2023 26th International Conference on Computer and Information Technology (ICCIT)
13-15 December, Cox's Bazar, Bangladesh

Contrail Analysis through Advanced Neural Network Architectures: Image Segmentation and Classification

Ashraful Alam Nirob, Shahriar Ahmed, Tahmidul Karim Takee, Adnan Rahman Eshan, Shadik Ul Haque, Ehsanur Rahman Rhythm and Annajiat Alim Rasel
Department of Computer Science and Engineering (CSE)
School of Data and Sciences (SDS)
Brac University
66 Mohakhali, Dhaka - 1212, Bangladesh
{ashraful.alam.nirob, shahriar.ahmed, tahmidul.karim.takee, adnan.rahman.eshan, shadik.ul.haque, ehsanur.rahman.rhythm}@g.bracu.ac.bd, annajiat@gmail.com

***Abstract*—The aviation industry's immense expansion is having an impact on global warming and has resulted in some significant environmental issues. When an airplane passes directly over them, tiny crisscross patterns, often known as contrails, may be visible. They are to blame for this effect. Contrails are really just airborne particles that have been compressed with water. They are uncommon since ice can only form under particular climatic circumstances, such as extremely cold, hot, humid, and saturating air. Even worse, because of the cooler climate at night, it is more dangerous because it has more time to live. They gather heat from the sun and store it, then release it into the atmosphere. Some experts have also warned the public that the radiation these contrails produce may be more damaging to the atmosphere than previously predicted. For this reason, scientists are looking for methods to reduce these contrails by comprehending their behaviors and patterns. Now, the proposed study segments and classifies images of contrails acquired from satellite data. In this study, complex neural network architectures, including U-Net, DeepLab, Attention Mechanism, and ResNet-50 with CNN, are used to segment and binary classify those photos. These architectural frameworks will aid this research in effectively classifying and segmenting those contrails from the satellite images so that further research can comprehend and observe their patterns and behaviors.**

***Index Terms*—Contrails, U-Net, DeepLab, ResNet-50, CNN, Aviation, Global Warming**

I. Introduction

Modern science and technology have exacerbated environmental issues, including climate change. Commercial air travel interconnects people worldwide but also raises environmental concerns such as condensation trails from aircraft, known as contrails. These were first scientifically observed in the 1920s [1]. Contrails comprise icy atmospheric air mixed with hot aircraft engine exhaust, developing at high altitudes in cold, humid atmospheric conditions. Sunlight passing through persistent contrails can seriously impact climate by retaining heat, altering weather patterns, and disturbing atmospheric dynamics. Individual contrail lifecycles differ substantially based on ambient factors like temperature and humidity. Atmospheric science experts advise reducing overall contrail formation and iveriexposure to mitigate their negative climatic effects. This study utilizes advanced neural network architectures, including U-Net, DeepLab, Attention Mechanism, and ResNet-50 with CNN to accurately segment and classify contrails from an extensive dataset of high-resolution multi-spectral satellite imagery. Precisely analyzing contrail distribution patterns and atmospheric behaviors will assist future research to better understand and potentially devise technological or operational solutions to curb their disproportionate contribution to climate change. Classifying and segmenting contrails forms an integral foundation to enable this critical environmental pattern analysis.

II. Literature Review

The paper by Hoffman et al. [2] first discussed contrails, aviation-induced cirrus clouds formed by aircraft emissions in saturated atmospheric conditions. Contrails impact climate through radiation forcing, which depends 35–81% on earth's opacity and radiation intensity. As contrails rapidly fluctuate in shape and volume, they are challenging to identify in satellite imagery. The research focuses on a contrail identification technique using contrail photos as a training, testing, and validation dataset. The strategy utilizes U-Net for image segmentation on photo temperature difference images with 11 x 12 x 11 channels, based on the same artificial intelligence technology used to quantify lead in sea ice. The 10,000 image dataset had a 70%:20%:10% cross-contamination ratio. The U-Net model can accurately detect contrails in images.

