RESEARCH ARTICLE

Utilization of Improved Annotations From Object-Based Image Analysis as Training Data for DeepLab V3+ Model: A Focus on Road Extraction in Very High-Resolution Orthophotos

SUSSI[1,2], EMIR HUSNI[1], RAHADIAN YUSUF[1], (Member, IEEE), AGUNG BUDI HARTO[3], DENI SUWARDHI[3], AND ARTHUR SIBURIAN[1]

[1]School of Electrical Engineering and Informatics, Institut Teknologi Bandung, Bandung 40132, Indonesia
[2]Department of Telecommunication, Faculty of Electrical Engineering, Telkom University, Bandung 40257, Indonesia
[3]Faculty of Earth Science and Technology, Institut Teknologi Bandung, Bandung 40132, Indonesia

Corresponding author: Emir Husni (ehusni@itb.ac.id)

ABSTRACT Road extraction plays a crucial role in various sectors, including transportation systems, disaster relief distribution, and urban planning, necessitating a more efficient method than the current labor-intensive approach. The conventional process, reliant on operator efforts, proves costly, time-consuming, and energy intensive. To address this, the employment of deep learning models such as DeepLab V3+ are done, leveraging encoders like ResNet 50, ResNet 101, and MobileNet V2, known for their effectiveness in deep learning applications. Although Deep Learning has demonstrated faster and more automated road extraction, the bottleneck persists in the manual creation of training data through road annotations by operators. Our research focuses on accelerating road extraction by incorporating enhanced annotations from Object-Based Image Analysis (OBIA) along with organic annotations into the training data to expedite dataset creation while ensuring extraction model accuracy. Specifically, we investigate the optimal ratio of synthetic to organic annotations that yields the highest road extraction accuracy. Moreover, we enhance OBIA-derived road annotations and regulate their integration into the training data. Our findings reveal an optimal composition of 25% for OBIA annotations and 50% for improved OBIA annotations, as exceeding these numbers results in diminished model performance. Significantly, the further improvement of OBIA annotations substantially boosts model performance metrics exemplified in the use of 100% composition in training data. On average, each model produces Pixel Accuracy of 0.942, IoUr of 0.012, mean IoU of 0.477, and Dice Score of 0.495 for every use of 100% OBIA annotations in the training data. Improvement in model performance evaluation metrics occurs when using 100% improved OBIA annotations in the training data where on average each model produces Pixel Accuracy values of 0.954, IoUr of 0.433, mean IoU of 0.692, and Dice Score of 0.771. The experimental results demonstrate the advantages of our proposed method, indicating a reduction in the time required to prepare Deep Learning datasets, reducing the number of organic annotations required to as little as 50% while maintaining model performance by leveraging OBIA road annotations as training data.

INDEX TERMS Deep learning, object-based image analysis, deeplab v3+, pixel accuracy, mean IoU.

The associate editor coordinating the review of this manuscript and approving it for publication was Geng-Ming Jiang.

I. INTRODUCTION

Road extraction has emerged as a preeminent concern within the realm of remote sensing investigation over the past ten

years [1], [2], [3] primarily due to its extensive utilization in various aspects of daily existence, including but not limited to, intelligent transportation system path determination, disaster navigation, cartography, goods distribution, and urban planning [4], [5], [6]. Nonetheless, the process of road extraction frequently relies on manual methods, requiring operators to directly delineate roads by visually examining orthophotos and in instances of complications must physically visit the location, introducing inefficiencies and time constraints to the road extraction process [7], [8], [9]. Therefore, researchers are conducting various studies to speed up the process of making road extractions [10], [11].

The choice of employing semiautomatic or automatic techniques with Deep Learning (DL) has been favored in the pursuit of expedited road extraction in the realm of research. Various investigations have explored the employment of semi-automatic methods in road extraction. For instance, Jiguang Dai devised a multiscale line segment orientation histogram (MLSOH) to accentuate road curvature [12], [13]. You Wu employed the Clustering Point Process (CPP) algorithm [14], while Jinming Zhan utilized graph theory to connect edges and nodes for road extraction. [15]. Some research studies have investigated the application of deep learning techniques, such as VNet [16], Dual-Attention Network-RoadNet (DA-Roadnet) [17], Direction-Aware Residual Network (DiResNet) [18], Multiple Parameter Guided Squeeze and Excitation Integrated D-LinkNet (MPGSE-D-LinkNet) [19], DeepLab V3+ [20], and Fully Convolutional Network (FCN) [21]. Semi-automated methodologies encounter limitations in road extraction, as human intervention is still required to set algorithmic parameters, and the applicability is constrained to specific study locations, limiting generalization [22], [23], [24]. In contrast, deep learning (DL) excels in road extraction generation; however, it mandates road annotations during training, a task performed manually by operators, demanding significant time and labor resources for the provision of essential training data [25], [26], [27], [28].

In our investigation, we observed road extraction results from the DeepLab V3+ model trained with data derived from both operator annotations and annotations generated through semi-automated methods with the primary objective of streamlining the provision of training data through the incorporation of synthetic data while preserving the precision of the road extraction model's performance. Our study holds significant utility, potentially reducing training data provision time by operators by 25-50%. We opt for the semi-automated approach of Object-Based Image Analysis (OBIA), leveraging its proficiency in delineating objects on very high-resolution orthophotos based on color proximity and object distance [29]. Furthermore, we enhance the annotations of OBIA results through morphology image processing to obtain optimal road

annotations for effective integration into the model training process.

The novelty of our research encompasses several key aspects, including:

1) The novelty inherent in our research is multifaceted. Firstly, we distinguish ourselves by utilizing a dataset not readily accessible through open-access channels. While prevalent studies on road extraction commonly leverage datasets like Massachusetts and Deep Globe, we diverge by employing an intricately detailed orthophoto of Cimahi City, Indonesia. This unique dataset not only boasts higher resolution but also serves as a representation of a developing nation within Southeast Asia. In contrast to the developed counterparts prevalent in the latter datasets, the heightened intricacy of road extraction challenges in developing countries introduces an additional layer of complexity to our research.

2) our research extends to the training data employed for the DeepLab V3+ model. Departing from conventional methodologies reliant solely on annotations by human operators, we integrate road annotations derived from Object-Based Image Analysis (OBIA). Furthermore, we introduce additional annotations obtained through a meticulous refinement of the original OBIA annotations. Striking a delicate balance, we carefully adjust the composition of both OBIA annotations and improved OBIA annotations to optimize the creation of a highly accurate road extraction model.

3) To augment the quality of OBIA annotations, we employ morphology image processing techniques to fill road pixel gaps in the OBIA annotations, thus obtaining road annotations that closely align with the true values of the road. Moreover, we then trained two different models, one trained on fully organic annotations, and one trained on a mixture of OBIA-annotated and human-annotated dataset. The road extraction results of both models are then compared, showcasing the performance of each model.

We organize this document systematically into four distinct sections. The introduction section functions as an introduction, providing a comprehensive overview of the contextual background, research objectives, and the unique aspects of the investigation. Subsequently, the Methodology section delves into intricate details concerning the dataset, Object-Based Image Analysis (OBIA), enhancement of OBIA annotations, the implementation of DeepLab V3+, and the subsequent evaluation of model performance. Following this, the third section meticulously outlines the various experiments conducted and presents the corresponding outcomes of these experiments. The last section comprises an in-depth analysis of the research's conclusions, complemented by suggestions for future research initiatives aimed at advancing the current study.

II. METHODOLOGY

Our research, centered on Deep Learning (DL), aimed to expedite the creation of accurate road extraction models, is conducted as outlined in Figure 1. The road annotations for training the model are retrieved from a variety of sources, including human operator, Object-Based Image Analysis (OBIA) annotations, and refined OBIA annotations, each of which we observed for their impact. The utilization of OBIA for annotation purposes is implemented to expedite the furnishing of training data; however, the results of road annotation remain imperfect, necessitating the need for enhancements.

To train the road extraction models, the aforementioned very high-resolution orthophotos and their road annotations are employed as training data. The VHR orthophoto and manual annotation are utilized as test data to assess the performance of the model by means of Pixel Accuracy (PA), mean Intersection Over Union (mIoU), IoU Road (IoUR), and Dice Score. This methodology section is subdivided into detailed discussions covering the dataset, OBIA, the refinement of OBIA annotations, DeepLab V3+, and the assessment of model performance.
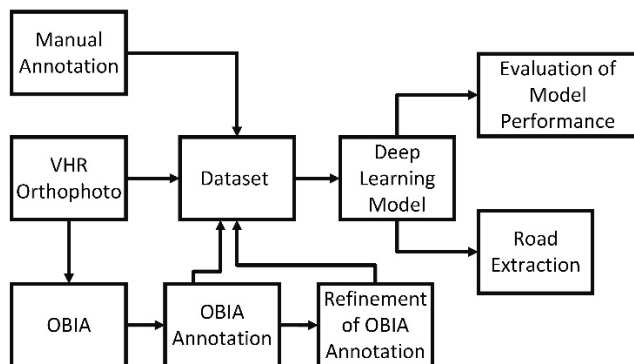

FIGURE 1. Overall research methodology.

