19	78.46	67.78	20.64	52.65	83.64	65.75
PSPNet(101) (bpc)	81.96	87.57	71.22	88.75	83.42	90.88	64.21	74.66	92.73	81.71
PSPNet(152)	12.81	53.19	30.45	43.01	48.90	65.48	11.44	24.21	72.75	40.25
PSPNet(152) (bp)	79.25	76.29	68.85	72.55	74.52	77.36	39.27	57.67	86.56	70.26
PSPNet(152) (bpc)	84.55	88.83	75.56	88.77	86.07	90.89	63.94	77.69	93.77	**83.34**

Based on the test results, it is also proven that the PSPNet(152) (bpc) model can detect and identify various objects with irregular shapes and sizes, can detect and identify various important objects affected by natural disasters, such as buildings and roads that are flooded, and can detect and identify objects with small shapes such as vehicles and pools, which is the most challenging task for semantic segmentation network models. This ability can be seen from the fairly high value of IoU for each object class and mIoU. The test results in this study prove that there is an increase in DCN's performance in producing aerial imagery semantic segmentation of natural disaster-affected areas accurately.

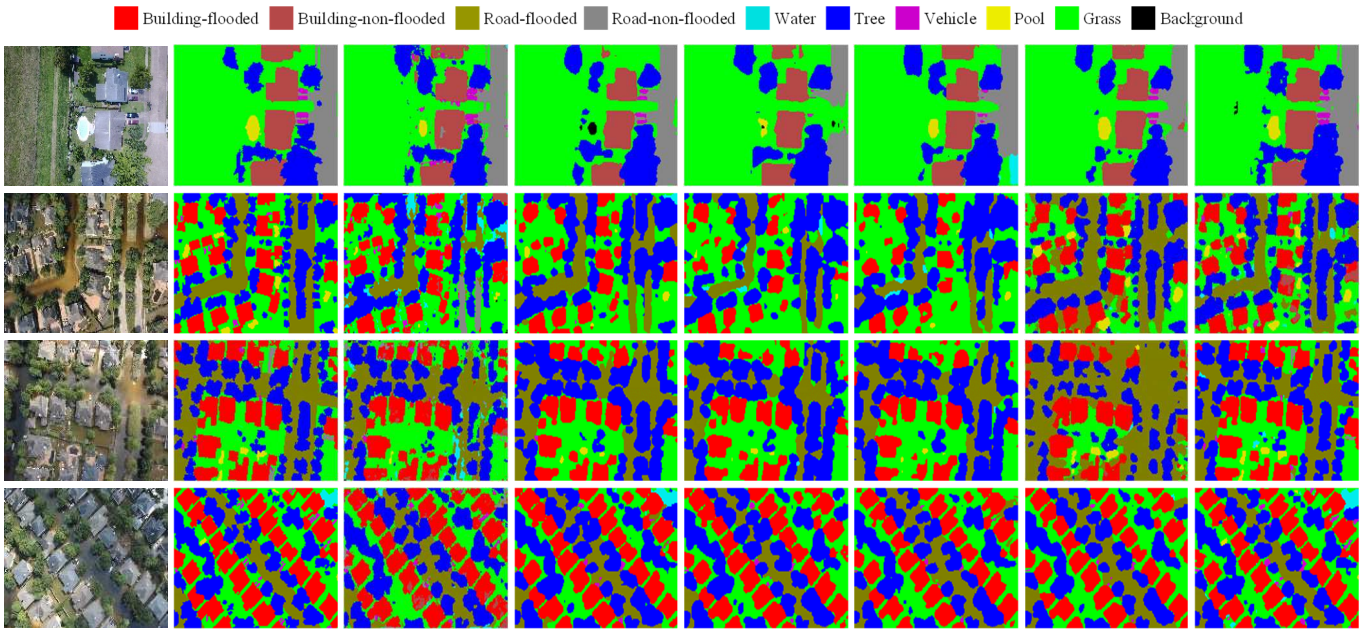
Based on the test results, we learned that using the best parameters, appropriate data augmentation criteria, and suitable pre-training dataset criteria can significantly improve DCN performance in aerial imagery semantic segmentation of natural disaster-affected areas, compared to using only default parameters and criteria (baselines). In addition, in scenarios that use the best parameters and scenarios that use the best parameters and criteria, the effect of increasing the number of layers in the PSPNet-(18, 34, 50, 101, 152) model results in an increase in the performance of the network model, which can be seen from the rise in mIoU value.

