Exploring Models and Band Selection for Improved Contrail Detection with Deep Learning

Alam Rahmatulloh [1], Virra R. A'izzah [1], Irfan Darmawan [2,*], and Muhammad Al-Husaini [1]

[1] Department of Informatics, Faculty of Engineering, Siliwangi University, Tasikmalaya, Indonesia
[2] Department of Information Systems, Faculty of Industrial Engineering, Telkom University, Bandung, Indonesia
Email: alam@unsil.ac.id (A.R.); 207006020@student.unsil.ac.id (V.R.A.);
irfandarmawan@telkomuniversity.ac.id (I.D.); alhusaini@unsil.ac.id (M.A.-H.)
*Corresponding author

***Abstract*—The consequences of climate change are becoming increasingly urgent, with contrails emerging as a potential contributing factor to this phenomenon. Consequently, there is an urgent need for precise techniques to detect them in satellite imagery. This study uses deep learning models and band selection to improve contrails detection in geostationary satellite imagery, using the Landsat-8 dataset with human-labeled contrails sourced from the GOES-16 Advanced Baseline Imager. By comparing different deep learning model methods such as DeepLabV3, U-Net, Fully Convolutional Network (FCN), Pyramid Scene Parsing Network (PSPNET2), ensemble deep learning, and different bands such as ash color scheme using Bands 11, 14, 15, and 08, this study investigates their collective impact in improving contrail identification. Results found that the selected deep learning model significantly affected the detection process, but incorporating band 08 into the input channel did not significantly improve model performance. The most effective model was the FCN equipped with Band 08, with the lowest average loss during training (0.032591) and validation (0.013321). This research is expected to improve contrail detection in satellite imagery by using deep learning models and band selection to assist policymakers and researchers in developing strategies to reduce aviation climate impacts.**

***Keywords*—climate change, contrails, Fully Convolutional Network (FCN), DeepLab, deep learning model, satellite imagery, Pyramid Scene Parsing Network (PSPNET)**

I. Introduction

Artificial cirrus clouds, often called contrails, are condensation trails that form behind airplanes as they pass through the upper atmosphere. These trails, while they may seem harmless, have a significant impact on global warming. In their contribution to global climate change, according to Intergovernmental Panel on Climate Change (IPCC) [1], contrails are thought to be responsible for about 35% of the total environmental impact stemming from aviation activities.

Reducing the environmental impact of contrails is an important challenge in dealing with global climate change. One approach that can be taken is to change the flight altitude of aircraft away from high-humidity areas, which is one of the main factors in the formation of contrails [2, 3]. This research utilizes image recognition technology, satellite data analysis, and a large dataset of labeled contrail images to achieve this goal. Through data processing, this research attempts to predict the potential locations of contrails with a high degree of accuracy.

Several deep-learning algorithms have been created to identify contrails in satellite images, including Convolutional Neural Networks (CNN), U-Net, Fully Convolutional Network (FCN), Pyramid Scene Parsing Network (PSPNET), and DeepLabV3. Convolutional Neural Networks (CNN) are a popular deep learning model for object identification and image categorization applications. One type of deep learning artificial neural network often used for image recognition and classification applications is the convolutional artificial neural network [4]. These networks have been used in hyperspectral image classification and are excellent at extracting spatial characteristics from images [5]. U-Net is a type of CNN commonly applied to jobs requiring segmentation or dividing an image into parts; U-Net can produce accurate segmentation results and is designed to operate with relatively few datasets [6, 7].

Fully Convolutional Network (FCN) excels in semantic segmentation by using locally connected layers like convolution, pooling, and upsampling. Its architecture is designed to reduce parameters and speed up training by streamlining the process. The network comprises a downsampling path for context extraction and an upsampling path for localization [8, 9]. High-accuracy object detection in images is achieved by integrating CNNs with additional approaches, as demonstrated by advanced image segmentation models such as PSPNET and DeepLabV3 [5, 10]. The DeepLabV3 model is an enhanced version of the DeepLabV2 model that considers objects at multiple scales and segments with much better accuracy thanks to the Atrous Convolution and Atrous Spatial Pyramid Pooling (ASPP) modules. Unlike its predecessors, the V1 and V2 models, the DeepLabV3 model does not include the Conditional Random Field (CRF) [11]. Another architecture for semantic segmentation is PSPNet. For improving semantic segmentation accuracy, PSPNet uses a pyramid pooling

Manuscript received December 12, 2023; revised January 15, 2024; accepted February 22, 2024; published June 5, 2024.

module to extract global context information from the input image [12, 13]. This approach finds application in various image segmentation tasks, including object detection and picture segmentation. These models are used, for example, in satellite photography to detect contrails.

Band ash color contrail detection uses particular color channels in satellite imagery to detect contrails. For this purpose, the Ash RGB color band is accommodating since it monitors and detects sulfur dioxide gas and volcanic ash day and night using infrared window channels [14].

In a study conducted by Meijer *et al.* [15], the objective was to enhance the capacity to distinguish distinct clouds by utilizing the Ash RGB color scheme. The Ash RGB product is a false-color composite that combines four GOES-16 Advanced Baseline Imager (ABI) infrared bands. Utilizing this technology enhances the separation of clouds and better distinguishes between them. [16]. This method enabled the development of a strategic approach to analyzing satellite images, which could enhance the efficiency of contrail detection and monitoring.

Recent studies [17, 18] have attempted to use deep learning to detect contrails, but the accuracy remains to be improved. Additionally, the model is solely based on CNN, U-Net, PSPNet, DeepLabV3, and DeepLabV3+.

Based on the problems described, this research contribution highlighted exploring machine learning models and optimal bands for contrail detection in satellite images. The goal is to find the most effective band and model for detecting contrails and assess its precision using the dice loss metric. This study is expected to help develop strategies to reduce contrails' impact on climate change by providing an accurate and dependable method for detecting contrails on satellite imagery. The outcomes of this research can also contribute to the growing research on contrail detection and offer an understanding of the potential advantages and disadvantages of using deep learning models for contrail detection.

The rest of this article is structured as follows: Section II briefly overviews previous research. Section III outlines the suggested research approach. Section IV discusses the experiments, the data sets used, the models and parameters evaluated, and the comparison results. Finally, Section V concludes this research and outlines potential future directions.

II. Literature Review

Bhandari *et al.* [17] comprehensively explores contrail detection methodologies and their implications on climate change. It presents a meticulous benchmarking study involving cutting-edge semantic segmentation models tailored for discerning contrails within low-orbit satellite imagery. Delving deeper, it elucidates the formation of contrails arising from aircraft engine emissions and their consequential impact on global warming by intensifying cloud cover, thereby trapping heat. Offering potential remedies to mitigate contrail formation and diminish their environmental influence, the paper showcases a human-labeled Landsat-8 contrails dataset comprising false color images and corresponding contrail masks. Through meticulous experimentation with diverse segmentation models like U-Net, PSPNet, DeepLabV3, and DeepLabV3+, coupled with varying loss functions and encoder backbones, the research scrutinizes their efficacy. The results divulge nuanced insights, with U-Net leveraging the Xception 71 backbone exhibiting the most promising IoU score of 0.4395. Additionally, the paper illustrates train and test IoU scores for all models and presents visual examples of predicted masks. However, it underscores existing limitations and complexities within the study, paving the way for future research directions and potential enhancements in contrail detection methodologies.