A. DATASET

The dataset employed in the investigation consists of orthophotos, which are photographs captured perpendicular to the earth's surface, possessing an exceptionally high resolution, with a spatial resolution value of 0.05 meters. The utilization of orthophotos with such high resolution enables the clear and detailed depiction of objects, thereby contributing to the complexities encountered in the realm of road extraction research [30], [31].

The investigation focuses on the urban landscape of Cimahi City, a municipality situated in Indonesia with coordinates latitude -6.899541 and longitude 107.533867. Cimahi is characterized by its densely populated residential areas, industrial zones, rivers, vacant land, and extensive rice fields and tree-covered expanses. Notably, several geometric entities within the Cimahi dataset share similar coloration with roads, including building rooftops, rivers, and cross-sections of rice fields. This similarity poses an additional layer of complexity in the challenges encountered during the road extraction study.

The datasets employed in each research scenario amounted to a total of 4400 images with dimension size of 1008×1008, comprising orthophotos and corresponding road annotations. The allocation of training, validation, and test data followed a distribution ratio of 8:1:1. The selection of this tiling size in particular was made to conform with the parameter settings for segmentation in the process of OBIA annotations generation while ensuring enough quantity of images for model training. A larger area will be encompassed given larger dimension; however, the quantity of training images would be even lower than 4400.

The orthophoto employed are RGB (Red-Green-Blue) images. The preference for RGB datasets is also as evidenced in various studies, showed them to be more favorable for land use and land cover classification (LULC) in comparison to alternative band combinations like short-wave infrared (SWIR) and color-infrared (CI). The presence of RGB bands within the visible spectrum renders them a desirable option as they offer adequate data to differentiate between diverse land cover categories [32]. The resulting road annotations encompassed manual annotations performed by an operator, road annotations derived from OBIA (Object-Based Image Analysis), and enhancements made to road annotations derived from OBIA. The composition of operator annotations, OBIA annotations, and improved OBIA annotations as road annotations within the dataset was adjusted in accordance with the research scenario, with test data consisting of only manual annotations.

The software and hardware employed in this study included Global Mapper 22.0, QGIS Desktop 3.24.1, and NVIDIA GeForce RTX 3090 Ti. The tiling of the Cimahi City orthophoto was executed using Global Mapper 22.0, while the training process to generate the road extraction model was carried out using the NVIDIA GeForce RTX 3090 Ti. QGIS Desktop 3.24.1 played a role in converting the orthophoto dataset into RGB bands for OBIA.

B. OJECT-BASED IMAGE ANALYSIS

Object-Based Image Analysis (OBIA), a semi-automated method we utilized to generate street annotations serving as train data for the training and validation stages of the Deep Learning model. Within the OBIA framework, two distinctive processes, segmentation and classification [33], [34], [35], contribute to the production of road annotations. The segmentation stage is particularly critical in determining the quality of OBIA's road annotations. Therefore, it is crucial to appropriately configure the segmentation parameters of OBIA to achieve accurate segmentation [36], [37], [38]. Figure 2 provides a visual representation of the OBIA architecture.

To achieve this, we employ a segmentation algorithm known as Simple Linear Iterative Clustering (SLIC) for the segmentation part. SLIC, a K-means Clustering-based algorithm, generates multiple clusters by grouping pixels

with similar color and position [39], [40]. The application of the SLIC algorithm makes OBIA well-suited for processing high-resolution orthophotos. Nonetheless, the quality of OBIA's road annotations is significantly influenced by the meticulous configuration of the SLIC algorithm parameters.
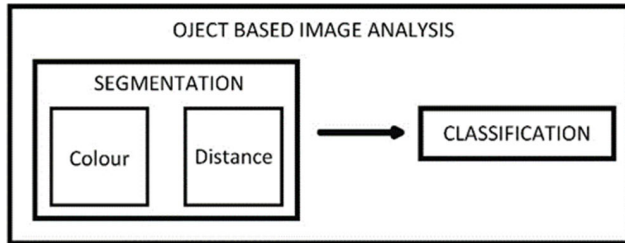

FIGURE 2. Architecture of OBIA.

We establish the parameters of the SLIC algorithm, including the number of segments, compactness, maximum number of iterations, connectivity enforcement, and the minimum and maximum size factors. The configuration of the SLIC algorithm's segmentation parameters is iteratively performed to obtain the optimal parameters that yield road annotations closely aligned with the ground truth value, while keeping undesired noise minimal. The optimal parameter settings for the SLIC algorithm we use in this study entails 2000 segments, a compactness value of 0.1, 50 iterations, a minimum size factor of 0.6, a maximum size factor of 3, and the enforcement of connectivity being set to true.

The classification phase within the Object-Based Image Analysis (OBIA) process employs the Histogram Gradient Boosting model. This classification process involves road and background classes. Initially, labels are randomly assigned to road and background areas in the orthophoto. The road labelling results derived from the orthophotos, and segmentation process are integrated into the classification phase, producing road annotations through the OBIA process resulting in annotations.

C. REFINEMENT OF OBIA ANNOTATION

The refinement of annotations obtained through Object-Based Image Analysis (OBIA) is imperative to ensure the generation of road annotations closely aligned with the actual ground truth. Objects such as vehicles, trees, or buildings obstructing the roads often introduce noise into OBIA annotations, leading to unconnected road segmentation. As a result, the road annotations may appear fragmented and deviate from a straight path. To mitigate this issue, we employ Morphology Image Processing, specifically utilizing the dilation operation, as expressed in the following equation,

$$S \oplus T = \{Z \mid [(Y_z) \cap S] \subseteq S\} \qquad (1)$$

where S represents the image, T denotes the kernel, Z signifies the number of pixels in the image, and Y represents the symmetry of T.

Dilation is a process that expands the boundaries of objects within the image [41] by configuring the kernel [42]. In our approach, a 3×3 kernel is employed for dilation to enhance the OBIA annotation. This dilation procedure involves 20 iterations, leading to the addition of extra pixels to the boundaries of road objects. Consequently, the gaps between pixels are filled, resulting in connected road segments. Figure 3 provides a visual comparison between the improved OBIA annotation, the original OBIA annotation, and the ground truth derived from the manually annotated orthophoto.
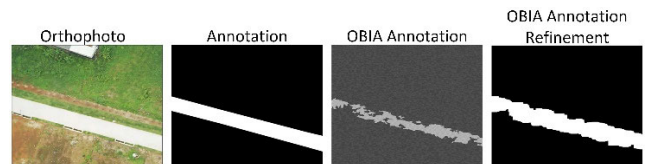

FIGURE 3. Comparison of OBIA Annotation Refinement and OBIA annotation with ground truth of an orthophoto.

D. DeepLab V3+

The road extraction process was executed using the Deeplab V3+ model, consisting of two fundamental components: the encoder and the decoder [43], [44]. The architectural configuration of the model employed is illustrated in Figure 4. We incorporate ResNet 50, ResNet 101, and MobileNetV2 as encoders of the model and compare the evaluation performance of each encoder.

The encoder in Deeplab V3+ employs the concept of atrous convolution and atrous spatial pyramid pooling (ASPP), resulting in improved image classification [45]. While traditional filters typically apply convolution continuously to the input [46], leading to significant time consumption when filtering a larger area [47], atrous convolution addresses this challenge by implementing a skip filter [44].

ASPP, also known as dilated convolution [48], was initially introduced by George [49] to acquire multiscale information [45], [50]. ASPP employs multiple atrous convolutions in parallel with various sampling rates of 6, 12, and 18 [51]. The extracted features from the multiple atrous convolutions in the encoder are combined and subsequently bilinearly upsampled by a factor of 4. The Deeplab V3+ decoder incorporates a 1×1 convolution before merging with encoder feature results. Following the merging process, the features undergo a 3×3 convolution and are bilinearly upsampled by a factor of 4 to produce the road extraction.

