Sky Pixel Detection in Outdoor Urban Scenes: U-Net with Transfer Learning

Athar Ibrahim Alboqomi, Rehan Ullah Khan
Department of Information Technology, College of Computer, Qassim University, Buraydah, Saudi Arabia

***Abstract*—The sky depicts a high visual importance in outdoor scenes, often appearing in video sequences and photos. Sky information is crucial for accurate sky detection in several computer vision applications, such as scene understanding, navigation, surveillance, and weather forecasting. The difficulty of detecting is clarified by variations in the sky's size, weather and lighting conditions, and the sky's reflection on other objects. This article presents a new contribution to address the challenges facing sky detection. A unique dataset was built that includes scenes of distinct lighting and atmospheric phenomena. Additionally, a modified U-Net architecture was proposed with pre-trained models as encoder VGG19, EfficientNetB4, InceptionV3, and DenseNet121 for sky detection to solve outdoor image limitations and evaluate the influence of different encoders when integrated with the U-Net, aiming to identify which encoder describes features of the sky accurately. The proposed approach shows encouraging results; as it presents improved performance over the adjusted U-Net architecture with inceptionv3 on the proposed dataset, achieving mean Intersection over union, dice similarity coefficient, recall, precision, and accuracy of 98.57 %, 99.57 %, 99.41 %, 99.73%, and 99.40 %, respectively. At the same time, the best loss was achieved in U-Net with VGG19 equivalent of 0.09.**

Keywords*—Computer vision; transfer learning; semantic segmentation; sky detection; U-Net; machine learning

I. Introduction

Sky has received remarkable interest over the past few years as a robust indicator of outdoor scenes. The scene’s environmental information the sky provides is more significant than other scenes’ components. Therefore, sky detection is considered a crucial preprocessing step in various vision applications, such as weather classification [1], image or video editing [2], and navigation [3]. Moreover, the sky mask can be used to evolve the accuracy of object detection and tracking algorithms. Given its significance, sky detection research became one of the active topics in the computer vision field. This segmentation task is dedicated to identifying and isolating the sky region within a scene from other objects. However, the complexity of sky regions poses a significant challenge, owing to the vast range of pixel intensities and the notable variations in sky tone across different weather conditions and times of day. Semantic segmentation techniques represent the ideal solution for this task.

Semantic segmentation is one of the leading computer vision tasks where the object boundaries are delineated precisely. Segmentation tasks usually need complex, advanced techniques and high-quality data.

In addressing these challenges, semantic segmentation techniques have emerged as the preferred solution for accurate sky detection. By employing advanced algorithms capable of understanding the semantic meaning of image pixels.

The research community proposed two different approaches to solve the problem of sky segmentation. The first approach was the traditional approach where researchers tend to use certain methods like color-based, edge-based and region-growing.

Another approach is the usage of Machine learning. In this approach, some researchers tended to use traditional machine learning models e.g., Support vector Machine (SVM), K-means, and Logistic regression (LR). Others used more advanced techniques such as Deep learning (DL), e.g., CNN. More details about these methods will be discussed in the literature review section.

The article is arranged as follows: Section II is literature review Section III details the proposed method and dataset followed by preprocessing; Section IV explain network architecture. Section V discusses the experimental results, Section VI is discussion and finally, the paper concludes in Section VII.

II. Literature Review

Two main approaches for semantic segmentation techniques are used in sky detection tasks: the traditional-based approach and the deep learning-based approach. Firstly, traditional methods, such as edge, colour, and region-based techniques, have been introduced [4]. These traditional-based methods mainly rely on manually engineered features, such as color, texture, edges, or shape information, to identify the objects in images. The traditional methods are simple, fast, and computationally effective; however, their dependency on low-level hand-crafted features leads to low segmentation performance. On the other hand, deep learning-based methods such as U-Net [5], FCN[6], or Mask R-CNN [7] are considered end-to-end techniques [8]. These methods utilize the convolutional neural network (CNN) to extract features automatically. Although deep learning techniques need powerful hardware and extensive data, they are more robust to noise originating in sky regions from weather variations.