Hoffman *et al.* [18] introduced a new method using the convolutional neural network U-Net to detect contrails in satellite images. The research used the GOES ABI 11 μm and 12 μm channels, Band 14 and Band 15, for training, testing, and validation purposes. Contrails have a complex impact on climate change, with a net warming effect due to their influence on longwave radiation. The efficiency of the U-Net model was demonstrated by its study findings, which included a detection probability of 0.51, a false alarm ratio of 0.46, and an F1-Score of 0.52. Further research can incorporate atmospheric conditions and flight path data to improve contrail detection accuracy, especially in regions conducive to contrail formation.

Agung *et al.* [19] explores the significance of band selection in identifying possible geothermal prospect locations within Songgoriti Batu and its surrounding areas using Landsat 8 satellite images. Employing an array of precise methodologies, including using Landsat 8 Bands 4, 5, 7, and 11, the researchers meticulously mapped surface structures, temperatures, and rock formations. Through extensive corrections and conversions, they produced comprehensive maps that could serve as pivotal data for identifying geothermal prospects. The findings were compelling, revealing dominant fault lines, heightened temperature anomalies, and altered rocks near volcanic activity and hot spring occurrences in Songgoriti. The research, poised for further development, proposes refining geophysical measurement techniques to enhance the identification of potential geothermal sites in the area. Crucially, the researchers underscored the critical impact of band selection, emphasizing its role in accurately extracting lithological, structural, and thermal information from the satellite imagery, thereby highlighting its pivotal importance in this investigative process.

Zheng *et al.* [20] investigated the effectiveness of band selection methods for hyperspectral images. The study used an attention mechanism-based convolutional network to select the most effective bands and compared the results with traditional band selection methods such as Successive Projection Algorithm (SPA), Genetic Algorithm (GA), and the latest Two-Branch Convolutional Neural Network (2B-CNN) algorithm. The research found that the effective bands selected by the proposed attention-based model achieved higher regression R^2 values and classification accuracies compared to the full-spectrum data and the comparative band selection methods. Based on this study, choosing representative and effective spectral bands is crucial to remove unnecessary data and lighten the

computational load for prospective real-time hyperspectral imaging applications.

Siddiqui [21] conducted a study that examined the application of deep learning algorithms for the detection of contrails in the atmosphere. This study used a Convolutional Neural Network (CNN) to classify pixels as contrail or non-contrail from images collected by the Total Sky Imager (TSI) from March 2017. The CNN was trained with 1600 images and validated with 400 images. The results show that the CNN achieved 97.5% accuracy on the training set and 98.5% on the validation set. This study also suggested some possible improvements and extensions to this research, such as increasing the size and diversity of the dataset, applying different CNN architectures and hyperparameters, and evaluating the impact of contrails on climate change.

Zhang *et al.* [22] proposed a new approach for contrail detection using a Convolutional Neural Network (CNN) model, ContrailMod, which classifies contrails from Himawari-8 satellite images and outperforms the conventional Contrails Detection Algorithm (CDA). This study also estimated Precipitated Calcium Carbonate (PCC) using specific temperature and humidity from ECMWF reanalysis data in South China. The CNN model achieved an accuracy of 0.97 and an F1-Score value of 0.94 on the test set. Future research can be done by applying the CNN model to other regions and satellites, incorporating more features and channels into the CNN model, and studying the contrail's radiation effect and climate impact.

Table I summarizes the research gaps based on the objectives and literature support.

TABLE I. RESEARCH GAP MATRIX

Research	Contrail Detection	Band Selection	Model Comparison	Model Used	Dataset Used
Bhandari *et al.* [17]	✓	×	✓	U-Net, PSPNet, DeepLabV3, DeepLabV3+	Landsat-8 Contrails Dataset
Hoffman *et al.* [18]	✓	×	×	U-Net	Satellite Images (Specifics not given)
Agung *et al.* [19]	×	✓	×	N/A	Landsat-8
Zheng *et al.* [20]	×	✓	✓	Attention-based CNN, SPA, GA, 2B-CNN	Hyperspectral Images
Siddiqui [21]	✓	×	×	CNN	Total Sky Imager Images
Zhang *et al.* [22]	✓	×	×	ContrailMod (CNN)	Himawari-8 Satellite Images
Current Research	✓	✓	✓	DeepLabV3, U-Net, FCN, PSPNET2, Ensemble Deep Learning	Landsat-8 Contrails Dataset sourced from GOES-16 Advanced Baseline Imager

III. RESEARCH METHODS

In this study, a model will be trained to detect contrails using several deep learning methods and distinct ash band schemes, serving as a comparative analysis. The use of different bands aims to use various infrared channels at different wavelengths to detect contrails. These bands play an important role in capturing and representing different aspects of the atmosphere. The "ash" color scheme, consisting of bands 11, 14, and 15, initially designed to observe volcanic ash, proved very useful in identifying contrails in GOES imagery. The unique characteristics of each band, determined by its specific wavelength and calibrated brightness temperature, allow for comprehensive analysis. Therefore, adding other bands to make it more optimal was investigated in addition to using the ash color scheme.

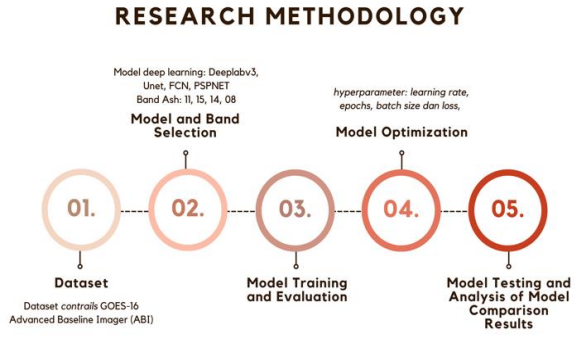


Fig. 1. Research methodology.

The stages in this research include data collection, model and band selection, model training and evaluation, model optimization, model testing, and analysis of model comparison results. The research methodology diagram can be seen in Fig. 1.

A. Dataset

The dataset used in this study consists of a collection of satellite images with aircraft contrails or condensation trails created by Joe Ng *et al.* [23, 24]. The authors state that they have built their dataset using GOES-16 ABI satellite imagery, which provides high spatial-temporal coverage and high-quality contrail labels by ensuring annotation consistency and with the same format [24]. Therefore, the dataset-cleaning stage was not performed in this study. This dataset uses geostationary satellite images to identify aircraft condensation trails. The original satellite images were obtained from the GOES-16 Advanced Baseline Imager (ABI) [14, 25], which is publicly available on Google Cloud Storage. The dataset contains data for training, testing, and validating the detection method. Since condensation traces are easier to identify with temporal context, this dataset provides image sequences at 10-minute intervals [24]. Each example (record_id) contains exactly one labeled frame. The training and validation sets contain folders representing the record_id and data such as band {08−16}.npy, human_individual_masks.npy, and human_pixel_masks.npy. The training set is used to train the deep learning model, while the validation set is used to validate the model. The test set contains data copies of the first two

records of the validation data (unlabeled) that are used to test the trained deep learning model. The GOES-16 Advanced Baseline Imager (ABI) dataset consists of 20,529 data for training, 1956 data for validation, and 2 data for testing.

The dataset comprises nine bands, each representing multiple images captured at different wavelengths of light. Only a partial composition will be considered as a sample for training, validation, and testing. A subset of the data will chose, which includes approximately 5% of the training data, 15% of the validation data, and all the testing data. This decision was mainly driven by research limitations and the need to balance computational resources and training time. It also ensures that the selected samples represent the underlying data distribution.

Fig. 2 is a visualization of the record 1000603527582775543; it can be seen in the image from left to right that various spectral bands (08−16) are visible, and from top to bottom, it is a contrail image taken with an interval of 10 min.