Our test results have advantages compared to the results of tests carried out by several studies in the literature review, which also use FloodNet as an evaluation dataset for training, validation, and testing in recognizing aerial images of natural disasters with semantic segmentation. In the study [32], it produced a mIoU value of 80.35%, which is the highest mIoU value in the study for the PSPNet(101) model, while in our study, it produced a higher mIoU value of 81.71% for the same model PSPNet(101). In the study [35], the mIoU value for the U-Net model was 23.9%, and the PSPNet(101) model was 46.65%, while in our study, the higher mIoU value was 55.22% for the U-Net model and 81.71% for the PSPNet(101) model. In the study [36], the highest mIoU value for the PSPNet(152) model was 56%, while in our study, the highest mIoU value for the PSPNet(152) model was 83.34%.

The results of the most optimal DCN model using the best parameters and criteria, namely the PSPNet(152) (bpc) model, were validated using the k-fold cross-validation method to evaluate the performance and validate the accuracy of the model. The validation results are shown in Table V. To visually see the accuracy of the DCN model in displaying the results of the aerial imagery semantic segmentation of natural disaster-affected areas. We present a visual comparison of the DCN model using the best parameters and criteria in Fig. 6.

TABLE V
RESULTS OF K-FOLD CROSS-VALIDATION ON DEEP CONVOLUTIONAL NETWORKS OPTIMAL MODEL

K-fold	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	mIoU (%)
1	99.09	89.49	91.40	90.37	82.93
2	98.99	89.70	90.37	90.00	82.35
3	99.03	90.64	90.88	90.74	83.47
4	99.08	91.14	91.46	91.28	84.32
5	99.08	91.24	91.34	91.28	84.36
6	99.05	91.26	90.80	91.02	83.94
7	99.05	91.11	91.02	91.06	83.99
8	99.02	91.13	90.80	90.96	83.84
9	99.03	91.18	90.69	90.93	83.79
10	98.99	90.84	90.47	90.65	83.34
Average	**99.04**	**90.77**	**90.92**	**90.83**	**83.63**



(a) Aerial imagery of natural disasters-affected areas
(b) Ground truth
(c) U-Net (bpc)
(d) PSPNet(18) (bpc)
(e) PSPNet(34) (bpc)
(f) PSPNet(50) (bpc)
(g) PSPNet(101) (bpc)
(h) PSPNet(152) (bpc)

Fig. 6 Visual comparison of deep convolutional networks model for aerial imagery semantic segmentation of natural disaster-affected areas using the best parameters and criteria

IV. Conclusion

This study revealed an effective way of improving the performance of Deep Convolutional Networks (DCN) for aerial imagery semantic segmentation of natural disaster-affected areas. An experimental study was conducted using the parameters in DCN, the appropriate data augmentation criteria, and the suitable dataset criteria in pre-training to produce the most optimal performance. In this study, we have integrated the Grid Search algorithm and DCN, and validated the results with the k-fold cross-validation method. The Grid Search algorithm performs parameter tuning on DCN, data augmentation criteria tuning, and dataset criteria tuning for pre-training.

This study uses U-Net and PSPNet as DCN models for semantic segmentation. The results of this study indicate that the Grid Search algorithm obtained the best parameters and criteria and improved the performance of the two models used. The most optimal DCN performance is achieved by the PSPNet(152) (bpc) model, which fully uses the best parameters and criteria, with an accuracy value of 98.99%, precision of 90.84%, recall of 90.47%, f1-score of 90.65%, and mean intersection over union (mIoU) of 83.34%. The validation results using the k-fold cross-validation method on the most optimal DCN model, namely PSPNet(152) (bpc), resulting in an average accuracy of 99.04%, precision of 90.77%, recall of 90.92%, f1-score of 90.83%, and mIoU of 83.63%. Significantly mIoU increased by 43.09% (83.34% − 40.25%) in the PSPNet(152) (bpc) model compared to only using default parameters and criteria (baselines). Likewise for other models, namely U-Net (bpc), PSPNet(18) (bpc), PSPNet(34) (bpc), PSPNet(50) (bpc), and PSPNet(101) (bpc), results in increased mIoU compared to only using default parameters and criteria (baselines).

The PSPNet(152) (bpc) model is able to detect and identify various objects with irregular shapes and sizes, is able to detect and identify various important objects affected by natural disasters such as flooded buildings and roads, and is able to detect and identify objects with small shapes such as vehicles and pools, which is the most challenging task for semantic segmentation network models. This capability can be seen from the results of a fairly high value in IoU for each object class, the mIoU value, and the visual display results. This study also proves that the effect of increasing the number of layers in the PSPNet-(18, 34, 50, 101, 152) model results in an increase in the model's performance. The results of this study prove that the proposed framework contributes to improving DCN performance to accurately produce aerial imagery semantic segmentation of natural disaster-affected areas.