The residual network (ResNet) model is one of the suitable models for training deep data, mitigating the Vanishing Gradient Problem frequently encountered in deep neural networks [52]. Several studies have shown the superior accuracy of ResNet in object classification tasks in comparison to VGG16, VGG19, Random Forest, and GoogleNet [32]. ResNet enables gradient flow through skip connections and
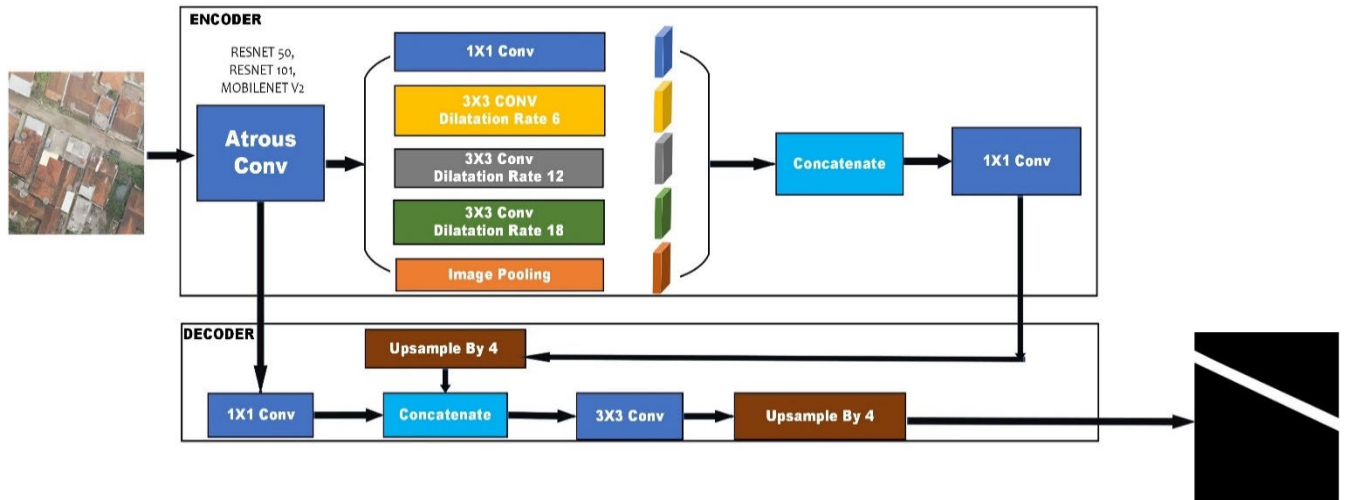
FIGURE 4. Deeplab V3+ architecture used in the research.

facilitates backpropagation in the nearest layer. The nomenclature of the ResNet model corresponds to its layer count, such as ResNet 50 with 50 layers and ResNet 101 with 101 layers. Greater layer numbers allow ResNet to extract deeper features.

MobileNetV2, an evolution of MobileNetV1, integrates inverted residuals to preserve information [53], [54]. MobileNetV2 aims to facilitate deep learning training on resource-constrained devices. Achieved through Depth-wise Separable Convolution and Linear Bottleneck, it reduces parameters and computational operations without compromising model performance.

E. MODEL'S PERFORMANCE EVALUATION
We assess the performance of the model using evaluation metrics such as Pixel Accuracy (PA), mean Intersection Over Union (mean IoU), IoU of the road class (IoUr), and Dice Score. Higher values on these metrics signify a robust predictive capability, whereas lower values suggest suboptimal performance. These evaluation metrics are instrumental in establishing thresholds for determining the quantity of training data utilizing the vanilla OBIA or the enhanced OBIA annotations, ensuring the performance of the trained model. PA, which measures the model's overall ability to predict correct pixels serves as a pivotal indicator of model performance [55], [56], [57]. Recognizing PA's limitations in accurately measuring accuracy amidst data imbalance, particularly in the context of road extraction, we conducted additional evaluations for this specific problem. PA is determined through the following equation.

$$PA = \frac{\text{Numbers of Correctly Classified Pixels}}{\text{Total Numbers of Pixels}} \quad (2)$$

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

IoU describes the conformity of the road extraction results with the ground truth, that is a heightened IoU value indicates

enhanced model performance [5], [58]. Mean IoU, derived from the average IoU across all classes during the segmentation process [59], represents the average of both road IoU and background IoU in our study. The Mean IoU calculation is expressed by the following equation:

$$\text{meanIoU} = \frac{\text{IoU}_{\text{Road}} + \text{IoU}_{\text{Background}}}{2} \quad (4)$$

Road IoU represents the ratio between the intersection of the road and the combination of road, which indicates the accuracy of road prediction outcomes relative to the actual road in the ground truth. Road IoU is computed by comparing the True Positive value to the sum of True Positive (TP), False Negative (FN), and False Positive (FP). True positive in terms of IoU road signifies that the model's predictions accurately depict the road, aligning with the ground truth. False negative in IoU road indicates that the model incorrectly categorizes the background, despite the presence of the road in the ground truth. False positive in IoU road clarifies that the model's prediction erroneously represents the road when the background is instead present in the ground truth.

The IoU background serves as a metric to evaluate the congruence between the predicted background and the background present in the ground truth. It quantifies the extent of agreement between the two backgrounds. The IoU background value is determined by comparing the True Positive value with the sum of True Positive (TP), False Negative (FN), and False Positive (FP). True positive in IoU background signifies that both the model's predicted results and the ground truth correctly indicate the presence of background. False negative in IoU background occurs when the model incorrectly identifies the road as background based on the ground truth. False positive in IoU background arises when the model correctly identifies the background, but the ground truth indicates the presence of the road. The calculation of both road IoU and Background IoU values adheres to

the provided equation.

$$IoU_{Road/Background} = \frac{TP}{TP + FP + FN} \quad (5)$$

Dice score is used to evaluate the performance of the model which indicates the conformity of the prediction results with the ground truth [3], [60], [61]. It is derived from the ratio of double the intersection size to the total area, where the total area comprises the sum of the predicted area and the ground truth area. Dice score is calculated using the equation below.

$$DiceScore = \frac{2x Intersection}{TotalArea} \quad (6)$$

III. EXPERIMENT AND RESULT

We divide this section into three subsections. Subsection A offers a comprehensive portrayal of the experimental scenario and its relation to the research objectives. Subsection B provides a visual representation of the results, featuring graphs to illustrate the evaluation metrics used to measure the model's performance. Finally, subsection C delves into an in-depth analysis and discussion regarding the outcomes presented in the previous subsections.

A. EXPERIMENT

The incorporation of OBIA annotations in the training data was set with varying percentages of 0%, 25%, 50%, 75%, and 100%. In instances where OBIA annotations did not constitute 100% of the training data, the remaining portion of the training data incorporated manually annotated data. A similar approach was taken for the enhanced OBIA annotations in the training data, resulting in a total of nine training iterations for each encoder deployed in DeepLab V3+ model. The encoders employed in the model include ResNet 50, ResNet 101, and MobileNetV2, resulting in a total of twenty-seven training sessions. The model utilizing 100% manual annotations underwent a single iteration for each DeepLab V3+ model, as the scenario utilizing 0% OBIA annotations is identical to that of 0% enhanced OBIA annotations. Table 1 succinctly outlines the research scenario.

The road prediction outcomes of each model were visually assessed and subjected to comparative analysis. Ten orthophotos depicting road conditions in rice fields, vacant land, and densely populated residential areas were utilized. These ten orthophotos feature objects of similar or distinct colors to the road. Performance evaluation metrics of all models are presented in graphs, with four graphs displaying the performance metrics of each individual model.

B. EXPERIMENT RESULT

The results of road extraction utilizing the DeepLab V3+ model using ResNet 50 as encoder, trained with various combinations of Object-Based Image Analysis (OBIA) training data, are depicted in Figure 5. The road extraction performance of the model was encouraging when trained with 100% or 75% (and 25% annotations from OBIA) manually annotated training data from annotations. Interestingly,

TABLE 1. Research scenario.

Model DL (Training Data Used)	Composition of Training Data (%)
Deeplab V3+ Encoder ResNet 50 (OBIA Annotation)	0, 25, 50, 75, 100
Deeplab V3+ Encoder ResNet 101 (OBIA Annotation)	0, 25, 50, 75, 100
Deeplab V3+ Encoder MobileNetV2 (OBIA Annotation)	0, 25, 50, 75, 100
Deeplab V3+ Encoder ResNet 50 (Refinement of OBIA Annotation)	0, 25, 50, 75, 100
Deeplab V3+ Encoder ResNet 101(Refinement of OBIA Annotation)	0, 25, 50, 75, 100
Deeplab V3+ Encoder MobileNetV2 (Refinement of OBIA Annotation)	0, 25, 50, 75, 100

when the model was trained with 25% training data from OBIA annotations, orthophoto 10, featuring road conditions akin to the color of surrounding building roofs, displayed a more accurate road extraction compared to the model trained with 100% operator annotations. The visualization of road extraction results from the model trained with 50% synthetic training data showed disjointed and blurred road extractions; models trained with 75% and 100% training data from OBIA annotations did not yield any road extractions.

Figure 6 showcases the road extraction achieved by the DeepLab V3+ encoder ResNet 50 model trained using training data derived from enhanced OBIA annotations. Models trained with 0%, 25%, and 50% compositions of the enhanced OBIA annotations exhibited satisfactory road extraction. However, similar to the former results, models trained with datasets from the improved OBIA annotations with 75% and 100% compositions yielded indistinct and fragmented road extractions. Notably, the model trained with a composition of 25% and 50% of the refined OBIA annotations demonstrated superior road extraction for orthophoto 10 compared to the model trained using entirely manual annotations.

The road extraction outcomes produced by the DeepLab V3+ model utilizing ResNet 101 as encoder, trained using the OBIA annotated dataset, are illustrated in Figure 7. The clarity of the resulting road extraction diminishes with an increasing proportion of OBIA annotated data. Models trained with 75% and 100% OBIA annotation datasets did not yield any road extractions, while those trained with 50% OBIA annotations exhibited indistinct and discontinuous road extractions. Similar to the above, the road extraction of orthophoto 10 from model trained with 25% OBIA annotations showed superior results when using the compared to using 100% human-annotated annotations.