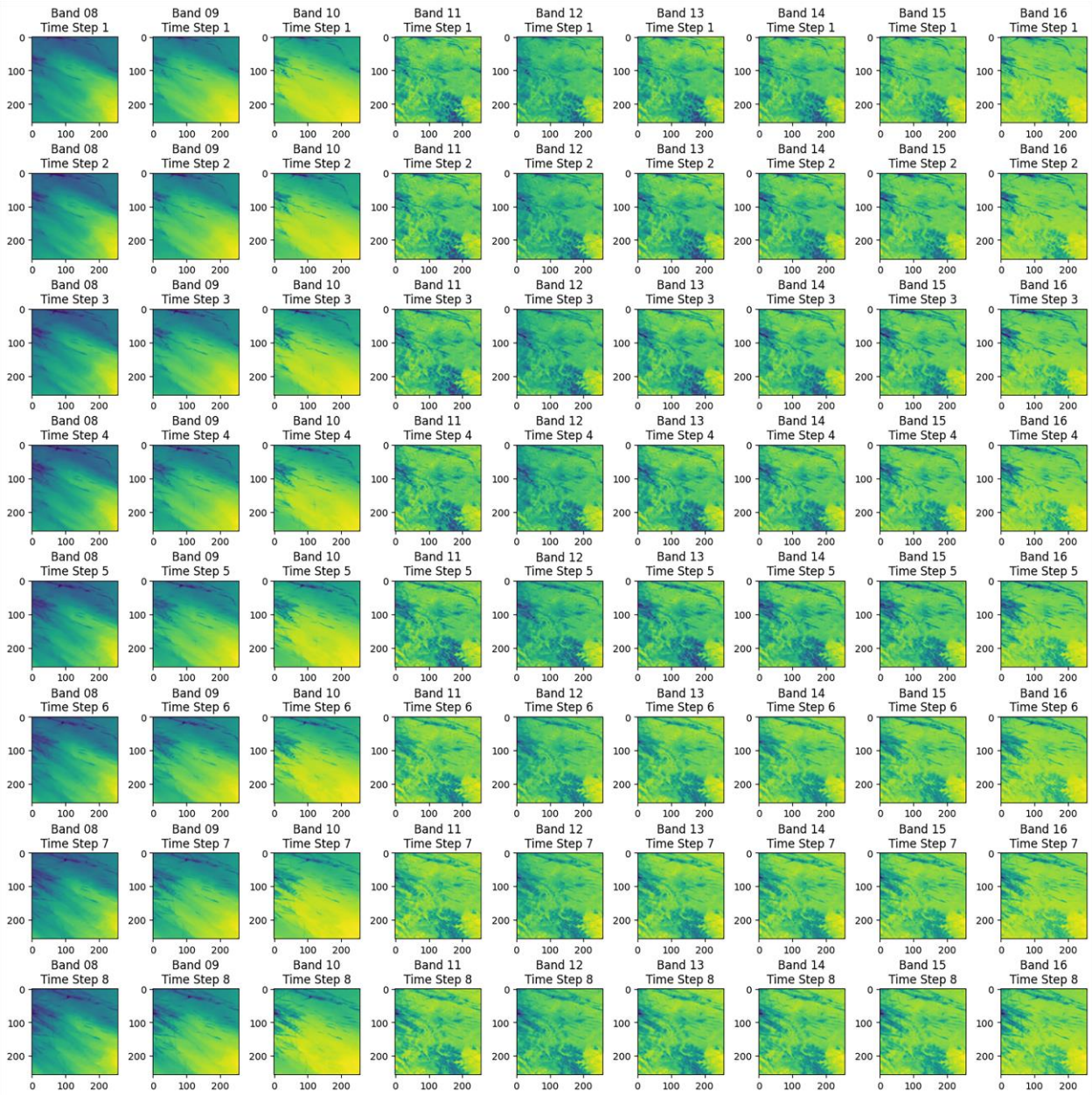


Fig. 2. Visualization of record 100603527582775543.

B. Model and Band Selection

Several deep learning models, such as DeepLabV3, U-Net, FCN, PSPNET2, and Ensemble Deep Learning (DeepLabV3, PSPNET, and U-Net), are selected and implemented in this step. These models were designed to learn the input data and make predictions. In addition, optimal bands were explored using the Ash color scheme, which included Bands 11, 14, and 15 and the addition of band 08. This step aims to understand and select the appropriate spectral bands to improve the accuracy of segmenting the condensation traces in the image by comparing them with the rest of the image.

This approach is based on deep learning and spectral analysis [26]. Using multiple models and optimal band exploration shows a comprehensive approach to improve image segmentation accuracy [27].

C. Model Training and Evaluation

The models used in this study were trained on the training set and evaluated on the validation set. The evaluation is done by comparing the performance of the models using a metric known as "dice loss" on two different tests, where both tests have different Ash color bands. The dice loss metric is used to measure the extent to which the models can produce results that match the target [28, 29] or ground truth, and the comparison is done to understand the extent to which the use of different Ash color bands and different models can affect the quality of contrails identification by the models.

The investigation is grounded on the configuration depicted in Table II. The seed parameter was set to a commonly used random value of 42 to ensure consistent reproducibility of the results. In addition, in model training, the batch size of data processed in each training iteration was set as 16 to streamline computation and memory usage without overloading the available resources. The images used for analysis were set at 256×256 pixels to ensure dimensional consistency across analyses. The number of data samples used was 1,000 for training (NUM_TRAIN_SAMPLES), 300 for validation (NUM_VAL_SAMPLES), and 2 for testing (NUM_TEST_SAMPLES). This number was selected due to limitations in this research and its computations. CUDA (GPU) devices are used in this research to maximize performance to speed up the training process.

TABLE II. MODEL CONFIGURATION

Configuration	Value
Seed	42
Batch Size	16
Image Size	256×256
Number Train Samples	1,000
Number Validation Samples	300
Number Test Samples	2
Device	CUDA (GPU)

D. Model Optimization

In the model optimization stage, this research involves fine-tuning the model's hyperparameters and optimizing them to improve model performance [30]. Hyperparameters are parameters determined before the learning process begins and affect the speed and quality of the learning process itself [31]. The fine-tuning of hyperparameters in the research includes learning rate, epochs, batch size, and loss to ensure that the models operate optimally.

Table III presents the detailed architecture and hyperparameters of the various models involved in this study. The models include DeepLabV3, PSPNET, U-Net, FCN, and an ensemble combining DeepLabV3, PSPNET, and U-Net. Each model is optimized using Stochastic Gradient Descent (SGD) and was trained for 45 epochs, with a learning rate of 0.001, except FCN, which was assigned a learning rate of 0.8. Input band refers to the number of input channels, with experiments conducted for three and four bands. The trainable parameters for each input band configuration varied across models. DeepLabV3 has 45,669,713 trainable parameters for three band inputs and 45,672,849 trainable parameters for four band inputs, PSPNET has 2,237,889 trainable parameters for three band inputs and 2,241,025 trainable parameters for four band inputs, U-Net has 51,513,233 trainable parameters for three band input and has 51,516,369 trainable parameters for four band input, and FCN has 64,673 trainable parameters for three band input and has 64,961 trainable parameters for four band input. The ensemble model consists of DeepLabV3, PSPNET, and U-Net, has 99,420,839 trainable parameters for three band inputs and 99,430,247 trainable parameters for four band inputs.

TABLE III. MODEL ARCHITECTURE

Model	Optimizer	Learning Rate	Epochs	Band Input	No. of Parameter Trained
DeepLabV3	SGD	0.001	45	3	45, 669, 713
				4	45, 672, 849
PSPNET	SGD	0.001	45	3	2, 237, 889
				4	2, 241, 025
U-Net	SGD	0.001	45	3	51, 513, 233
				4	51, 516, 369
FCN	SGD	0.08	45	3	64, 673
				4	64, 961
Ensemble (DeepLabV3, PSPNET, U-Net)	SGD	0.001	45	3	99, 420, 839
				4	99, 430, 247

E. Analysis of Model Comparison Results

At the stage of analyzing model comparison results, the performance of deep learning models trained using the same metrics is compared. The comparison results are expected to be used for recommendations in determining the most optimal model for contrail detection.