We obtained several knowledge findings in this study, namely: 1) the combined aerial imagery pre-training dataset originating from the UAV, which has high image resolution, good segmentation annotations, and multiple labels, is very precise and has great opportunities for transfer learning of the dataset aerial imagery of areas affected by natural disasters, provided that the number of images is large enough. So for future research, we suggest the need to utilize a special dataset from aerial imagery originating from UAVs at the pre-training stage for transfer learning in improving DCN performance; 2) increase the amount of data practically by simply combining multiple datasets and proving the combined datasets increase the level of performance compared to pre-training only one dataset. We suggest that for future research, it is necessary to combine multiple datasets containing complex images (real-world images + urban images + road images) and multiple labels to improve accuracy; 3) carefully annotated combined pre-training datasets, effectively training aerial imagery datasets for semantic segmentation tasks; and 4) using the best parameters, appropriate data augmentation criteria, and suitable pre-training dataset criteria can significantly improve DCN's performance in aerial imagery semantic segmentation of natural disaster-affected areas, compared to only using default parameters and criteria (baselines).

We also see opportunities for further research. With the advent of Transformers, it can also be explored for aerial imagery semantic segmentation of natural disaster-affected areas and compare the results with DCN to obtain the most optimal performance.

Acknowledgment

This research was funded by the Directorate of Research, Technology, and Community Service, Directorate General of Higher Education, Research, and Technology, Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia in the Doctoral Dissertation Research (*Penelitian Disertasi Doktor - PDD*) program.

References

[1] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image Segmentation Using Deep Learning: A Survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1–1, 2021, doi: 10.1109/tpami.2021.3059968.

[2] X. Yuan, J. Shi, and L. Gu, "A review of deep learning methods for semantic segmentation of remote sensing imagery," Expert Systems with Applications, vol. 169, p. 114417, May 2021, doi:10.1016/j.eswa.2020.114417.

[3] F. Sultana, A. Sufian, and P. Dutta, "Evolution of Image Segmentation using Deep Convolutional Neural Network: A Survey," Knowledge-Based Systems, vol. 201–202, p. 106062, Aug. 2020, doi:10.1016/j.knosys.2020.106062.

[4] A. Bakhtiarnia, Q. Zhang, and A. Iosifidis, "Efficient High-Resolution Deep Learning: A Survey," *arXiv e-prints*, Jul. 2022, doi:10.48550/arXiv.2207.13050.

[5] M. Kampffmeyer, A.-B. Salberg, and R. Jenssen, "Semantic Segmentation of Small Objects and Modeling of Uncertainty in Urban Remote Sensing Images Using Deep Convolutional Neural Networks," 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Jun. 2016, doi: 10.1109/cvprw.2016.90.

[6] D. Marmanis, J. D. Wegner, S. Galliani, K. Schindler, M. Datcu, and U. Stilla, "SEMANTIC SEGMENTATION OF AERIAL IMAGES WITH AN ENSEMBLE OF CNNS," ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. III–3, pp. 473–480, Jun. 2016, doi: 10.5194/isprsannals-iii-3-473-2016.

[7] Y. Chen et al., "LightFGCNet: A Lightweight and Focusing on Global Context Information Semantic Segmentation Network for Remote Sensing Imagery," Remote Sensing, vol. 14, no. 24, p. 6193, Dec. 2022, doi: 10.3390/rs14246193.

[8] Z. Fan, J. Hou, Q. Zang, Y. Chen, and F. Yan, "River Segmentation of Remote Sensing Images Based on Composite Attention Network," Complexity, vol. 2022, pp. 1–13, Jan. 2022, doi:10.1155/2022/7750281.

[9] A. Abdollahi, B. Pradhan, and A. M. Alamri, "An ensemble architecture of deep convolutional Segnet and Unet networks for building semantic segmentation from high-resolution aerial images," Geocarto International, vol. 37, no. 12, pp. 3355–3370, Dec. 2020, doi:10.1080/10106049.2020.1856199.

[10] C. Sebastian, R. Imbriaco, E. Bondarev, and P. H. N. de With, "Adversarial Loss for Semantic Segmentation of Aerial Imagery," in

Proc. IEEE Symposium on Information Theory and Signal Processing, Benelux: IEEE, 2020, pp. 1–5. doi: 10.48550/arXiv.2001.04269.

