Contextual Segmentation of Fire Spotting Regions Through Satellite-Augmented Autonomous Modular Sensor Imagery

Nikhil Behari*, Henry Holbrook† and Paris Garrett ‡
NASA Langley Research Center, Hampton, VA 23681

Corey Ippolito §
NASA Ames Research Center, Moffett Field, CA 94035

Chester Dolph ¶
NASA Langley Research Center, Hampton, VA 23681

Globally, forest fires remain a significant threat to human and environmental wellbeing. Towards mitigating the impacts of forest fires, it is critical that accurate and updated information regarding not only the fire line, but also nearby human settlements, vegetation, and water sources is reported quickly to emergency services. However, while existing UAS-based fire detection methods are effective, they largely do not report the contextual environmental information necessary to best serve nearby communities in disaster response. Additionally, modern advancements in deep learning offer new approaches for image segmentation which may improve classification accuracy beyond current pixel-wise indices. In this work, we benchmark the performance of these modern segmentation techniques in locating both fire lines and environmental features in historical Autonomous Modular Sensor imagery. Furthermore, we augment these outputs with satellite imagery segmentation towards developing a robust contextual mapping tool for rapid emergency fire response and decision making.

I. Nomenclature

AMS	=	Autonomous Modular Sensor
CNN	=	Convolutional Neural Network
FCN	=	Fully Convolutional Network
IR	=	Infrared
P, R, IoU	=	Precision, Recall, Intersection over Union
RCNN	=	Region-Based Convolutional Neural Network
RGB	=	Red, Green, Blue
TM	=	Thematic Mapper
UAS	=	Unmanned Aerial Systems
UAV	=	Unmanned Aerial Vehicle
VIIRS	=	Visible Infrared Imaging Radiometer Suite
A	=	predicted mask
A_{gt}	=	ground truth mask
FN	=	false negative
FP	=	false positive
TP	=	true positive

*Research Intern, NASA Langley Research Center, AIAA Member
†Research Intern, NASA Langley Research Center, AIAA Member
‡Research Intern, NASA Langley Research Center, AIAA Member
§Aerospace Scientist, NASA Ames Research Ceenter, AIAA Member
¶Aerospace Engineer, NASA Langley Research Center, AIAA Member

II. Introduction

Wildfires, defined as uncontrolled fires in natural areas, remain highly prevalent in the United States, with more than 59 thousand reported fires and 10 million acres burned in 2020 alone [1]. Ongoing climate change, in particular, has made these fires more unpredictable [2, 3], leading to devastating impacts in nearby communities and ecosystems [1, 4]. To best mitigate the impacts of uncontrolled forest fires and improve situational awareness, emergency response teams require information regarding not only the fire line, but also nearby human settlements, vegetation, and water sources [5, 6]. However, existing methods and data sources for active fire spotting are insufficient for providing all of these products in a timely and accurate manner. For example, satellite images may be readily available for situational mapping, but provide an outdated view of current land features. Alternatively, UAS-based fire detection methods are efficient and accurate, but largely do not report the contextual environmental information necessary to best serve nearby communities in disaster response [7]. In addition to these shortcomings, current approaches for segmentation of fire hotspots primarily use index-based techniques [8]. These techniques are effective, but could be further verified and validated using additional labels generated through recently-developed deep learning models, which have been applied to numerous remote sensing and vision tasks in prior literature [9]. In this work, we improve upon prior research in wildfire spotting by benchmarking the performance of modern deep learning segmentation techniques in locating both fire lines as well as environmental features in UAS-mounted AMS imagery [10]. This imagery, primarily capturing historical wildfires in California, is augmented with remote sensing products to train deep learning networks for wildfire and land feature segmentation, and the results of the analysis are used to develop a robust contextual mapping tool for rapid emergency fire response and decision making.

III. Related Work

Research in vision-based fire spotting broadly falls under either risk prediction or active monitoring methodologies [11]. Active UAS-based monitoring approaches, in particular, have been extensively studied with a variety of onboard sensor instruments [12]. Visible, IR, and multispectral sensors have been commonly utilized, each with unique benefits and constraints [13]. IR sensors, for example, can accurately detect heat, but are prone to false positives from sunlight reflection. Alternatively, multispectral sensors utilized in AMS imagery capture a comprehensive 16 spectral bands, but are limited in deployment by their physical weight. The choice of sensor typically guides the visual processing methods used to detect wildfires from collected data.