IV. RESULT AND DISCUSSION

A. Optimal Band Exploration

The band selection for the input channel is taken from the band with a low correlation in the ash color scheme. The correlation matrix provides insights into how strongly pairs of bands are related. Understanding the correlation between bands in remote sensing is crucial for feature selection, as highly correlated bands may contain redundant information. The correlation matrix aids in identifying and justifying the choice of optimal bands by ensuring that selected bands are diverse and offer unique information, avoiding multicollinearity issues in subsequent analyses.

Fig. 3 is the correlation matrix with the ash color scheme. The matrix shows that Band 08 has less correlation with the Ash colors but is quite well-correlated with the contrails, indicating that it could provide new and valuable information.

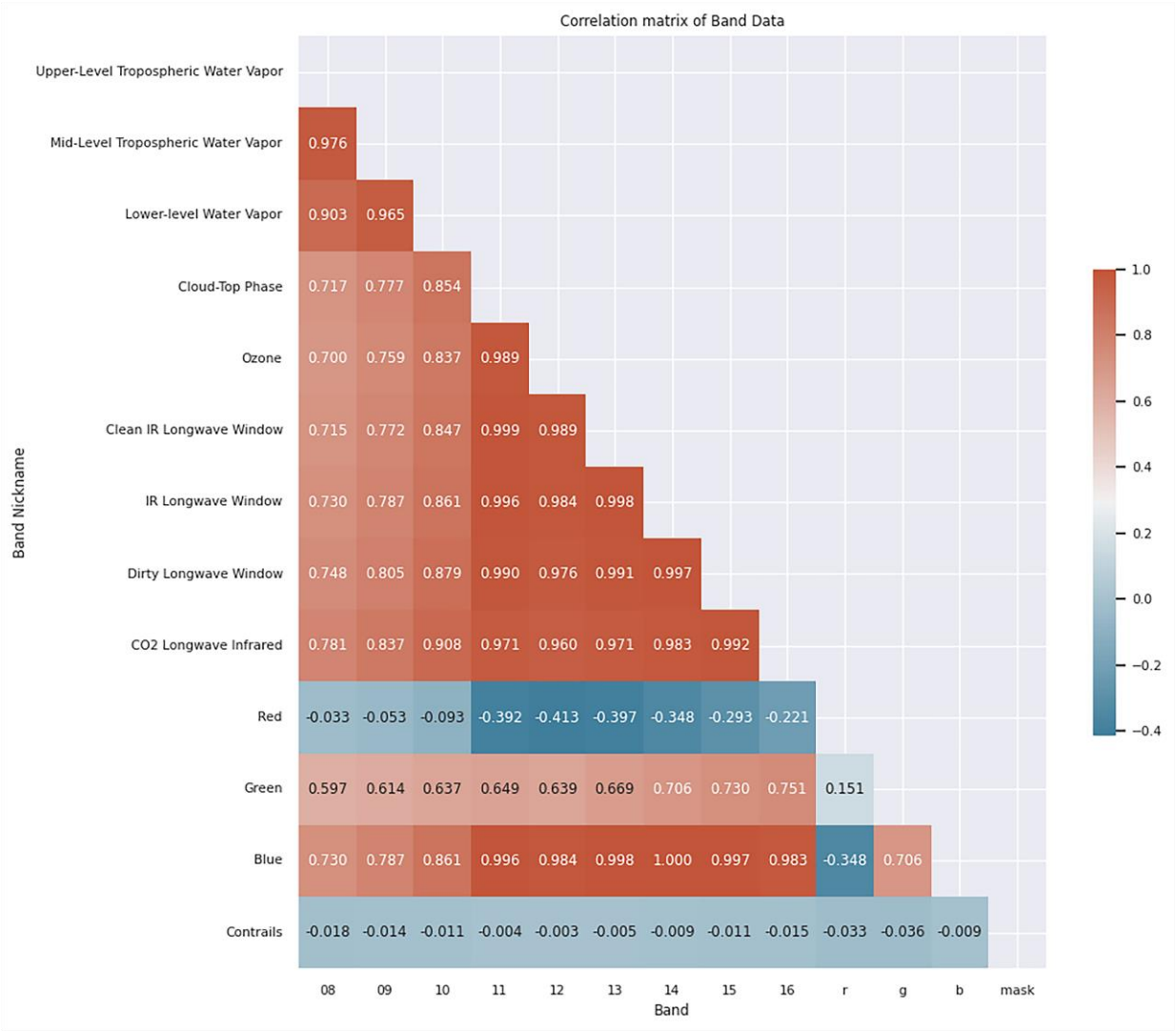


Fig. 3. Correlation matrix of Ash color scheme.

TABLE IV. BAND TESTING

Experiment	Band08	Band11	Band14	Band15
1	×	✓	✓	✓
2	✓	✓	✓	✓

Therefore, Table IV represents a categorization of experiments. Experiment 1 uses Ash false color images with Bands 15, 14, and 11. In Experiment 2, Band 08 is added as an input channel to use four channels: Bands 08, 11, 14, and 15.

Figs. 4 and 5 are the visualization results of the loaded and augmented training images. Fig. 4 shows the utilization of input Bands 11, 14, and 15, whereas Fig. 5 depicts the utilization of input Bands 08, 11, 14, and 15.

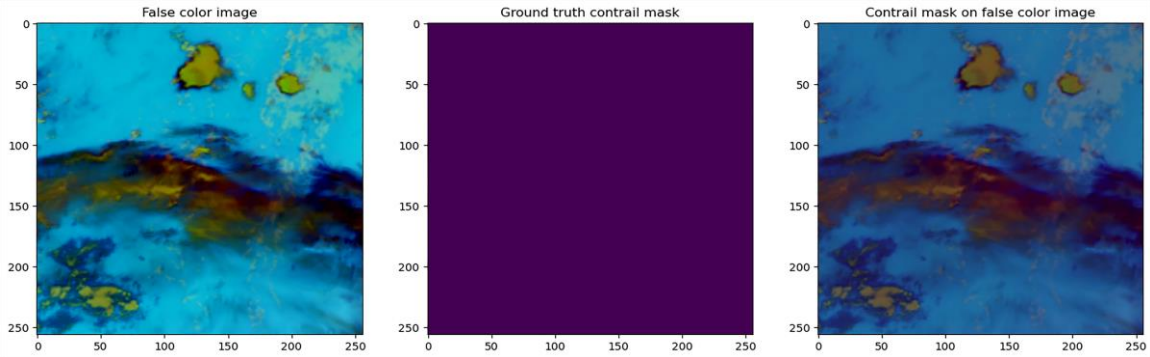


Fig. 4. Visualization of random samples Bands 11, 14, 15.