Image segmentation has bimodal effects with maximal solutions at 0 or near 250. With a conservative detection threshold, metrics are satisfactory. This technique enables continuous day/night coverage every 10 minutes over the Western Hemisphere at 2 km resolution. The GOES ABI Cloud Top Height ("ACHA") product sorts contrails by color, initially introduced to match the 2 km resolution. "ACHA" height is converted to feet for aviation users, providing the best peak retrieval within each contrail. Height is given in kilometers.

The GOES contrail detection case indicates that contrails are barely distinguishable in true color imagery but are conveniently apparent in thermal IR. The detection technique uses strategies that are powerful in satellite remote sensing packages, and it's easier to run. However, non-cloud features with similar spatial characteristics and brightness temperature differences can cause false detections. Contrail detections are often longer and wider than those detected in cloud masks or peak products. Future paintings can even include atmospheric situations, flight paths, and industrial air visitors to provide extra self assurance in contrail detection.

For different models of image segmentation, various papers and journals have been reviewed. Among those,the paper entitled "U-Net: Convolutional Networks for Biomedical Image Segmentation" by Olaf Ronneberger et al. [3] published in 2015, shows the challenges of using deep learning techniques in the case of biomedical image segmentation. U-Net architecture, a novel convolutional neural network (CNN) design, and these two methods have been used in this paper for doing image segmentation. The architecture's unique design and its application to different perspectives of biomedical imaging have accelerated subsequent research in various image segmentation tasks.

According to the authors, traditional image segmentation methods often relied on handcrafted features, which struggled to capture complex spatial relationships and variations in biomedical images. In this regard, deep learning, particularly convolutional neural networks, showed promise in automatically learning time relevant features for image analysis, although adaptation of these methods to biomedical images did require addressing the specific challenges posed by these data.

The U-Net architecture utilizes an innovative encoder-decoder structure. The contracting path applies convolutional and max-pooling layers to reduce spatial dimensions while capturing contextual and high-level visual features. The expansive path then upsamples the compressed feature maps back to the original image resolution using transposed convolutions, gradually recovering the lost spatial information. Critically, skip connections between corresponding layers in the contracting and expansive paths incorporate both fine-grained local and high-level global features. This dual incorporation across scale enables U-Net to produce accurate pixel-level image segmentations. These key architectural innovations laid the groundwork for U-Net's subsequent widespread success.

The article by Xiea et al. explores attention mechanisms in medical image segmentation [4]. Segmentation distinguishes anatomical or pathological structures in medical images like CT, PET, MRI, X-ray and microscopy. The authors first reviewed core principles and formulations of attention mechanisms. Through surveying over 300 articles, they analyzed the effects of non-Transformer and Transformer attention in segmentation. Non-Transformer attention works by refining feature maps while Transformer attention models long-range global contexts. Both prove effective for medical images with complex textures, shapes and pixel-level details. Further mechanisms and integrations with neural networks may improve segmentation performance.

The authors define attention mechanisms as methods to weight the importance of different image regions. They surveyed 300 papers and grouped them into non-Transformer and Transformer attention categories. Non-Transformer attention refines feature maps to process critical information and is divided into channel, spatial, and temporal types. Channel attention reweights network channels representing different objects. Spatial attention identifies important image regions by scoring locations. Temporal attention selects informative timeframes from temporal data. Transformer attention models long-range dependencies and global context. It can form encoder-decoder networks where the encoder compresses inputs into a latent representation decoded into outputs. For medical images with complex shapes and textures, both attention types improve segmentation, though spatial attention is more popular for delineating blurry boundaries. Experiments showed spatial attention worked well for segmenting polygons, skin lesions, prostates, etc. Attention mechanisms show promise for medical image segmentation, with further development of architectures and integration with neural networks potentially improving performance.

The next part of the literature review contains another study by CNN on satellite imagery by Guoyu Zhang et al. which gives a brief implementation of contraill classification, which was done by using the Himawari-8 stationary satellite [5]. This process outperforms the conventional algorithms that we use. Contrails are vapor from airplane engine fuels that affect the earth's energy balance and are responsible for climate change. The study was done in the South-China region. The result shows a correlation between contrail occurrence, persistence, and potential contrail coverage. ContrailMod, a CNN detection model, is utilized with the Himawari-8 geostationary satellite to identify contrails, which are highly dependent on atmospheric temperature and humidity conditions. Contrails induce considerable radiative forcing that exacerbates climate change more than aviation's carbon dioxide emissions alone. Leveraging the extensive ECMWF global atmospheric reanalysis dataset since 1979, ContrailMod estimates historical contrail coverage over South China, further advancing scientific understanding of contrail recognition and quantifying the relationship between persistent contrails and aviation's effect on climate change.