In the past few years, extensive research has been focused on sky and ground segmentation. Yehu Shen and Qicong Wang [9] proposed a technique based on gradient information to detect the horizon line. This method defined the border point in each column all over the image and then defined the region above these border points as the sky region. The previous method didn't detect the sky regions occluded by foreground objects. Zhao Zhijie et al. handled this challenge [10] by using color and gradient features to detect multiple border points in each column. The horizon-based approaches lost their detection efficacy as the complexity of scenes increased. Subsequently, classification-based approaches were introduced where the classifiers depend on the handcrafted features to detect the sky. Xiuzhuang Zhou et al. [11] proposed a novel technique that combines the advantages of superpixels and context inference. This method used features like lines, texture, color, position, and shape to train a Support Vector Machine (SVM) as a local superpixel classifier. Then, the conditional random field (CRF) was implemented as a contextual inference model to refine the segmentation. Additionally, Fl´avia de Mattos et al. [12] utilized eleven whiteness indexes as extracted features to feed (SVM) classifier. Yingchao Song et al. [13] proposed a novel model with two imbalanced SVM classifiers trained on several haze-relevant features. This model was trained on a hazy sky dataset with 500 annotated hazy images and divided the image into three areas: high confidence of being the sky regions with high confidence of not being the sky, and uncertain regions. In addition to the supervised traditional approach, cluster-based methods can be used in sky segmentation. Chao Fang et al. [14] deploy the K-mean clustering method to segment the sky regions based on pixels' brightness. Additionally, Yin et al. presented an innovative method called Sky-GVINS for achieving precise positioning in densely built environments and open sky areas with GNSS measurements [15]. This method relied on a lightweight sky segmentation, utilizing a global threshold technique to distinguish sky and non-sky regions in fish-eye sky-pointing imagery. The experimental dataset comprised 500 images representing diverse conditions, such as occlusions in the sky presented by buildings and trees.

Traditional machine learning techniques based on hand-crafted features adapted poorly to the variational complex sky appearance. Therefore, computer vision scientists have directed their attention to end-to-end deep learning techniques that extract features automatically using CNN to handle sky segmentation tasks. Yi-Hsuan Tsa [16] proposed a sky segmentation model based on FCN. This model was trained on 15,000 images from the LMSun dataset and achieved 94 %-pixel accuracy. Radu P. Mihail et al. [17] created a new dataset called Sky Finder and evaluated three approaches for sky segmentation in natural outdoor scenes. The results argued poor performance due to local lighting and weather conditions. Then, a new deep ensemble method that combined the output of existing methods with raw image data using rCNN was proposed and shown to improve performance with an MCR of 12.96%. Zou et al.'s study presented a novel approach to sky segmentation, combining computer vision and deep learning [18]. They proposed a new computer vision-based "flow propagation" method for robust background motion and feature estimation. These features were fed into a customized deep CNN model ResNet-50 based for training. The networks can be effectively trained on videos without using external data annotations. The proposed method was tested on BDD100k datasets. This innovative blend of handcrafted and CNN features demonstrates a unique strategy in sky segmentation research. The method is designed to operate on trained data; therefore, it does not work on different datasets. Wang et al. introduced a real-time sky segmentation method formulated for mobile augmented reality based on a deep semantic network called FSNet [19]. The authors designed the method for efficient segmentation under varied weather conditions, validated through extensive testing on a substantially large dataset. For refining the segmented regions, sky-aware constraints were included, which considered factors such as color, the sky's position, and temporal coherence across neighbouring frames. Extensive qualitative and quantitative analysis testing demonstrated that the proposed method surpasses other leading methods in real-time performance. The result's accuracy was gauged using the mean intersection over union (mIOU) metric, achieving 90.17%. However, the method showed limitations and did not perform efficiently for heterogeneous skies, such as during sunsets. Recently, U-Net has been one of the most commonly developed deep learning algorithms, especially for biomedical segmentation tasks. Due to its efficiency, U-Net architecture was widely implemented in all segmentation applications, including sky segmentation. Liba and colleagues introduced a precise method for sky optimization aimed at enhancing the sky's appearance in images, including sky segmentation [20]. They constructed a dataset of sky masks utilizing partially annotated images that were painted and refined using a modified weighted guided filter. Moreover, they trained a U-net neural network to conduct sky segmentation on RGB images by predicting the sky probability for each pixel. The Morph-Net method was employed to optimize performance and minimize network size. In their work, Kuang et al. proposed an innovative framework for segmenting sky and ground in the visual navigation of planetary rovers [21]. The study introduced a U-shaped neural network entitled NI-U-Net and incorporated a conservative annotation method to minimize human interference. Augmented results were exhibited through a pre-training process across complex scenarios using the Skyfinder dataset, a well-acknowledged benchmark. The framework was evaluated based on seven metrics, achieving high results.

Although deep learning-based segmentation models such as U-Net have achieved high performance, the requirement of huge high-quality labelled data and large costly computation power for model training limit their implementation in practical systems. Training CNN-based models from scratch is an impractical time-consuming technique as it takes a long time for the model to converge. The transfer learning approach was introduced as an ideal solution to overcome these challenges where the model uses prior information in a new task. The pre-trained weights learned from tasks that are not completely relevant to new tasks are more useful than randomly initialized weights.