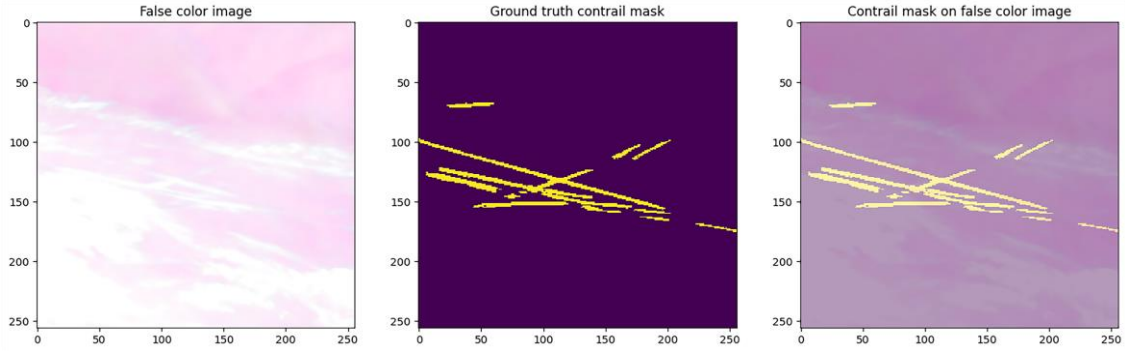


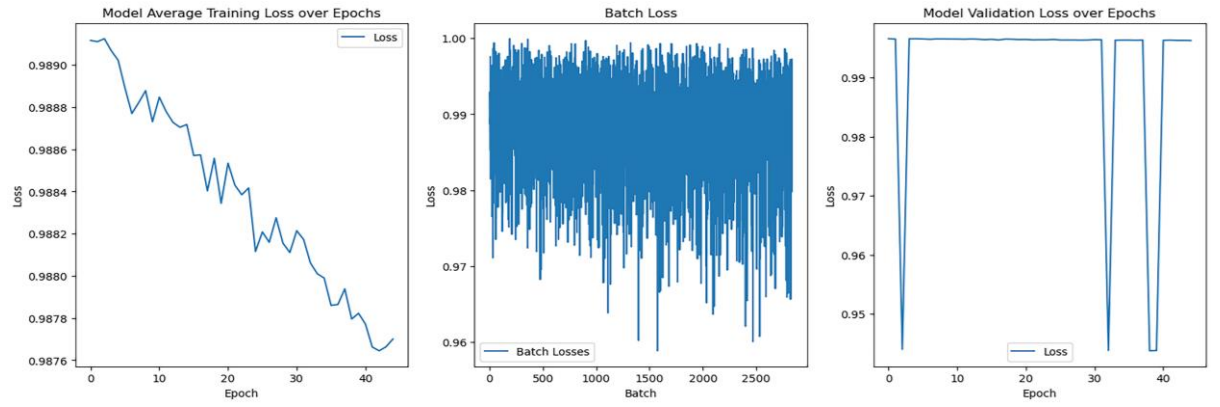
Fig. 5. Visualization of random samples Bands 08, 11, 14, 15.

B. Comparison Results

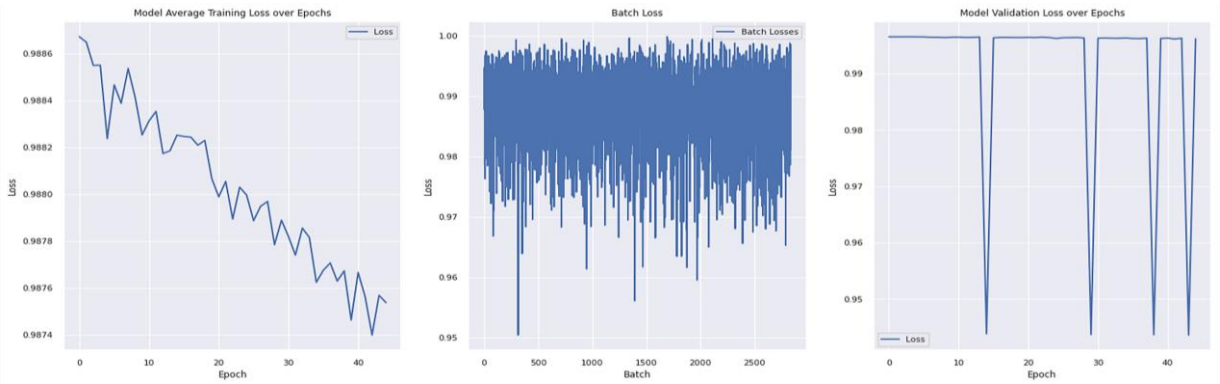
After the training and validation process, the performance of each model is visualized in three graphs or charts related to the machine learning model's performance over time. These graphs represent the average training loss, batch loss, and validation loss over time (epochs). These graphs can provide insight into how well the model learns and improves over time.

Fig. 6 illustrates the performance metrics of DeepLabv3 using 4-band and 3-band inputs that are similar. The first graph tracks the average training loss of the model across epochs, illustrating its learning curve. A sharp drop in the loss rate early indicates that the model converges to a satisfactory solution. Meanwhile, the second graph displays batch loss fluctuations in mini-batches, indicating the model's sensitivity to the varying data in these batches. Finally, the third graph shows the model's validation loss across epochs, reflecting its ability to generalize to unseen data; initially decreasing, the loss eventually shows a slight increase, indicating a potential risk of overfitting, especially in models with the addition of Band 08.

Fig. 7 is a graphical comparison of the PSPNET model performance on the two forms of testing. The model using Bands 11, 14, and 15 emerges as the best-performing model, showing the lowest training and validation losses, indicating superior performance. However, it is interesting that the model incorporating the additional Bands 08, 11, 14, and 15 showed higher validation losses, indicating potential overfitting on the training data. This issue highlights the necessary balance between learning from the data and avoiding over-reliance on certain patterns that may not generalize well.

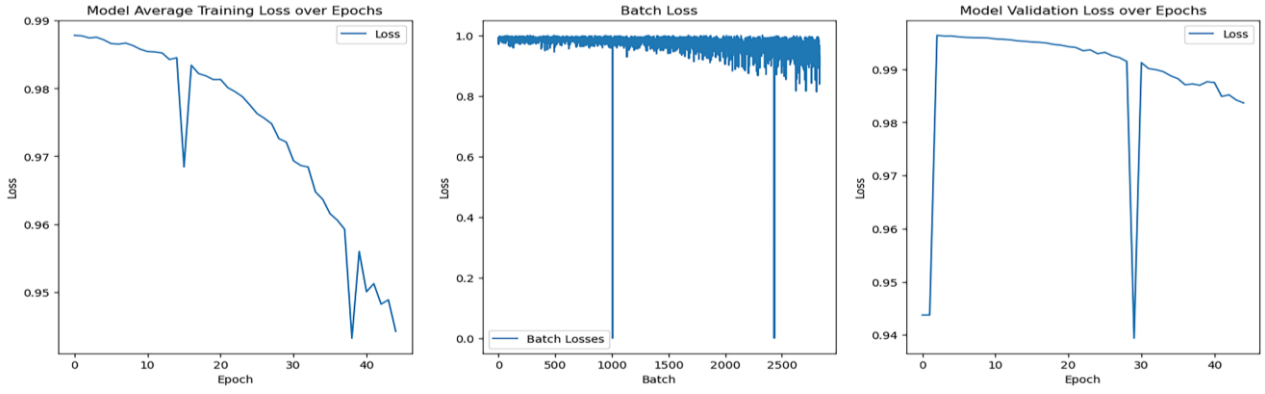


(a)