As several datasets have been discussed for the analysis of contrail detection, another study that was useful was the study by Hermann Manstein et al. [6] in which the Advanced Very High Resolution Radiometer (AVHRR)'s infrared channels are used to demonstrate an automated contrail detection method. Given the difficulties of identifying contrails using infrared data, this algorithm was created to reduce false detections while maintaining a low contrail detection rate.

The objective of the study is to examine contrail cloudiness patterns in different parts of Western and Central Europe. Contrail coverage values over those two parts were roughly

0.5%, with regional variations reaching a maximum of 1.2%, according to the authors' analysis of daily AVHRR data from 1996.

Not only that, but the contrast between daytime and night-time contrail coverage is also highlighted in the paper. It was discovered that the coverage of contrails at night, when greenhouse forcing may be more strongly affected, was roughly one-third that seen during the day.

Furthermore, in order to identify contrails in the data from the Advanced Very High Resolution Radiometer (AVHRR), the paper introduces a sophisticated algorithm. This algorithm is the most recent in a series of advancements made to improve contrail detection.

The new algorithm makes use of a number of tests to reduce the possibility of false contrail detection. By avoiding scene-specific operators, for example, the Hough transform, and normalizing data mostly on the scale of a region, it highlights the significance of achieving independence from the characteristics of specific scenes [7]. The dataset AVHRR has been previously preprocessed using the APOLLO system [8]. These data consist of high-resolution images with dimensions of 1440 by 2048 pixels.

The efficiency of the algorithm's application is one noteworthy feature. On an Ultra Sparc 2 machine, the contrail detection procedure currently takes less than 30 minutes to complete for an image of this size. Additionally, the algorithm has been continuously operating since January 1996, with a focus on the noontime overpass of the NOAA-14 satellite. Moreover,this study's contrail detection algorithm shows itself to be extremely resistant to incorrectly identifying surface features. Only 25 of the total image's pixels produce counts of more than 10, which shows that very few pixels are vulnerable to misclassification. These contrail-detected pixels are regarded as being completely obscured by contrails.

This results in an annual average local contrail coverage of about 3% in the study area, which is a small 25 km2 region. The highest count recorded is 12 (out of a possible 357), which equates to about 4% of local coverage and was recorded above Hungary's Lake Balaton. Along the elongated lake, this specific pixel and a few other outliers form a recognizable pattern. The substantial air traffic in the Balaton region is considered in the algorithm. Then, these data are smoothed using a circular Gaussian kernel with a 50-pixel radius (roughly 50 km full width and half maximum).

Using information from AVHRR channels 4 and 5, the algorithm discussed in this paper shows its capacity for the quick and practical detection of persistent contrails with linear shapes. This pattern recognition-based method offers adaptability to other data types and makes it easier to compare results across various datasets.

III. Methodology

The field of remote sensing has witnessed remarkable advancements with the proliferation of satellite imagery, offering valuable insights into various atmospheric phenomena. One such phenomenon of critical environmental and scientific interest is the formation of contrails—thin cloud-like trails that result from the interaction of aircraft engines with the upper atmosphere. Accurate identification and segmentation of contrails within satellite imagery hold significant potential for climate research, aviation studies, and environmental impact assessment.

In pursuit of enhancing contrail analysis, this paper delves into the application of advanced neural network architectures for image segmentation and classification. The central objective is to ascertain which image segmentation model among U-Net, Attention Mechanism, RestNet50, and DeepLab demonstrates superior efficacy in detecting contrails within a vast satellite imagery dataset.