The road extraction results generated by the DeepLab V3+ model utilizing ResNet 101 as encoder, trained using the enhanced OBIA annotations, are illustrated in Figure 8. It can be seen that the clarity and road connectivity of the extraction result plummeted when utilizing models trained with dataset comprised of 75% and 100% enhanced OBIA annotations. Nevertheless, the model trained with improved

OBIA annotation demonstrates superior road extraction on orthophoto 10 compared to the model trained with an entirely manually annotated dataset.
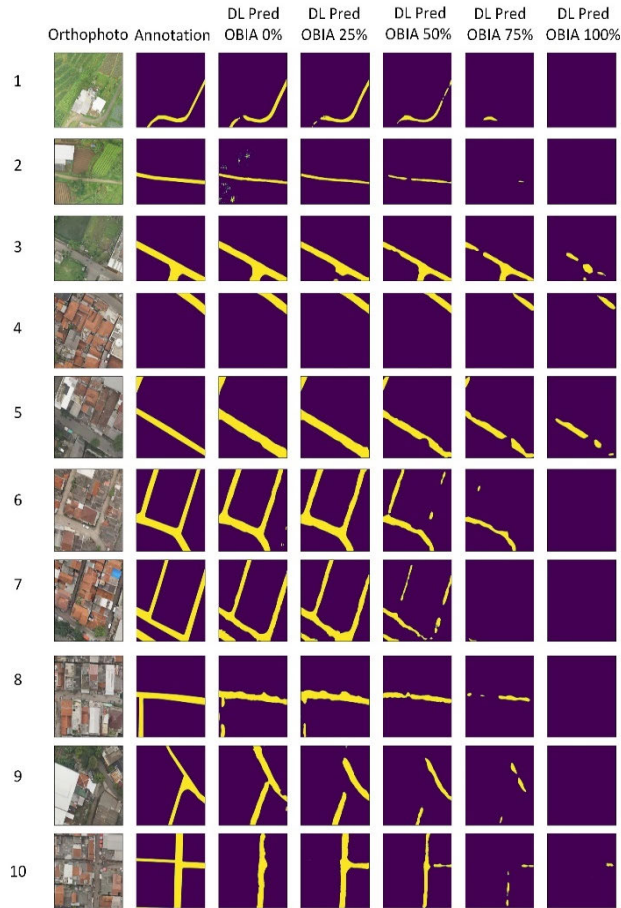

FIGURE 5. Visualisation of road extraction from DeepLab V3+ encoder ResNet 50 model using various compositions of OBIA annotations as training data.

The road extraction results produced by the DeepLab V3+ encoder MobileNetV2 model are presented in Figures 9 and 10. Figure 9 displays the road extraction results from the model trained using OBIA annotations, while Figure 10 showcases the road extraction results from the model trained using enhanced OBIA annotations. Similar to the DeepLab V3+ with ResNet model variants, the DeepLab V3+ model with MobileNetV2 encoder trained using the improved OBIA dataset exhibits road extraction for all training data compositions. This contrasts with the model trained using vanilla OBIA annotations, where road extraction is only evident in models trained using 0% and 25% training data. Also similar to the above, the road extraction results from orthophoto 10 were superior for models trained using both OBIA annotations and refined OBIA annotations when compared to those trained using human-annotated data.

The performance evaluation outcomes of each model are showcased in Figure 11. Furthermore, to offer a more profound insight into the efficacy of each model, the mean values
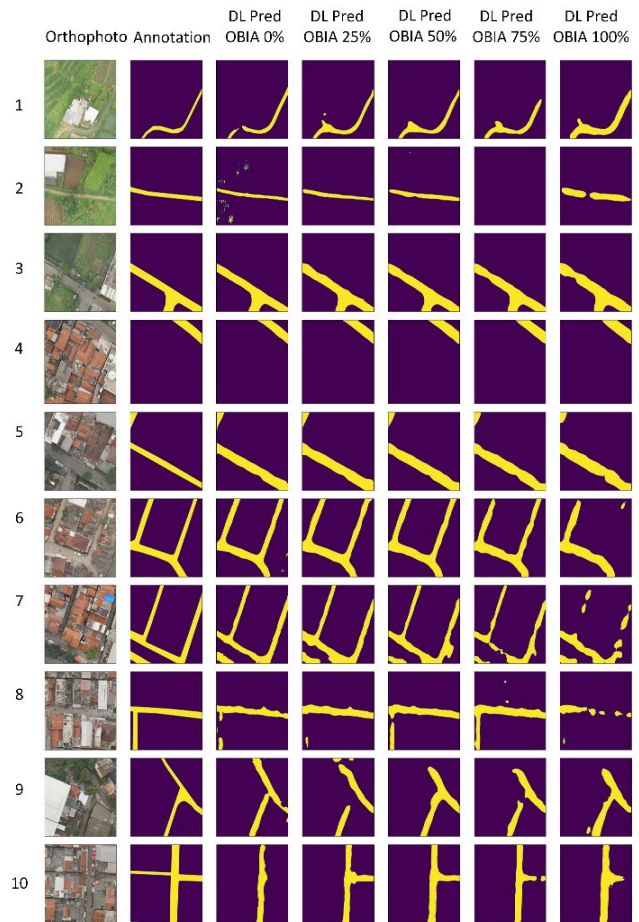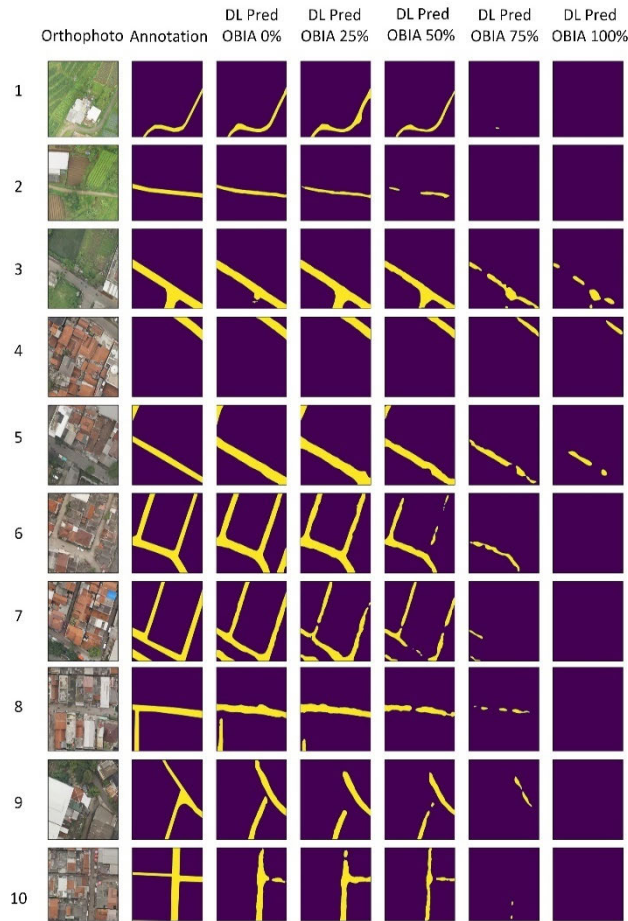

FIGURE 6. Visualisation of road extraction from DeepLab V3+ encoder ResNet 50 model using various compositions Refinement of OBIA Annotation as training data.

of the assessment criteria for each configuration, utilizing both OBIA annotations and improved OBIA annotations, have also been computed and succinctly outlined in Table 2. The Pixel Accuracy (PA) values of models trained with improved Object-Based Image Analysis (OBIA) annotations demonstrate a higher performance compared to models trained with standard OBIA annotations. Figure 10 provides a visual comparison of PA values across models trained with different compositions of training data derived from OBIA and its refined version as annotations. The model that undergoes training using fully organic annotations showcases the most significant PA, boasting an average of 0.966. Conversely, the model trained with OBIA annotations achieves its peak PA when the composition of improved OBIA annotations amounts to 25% of all training data, reaching an average value of 0.961. Notably, the model trained with training data derived from refined OBIA annotations attains the highest PA value of 0.963, when the composition of improved OBIA annotations constitutes 25% of all training data. The utilization of enhanced OBIA annotations presents a significant increase in PA compared to using OBIA only. Models trained with training data comprising 75% and 100% OBIA

annotations exhibit PA values of 0.947 and 0.942, respectively. On the other hand, models trained with training data comprising 75% and 100% improved OBIA annotations showcase PA values of 0.957 and 0.954, respectively.

for each composition of the training data. Notably, a substantial increase in IoU Road values is evident in models trained with 75% and 100% improved OBIA annotations in the training data, achieving average values of 0.450 and 0.432, respectively, as compared to original OBIA annotations.

FIGURE 7. Visualisation of road extraction from DeepLab V3+ encoder ResNet 101 model using various compositions of OBIA Annotation as training data.