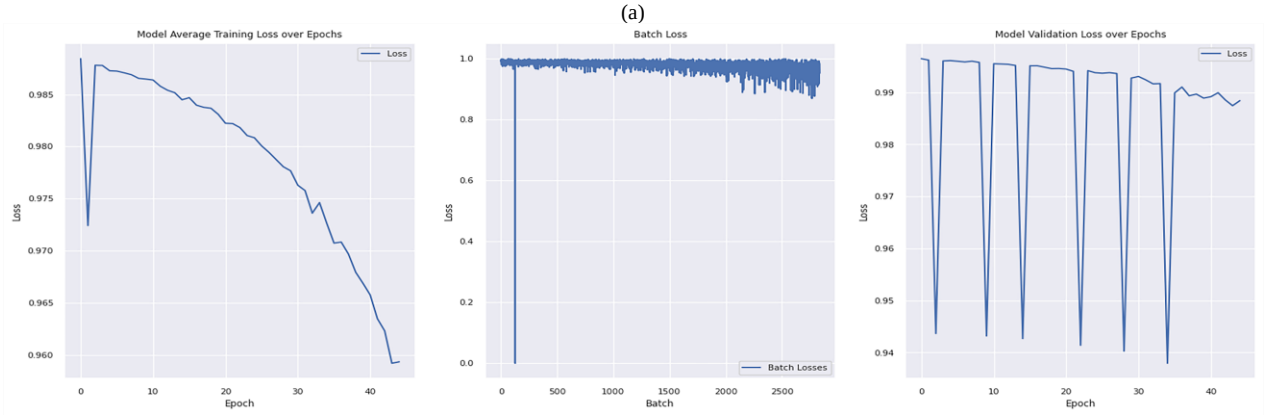


(b)

Fig. 6. Model DeepLabv3 (a) with Band 11, 14, 15; (b) with Band 08, 11, 14, 15.



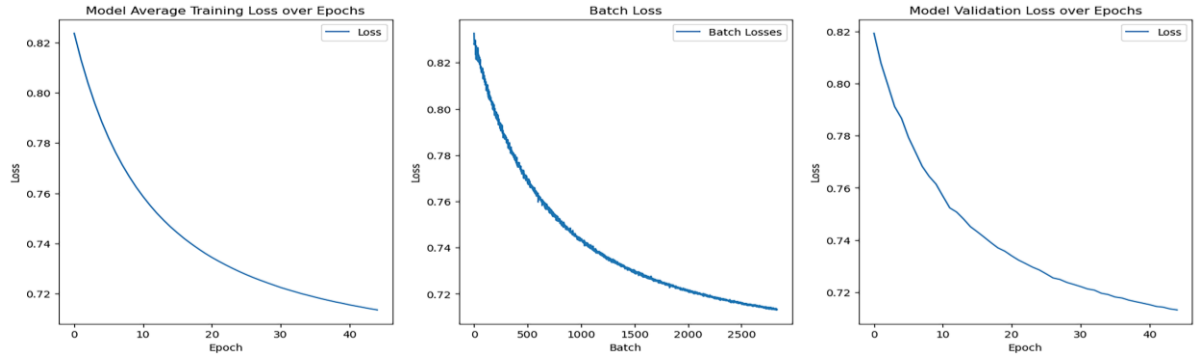
(a)



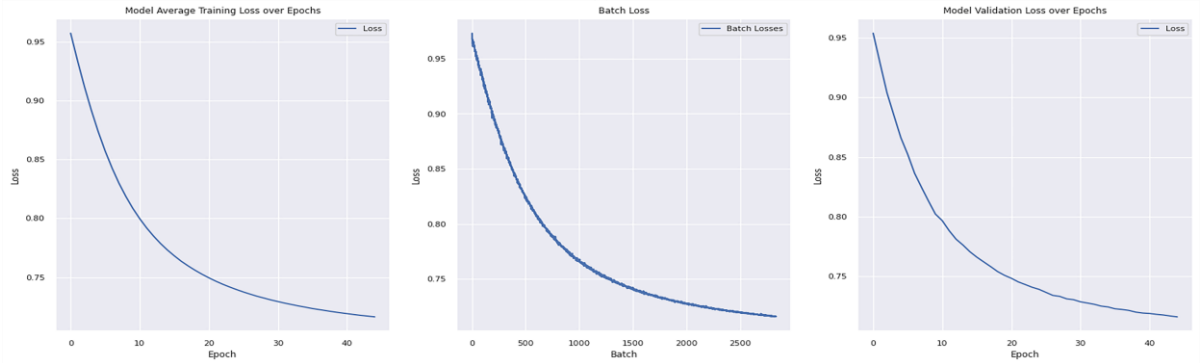
(b)

Fig. 7. Model PSPNET (a) with Band 11, 14, 15; (b) with Band 08, 11, 14, 15.

Fig. 8 illustrates the average training and validation losses across batches of U-Net model test results, which do not appear to have significant differences. The U-Net model with Bands 11, 14, and 15 or with the addition of Band 08 has similar performance results.



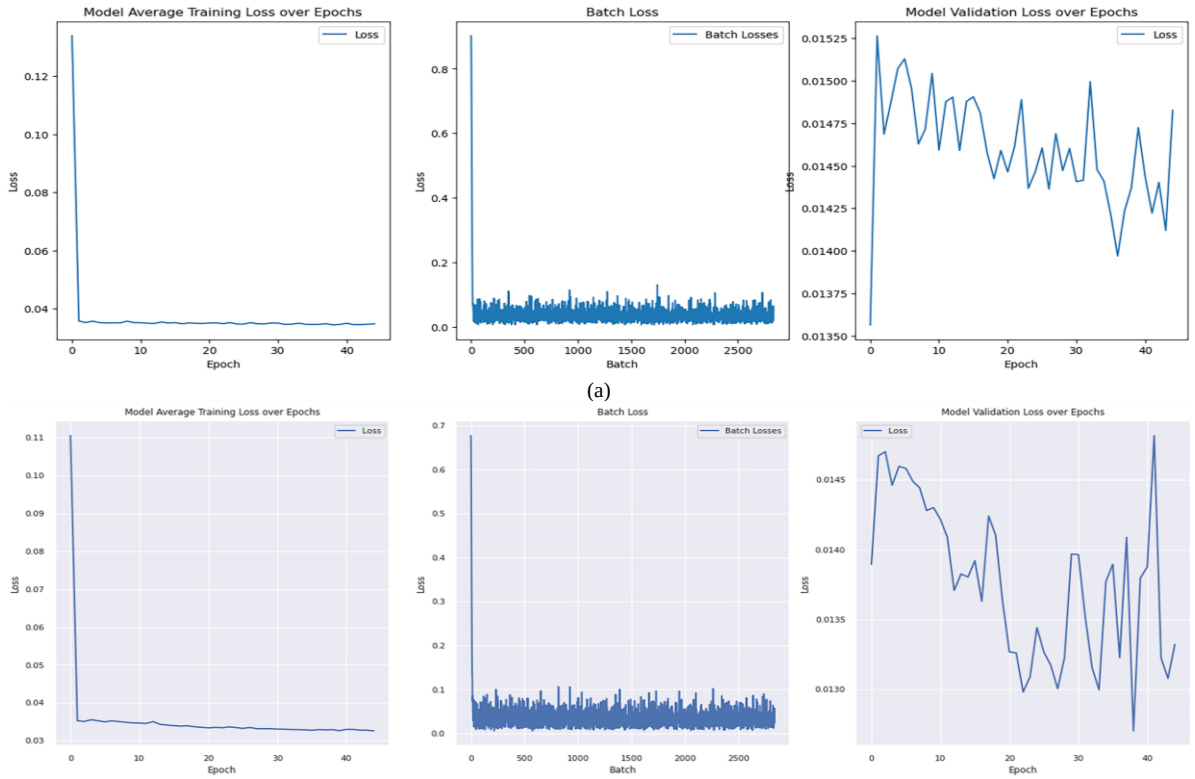
(a)



(b)

Fig. 8. Model U-Net (a) with Band 11, 14, 15; (b) with Band 08, 11, 14, 15.