Several studies aim to develop methods for wildfire segmentation using UAV for IR surveillance. Martínez-de Dios et al. [14] develop an IR-based fire detection approach through pixel thresholding of high intensity regions, but find that the approach risks false positives from other heated objects within view. Yuan et al. [15] improve upon this IR-based approach through both histogram-based segmentation and optical flow methods to filter candidate fire regions. Alternatively, fire segmentation in UAV-mounted visible spectrum cameras has been studied using thresholding of RGB regions, which has similar issues in false positive rates [16]. Generally, these color segmentation methods, while effective in isolated settings, have similar limitations in deployment that limit their use. Most notably, they lack robustness across a variety of environments, such as the type of fire, the time of day, and the spatial movement, leading to false positives or poor generalizability [13, 17]. Additionally, these methods typically do not apply to contextual segmentation outside of the fire region of interest.

More recently, deep learning-based segmentation methods have been used to improve the robustness of fire spotting in unstructured settings. Akhloufi et al. [18], for example, propose a deep CNN for wildfire segmentation, while Yuan et al. [19] and Frizzi et al. [20] find accurate segmentation performance from developed networks for smoke segmentation, albeit in simpler image settings. Alternatively, Zhao et al. [21] compare performance of fire segmentation across various CNN models in UAV-mounted wildfire imagery, and achieve up to 98% validation accuracy. However, while these deep learning-based approaches show good results in fire and smoke segmentation, they largely fail to incorporate segmentation of other contextual environmental features. Studies in satellite-based segmentation, on the other hand, have focused on environmental features such as human development, vegetation, and waterways [22, 23], but no such study has aimed to benchmark performance of modern segmentation techniques in UAS imagery for use in wildfire spotting.

IV. Methodology

The proposed methodology detailed below was comprised of four main steps. The first step was to extract, resample, and composite the original AMS imagery to use with remote sensing imagery in the image segmentation architectures.

The second step was to generate this corresponding remote sensing imagery, which served as data labels for areas of active fire and contextual land features. Third, we implemented three popular deep image segmentation architectures, as well as two loss functions and three performance metrics to test the efficacy of various model configurations. Finally, we trained each model under a standard training configuration to generate performance results.

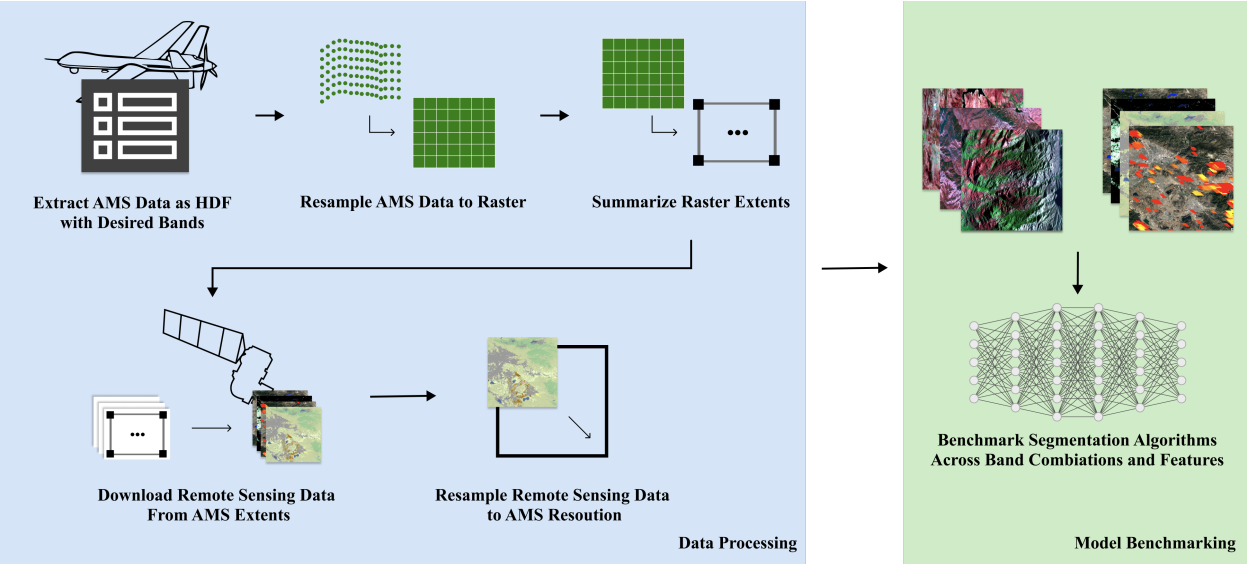

Fig. 1 Proposed methodology for contextual segmentation of fire spotting regions in AMS imagery.