In this work, the main objective is to remarkably enhance the sky segmentation task in adverse weather and lighting conditions by a modified U-Net architecture with pre-trained models as encoder VGG19, EfficientNetB4, InceptionV3, and DenseNet121 for sky detection to solve outdoor image limitations and evaluate the influence of different encoders when integrated with the U-Net, aiming to identify which encoder describes features of the sky accurately. The reason behind choosing the U-Net architecture is that it outperformed other architectures used in most of the research works. The integration between UNet architecture and transfer learning allows us to handle sky segmentation tasks effectively with high segmenting performance while saving computation power.

III. Proposed Methodology

The goal in this proposed approach is to simplify image segmentation and develop efficient, robust algorithms for sky segmentation. The methodology showed in Fig. 1. forms the basis of our approach, ensuring that the outcomes are trustworthy and valid. The widely used U-Net architecture was modified by adapting different backbones as an encoder path, which improved the depth of the network and produced better results. The model was tested using collected dataset and found that our approach outperformed existing methods in terms of capabilities. The overall approach is given in Fig. 1.

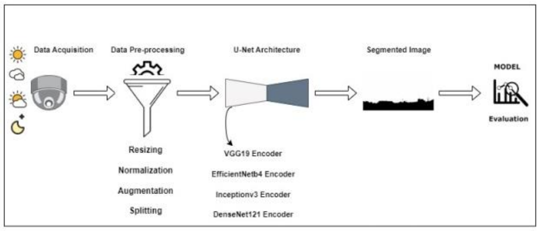

Fig. 1. Proposed methodology.

A. DataSet Acquisition

Data was systematically collected to capture different aspects of the sky. This data was collected at various times and in diverse weather conditions. To ensure comprehensive collection, stationary outdoor cameras were strategically placed in 11 specific urban locations in the Kingdom of Saudi Arabia. These regions were chosen based on their varied urban landscapes, allowing for expansive sky views to be captured.

The dataset incorporates different periods of the day (morning, midday, and evening) as well as various weather conditions (sunny, cloudy, partly cloudy). This comprehensive dataset allows for a wider range of image variations. In this research work, special care was taken to ensure that the photos collected were high quality and free from any unwanted elements or issues such as artifacts, noise, repeated images, or spots on the lens. The resulting dataset consists of RGB images that brightly represent the dynamic nature of the sky in these areas. In total, the dataset consists of 1691 diverse images captured. It's important to note that all images in the dataset contain both sky and non-sky areas. Sample images from each location are presented in Fig. 2.

B. Ground Truth

To enhance the accuracy of the dataset even further, a specialized computer vision annotation tool called CVAT was utilized. Manual annotations were made for each image through this tool by creating binary mask segmentations. These masks specifically separate regions into two categories: sky and non-sky. The definition of "sky" includes sky and other elements commonly found in skies, like clouds, sun, or moon. Conversely, "non-sky" consists of all other areas that do not fall under this sky category. These masks are ground truth.