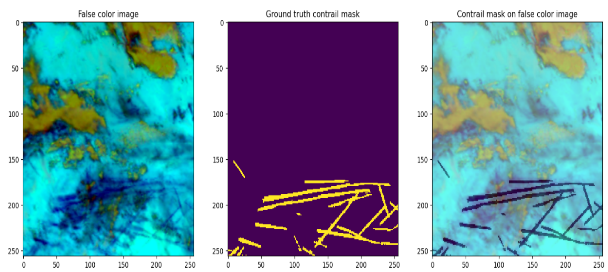

Fig. 1: Data Visualization

A. Data Preparation

The success of any machine learning endeavor heavily relies on the quality and suitability of the dataset. In this study, we utilize a geostationary satellite image dataset obtained from the GOES-16 Advanced Baseline Imager (ABI), which is renowned for its comprehensive coverage and temporal granularity. The dataset is specially curated for contrail detection and contains sequences of images captured at 10-minute intervals, enabling the exploitation of temporal context for improved accuracy.

To facilitate model training and evaluation, several data preprocessing steps are applied:

1) Normalization: The images are normalized to the range $[0, 1]$ using the `normalize_range` function. This normalization procedure ensures that all image data falls within a consistent numerical range, contributing to stable model convergence.

2) False Color Composition: A vital component of the data preparation process is the creation of false color images. These images are synthesized by combining brightness temperatures from distinct bands (`band_15`, `band_14`, and `band_11`) using predetermined bounds (`_T11_BOUNDS`, `_CLOUD_TOP_TDIFF_BOUNDS`, and `_TDIFF_BOUNDS`). The resulting false color images offer enhanced visual cues for identifying contrails.

B. Model Architectures

The U-Net architecture is designed specifically for biomedical image segmentation, comprised of contracting and expansive paths to capture context and precise localization, respectively [9]. This architecture has proven effective across various segmentation tasks, including with time series image data [10]. Mask R-CNN extends Faster R-CNN for parallel prediction of object masks alongside bounding boxes [11], simultaneously detecting and segmenting instances. Attention mechanisms guide models to focus on the most relevant input regions and features [12], [13]. DeepLab utilizes dilated convolutions for semantic segmentation, refines output with conditional random fields [14], and shows promise in detecting contrails by training directly on satellite imagery rather than hand-engineered features [15]. DeepLab's segmentation capabilities enable estimating contrail coverage over large areas [14].

1) U-Net for Contrail Segmentation: We use a U-Net architecture to tackle the challenge of contrail segmentation inside satellite pictures. A well-known deep convolutional neural network architecture that is excellent at segmenting images is called U-Net.

An encoder and associated decoder make up the architecture. The encoder progressively downsamples the input image while capturing relevant features. Conversely, the decoder aims to upsample the encoded features to generate a segmentation mask.

We introduce several key components within the U-Net architecture:

a) Convolutional Layers: Convolutional layers are pivotal in extracting hierarchical features from the input data. We use a series of convulational layers, with each layer being followed by batch normalization and an activation function called rectified unit (ReLU). The function `build_convolutional_layers` encapsulates this process, and the resulting feature maps are vital for accurate contrail segmentation.

b) Max-Pooling and Upsampling Layers: Max-pooling layers facilitate spatial downsampling in the encoder, while upsampling layers recover the spatial resolution in the decoder. The function `build_maxpooling_layer` encapsulates max-pooling, and `build_upsampling_layer` encapsulates upsampling.

c) Skip Connections: To enhance feature propagation between the encoder and decoder, we introduce skip connections. These connections enable the decoder to leverage features from earlier layers, aiding in capturing finer details. The function `build_skip_connection` implements this mechanism.

d) Classification Layer: The final layer of the decoder serves as the classification layer. This layer is responsible for generating the binary segmentation mask, indicating contrail presence. The `build_classification_layer` function generates the classification output using a sigmoid activation function.

The complete U-Net architecture is constructed by connecting these components in a symmetrical manner. In this study, we incorporate 10 layers in both the encoder and decoder sides of the U-Net.