FIGURE 8. Visualisation of road extraction from DeepLab V3+ encoder ResNet 101 model using various compositions Refinement of OBIA Annotation as training data.

The models trained exclusively with human-made annotations as the training data exhibit the highest IoU Road value, averaging 0.625, while the smallest IoU Road value is observed when each model is trained using 100% OBIA annotations, yielding an average value of 0.011. In the context of models trained with standard OBIA annotations, the highest Road IoU is obtained from the model trained with 25% synthetic annotation composition, achieving an average value of 0.480. Models trained with improved OBIA annotations also achieve their highest Road IoU value at 25% composition ratio of the training data, averaging at 0.530. Additionally, at 50% composition, the models trained with improved OBIA annotations produce IoU Road with an average value of 0.516, which is comparatively similar to 25% composition, even at a lesser amount of organic training annotations. Moreover, models trained with refined OBIA annotations consistently show increased Road IoU values compared to models trained with standard OBIA annotations
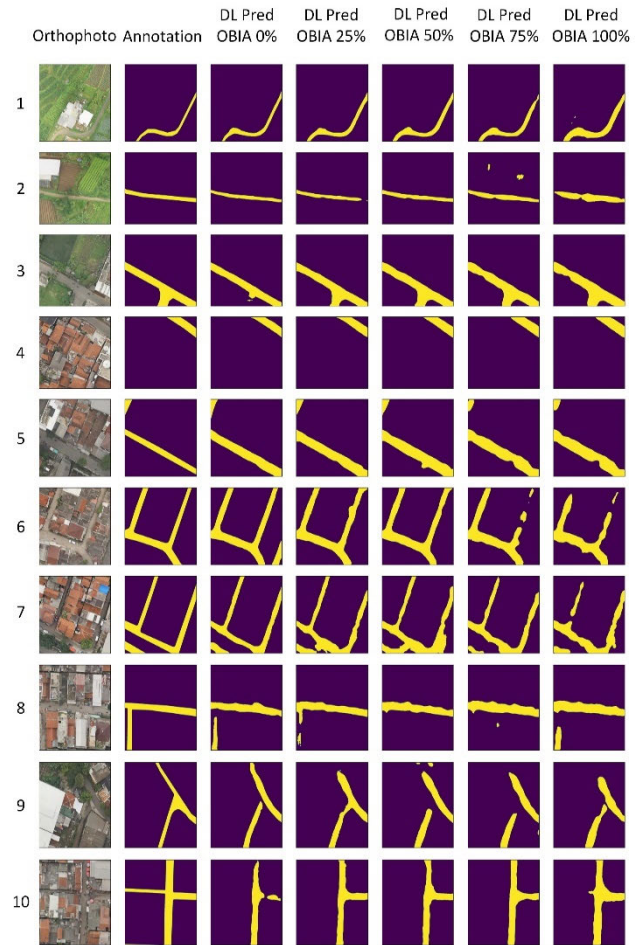
Similarly, the highest mIoU is also achieved by models trained exclusively with 100% human-annotated training dataset, resulting in an average value of 0.794. Conversely, the lowest mIoU is obtained when the model is trained using both 100% OBIA annotations and improved OBIA annotations, yielding averages of 0.467 and 0.672, respectively. Analogous to previous results, each model demonstrates an increase in mIoU value when trained with enhanced OBIA annotations as compared to the vanilla ones, with the most significant increase in mIoU value is observed when each model is trained with 100% enhanced OBIA annotations, resulting in an average mean IoU increase of 0.215 compared to models trained with 100% OBIA annotations as the training dataset.

The highest Dice Score is also achieved by models trained exclusively with 100% human-annotated training dataset,
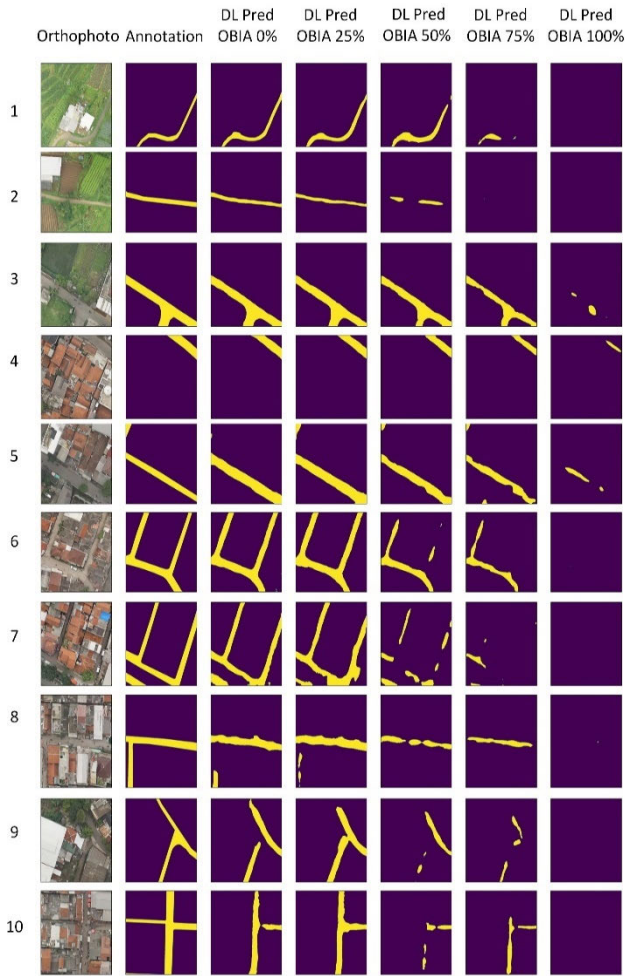
FIGURE 9. Visualisation of road extraction from DeepLab V3+ encoder MobileNetV2 model using various compositions of OBIA Annotation as training data.

FIGURE 10. Visualisation of road extraction from DeepLab V3+ encoder MobileNetV2 model using various compositions Refinement of OBIA Annotation as training data.

yielding an average value of 0.859, with the lowest Dice Score is observed when each model is trained using both 100% OBIA annotations and 100% improved OBIA annotations, resulting in average Dice Scores of 0.494 and 0.770, respectively. Increases in Dice Scores are also evident in models trained with refined OBIA annotations as compared to vanilla OBIA, with the most substantial increase occurring when the model is trained with 100% improved OBIA annotations, leading to a 0.276 increase from the average Dice Score of the model trained with 100% vanilla OBIA annotated dataset. The complete model performance evaluation results for each model can be found in Figure 11, and the average values of the model performance evaluation metrics for each composition using OBIA annotations its refined version are presented in Table 2.

C. ANALYSIS AND DISCUSSION

The clarity of road extraction visualizations diminishes as the ratio of standard OBIA annotations utilized in the training data increases. This degradation is attributed to the noise
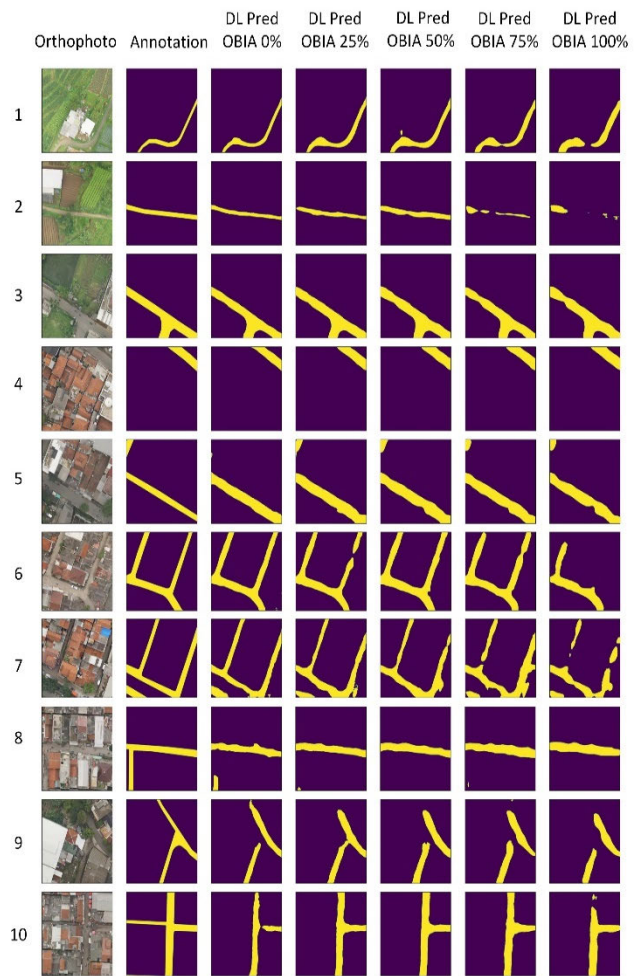
present in OBIA annotations, resulting in more interference to the model weights during training, and, subsequently, the quality of road extraction as the amount of OBIA annotations in the training data rises. In contrast, models trained with improved OBIA annotations exhibit consistent road extractions across various composition variations. This is attributed to the lesser amount of noise in the OBIA annotations, as some obvious incorrect segmentations are cleared and disjointed road nodes are connected, resulting in more robust road extractions compared to models trained with standard OBIA annotations. Furthermore, each model trained with improved OBIA annotations consistently outperforms its counterpart trained solely with OBIA annotations, indicating superior road extraction capabilities as compared to the former.