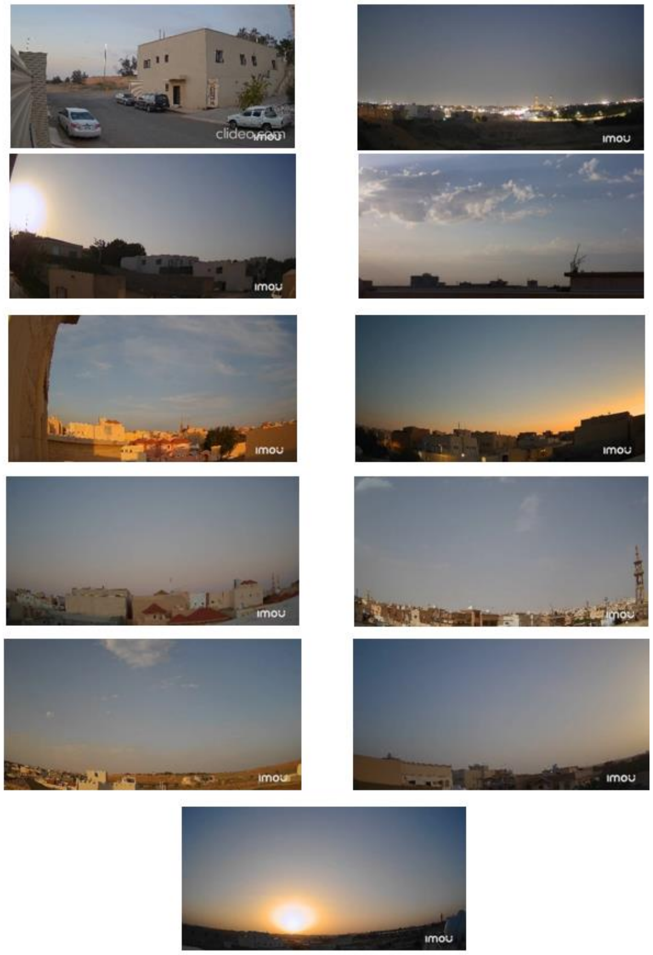

Fig. 2. Sample images from different locations.

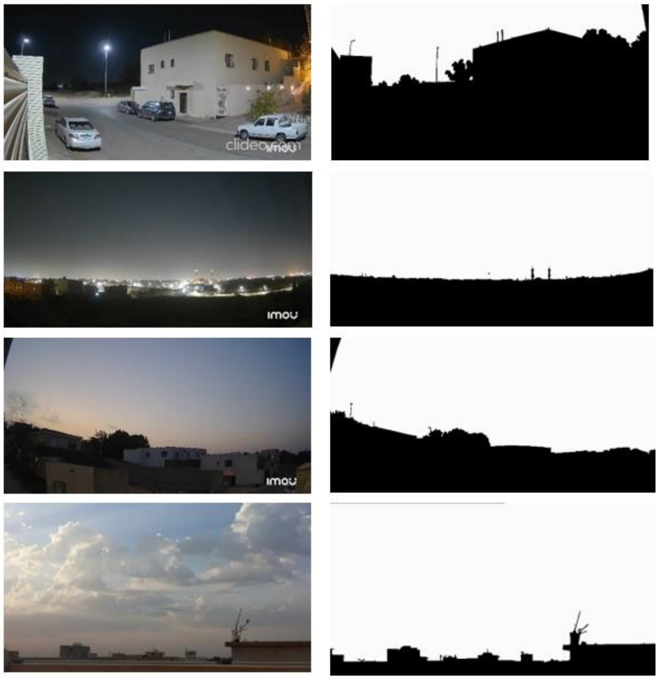

Fig. 3. Some of the data samples with ground truth.

Ground truth masks are an essential element for any machine learning application. They provide an essential reference or benchmark for algorithm training in image processing tasks, hence the term 'ground truth' [22], as they provide a standard against which the outcomes of the algorithms can be measured. Ground truth masks were generated for each image. These are binary masks, where 0 represents the sky region, and 1 represents the non-sky regions. Fig. 3 provides samples from dataset and their corresponding ground truth masks.

C. Data Preprocessing

Data pre-processing was carried out to increase computational performance and have efficient processing. First, the images were resized to 256×256 pixels. Additionally, the data normalization was also carried out by normalizing each pixel value of the data. By dividing each pixel value by 255, all data values fall within a range from 0 to 1. This normalization process is beneficial as it improves both the speed and accuracy of convergence during further calculations. Furthermore, masks (which indicate specific areas) were converted into binary format for more accessible analysis and understanding. To ensure accurate and reliable results, the images in the dataset were carefully divided into two separate datasets: the training dataset and the validation dataset. The training dataset accounts for 80% of collected data, while the remaining 20% is allocated to the validation dataset, allowing the algorithm to familiarize itself with various patterns and features within the images.

To address the issue of insufficient data and overcome hardware constraints, a solution was implemented using the 'ImageDataGenerator' class from the Keras framework, augmented images were generated on the fly during each epoch of training. This ensured that the model received diverse new variations in each iteration, effectively enhancing its ability to learn and generalize patterns. This strategy helps mitigate overfitting issues that often arise when working with small datasets. The augmentation techniques for this research include random rotation, horizontal and vertical shifts, shear transformation, and zoom. Fig. 4 presents a selection of samples that demonstrate the effects of these augmentation techniques on images in dataset.

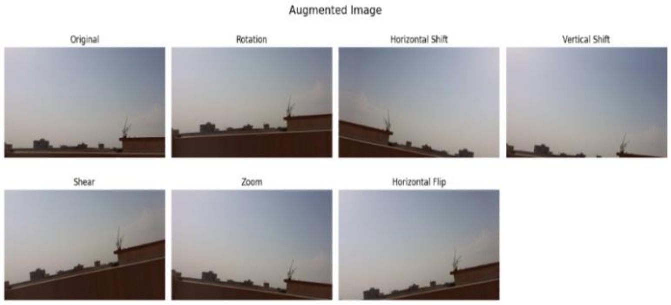

Fig. 4. Samples for augmentation techniques.