A. AMS Imagery

Historical AMS data was used in this study to investigate the performance of deep learning models in segmenting wildfires in the western United States from aerial imagery. The AMS dataset consists of a total of 16 flights from 2006 to 2011, primarily covering wildfires in California and surrounding western states. Each flight image set within the AMS data consists of a number of separate and distinct flight lines, which together form the complete UAS flight path for the given fire. Flight lines are comprised of sequential readings from the AMS whiskbroom sensor, and may be subject to arbitrary directional and altitude changes along the data collection path. 13 such image sets within the AMS data were taken during the day, 2 were taken at night, and 1 image set spanned both day and night. The AMS sensor used for capturing the ground images is a 16-band multispectral scanner capable of capturing both visible and IR spectrum wavelengths. The specifications for these bands are shown in Table 1 [24]. Due to the variable flight altitude of the deployed UAS in each setting, the AMS imagery was captured at a variety of resolutions, ranging from 7.75m to 33.75m. All but one of these images were captured at a resolution of 20m or smaller. To account for these different resolutions, AMS images were each resampled to a consistent 20m resolution, using spatial interpolation when required to upsample images with larger native resolutions. Finally, these images were exported to raster format for compatibility with remote sensing imagery.

The deep learning architectures used in this study are typically applied to standard 3 or 4-channel imagery, most commonly from red, green, and blue channels, sometimes including an alpha value. Therefore, to use the AMS imagery with these desired architectures, we select three common types of band combinations used in remote sensing research, in order to test which composite image performs best for feature segmentation. The three band combinations and the respective bands used are: Visible spectrum (5, 3, 2), Color infrared (7, 5, 3), and Fire Emphasizing (10, 7, 3). A sample AMS image using each of these band combinations is shown in Figure 2. Each band combination may be used to emphasize different features of the terrain; for example, while the visible spectrum combination shows the terrain as perceived by the human eye, color infrared emphasizes healthy vegetation in red, while the “fire emphasizing” band combinations shows active fires in bright red and vegetation in green.

Table 1 AMS imagery band specifications. TM and VIIRS wavelengths correspond to Landsat-equivalent satellite bands.

Band	Wavelength, μm
1	0.42–0.45
2	0.45–0.52 (TM1)
3	0.52–0.60 (TM2)
4	0.60–0.62
5	0.63–0.69 (TM3)
6	0.69–0.75
7	0.75–0.90 (TM4)
8	0.91–1.05
9	1.55–1.75 (TM5) (high gain)
10	2.08–2.35 (TM7) (high gain)
11	3.60–3.79 (VIIRS M12) (high gain)
12	10.26–11.26 (VIIRS M15) (high gain)
13	1.55–1.75 (TM5) (low gain)
14	2.08–2.35 (TM7) (low gain)
15	3.60-3.79 (VIIRS M12) (low gain)
16	10.26-11.26 (VIIRS M15) (low gain)

B. Remote Sensing Data Labels

Remote sensing images, captured from earth observation systems, were used in this study to provide data labels for the historical AMS imagery. These data labels, classifying the type of pixel at a specific location, were then used to train the deep learning models, and help identify locations of active fire, as well as other contextual land features. First, information on contextual land features including forests, water, and human settlements was collected using the USGS National Land Cover Database [25]. This remote sensing database contains land cover classifications in roughly 2-3 year intervals from 1992 to 2016, derived from decision-tree classification methods on Landsat satellite data. There are two assumptions made in using this dataset to label the AMS imagery. First, the assumption is made that the USGS Land Cover dataset provides an accurate classification of pixels within the AMS imagery, and is roughly representative of the land cover during the collected time period. In reality, there are small differences in the time of the collected AMS image and the USGS land cover set, as well as in the methods of classification used for determining land cover. However, we assume that the data, regardless, serves as a useful approximation of the true land cover classification. Second, we make the assumption that 30 meter resolution, the resolution of the USGS land cover data, is sufficient to test the efficacy of the deep learning models. To match the resolution of the AMS imagery, this data is resampled using nearest-neighbor resampling. While this may introduce error in actual pixel classification, we assume that this error is reasonably small given the large capture area of the AMS imagery, and may be further refined in the future if desired.