To analyze variations in model performance and establish a threshold for the utilization of OBIA annotations and its refined annotations in the training data, trendlines were constructed for each model's performance evaluation score. The trendlines for all metrics measured for the models trained

TABLE 2. Average performance metrics of models for each composition in the training data.

Composition in Training Data (%)	Obia Annotation				Refinement Obia Annotation			
	PA	IoUR	mIoU	Dice Score	PA	IoUR	mIoU	Dice Score
0	0,966	0,625	0,794	0,859	0,966	0,625	0,794	0,859
25	0,961	0,480	0,720	0,785	0,963	0,531	0,746	0,815
50	0,957	0,329	0,642	0,709	0,958	0,516	0,736	0,810
75	0,947	0,096	0,521	0,559	0,957	0,450	0,703	0,778
100	0,942	0,012	0,477	0,495	0,954	0,433	0,692	0,771
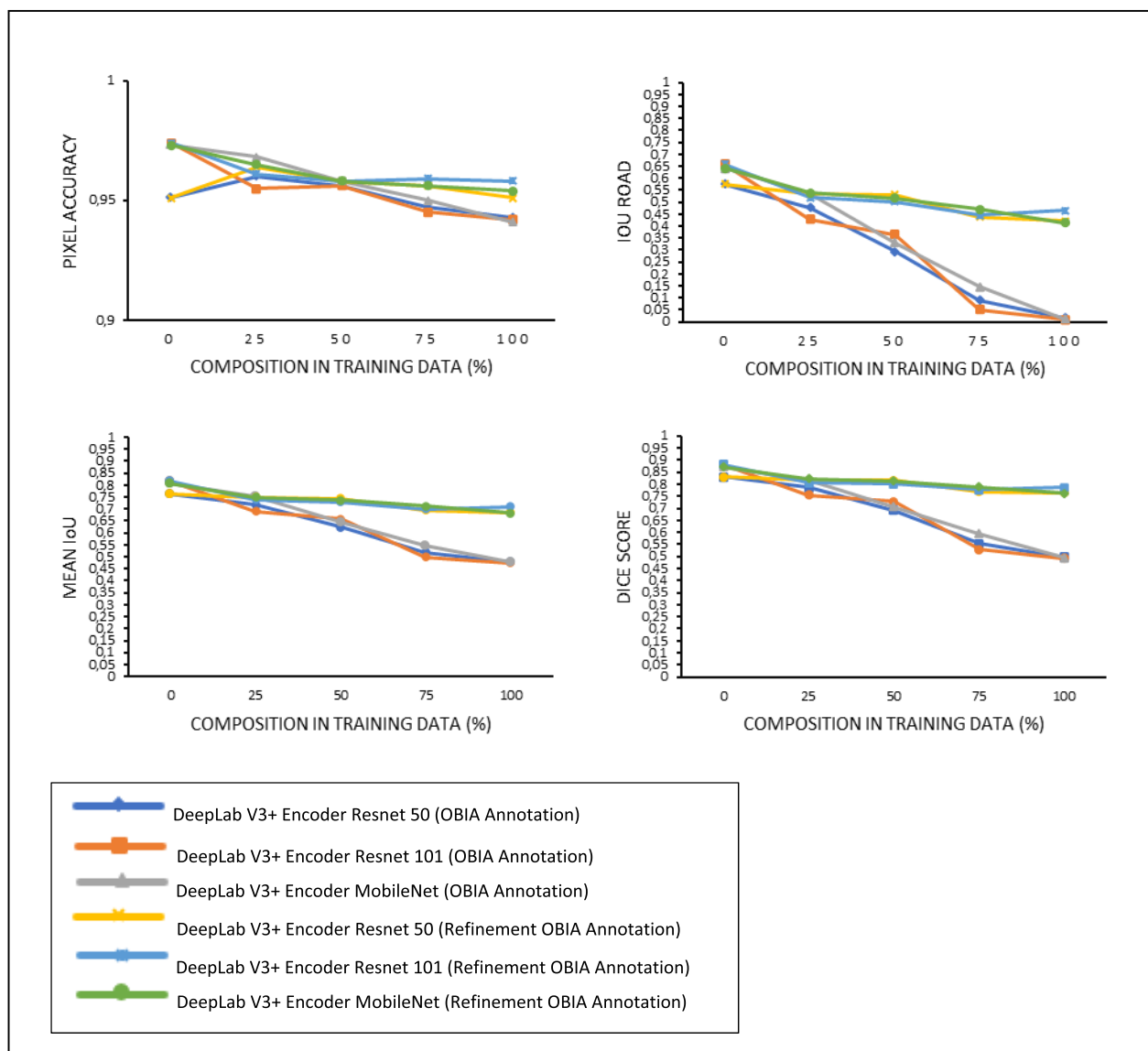
FIGURE 11. Model performance evaluation metric.

with synthetic (vanilla OBIA and refined OBIA) annotations are illustrated in Figure 12. Overall, there is a decline in model performance evaluation metrics with the increasing ratio of OBIA annotations and its enhanced counterpart in the training data. However, a significant drop in performance does not occur until the composition of OBIA annotations exceeds
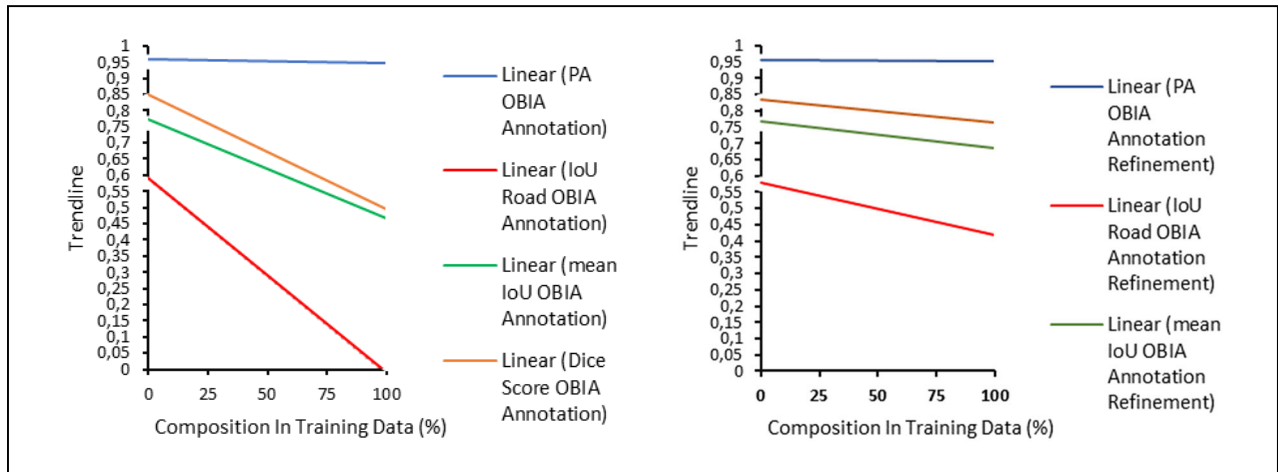
FIGURE 12. Trendline of each model's performance evaluation metrics.

25% of the training data. In a similar fashion, a noticeable decrease in the performance of the model trained with improved OBIA annotations occurs when the composition of improved OBIA annotations surpasses 50% of the training data.

Models trained with either OBIA annotations or improved OBIA annotations demonstrate suitability for generating road extractions in densely populated areas where the color of the roofs closely resembles that of the roads. This is due to OBIA's capacity to produce detailed annotations based on color variations, making it effective in scenarios such as orthophoto 10, where models trained with OBIA annotations or improved OBIA annotations consistently outperform models trained with manually annotated images comprising the dataset.

IV. CONCLUSION

The dataset provisioning time for the training process can be significantly reduced by incorporating both road annotations obtained from Object-Based Image Analysis (OBIA) and enhanced OBIA annotations. The recommended upper limit for including OBIA annotations in the training data is 25%, while for improved OBIA annotations, it is 50%. This ensures that road extraction remains effective with satisfactory visualization and high model performance evaluation metrics. Consequently, the use of OBIA annotations and improved OBIA annotations results in a 25% to 50% reduction in dataset provisioning time compared to obtaining training data solely from operator annotations. Moreover, the integration of annotations derived from OBIA and improved OBIA outcomes enhances the model's performance in generating accurate road extractions, particularly in areas with high-resolution photos that exhibit color similarities with road colors. In future investigations, alternative semi-automatic methods can be explored to provide training data for road extraction in deep learning.