[11] D. Q. Tran, M. Park, D. Jung, and S. Park, "Damage-Map Estimation Using UAV Images and Deep Learning Algorithms for Disaster Management System," Remote Sensing, vol. 12, no. 24, p. 4169, Dec. 2020, doi: 10.3390/rs12244169.

[12] J. Wang, X. Fan, X. Yang, T. Tjahjadi, and Y. Wang, "Semi-Supervised Learning for Forest Fire Segmentation Using UAV Imagery," Forests, vol. 13, no. 10, p. 1573, Sep. 2022, doi:10.3390/f13101573.

[13] S. Muksimova, S. Mardieva, and Y.-I. Cho, "Deep Encoder–Decoder Network-Based Wildfire Segmentation Using Drone Images in Real-Time," Remote Sensing, vol. 14, no. 24, p. 6302, Dec. 2022, doi:10.3390/rs14246302.

[14] M. Yandouzi *et al.*, "Review on Forest Fires Detection and Prediction Using Deep Learning and Drones," *J Theor Appl Inf Technol*, vol. 100, no. 12, pp. 4565–4576, Jun. 2022.

[15] K. K. Eerapu, S. Lal, and A. V. Narasimhadhan, "O-SegNet: Robust Encoder and Decoder Architecture for Objects Segmentation From Aerial Imagery Data," IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 6, no. 3, pp. 556–567, Jun. 2022, doi:10.1109/tetci.2020.3045485.

[16] M. S. Iqbal, H. Ali, S. N. Tran, and T. Iqbal, "Coconut trees detection and segmentation in aerial imagery using mask region‐based convolution neural network," IET Computer Vision, vol. 15, no. 6, pp. 428‐439, Apr. 2021, doi: 10.1049/cvi2.12028.

[17] M. Kakooei and Y. Baleghi, "Fusion of satellite, aircraft, and UAV data for automatic disaster damage assessment," International Journal of Remote Sensing, vol. 38, no. 8–10, pp. 2511–2534, Feb. 2017, doi:10.1080/01431161.2017.1294780.

[18] N. S. Ibrahim, S. M. Sharun, M. K. Osman, S. B. Mohamed, and S. H. Y. S. Abdullah, "The application of UAV images in flood detection using image segmentation techniques," Indonesian Journal of Electrical Engineering and Computer Science, vol. 23, no. 2, p. 1219, Aug. 2021, doi: 10.11591/ijeecs.v23.i2.pp1219-1226.

[19] O. Ghorbanzadeh, T. Blaschke, K. Gholamnia, S. Meena, D. Tiede, and J. Aryal, "Evaluation of Different Machine Learning Methods and Deep-Learning Convolutional Neural Networks for Landslide Detection," Remote Sensing, vol. 11, no. 2, p. 196, Jan. 2019, doi:10.3390/rs11020196.

[20] S. N. K. B. Amit and Y. Aoki, "Disaster detection from aerial imagery with convolutional neural network," 2017 International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC), Sep. 2017, doi: 10.1109/kcic.2017.8228593.

[21] A. Fujita, K. Sakurada, T. Imaizumi, R. Ito, S. Hikosaka, and R. Nakamura, "Damage detection from aerial images via convolutional neural networks," 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA), May 2017, doi:10.23919/mva.2017.7986759.

[22] N. Attari, F. Ofli, M. Awad, J. Lucas, and S. Chawla, "Nazr-CNN: Fine-Grained Classification of UAV Imagery for Damage Assessment," 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Oct. 2017, doi:10.1109/dsaa.2017.72.

[23] M. Rahnemoonfar, R. Murphy, M. V. Miquel, D. Dobbs, and A. Adams, "Flooded Area Detection from Uav Images Based on Densely Connected Recurrent Neural Networks," IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, Jul. 2018, doi: 10.1109/igarss.2018.8517946.

[24] D. Popescu, L. Ichim, and F. Stoican, "Flooded Area Segmentation from UAV Images Based on Generative Adversarial Networks," 2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV), Nov. 2018, doi: 10.1109/icarcv.2018.8581341.

[25] A. Gebrehiwot, L. Hashemi-Beni, G. Thompson, P. Kordjamshidi, and T. Langan, "Deep Convolutional Neural Network for Flood Extent Mapping Using Unmanned Aerial Vehicles Data," Sensors, vol. 19, no. 7, p. 1486, Mar. 2019, doi: 10.3390/s19071486.