In addition to labels regarding contextual land features such as forests, water, and human settlements, remote sensing imagery was used to develop labels regarding the pixels of active fire. For this, the differenced Normalized Burn Ratio (dNBR) index was used to compare burn severity in a certain location before and after an active fire. This index is commonly used to classify fire regions in Landsat 7 imagery, and is derived by comparing the pre and post-fire normalization of the near infrared and shortwave infrared wavelengths [26–28]. This is shown in Eq. 1 and Eq. 2. dNBR values can theoretically range from -2 to 2, and although the index is continuous and can be subject to local and context-specific variability, values above 0.270 may be generally classified as moderate to high levels of fire severity. To compute this index, we use a Landsat 7-derived NBR index from a year before the respective AMS image sets, and difference this with an NBR index computed from the AMS imagery itself. A visualization of this index-based classification for a sample AMS image is shown in Figure 2.

$$NBR = \frac{NIR - SWIR}{NIR + SWIR} \tag{1}$$

$$dNBR = PreNBR - PostNBR \tag{2}$$

Where NIR is the near-infrared (NIR) band reflectance value, and SWIR is the shortwave-infrared (SWIR) band reflectance value. PreNBR is the NBR index measurement before the occurrence of fire, while PostNBR is the measurement during or after the fire.

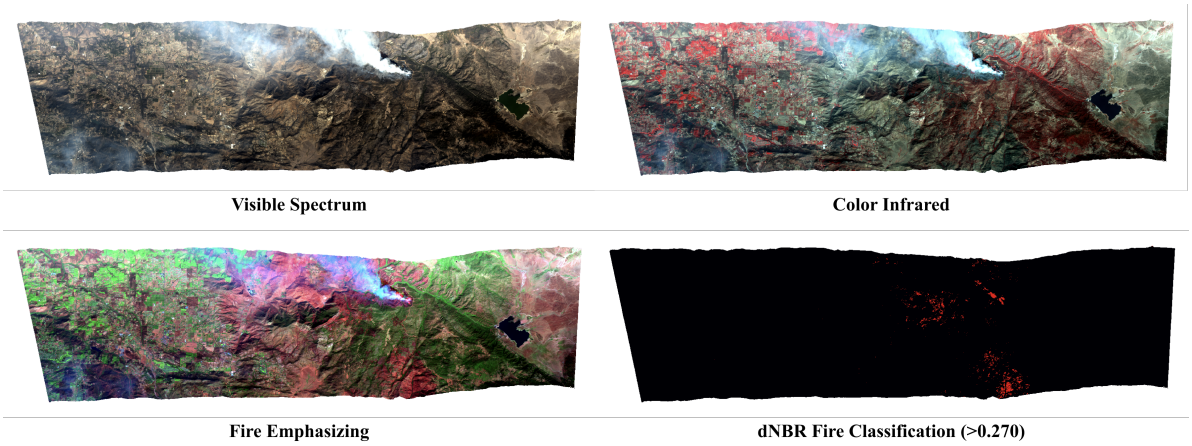

Fig. 2 Sample visualization of AMS imagery in three band combinations, and the respective active fire labels using the dNBR index with threshold value 0.270.

C. Deep Learning Model Benchmarking

We investigate the performance of three CNN-based models in fire spotting region segmentation: U-Net [29], Feature pyramid networks (FPNs) [30], and DeepLab [31] architectures. The distinct architectures across these models lead to various reported strengths and weaknesses in prior segmentation literature [32]. U-Net, for example, uses a FCN-based encoder-decoder architecture, which has proven particularly effective in remote sensing settings [33]. FPNs utilize feature pyramids built on image pyramids, which are intended to help detection of objects at different scales. Alternatively, DeepLab extends an FCN architecture with atrous convolution to incorporate larger context into filters. A common challenge in semantic segmentation is generalizing the performance of distinct architectures in unknown settings [34]; this benchmarking will therefore provide a useful baseline for future research in deep learning-based segmentation of AMS imagery. Binary (two-class) segmentation performance across features of interest is is independently benchmarked, as features may overlap in the AMS imagery. The features of interest include wildfire regions, forest cover, water bodies, and human settlements. Two loss functions are compared in benchmarking: intersection-over-union (IoU) loss (Eq. 3, and dice (Eq. 4) loss [35].

$$\mathcal{L}_{IoU} = -\ln \frac{|A \cap A_{gt}|}{|A \cup A_{gt}|} \tag{3}$$

Where A is the mask set as predicted by the segmentation model, and A_{gt} is the ground truth mask from the remote sensing data labels. Then, the numerator of the loss function is defined as the intersection of the two masks, while the denominator is the union of the two masks.

$$\mathcal{L}_{D} = 1 - \frac{2|A \cap A_{gt}|}{|A| + |A_{gt}|} \tag{4}$$

Similarly, Dice loss is defined with the area of intersection between the predicted and ground truth masks, but is divided instead by the sum of the separate mask areas.