The time required to prepare the road extraction dataset for training can be reduced by employing synthetic road annotations, such as those obtained from OBIA and enhanced OBIA annotations. In this study, we determined an upper limit of 25% and 50%, respectively, for the percentage of OBIA annotations and improved OBIA annotations in the training data. Empirically, we show the stated amount ensures that road extraction is still achieved with satisfactory quality and boasts a high model performance evaluation metrics score. This decrease in the amount of needed organic data, consequently, also decreases the dataset provisioning time by 25% to 50% compared to the time taken by human annotators to prepare the training data. Furthermore, the implementation of annotations derived from OBIA and improved OBIA can also enhance the model's performance in generating road extractions, particularly in areas that exhibit color similarities with road colors. We encourage future studies to investigate how other semi-automatic methods can be employed to supply synthetic training data in a timely manner with only a slight effect in model's performance for road extraction using deep learning. In addition, we also encourage future investigations to further analyze the effect of different band visible layers to extraction results.

REFERENCES

[1] Z. Yang, D. Zhou, Y. Yang, J. Zhang, and Z. Chen, "Road extraction from satellite imagery by road context and full-stage feature," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023, doi: 10.1109/LGRS.2022.3228967.

[2] X. Zhang, X. Ma, Z. Yang, X. Liu, and Z. Chen, "A context-aware road extraction method for remote sensing imagery based on transformer network," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023, doi: 10.1109/LGRS.2023.3324644.

[3] L. Qiu, D. Yu, C. Zhang, and X. Zhang, "A semantics-geometry framework for road extraction from remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023, doi: 10.1109/LGRS.2023.3268647.

[4] L. Dai, G. Zhang, and R. Zhang, "RADANet: Road augmented deformable attention network for road extraction from complex high-resolution remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5602213, doi: 10.1109/TGRS.2023.3237561.

[5] A. Abdollahi, B. Pradhan, and A. Alamri, "SC-RoadDeepNet: A new shape and connectivity-preserving road extraction deep learning-based network from remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5617815, doi: 10.1109/TGRS.2022.3143855.

[6] S. M. Nasution, E. Husni, K. Kuspriyanto, and R. Yusuf, "Personalized route recommendation using F-AHP-express," *Sustainability*, vol. 14, no. 17, p. 10831, Aug. 2022, doi: 10.3390/su141710831.

[7] C. Wang, R. Xu, S. Xu, W. Meng, R. Wang, J. Zhang, and X. Zhang, "Toward accurate and efficient road extraction by leveraging the characteristics of road shapes," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4404616, doi: 10.1109/TGRS.2023.3284478.

[8] Y. Li, L. Xiang, C. Zhang, F. Jiao, and C. Wu, "A guided deep learning approach for joint road extraction and intersection detection from RS images and taxi trajectories," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 8008–8018, 2021, doi: 10.1109/JSTARS.2021.3102320.

[9] Q. Xu, C. Long, L. Yu, and C. Zhang, "Road extraction with satellite images and partial road maps," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4501214, doi: 10.1109/TGRS.2023.3261332.

[10] M. Zhou, H. Sui, S. Chen, X. Chen, W. Wang, J. Wang, and J. Liu, "UGRoadUpd: An unchanged-guided historical road database updating framework based on bi-temporal remote sensing images," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 11, pp. 21465–21477, Nov. 2022, doi: 10.1109/TITS.2022.3181095.

[11] R. Lian, W. Wang, N. Mustafa, and L. Huang, "Road extraction methods in high-resolution remote sensing images: A comprehensive review," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5489–5507, 2020, doi: 10.1109/JSTARS.2020.3023549.

[12] J. Dai, R. Ma, and H. Ai, "Semi-automatic extraction of rural roads from high-resolution remote sensing images based on a multifeature combination," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, doi: 10.1109/LGRS.2020.3026674.

[13] J. Dai, T. Zhu, Y. Wang, R. Ma, and X. Fang, "Road extraction from high-resolution satellite images based on multiple descriptors," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 227–240, 2020, doi: 10.1109/JSTARS.2019.2955277.

[14] Y. Wu, Q. Zhao, Z. Shen, and Y. Li, "Clustering point process based network topology structure constrained urban road extraction from remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 2087–2098, 2022, doi: 10.1109/JSTARS.2022.3151757.

[15] J. Zhang, X. Hu, Y. Wei, and L. Zhang, "Road topology extraction from satellite imagery by joint learning of nodes and their connectivity," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5602613, doi: 10.1109/TGRS.2023.3241679.

[16] A. Abdollahi, B. Pradhan, and A. Alamri, "VNet: An end-to-end fully convolutional neural network for road extraction from high-resolution remote sensing data," *IEEE Access*, vol. 8, pp. 179424–179436, 2020, doi: 10.1109/ACCESS.2020.3026658.

[17] J. Wan, Z. Xie, Y. Xu, S. Chen, and Q. Qiu, "DA-RoadNet: A dual-attention network for road extraction from high resolution satellite imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6302–6315, 2021, doi: 10.1109/JSTARS.2021.3083055.

[18] L. Ding and L. Bruzzone, "DiResNet: Direction-aware residual network for road extraction in VHR remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10243–10254, Dec. 2021, doi: 10.1109/TGRS.2020.3034011.

[19] J. Ai, S. Hou, M. Wu, B. Chen, and H. Yan, "MPGSE-D-LinkNet: Multiple-parameters-guided squeeze-and-excitation integrated D-LinkNet for road extraction in remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023, doi: 10.1109/LGRS.2023.3306725.

[20] E. Husni, A. Siburian, R. Yusuf, A. B. Harto, and D. Suwardhi, "Automatic road extraction from very high resolution orthophoto using DeepLab V3+," in *Proc. Int. Conf. Inf. Technol. Syst. Innov. (ICITSI)*, Nov. 2022, pp. 264–268, doi: 10.1109/ICITSI56531.2022.9970810.

[21] A. Buslaev, S. Seferbekov, V. Iglovikov, and A. Shvets, "Fully convolutional network for automatic road extraction from satellite imagery," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 1–12, doi: 10.1109/CVPRW.2018.00035.

[22] J. Li, Y. Meng, D. Dorjee, X. Wei, Z. Zhang, and W. Zhang, "Automatic road extraction from remote sensing imagery using ensemble learning and postprocessing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 10535–10547, 2021, doi: 10.1109/JSTARS.2021.3094673.

[23] C. Avci, E. Sertel, and M. E. Kabadayi, "Deep learning-based road extraction from historical maps," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, doi: 10.1109/LGRS.2022.3204817.

[24] A. Abdollahi, B. Pradhan, G. Sharma, K. N. A. Maulud, and A. Alamri, "Improving road semantic segmentation using generative adversarial network," *IEEE Access*, vol. 9, pp. 64381–64392, 2021, doi: 10.1109/ACCESS.2021.3075951.

[25] Q. Guo and Z. Wang, "A self-supervised learning framework for road centerline extraction from high-resolution remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4451–4461, 2020, doi: 10.1109/JSTARS.2020.3014242.

[26] R. Liu, Q. Miao, Y. Zhang, M. Gong, and P. Xu, "A semi-supervised high-level feature selection framework for road centerline extraction," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 5, pp. 894–898, May 2020, doi: 10.1109/LGRS.2019.2931928.

[27] Y. Wei and S. Ji, "Scribble-based weakly supervised deep learning for road surface extraction from remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5602312, doi: 10.1109/TGRS.2021.3061213.

[28] A. Yu, Y. Quan, R. Yu, W. Guo, X. Wang, D. Hong, H. Zhang, J. Chen, Q. Hu, and P. He, "Deep learning methods for semantic segmentation in remote sensing with small data: A survey," *Remote Sens.*, vol. 15, no. 20, p. 4987, Oct. 2023, doi: 10.3390/rs15204987.

[29] Z. Ye, K. Yang, Y. Lin, S. Guo, Y. Sun, X. Chen, R. Lai, and H. Zhang, "A comparison between pixel-based deep learning and object-based image analysis (OBIA) for individual detection of cabbage plants based on UAV visible-light images," *Comput. Electron. Agricult.*, vol. 209, Jun. 2023, Art. no. 107822, doi: 10.1016/j.compag.2023.107822.

[30] M. Li, X. Liu, X. Wang, and P. Xiao, "Detecting building changes using multimodal Siamese multitask networks from very-high-resolution satellite images," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5404322, doi: 10.1109/TGRS.2023.3290817.

[31] Z. Zhang, T. Jiang, C. Liu, and L. Zhang, "An effective classification method for hyperspectral image with very high resolution based on encoder–decoder architecture," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1509–1519, 2021, doi: 10.1109/JSTARS.2020.3046245.

[32] R. Naushad, T. Kaur, and E. Ghaderpour, "Deep transfer learning for land use and land cover classification: A comparative study," *Sensors*, vol. 21, no. 23, p. 8083, Dec. 2021, doi: 10.3390/s21238083.

[33] H. Mao, X. Chen, Y. Luo, J. Deng, Z. Tian, J. Yu, Y. Xiao, and J. Fan, "Advances and prospects on estimating solar photovoltaic installation capacity and potential based on satellite and aerial images," *Renew. Sustain. Energy Rev.*, vol. 179, Jun. 2023, Art. no. 113276, doi: 10.1016/j.rser.2023.113276.