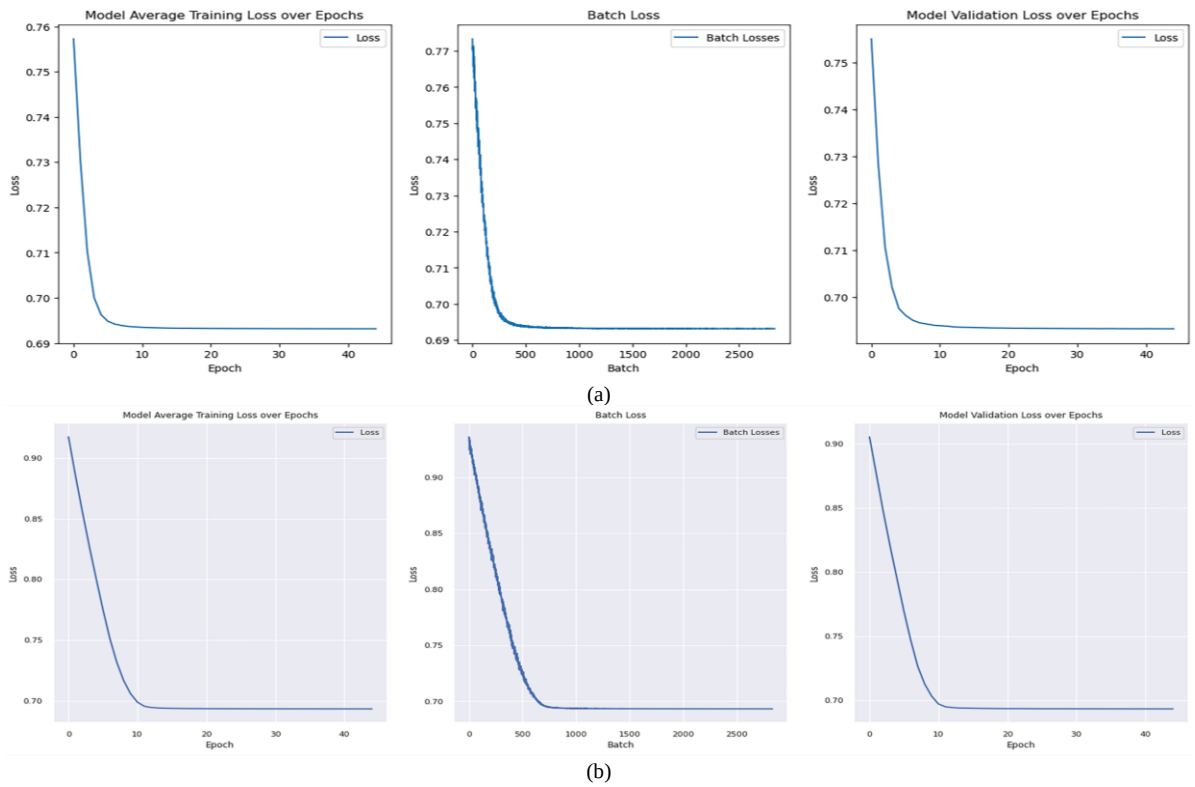
Fig. 9 displays the performance of the FCN model with various test bands. Models using Bands 11, 14, and 15 show low training and validation losses, indicating strong data fit and generalization. While in FCN, with the addition of Band 08, the training loss decreased, the validation loss remained relatively high and in some epochs, there were significant fluctuations. However, the last epoch had a lower loss than the first test.



(a)

(b)

Fig. 9. Model FCN (a) with Band 11, 14, 15; (b) with Band 08, 11, 14, 15.



(a)

(b)

Fig. 10. Model Ensemble (DeepLabV3, PSPNET, U-Net) deep learning (a) with Band 11, 14, 15; (b) with Band 08, 11, 14, 15.

Based on Fig. 10, the graph of the average training loss of the model and the average validation loss against the epoch of the Ensemble model of both tests shows that the loss decreases steadily as the number of epochs increases. However, the Ensemble model showed the first decrease in the first test with Bands 11, 14, and 15. The test comparison results can be seen in Table V.

TABLE V. COMPARISON RESULTS

Model	Backbone	Input Band	Train Avg Loss	Val Avg Loss
DeepLabV3	Resnet101	3	0.987700	0.996298
		4	0.987538	0.996168
PSPNET	Resnet101	3	0.944208	0.983661
		4	0.959346	0.988406
U-Net	Resnet101	3	0.713500	0.713250
		4	0.716402	0.716033
FCN	Resnet101	3	0.034756	0.014825
		4	0.032591	0.013321
Ensemble (DeepLabV3, PSPNET, U-Net)	Resnet101	3	0.693173	0.69313
		4	0.693234	0.693248

Based on Table V and Figs. 6−10, the FCNN Model demonstrates superior performance among all models, exhibiting the lowest average loss for training and validation, irrespective of the number of input channels. This indicates that it is highly efficient in fitting the data effectively and generalizing well to new instances. In contrast, the DeepLabV3 and PSPNET models display the poorest performance, with the highest average loss for training and validation. This suggests they tend to overfit and cannot adapt to new data.

The U-Net model performs moderately, showing a lower average loss than DeepLabV3 and PSPNET but a higher average loss than FCN for training and validation implying a degree of overfitting that is less severe than observed in DeepLabV3 and PSPNET. The Ensemble Deep Learning model, which integrates DeepLabV3, PSPNET, and U-Net, performs similarly to U-Net, with a slightly lower average loss for training but a slightly higher average loss for validation, indicating a moderate level of overfitting compared to DeepLabV3 and PSPNET.

Considering the information in Table V, the number of input channels does not significantly impact the performance of any model, as the average loss values remain very close for both three (Band 11, 14, 15) and four input channels (Band 08, 11, 14, 15). Adding Band 08 as extra channels does not substantially enhance or degrade the model's capacity to learn from the data.

When evaluating these differences, it is important to consider various factors outlined in Table II, such as model architecture, optimization techniques, learning rates, epochs, and neural network structure. The variation in these elements contributes to the observed performance differences among the models. Notably, the number of parameters trained in each model also plays a crucial role, with FCN efficiently utilizing fewer parameters while achieving superior performance. Examining training comparison results in Table V further highlights the consistent superiority of FCN, emphasizing its robust ability to fit data, and generalize effectively. Conversely, the higher average losses of DeepLabV3 and PSPNET underscore their challenges in overcoming overfitting and adapting to new data.

The moderate performance of U-Net and the Ensemble model's similarities to U-Net suggest a balance between fitting the data and avoiding overfitting. Lastly, the impact of learning rates and optimizer choices on training dynamics is crucial, with the higher learning rate of 0.08 for FCN potentially contributing to its ability to escape local minima and converge to optimal solutions, in contrast to the shared usage of Stochastic Gradient Descent (SGD) optimizer with a learning rate of 0.001 for DeepLabV3, PSPNET, and U-Net, which may contribute to their tendencies towards overfitting.

Interesting future research could delve deeper into aspects such as model selection, model configuration, and selection of other bands used in the image, testing on diverse data sets, and applying advanced techniques to reduce overfitting.

V. CONCLUSION

Based on the results of this study, it was found that the selection of the deep learning model proved to have a significant impact on the contrail detection performance in satellite images. However, the addition of the input channel, band 08 in ash color, did not significantly impact the model performance. The comparison results show that the FCNN model emerged as the best model with the lowest average loss in training and validation in both input channel tests. On the other hand, models like DeepLabV3 and PSPNET suffered from overfitting, which negatively impacted their ability to process unseen data. The U-Net model also showed good results with little overfitting. As a suggestion, this study highlights the importance of further exploration regarding model selection, configuration, and the choice of bands used in the imagery. It is also possible to explore additional factors that affect image segmentation, test on more diverse datasets, and apply techniques such as cross-validation or others to models that tend to overfit. The results of this study are expected to provide insight for policymakers and researchers in developing strategies to minimize climate impacts associated with aviation.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Alam Rahmatulloh conceived and designed the research, as well as contributed to writing the manuscript. Virra Retnowati A'izzah collected data, conducted experiments, analyzed data, and drafted the initial version of the manuscript. Irfan Darmawan contributed to writing the paper and conducted data analysis. Muhammad Al-Husaini contributed to writing the manuscript and provided critical feedback. All authors have approved the final version.