IV. Network Architecture

The illustration in Fig. 5 demonstrates the working of the proposed architecture for an RGB input image with dimensions 256×256×3. The segmented output map with dimensions 256×256×1 using the U-Net [5] network is received at the output. It is observed that there is no reduction in size between the input and output.

Four different deep learning-based networks are proposed as alternatives to the contracting path for the U-Net. These encoders are VGG19 [23], EfficientNetb4 [24], InceptionV3 [25], and DenseNet121 [26] to extract deep features, both height and width progressively decreasing while channel numbers increase. This channel augmentation enables capturing higher-level features as information flows through this pathway. The model undergoes a final convolution operation at the bottleneck, resulting in a feature map of size 16×16×1024. The expansive path then reconstructs an image of the exact dimensions as the original input from this feature map. Up-sampling layers are employed to increase spatial resolution while reducing channel count. The decoder layers utilize skip connections from the contracting path to locate and enhance features in the image. Ultimately, each pixel in the output image represents a label corresponding to the class in the input image. In this case, the output is a segmentation map, distinguishing between foreground and background regions for each pixel. The foreground represents the sky region, and the background represents the non-sky regions.

The crucial hyperparameters necessary for the convergence of the proposed models were identified. This includes batch size was set to 32 for all models, 100 epochs, AdamW optimizer selection, and learning rates set to 0.00001. The loss function applies the binary cross-entropy.

The framework for statistically evaluating the efficacy of the models for sky segmentation is one of the key points of emphasis in the research. A suite of metrics is selected to quantify the performance of the models. The work leverages mean Intersection over Union (mIoU), Dice Similarity Coefficient (DSC), precision, recall, and accuracy, as critical metrics for this study.

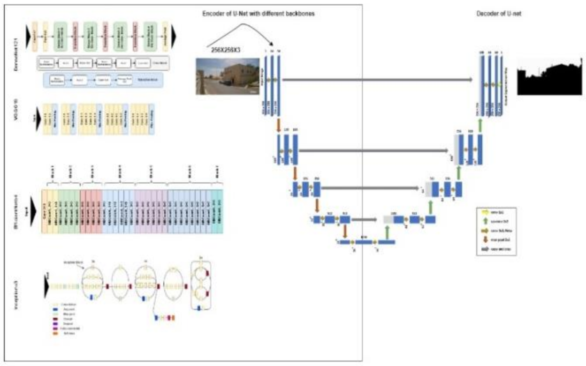

Fig. 5. Demonstratration of the working of the proposed architecture.

V. Results

The operating system of this work is Windows 11, the deep learning environment is Keras and TensorFlow, and the programming language is Python. The hardware configuration is an Intel Core i7-2.80 GHz CPU and 16.0 GB RAM.

In this research study, the proposed novel deep learning algorithm for sky segmentation was evaluated by utilizing various encoders within the U-Net architecture and subjecting it to critical analysis.

After several experiments, models that consist of the encoder employing VGG19, EfficientNetB4, InceptionV3, and DenseNet121 were compared, InceptionV3 U-Net showed the best perfomence, obtaining remarkable mIOU and DSC scores as high as 98.57% and 99.57% respectively. mIOU stands for mean Intersection over Union while DSC stands for Dice Similarity Coefficient which tells us how great the proposed model can perfectly draw the Sky region in the images. Such high scores are indicative of the model's robust performance in capturing the intricate details of the sky, laying the foundation for applications such as obstacle detection and path planning for autonomous vehicles.

Table I depicts the model’s performance concisely, measuring the Recall, precision, accuracy, and F1 score. Embodied in the table is not only the performance of the Inceptionv3 U-Net but also its other architectures, the VGG19 U-Net, the DenseNet121 U-Net, and the Efficientnetb4 U-Net. Though the VGG19 U-Net, the DenseNet121 U-Net, and the Efficientnetb4 U-Net all offer comparable results, their scores are slightly less in comparison to the Inceptionv3 U-Net. The high performance that is demonstrated by all models serves as a testament to the ability of different encoders to amplify the U-Net architecture where sky segmentation is concerned.