The model is trained using the Adam optimizer and binary cross-entropy loss, given by the equation [16]:

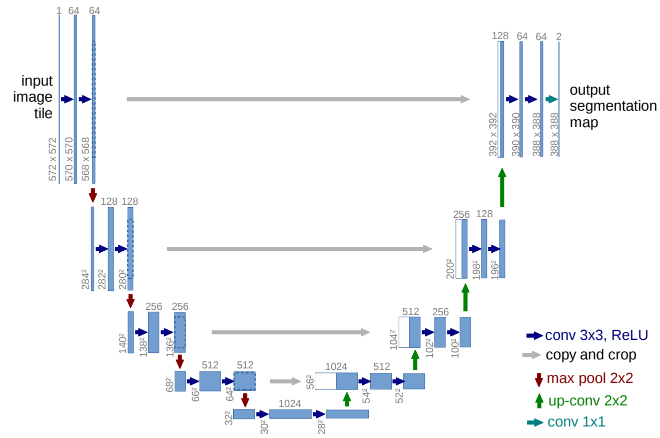

Fig. 2: U-net architecture (for 32x32 pixels in the lowest resolution) [9]

$$-\frac{1}{N}\sum_{i=1}^{N} y_i \cdot \log(p(y_i)) + (1-y_i)\cdot \log(1-p(y_i)) \qquad (1)$$

The trained model is assessed based on both accuracy and other suitable evaluation metrics, providing insight into its effectiveness in contrail segmentation within the satellite imagery dataset.

2) ResNet50 Model: Our model utilizes a ResNet50 encoder for feature extraction followed by a custom upsampling decoder for semantic segmentation. To forecast pixel-wise segmentation masks, the model is trained from beginning to end.

a) Encoder: We employ a pretrained ResNet50 network as the encoder backbone. The weights are initialized from imagenet classification training to leverage learned feature representations. To match the target image dimensions, we modify the first few layers of ResNet50:

- First 7x7 convolution replaced by 3x3 conv, stride 2
- First max pool layer removed
- Second 3x3 convolution's stride increased from 1 to 2

This adjusts the downsample rate to reach an output stride of 16. The later encoder features are taken from layers conv3 onwards of ResNet50.

b) Identity Mappings: A key aspect of ResNet is the use of identity mappings which allow the signal to skip layers and improve information propagation. We leverage identity connections in the encoder which helps training deeper models.

c) Bottleneck Modules: The ResNet50 encoder uses bottleneck modules which contain 1x1, 3x3, and 1x1 convolutions. The 1x1 layers reduce and restore dimensions while computing the 3x3 convolution on a lower-dimensional signal. This improves computational efficiency.

d) Decoder: The decoder module upsamples the encoder output to recover spatial resolution for dense prediction. We use transpose convolution layers to upsample by a factor of 2, concatenating with earlier encoder features to incorporate fine details. After upsampling we apply two 3x3 convolutions for smoothing.

3) Attention Mechanism Model: In this part, we outline the approach for our proposed Attention Mechanism architecture, which makes use of an encoder, decoder, bottleneck, positional encoding architecture enhanced with attention modules to perform semantic segmentation.

a) Encoder: Hierarchical feature representations from the input image are extracted by the encoder module. It comprises of three encoder blocks, each of which downsamples the spatial dimensions by two 3x3 convolutions and two 2x2 max pooling.

b) Bottleneck: The bottleneck module is a crucial part of the network. It employs dilated convolutions to capture wider contexts without further downsampling. The output of the bottleneck module serves as the final encoder output, summarized across spatial dimensions.

c) Decoder: The decoder module is tasked with recovering the spatial resolution using upsampling and incorporates skip connections from the encoder. We have implemented three decoder blocks, each featuring 2x upsampling using transpose convolutions followed by concatenation with the corresponding encoder features.

d) Attention Mechanism: To focus on relevant image regions and enhance localization, we have incorporated attention mechanisms within the decoder module. The attention block computes correlations between encoder and decoder features, generating an attention map. This map is applied to the encoder output, allowing the model to derive pertinent regions to pass to the decoder.

e) Positional Encoding: Given that convolutional networks lack an inherent notion of position, we inject positional information using fixed sinusoidal encodings. This augmentation enables the model to take spatial relationships into account.

The attention-based network effectively aggregates spatial contexts while emphasizing salient image regions. Model training is supervised end-to-end using ground truth segmentation masks.