[26] A. Gupta, S. Watson, and H. Yin, "Deep learning-based aerial image segmentation with open data for disaster impact assessment," Neurocomputing, vol. 439, pp. 22–33, Jun. 2021, doi:10.1016/j.neucom.2020.02.139.

[27] T. Chowdhury, M. Rahnemoonfar, R. Murphy, and O. Fernandes, "Comprehensive Semantic Segmentation on High Resolution UAV Imagery for Natural Disaster Damage Assessment," 2020 IEEE International Conference on Big Data (Big Data), Dec. 2020, doi:10.1109/bigdata50022.2020.9377916.

[28] Y. Pi, N. D. Nath, and A. H. Behzadan, "Detection and Semantic Segmentation of Disaster Damage in UAV Footage," Journal of Computing in Civil Engineering, vol. 35, no. 2, Mar. 2021, doi:10.1061/(asce)cp.1943-5487.0000947.

[29] X. Zhu, J. Liang, and A. Hauptmann, "MSNet: A Multilevel Instance Segmentation Network for Natural Disaster Damage Assessment in Aerial Videos," 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), Jan. 2021, doi:10.1109/wacv48630.2021.00207.

[30] T. Chowdhury and M. Rahnemoonfar, "Attention Based Semantic Segmentation on UAV Dataset for Natural Disaster Damage Assessment," in *Proc. 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, Brussels, Belgium: IEEE, 2021, pp. 1–5. doi: 10.48550/arXiv.2105.14540.

[31] H. S. Munawar, F. Ullah, S. Qayyum, S. I. Khan, and M. Mojtahedi, "UAVs in Disaster Management: Application of Integrated Aerial Imagery and Convolutional Neural Network for Flood Detection," Sustainability, vol. 13, no. 14, p. 7547, Jul. 2021, doi:10.3390/su13147547.

[32] M. Rahnemoonfar, T. Chowdhury, A. Sarkar, D. Varshney, M. Yari, and R. R. Murphy, "FloodNet: A High Resolution Aerial Imagery Dataset for Post Flood Scene Understanding," IEEE Access, vol. 9, pp. 89644–89654, 2021, doi: 10.1109/access.2021.3090981.

[33] R. Ünlü and R. Kiriş, "Detection of damaged buildings after an earthquake with convolutional neural networks in conjunction with image segmentation," The Visual Computer, vol. 38, no. 2, pp. 685–694, Jan. 2021, doi: 10.1007/s00371-020-02043-9.

[34] T. Chowdhury and M. Rahnemoonfar, "Self Attention Based Semantic Segmentation on a Natural Disaster Dataset," 2021 IEEE International Conference on Image Processing (ICIP), Sep. 2021, doi:10.1109/icip42928.2021.9506366.

[35] S. Khose, A. Tiwari, and A. Ghosh, "Semi-Supervised Classification and Segmentation on High Resolution Aerial Images," in *Proc. EARTHVISION 2021, Computer Vision and Pattern Recognition (CVPR) 2021 Conference*, Ithaca, New York: Cornell University, 2021, pp. 1–5. doi: 10.48550/arXiv.2105.08655.

[36] D. Hernández, J. M. Cecilia, J.-C. Cano, and C. T. Calafate, "Flood Detection Using Real-Time Image Segmentation from Unmanned Aerial Vehicles on Edge-Computing Platform," Remote Sensing, vol. 14, no. 1, p. 223, Jan. 2022, doi: 10.3390/rs14010223.

[37] Y. Zhan, W. Liu, and Y. Maruyama, "Damaged Building Extraction Using Modified Mask R-CNN Model Using Post-Event Aerial Images of the 2016 Kumamoto Earthquake," Remote Sensing, vol. 14, no. 4, p. 1002, Feb. 2022, doi: 10.3390/rs14041002.

[38] J. F. Guerrero Tello, M. Coltelli, M. Marsella, A. Celauro, and J. A. Palenzuela Baena, "Convolutional Neural Network Algorithms for Semantic Segmentation of Volcanic Ash Plumes Using Visible Camera Imagery," Remote Sensing, vol. 14, no. 18, p. 4477, Sep. 2022, doi: 10.3390/rs14184477.

[39] C. Fang, X. Fan, H. Zhong, L. Lombardo, H. Tanyas, and X. Wang, "A Novel Historical Landslide Detection Approach Based on LiDAR and Lightweight Attention U-Net," Remote Sensing, vol. 14, no. 17, p. 4357, Sep. 2022, doi: 10.3390/rs14174357.