ACKNOWLEDGMENT

The authors express their gratitude to the professors for their invaluable comments and assistance, which greatly contributed to the development and completion of this research.

REFERENCES

[1] Intergovernmental Panel on Climate Change (IPCC), Ed. "Transport," in *Climate Change 2022—Mitigation of Climate Change: Working Group III Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, 2023, pp. 1049–1160. doi: 10.1017/9781009157926.012

[2] R. Teoh, U. Schumann, and M. E. J. Stettler, "Beyond contrail avoidance: Efficacy of flight altitude changes to minimise contrail climate forcing," *Aerospace*, vol. 7, no. 9, 2020. doi: 10.3390/aerospace7090121

[3] Project contrails—Google research. [Online]. Available: https://sites.research.google/contrails/

[4] J. Ren and Y. Wang, "Overview of object detection algorithms using convolutional neural networks," *Journal of Computer and Communications*, vol. 10, no. 1, pp. 115–132, 2022.

[5] T. Hoeser and C. Kuenzer, "Object detection and image segmentation with deep learning on earth observation data: A review—Part I: Evolution and recent trends," *Remote Sensing*, vol. 12, no. 10, 2020. doi: 10.3390/rs12101667

[6] Y. Gu, J. Hao, B. Chen, and H. Deng, "Top-down pyramid fusion network for high-resolution remote sensing semantic segmentation," *Remote Sensing*, vol. 13, no. 20, 2021. doi: 10.3390/rs13204159

[7] R. A. Emek and N. Demir, "Building detection from SAR images using U-Net deep learning method," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 44, pp. 215–218, 2020. doi: 10.5194/isprs-archives-XLIV-4-W3-2020-215-2020

[8] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," arXiv reprint, arXiv:1605.06211, May 20, 2016.

[9] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object detection via region-based fully convolutional networks," arXiv reprint, arXiv:1605.06409, Jun. 21, 2016.

[10] F. Sultana, A. Sufian, and P. Dutta, "Evolution of image segmentation using deep convolutional neural network: A survey," *Knowledge-Based Systems*, vol. 201–202, 106062, Aug. 2020. doi: 10.1016/j.knosys.2020.106062

[11] DeepLabv3 and DeepLabv3+ The Ultimate PyTorch Guide. (December 2023). *Learn Open CV*. [Online]. Available: https://learnopencv.com/deeplabv3-ultimate-guide/

[12] W. Chen *et al.*, "An overview on visual SLAM: From tradition to semantic," *Remote Sensing*, vol. 14, no. 13, Jan. 2022. doi: 10.3390/rs14133010

[13] A. V. Kakade, S. Rajkuma, K. Suganthi, and L. Ramanathan, "Object detection in satellite images using modified pyramid scene parsing networks," *Sensor Data Analysis and Management: The Role of Deep Learning*, pp. 147–160, 2021. doi: 10.1002/9781119682806.ch9

[14] Geostationary satellites. (November 2023). *NESDIS*. [Online]. Available: https://www.nesdis.noaa.gov/current-satellite-missions/currently-flying/geostationary-satellites

[15] V. R. Meijer *et al.*, "Contrail coverage over the United States before and during the COVID-19 pandemic," *Environ. Res. Lett.*, vol. 17, no. 3, 034039, 2022. doi: 10.1088/1748-9326/ac26f0

[16] R. Kinkel. (2023). Google research contrail detection: A Strategic approach to satellite image analysis. [Online]. Available: https://www.linkedin.com/pulse/google-research-contrail-detection-strategic-approach-ralf-kinkel/

[17] A. Bhandari, S. Rallabandi, S. Singhal, A. Kasliwal, and P. Seth, "Performance evaluation of deep segmentation models for contrails detection," arXiv reprint, arXiv:2211.14851, Nov. 04, 2023.

[18] J. P. Hoffman, T. F. Rahmes, A. J. Wimmers, and W. F. Feltz, "The application of a convolutional neural network for the detection of contrails in satellite imagery," *Remote Sensing*, vol. 15, no. 11, Jan. 2023. doi: 10.3390/rs15112854

[19] B. H. Agung, A. Faizal, A. Anggi, A. S. Bahri, and W. Utama, "Application of Landsat 8 satellite imagery to identify geothermal prospect areas in the Songgoriti Batu area and its surroundings," *Geoscientific Journal*, vol. 3, no. 3, Dec. 2017. doi: 10.12962/j25023659.v3i3.3212 (in Indonesia)

[20] Z. Zheng, Y. Liu, M. He, D. Chen, L. Sun, and F. Zhu, "Effective band selection of hyperspectral image by an attention mechanism-based convolutional network," *RSC Adv.*, vol. 12, no. 14, pp. 8750–8759, Mar. 2022. doi: 10.1039/D1RA07662K

[21] N. Siddiqui, "Atmospheric contrail detection with a deep learning algorithm," *Scholarly Horizons: University of Minnesota, Morris Undergraduate Journal*, vol. 7, no. 1, Jul. 2020. doi: 10.61366/2576-2176.1087

[22] G. Zhang, J. Zhang, and J. Shang, "Contrail recognition with convolutional neural network and contrail parameterizations evaluation," *SOLA*, vol. 14, pp. 132–137, 2018. doi: 10.2151/sola.2018-023

[23] J. Ng, C. Elkin, A. Sarna, W. Reade, and M. Demkin. (May 2023). Google research—Identify contrails to reduce global warming. *Kaggle*. [Online]. Available: https://kaggle.com/competitions/google-research-identify-contrails-reduce-global-warming

[24] J. Y.-H. Ng *et al.*, "OpenContrails: Benchmarking contrail detection on GOES-16 ABI," arXiv reprint, arXiv:2304.02122, 2023.

[25] Geostationary operational environmental satellites—R series. NOAA/NASA. [Online]. Available: https://www.goes-r.gov/

[26] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, pp. 3523–3542, Jul. 2022. doi: 10.1109/TPAMI.2021.3059968

[27] A. Signoroni, M. Savardi, A. Baronio, and S. Benini, "Deep learning meets hyperspectral image analysis: A multidisciplinary review," *Journal of Imaging*, vol. 5, no. 5, 2019. doi: 10.3390/jimaging5050052

[28] Y. Zhang, S. Liu, C. Li, and J. Wang, "Rethinking the dice loss for deep learning lesion segmentation in medical images," *J. Shanghai Jiaotong Univ. (Sci.)*, vol. 26, no. 1, pp. 93–102, 2021. doi: 10.1007/s12204-021-2264-x

[29] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Proc. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, DLMIA ML-CDS 2017*, 2017, vol. 10553, pp. 240–248. doi: 10.1007/978-3-319-67558-9_28

[30] M. Wojciuk, Z. Swiderska-Chadaj, K. Siwek, and A. Gertych. (April 2022). The role of hyperparameter optimization in fine-tuning of CNN models. *SSRN*. [Online]. Available: https://ssrn.com/abstract=4087642

[31] X. Zhang, X. Chen, L. Yao, C. Ge, and M. Dong, "Deep neural network hyperparameter optimization with orthogonal array tuning," arXiv reprint, arXiv:1907.13359, Feb. 2020.