4) DeepLab for Contrail Segmentation: In this section, we go over the process for creating a deep convolutional neural network that can segment contrail images based on their semantic content. Our model's architecture resembles that of DeepLab, and it has been thoroughly trained to carry out pixel-wise categorization.

a) Encoder: The encoder part of the network is in charge of extracting features from the input image. In order to gradually downsample the spatial dimensions, max pooling is done after every two blocks and five blocks of convolutional layers, each followed by batch normalization. Two 3x3 convolution layers with 256 filters and ReLU activation make up each convolutional block. After max pooling, downsampling by a factor of two is carried out to lessen spatial resolution.

b) Decoder: The decoder module's purpose is to regain the spatial resolution by upsampling the encoder characteristics. To collect small details, it uses skip connections from the encoder. The decoder's five upsampling sections mirror the architecture of the encoder. The feature maps are upsampled using 2D transpose convolution layers. The appropriate encoder block activations and the upsampled features are concatenated, and additional convolution layers are used to further enhance the segmentation.

c) Atrous Convolutions: We used atrous (dilated) convolutions in our model to broaden the field of view and capture multi-scale context. To capture more context without drastically increasing the number of parameters, atrous convolutions apply convolution with upsampled filters. We incorporate atrous convolutions with various dilation rates into the encoder pipeline.

d) ASPP Module: Context is recorded at various scales using the Atrous Spatial Pyramid Pooling (ASPP) module. A parallel dilated convolution with various dilation rates is used to concatenate the produced features. The collecting of multi-scale contextual data is made possible by this method. Just before entering the decoder, we add an ASPP module on top of the encoder's features.

IV. Results

In our study, we aimed to achieve significant improvements in the performance of semantic segmentation models, with a particular focus on the U-Net model. Due to resource constraints and the need to reduce training time, we had to limit the size of our training and test datasets to 56X56 dimnesion, which inevitably had an impact on our results.

Despite the challenges posed by the reduced dataset size, the U-Net model outperformed all other models significantly. U-Net, which was our primary focus, exhibited a commendable performance, even in this constrained setting. This underscores the robustness of the U-Net architecture and its ability to handle such challenges effectively.

It is worth noting that the other models utilized in our study are also semantic segmentation models, tailored for identifying contrails using a binary classification approach. While U-Net demonstrated superior performance, DeepLab also exceeded our expectations with its segmentation results.

TABLE I: Evaluation Metrics for Different Models

Models	IoU Score	Dice Coefficient	Accuracy
U-Net	0.373	0.651	92%
DeepLab	0.138	0.154	76%
Attention	0.012	0.029	71%
ResNet50	0.112	0.218	81%

Analyzing the pixel-level similarity across all models, we observed a common strength in accurately identifying black pixels, indicating an effective detection of contrail segments. However, it is important to note that there were instances of misidentification of white pixels in the final segmentation results of all models. This suggests that there is room for further improvement, particularly in the differentiation of white pixels within the contrail regions.