TABLE I. PERFORMANCE EVALUATION FOR SKY SEGMENTATION MODELS

Evaluation	Models			
	VGG19 U-Net	***Densenet121 U-Net***	***Efficientnetb4 U-Net***	***Inceptionv3 U-Net***
mIoU	98.46 %	98.48 %	98.45 %	98.57 %
DSC	99.53 %	99.54 %	99.53 %	99.57 %
Recall	99.36 %	99.73 %	99.33 %	99.41 %
Precision	99.71 %	99.35 %	99.72 %	99.73 %
Accuracy	99.35 %	99.36 %	99.34 %	99.40 %
Loss	0.09	0.11	0.14	0.11

A further illustration of training and testing curves is shown in Fig. 6 and Fig. 7. In Fig. 6, the upper row corresponds to the InceptionV3 U-Net model, revealing its superior performance compared to the lower row representing the DenseNet121 U-Net model. Similarly, Fig. 7 presents the learning trajectories of the VGG19 U-Net and EfficientNetB4 U-Net models, offering nuanced perspectives on their adaptability and convergence. The models have performed consistently well over both the training and validation sets, hence no signs of overfitting. In addition to the numerical metrics, it can be seen that the loss value during training is consistently low, again reaffirming the robustness of training procedures. Moreover, it can be concluded the models with high accuracy scores can do pixel classification in the sky region very well, which indicates the models are good enough even in discerning subtle details.

The attainment of high learning due to the decrease in loss curves indicates the learning models’ saturation. Conversely, the rise in accuracy and IOU curves shows that the model is still learning, which means that the learning phase is not over. Both curves are proposed to be well-balanced so as to enable the learning process to be terminated.

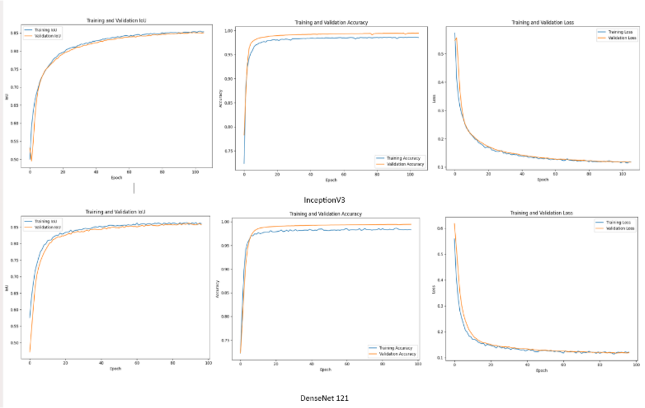

Fig. 6. The performance curves of the inceptionv3 U-Net and Densenet121 U-Net.

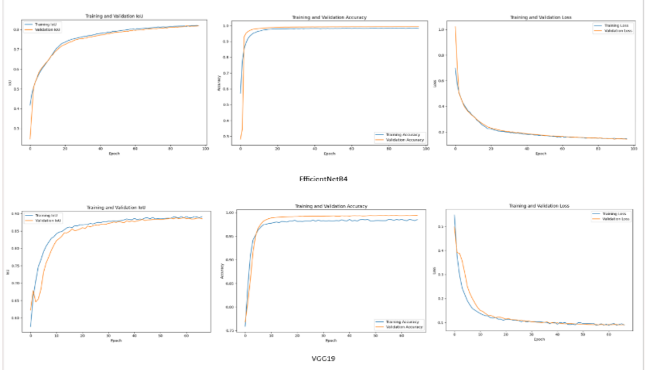

Fig. 7. The performance curves of the Efficientnetb4 UNet and VGG19 Unet.

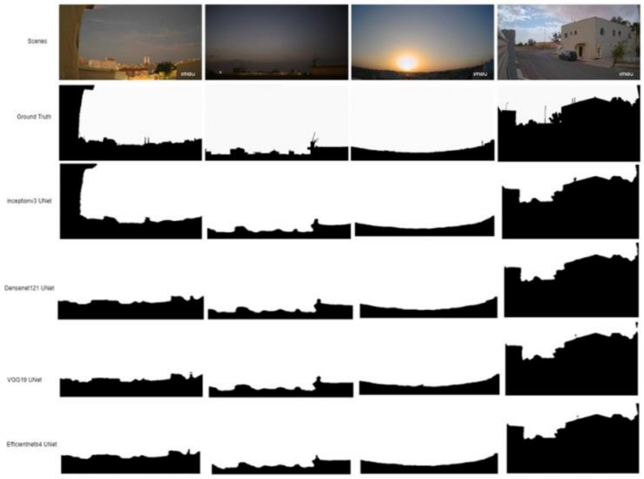

Fig. 8. Samples of prediction of the models.

The qualitative assessment of model predictions is presented next in Fig. 8. This includes visual samples of the models’ predictions on various scenes in different weather conditions and times of the day. The examples show that the models perform well in challenging setups such as back illumination or low light. This preliminary qualification holds true across the U-Net architecture and for the different backbone encoders, where it is evident that all models very accurately detect sky and ground pixels in various scenes.