[34] L. Ma, Z. Yang, L. Zhou, H. Lu, and G. Yin, "Local climate zones mapping using object-based image analysis and validation of its effectiveness through urban surface temperature analysis in China," *Building Environ.*, vol. 206, Dec. 2021, Art. no. 108348, doi: 10.1016/j.buildenv.2021.108348.

[35] M. K. Garajeh, B. Feizizadeh, Q. Weng, M. H. R. Moghaddam, and A. K. Garajeh, "Desert landform detection and mapping using a semi-automated object based image analysis approach," *J. Arid Environ.*, vol. 199, Apr. 2022, Art. no. 104721, doi: 10.1016/j.jaridenv.2022.104721.

[36] A. Chouhan, D. Chutia, and P. Raju, "Comparison of tree canopy extraction using object based image analysis and deep learning technique in UAV images," in *Proc. Int. Conf. Intell. Comput. Remote Sens. (ICICRS)*, Jul. 2019, pp. 1–5, doi: 10.1109/ICICRS46726.2019.9555767.

[37] C. Halliday, N. Stuart, and I. Cameron, "Mapping savanna wildfires in southern belize using Sentinel-1 SAR and object based image analysis," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2021, pp. 8420–8423, doi: 10.1109/IGARSS47720.2021.9553534.

[38] C. Yahui, G. Xiaoyue, L. Xiaoli, and D. Yihao, "Application of object based image analysis in earthquake emergency products," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2021, pp. 8594–8596, doi: 10.1109/IGARSS47720.2021.9553886.

[39] H. Hu, C. Feng, X. Cui, K. Zhang, X. Bu, and F. Yang, "A sample enhancement method based on simple linear iterative clustering superpixel segmentation applied to multibeam seabed classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4202115, doi: 10.1109/TGRS.2023.3247827.

[40] J. Yin, T. Wang, Y. Du, X. Liu, L. Zhou, and J. Yang, "SLIC superpixel segmentation for polarimetric SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5201317, doi: 10.1109/TGRS.2020.3047126.

[41] *2017 International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET)*, Inst. Electr. Electron. Engineers, Jakarta, Indonesia, 2017.

[42] A. Soni and A. Rai, "Automatic colon malignancy recognition using Sobel & morphological dilation," in *Proc. 3rd Int. Conf. Multimedia Process., Commun. Inf. Technol. (MPCIT)*, Dec. 2020, pp. 63–68, doi: 10.1109/MPCIT51588.2020.9350423.

[43] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam. *Encoder–Decoder With Atrous Separable Convolution for Semantic Image Segmentation*. Accessed: Jan. 5, 2024. [Online]. Available: https://github.com/tensorflow/models/tree/master/

[44] S. Das, A. A. Fime, N. Siddique, and M. M. A. Hashem, "Estimation of road boundary for intelligent vehicles based on DeepLabV3+ architecture," *IEEE Access*, vol. 9, pp. 121060–121075, 2021, doi: 10.1109/ACCESS.2021.3107353.

[45] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018, doi: 10.1109/TPAMI.2017.2699184.

[46] Z. Niu, W. Liu, J. Zhao, and G. Jiang, "DeepLab-based spatial feature extraction for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 2, pp. 251–255, Feb. 2019, doi: 10.1109/LGRS.2018.2871507.

[47] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, arXiv:1706.05587.

[48] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, arXiv:1511.07122.

[49] *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE Comput. Soc., Boston, MA, USA, Jun. 2015, pp. 7–12.

[50] K. Zheng, H. Wang, F. Qin, C. Miao, and Z. Han, "An improved land use classification method based on DeepLab V3+ under GauGAN data enhancement," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 5526–5537, 2023, doi: 10.1109/JSTARS.2023.3278862.

[51] X. Sun, Y. Xie, L. Jiang, Y. Cao, and B. Liu, "DMA-Net: DeepLab with multi-scale attention for pavement crack segmentation," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 18392–18403, Oct. 2022, doi: 10.1109/TITS.2022.3158670.

[52] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015, arXiv:1512.03385.

[53] M. Hassam, M. A. Khan, A. Armghan, S. A. Althubiti, M. Alhaisoni, A. Alqahtani, S. Kadry, and Y. Kim, "A single stream modified MobileNet v2 and whale controlled entropy based optimization framework for citrus fruit diseases recognition," *IEEE Access*, vol. 10, pp. 91828–91839, 2022, doi: 10.1109/ACCESS.2022.3201338.

[54] K. Dong, C. Zhou, Y. Ruan, and Y. Li, "MobileNetV2 model for image classification," in *Proc. 2nd Int. Conf. Inf. Technol. Comput. Appl. (ITCA)*, Dec. 2020, pp. 476–480, doi: 10.1109/ITCA52113.2020.00106.

[55] Y. Lin, F. Jin, D. Wang, S. Wang, and X. Liu, "Dual-task network for road extraction from high-resolution remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 66–78, 2023, doi: 10.1109/JSTARS.2023.3289217.

[56] W. Chen, G. Zhou, Z. Liu, X. Li, X. Zheng, and L. Wang, "NIGAN: A framework for mountain road extraction integrating remote sensing road-scene neighborhood probability enhancements and improved conditional generative adversarial network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5626115, doi: 10.1109/TGRS.2022.3188908.

[57] R. Wang, M. Cai, and Z. Xia, "A lightweight high-resolution RS image road extraction method combining multi-scale and attention mechanism," *IEEE Access*, vol. 11, pp. 108956–108966, 2023, doi: 10.1109/ACCESS.2023.3313390.

[58] A. Filho, M. Shimabukuro, and A. D. Poz, "Deep learning multimodal fusion for road network extraction: Context and contour improvement," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023, doi: 10.1109/LGRS.2023.3291656.

[59] Z. Luo, K. Zhou, Y. Tan, X. Wang, R. Zhu, and L. Zhang, "AD-RoadNet: An auxiliary-decoding road extraction network improving connectivity while preserving multiscale road details," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 8049–8062, 2023, doi: 10.1109/JSTARS.2023.3289583.

[60] X. Mi, B. Yang, Z. Dong, C. Chen, and J. Gu, "Automated 3D road boundary extraction and vectorization using MLS point clouds," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 6, pp. 5287–5297, Jun. 2022, doi: 10.1109/TITS.2021.3052882.

[61] C. Ye, H. Zhao, L. Ma, H. Jiang, H. Li, R. Wang, M. A. Chapman, J. M. Junior, and J. Li, "Robust lane extraction from MLS point clouds towards HD maps especially in curve road," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 2, pp. 1505–1518, Feb. 2022, doi: 10.1109/TITS.2020.3028033.

SUSSI is currently pursuing the Ph.D. degree with the School of Electrical Engineering and Informatics, Institut Teknologi Bandung, Indonesia. During the Ph.D. degree, her specific areas of research interests include deep learning, remote sensing, and geomatics. The focal point of the Ph.D. studies revolves around road extraction. In addition to her studies, she also holds a position as one of the lecturers at Telkom University Bandung. Her research focuses on the university, encompassing the fields of the Internet of Things, multimedia and games, deep learning, and cloud gaming.

EMIR HUSNI is currently a Professor with the School of Electrical Engineering and Informatics, Institut Teknologi Bandung. He actively participates in a wide array of academic activities, including collaborative efforts with numerous industries within his field of expertise. His research interests include digital twinning, metaverse, artificial intelligence, networks, satellite technology, wireless communication, and intelligent transportation systems.

RAHADIAN YUSUF (Member, IEEE) is currently a Lecturer with the School of Electrical Engineering and Informatics and the Institut Teknologi Bandung. He possesses deep knowledge and expertise in the realm of deep learning, which he acquired during the Ph.D. degree in Japan. His academic journey has been significantly influenced and guided by various mentors. During his career, he has collaborated with several industries on numerous research projects. His research interests include several areas, most notably affective computing, human–computer interaction, and artificial intelligence.

AGUNG BUDI HARTO is currently a Faculty Member with the Department of Geodetic and Geomatic Engineering, Faculty of Earth Science and Technology, Institut Teknologi Bandung, specializing in the research areas of remote sensing and GIS. As a highly regarded lecturer, he leads a productive laboratory with a significant number of students. He actively engages in collaborative efforts with various industries, leveraging his expertise in object-based image analysis. In addition, he is a vital supervisor for Sussi, a Ph.D. student across different faculties. To engage in fruitful discussions about research, interested individuals can reach out to him via e-mail.

ARTHUR SIBURIAN received the master's degree from the School of Electrical Engineering and Informatics, Institut Teknologi Bandung. His expertise lies in the domains of artificial intelligence, machine learning, deep learning, and remote sensing.

• • •

DENI SUWARDHI is currently a Lecturer with the Department of Geodetic and Geomatics Engineering, Faculty of Earth Science and Technology, Institut Teknologi Bandung, specializes in three-dimensional (3D) surveying and modeling and information systems. He actively participates in various industrial activities and has established research partnerships with organizations across Indonesia and abroad.