[40] H. Zhang, M. Wang, Y. Zhang, and G. Ma, "TDA-Net: A Novel Transfer Deep Attention Network for Rapid Response to Building Damage Discovery," Remote Sensing, vol. 14, no. 15, p. 3687, Aug. 2022, doi: 10.3390/rs14153687.

[41] L. P. Soares, H. C. Dias, G. P. B. Garcia, and C. H. Grohmann, "Landslide Segmentation with Deep Learning: Evaluating Model Generalization in Rainfall-Induced Landslides in Brazil," Remote Sensing, vol. 14, no. 9, p. 2237, May 2022, doi: 10.3390/rs14092237.

[42] V. Boehm *et al.*, "Deep Learning for Rapid Landslide Detection using Synthetic Aperture Radar (SAR) Datacubes," *ArXiv*, vol. abs/2211.02869, pp. 1–8, Nov. 2022, doi: 10.48550/arXiv.2211.02869.

[43] F. Zhang, Y. Shi, Q. Xu, Z. Xiong, W. Yao, and X. X. Zhu, "On the Generalization of the Semantic Segmentation Model for Landslide Detection," in *Proceedings of the Second Workshop on Complex Data Challenges in Earth Observation (CDCEO 2022)*, Vienna, Austria: CEUR Workshop Proceedings (CEUR-WS.org), Jul. 2022, pp. 97–101.

[44] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, pp. 234–241, 2015, doi: 10.1007/978-3-319-24574-4_28.

[45] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid Scene Parsing Network," in *Proc. IEEE Conference on Computer Vision and Pattern*

Recognition (CVPR), Honolulu, HI, USA: IEEE Computer Society, Dec. 2017, pp. 1–11. doi: 10.48550/arXiv.1612.01105.

[46] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2016, doi: 10.1109/cvpr.2016.90.

[47] O. Russakovsky et al., "ImageNet Large Scale Visual Recognition Challenge," International Journal of Computer Vision, vol. 115, no. 3, pp. 211–252, Apr. 2015, doi: 10.1007/s11263-015-0816-y.

[48] T.-Y. Lin *et al.*, "Microsoft COCO: Common Objects in Context," in *Proc. European Conference on Computer Vision (ECCV 2014)*, Zurich, Switzerland: Springer, Cham, Sep. 2014, pp. 740–755. doi:10.48550/arXiv.1405.0312.

[49] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes Challenge: A Retrospective," International Journal of Computer Vision, vol. 111, no. 1, pp. 98–136, Jun. 2014, doi: 10.1007/s11263-014-0733-5.

[50] M. Cordts *et al.*, "The Cityscapes Dataset for Semantic Urban Scene Understanding," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, United States: IEEE Computer Society, 2016, pp. 3213–3223. doi:10.48550/arXiv.1604.01685.

[51] H. Tung, "dira-simulator-road-segment," 2019. https://www.kaggle.com/datasets/phamthaihoangtung/dirasimulatorroadsegment (accessed Sep. 05, 2022).

[52] G. Neuhold, T. Ollmann, S. R. Bulo, and P. Kontschieder, "The Mapillary Vistas Dataset for Semantic Understanding of Street Scenes," 2017 IEEE International Conference on Computer Vision (ICCV), Oct. 2017, doi: 10.1109/iccv.2017.534.

[53] Y. Lyu, G. Vosselman, G.-S. Xia, A. Yilmaz, and M. Y. Yang, "UAVid: A semantic segmentation dataset for UAV imagery," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 165, pp. 108–119, Jul. 2020, doi: 10.1016/j.isprsjprs.2020.05.009.

[54] Graz University of Technology, "Semantic Drone Dataset," 2022. http://dronedataset.icg.tugraz.at (accessed Sep. 05, 2022).

[55] M. Abbaszadeh, S. Soltani-Mohammadi, and A. N. Ahmed, "Optimization of support vector machine parameters in modeling of Iju deposit mineralization and alteration zones using particle swarm optimization algorithm and grid search method," Computers & Geosciences, vol. 165, p. 105140, Aug. 2022, doi:10.1016/j.cageo.2022.105140.

[56] L. Yao, Z. Fang, Y. Xiao, J. Hou, and Z. Fu, "An Intelligent Fault Diagnosis Method for Lithium Battery Systems Based on Grid Search Support Vector Machine," Energy, vol. 214, p. 118866, Jan. 2021, doi:10.1016/j.energy.2020.118866.