VI. Discussion

The proposed models are able to perform with a relatively high degree of accuracy. In particularly hard scenarios, such as night scenes, the results are outstanding. They have done quite well in the task of distinguishing sky- and non-sky regions. And have done much better than other systems to avoid misclassifications such as white buildings as clouds, with a significantly greater precision.

However, the models' limitations become apparent in more complex scenarios, such as scenes with intricate structures like trees or poles on buildings. In these instances, the models struggle to accurately detect sky areas, revealing areas that might benefit from further refinement. This acknowledgement of limitations is crucial for guiding future iterations of the models.

This project opens avenues for future research, outlining potential directions to enhance the model's capabilities and address identified limitations. As a future work, various complex sites will be incorporated to train models robustly in the future by increasing dataset used in this project. Additionally, assessing how the model performs on different datasets to assess its adaptability to different scenarios and datasets. Moreover, this study will be expanded by focusing on a specific challenge, e.g., in weather phenomena, such as dust or rain and how to detect the sky.

VII. Conclusion

The finalization of this work throws light on the significant achievements that have been made in the field of Semantic Sky Segmentation. Initially, the main aim was to search for the most suitable encoder that could capture all the details of the sky, ideally during the segmentation process, to get very accurate results. The whole project was carried out in a series of different stages where alterations were made to the U-Net Architectures that employed several other encoders such as VGG19, EfficientNetB4, InceptionV3, and DenseNet121. The end-to-end binary segmentation model was a key stage in the proposed approach. The project's foundation rested on various steps, from comprehensive data preparation to the advanced image processing steps. The choice of the Keras framework facilitated a simplified model construction process, allowing for essential data augmentation to increase the dataset and enhance the model's overall performance.

Model evaluation was accomplished with the help of metrics like mIOU, Accuracy, Precision, Recall, DSC, Loss, etc. The results were as follows: the mean Intersection over Union, Dice similarity coefficient, recall, precision, and accuracy scores of 98.57%, 99.57%, 99.41%, 99.73%, and 99.40%, respectively. Additionally, it's noteworthy that the U-Net with VGG19 equivalent achieved the best loss of 0.09, underscoring its effectiveness in minimizing error.

This comprehensive approach of the evaluation process helps to understand the various aspects of the models. The InceptionV3 UNet model was identified as the most robust performer among the models tested over this dataset. Thus, the extended view of performance metrics for the different models validated the precision of the model's segmentation.

The success of this project lies not only in the numbers but the proof lies in models' improved perception of complex scenes and their ability to work more effectively with applications. These have done well with an ability to tell sky pixels away from the ground.

In conclusion, the outcomes of these experiments not only contribute to the growing body of knowledge in computer vision but also pave the way for practical applications where precise sky segmentation holds significant importance.

Acknowledgment

Researchers would like to thank the Deanship of Scientific Research, Qassim University for funding publication of this project.

References

[1] C. Lu, D. Lin, J. Jia, and C. K. Tang, "Two-Class Weather Classification," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 12, pp. 2510–2524, 2017, doi: 10.1109/TPAMI.2016.2640295.

[2] T. Halperin, H. Cain, O. Bibi, and M. Werman, "Clear Skies Ahead: Towards Real-Time Automatic Sky Replacement in Video," Comput. Graph. Forum, vol. 38, no. 2, pp. 207–218, 2019, doi: 10.1111/cgf.13631.

[3] T. Ahmad, E. Emami, M. Cadik, and G. Bebis, "Resource Efficient Mountainous Skyline Extraction using Shallow Learning," Proc. Int. Jt. Conf. Neural Networks, vol. 2021-July, pp. 1–9, 2021, doi: 10.1109/IJCNN52387.2021.9533859.

[4] S. Ghosh, N. Das, I. Das, and U. Maulik, "Understanding deep learning techniques for image segmentation," ACM Comput. Surv., vol. 52, no. 4, 2019, doi: 10.1145/3329784.

[5] T. B. Ronneberger, Olaf, Philipp Fischer, "U-Net: Convolutional Networks for Biomedical Image Segmentation," Med. Image Comput. Comput. Interv. 2015 18th Int. Conf., 2015, doi: 10.1109/ACCESS.2021.3053408.

[6] E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 4, pp. 640–651, 2014, doi: 10.1109/TPAMI.2016.2572683.

[7] He, Kaiming & Gkioxari, Georgia & Dollar, Piotr & Girshick, Ross. (2017). Mask R-CNN. 2980-2988. 10.1109/ICCV.2017.322.