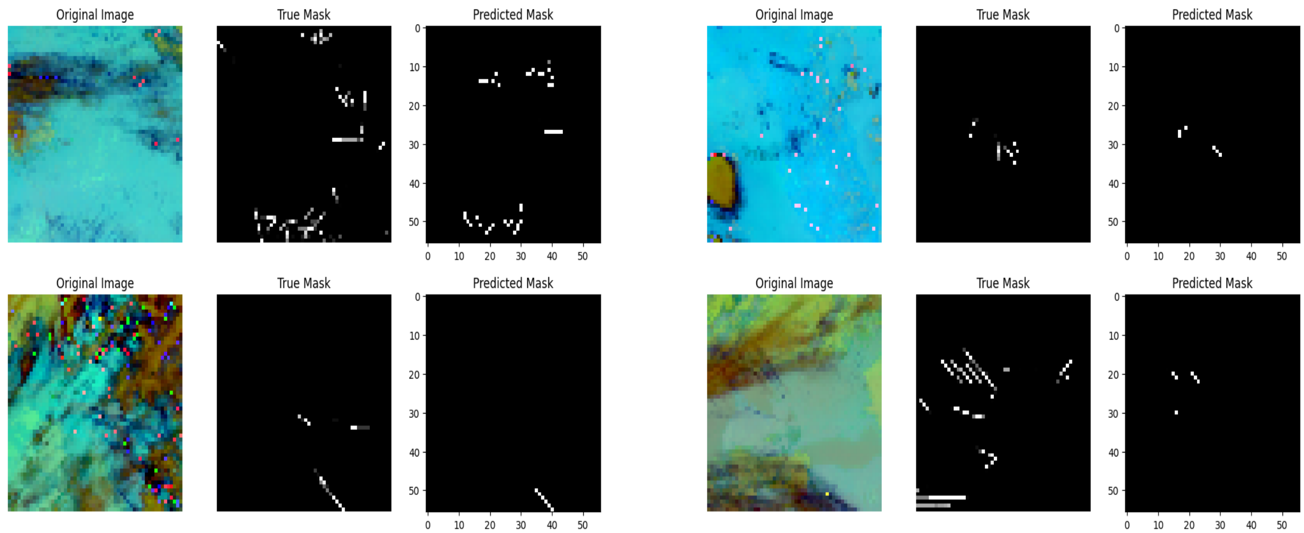

Fig. 3: Examples of Predicted Mask from Original Image

To quantify our models' performance, we calculated the Intersection over Union (IoU) scores. The U-Net model achieved the highest IoU score, reflecting its superior ability to accurately segment contrails. DeepLab also delivered a respectable IoU score, further validating its effectiveness.

V. Conclusion

The study shows a thorough evaluation and comparison of various neural network models for detecting contrails from satellite images, including U-Net, Attention Mechanism, Deep Lab and RestNet 50. In terms of identifying, segmenting, and classifying those contrails from the images, it demonstrates how accurate these models are. Due to the fact that U-Net is essentially an image segmentation model and uses individual pixels of an image to mask contrails, it is abundantly clear from evaluating the results that it has outperformed all other models in terms of accurately segmenting contrails. Other models used in this study also performed well, but they were semantic segmentation models that employed boundary line detection rather than U-Net. Because of this, even though they met expectations, DeepLab did not perform as well as U-Net. Overall, this study will significantly aid in the detection of contrails due to its precision and the insightful information it can offer.

References

[1] U. Schumann, "Formation, properties and climatic effects of contrails," *Comptes Rendus Physique*, vol. 6, no. 4-5, pp. 549–565, 2005.

[2] 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, p. 2854, 2023.

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.

[4] Y. Xie, B. Yang, Q. Guan, J. Zhang, Q. Wu, and Y. Xia, "Attention mechanisms in medical image segmentation: A survey," *arXiv preprint arXiv:2305.17937*, 2023.

[5] G. Zhang, J. Zhang, and J. Shang, "Contrail recognition with convolutional neural network and contrail parameterizations evaluation," *SOLA*, vol. 14, pp. 132–137, 2018.

[6] H. Mannstein, R. Meyer, and P. Wendling, "Operational detection of contrails from noaa-avhrr-data," *International Journal of Remote Sensing*, vol. 20, no. 8, pp. 1641–1660, 1999.

[7] W. K. Pratt, *Digital image processing: PIKS Scientific inside*. Wiley Online Library, 2007, vol. 4.

[8] K. Kriebel, R. Saunders, and G. Gesell, "Optical properties of clouds derived from fully cloudy avhrr pixels," *Beiträge zur Physik der Atmosphäre*, vol. 62, pp. 165–171, 1989.

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.

[10] M. Perslev, M. Jensen, S. Darkner, P. J. Jennum, and C. Igel, "U-time: A fully convolutional network for time series segmentation applied to sleep staging," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[11] K. He, G. Gkioxari, P. Doll'ar, and R. Girshick, "Mask r-cnn," in *IEEE International Conference on Computer Vision (ICCV)*, 2017.

[12] L. Liang and N. Xiang, "Image captioning based on an improved attention mechanism," S. X. Yang and D. A. Karras, Eds., Jul. 2023, p. 127173B.

[13] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[14] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2018.

[15] I. Koren and G. Feingold, "Aeronet-based inversion of cloud-free aerosol scattering and absorption in the atmosphere for granulometric aerosol size separation," *Meteorology and Atmospheric Physics*, vol. 117, no. 1-2, pp. 33–49, 2011.

[16] N. J. Hardenbrook and P. Zhang, "A structural view of the sars-cov-2 virus and its assembly," *Current opinion in virology*, vol. 52, pp. 123–134, 2022.