[8] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image Segmentation Using Deep Learning: A Survey," IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 7, pp. 3523–3542, 2022, doi: 10.1109/TPAMI.2021.3059968.

[9] Y. Shen and Q. Wang, "Sky region detection in a single image for autonomous ground robot navigation," Int. J. Adv. Robot. Syst., vol. 10, pp. 1–13, 2013, doi: 10.5772/56884.

[10] Z. Zhao, Q. Wu, H. Sun, X. Jin, Q. Tian, and X. Sun, "A Novel Sky Region Detection Algorithm Based On Border Points," Int. J. Signal Process. Image Process. Pattern Recognit., vol. 8, no. 3, pp. 281–290, 2015, doi: 10.14257/ijsip.2015.8.3.26.

[11] Y. Shang, G. Li, Z. Luan, X. Zhou, and G. Guo, "Sky detection by effective context inference," Neurocomputing, vol. 208, pp. 238–248, 2016, doi: 10.1016/j.neucom.2015.12.126.

[12] F. de Mattos, A. T. Beuren, B. M. N. de Souza, A. De Souza Britto, and J. Facon, "Supervised approach to sky and ground classification using whiteness-based features," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 10633 LNAI, no. February 2019, pp. 248–258, 2018, doi: 10.1007/978-3-030-02840-4_20.

[13] Y. Song, H. Luo, J. Ma, B. Hui, and Z. Chang, "Sky detection in hazy image," Sensors (Switzerland), vol. 18, no. 4, pp. 1–18, 2018, doi: 10.3390/s18041060.

[14] C. Fang, C. Lv, F. Cai, H. Liu, J. Wang, and M. Shuai, "Low Light Image Enhancement for Color Images Combined with Sky Region Segmentation," Proc. - 2022 Int. Conf. Mach. Learn. Knowl. Eng.

MLKE 2022, pp. 169–172, 2022, doi: 10.1109/MLKE55170.2022.00039.

[15] J. Yin, T. Li, H. Yin, W. Yu, and D. Zou, "Sky-GVINS: a sky-segmentation aided GNSS-Visual-Inertial system for robust navigation in urban canyons," Geo-Spatial Inf. Sci., 2023, doi: 10.1080/10095020.2023.2191649.

[16] Y. H. Tsai, X. Shen, Z. Lin, K. Sunkavalli, and M. H. Yang, "Sky is not the limit: Semantic-aware sky replacement," ACM Trans. Graph., vol. 35, no. 4, 2016, doi: 10.1145/2897824.2925942.

[17] R. P. Mihail, S. Workman, Z. Bessinger, and N. Jacobs, "Sky segmentation in the wild: An empirical study," 2016 IEEE Winter Conf. Appl. Comput. Vision, WACV 2016, 2016, doi: 10.1109/WACV.2016.7477637.

[18] Z. Zou, R. Zhao, T. Shi, S. Qiu, and Z. Shi, "Castle in the Sky: Dynamic Sky Replacement and Harmonization in Videos," IEEE Trans. Image Process., vol. 31, pp. 5067–5078, 2022, doi: 10.1109/TIP.2022.3192717.

[19] X. Wang et al., "MobileSky: Real-Time Sky Replacement for Mobile AR," IEEE Trans. Vis. Comput. Graph., vol. PP, no. X, pp. 1–17, 2023, doi: 10.1109/TVCG.2023.3257840.

[20] O. Liba et al., "Sky optimization: Semantically aware image processing of skies in low-light photography," IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work., vol. 2020-June, pp. 2230–2238, 2020, doi: 10.1109/CVPRW50498.2020.00271.

[21] Z. A. R. and Y. Z. Boyu Kuang, "Sky and Ground Segmentation in the Navigation Visions of the Planetary Rovers," Sensors, vol. Volume 21, no. issue 21, 2021, doi: https://doi.org/10.3390/s21216996.

[22] Scott Krig, "Ground Truth Data, Content, Metrics, and Analysis. In: Computer Vision Metrics," Springer, Cham, 2016.

[23] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., pp. 1–14, 2015.

[24] M. Tan and Q. V Le, "EfficientNet : Rethinking Model Scaling for Convolutional Neural Networks," 2019.

[25] C. Szegedy, V. Vanhoucke, and J. Shlens, "Rethinking the Inception Architecture for Computer Vision," pp. 2818–2826, 2016.

[26] G. Huang, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," Proc. IEEE Conf. Comput. Vis. pattern Recognit., pp. 4700–4708, 2017.