To assess segmentation performance for each model configuration, the precision (P), recall (R), and intersection over union (IoU) are reported (Eq. 3). A visualization of these performance metrics is shown in Figure 3. Each of these performance metrics reports a different strength or weakness of a given model. Precision, for example, reports the number of true pixels detected as a fraction of the total number of predicted labels, which may be desired if

avoiding false positives is critical. Alternatively, recall provides the fraction of the number of true pixels detected over the number of total true pixels; this may be a useful quantity when finding all true positives is of concern. Finally, IoU provides a general metric representing the accuracy in overlap of a predicted label, which can serve as a useful summarization of segmentation performance. In reporting all three metrics, the best performing models may then be selected given domain-specific constraints, which can help develop accurate, customized pipelines for fire spotting and region segmentation.

$$P = \frac{TP}{TP + FP} \qquad R = \frac{TP}{TP + FP} \qquad IoU = \frac{TP}{TP + FN + FP} \tag{5}$$

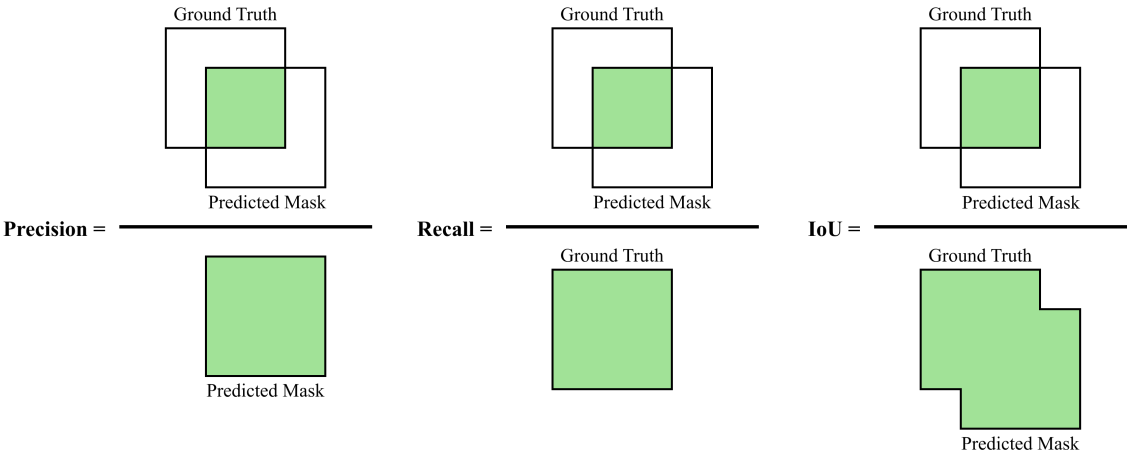

Fig. 3 Visualization of the metrics used to evaluate the learning performance of deep learning model architectures in segmenting AMS imagery.

D. Training Configuration

A review of the tested benchmarking configurations is shown in Figure 4. In summary, the performance of three deep learning models is tested with respect to 1) Three band combinations for AMS image visualization, 2) Four features of interest, including active fire and contextual region features, and 3) Two loss functions. Input AMS imagery and remote sensing-derived training labels were split into three distinct training, validation, and testing sets for model training using a 60-20-20 split. Minimal data augmentation was performed to avoid unrealistic representations of aerial imagery; for this reason, only flips and rotations were used in training inputs. For each training configuration, we use the standard model architectures pretrained on Imagenet [36]. Networks were trained using the Adam optimizer [37] with a learning rate of $5 \cdot 10^{-4}$, a batch size of 8, and 50 epochs.

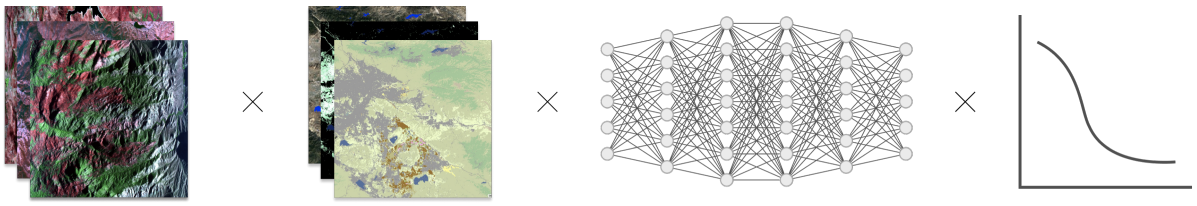

Fig. 4 Summary of the benchmarking configurations tested in this study.

V. Results

We present the benchmarking results for the various band combinations, segmentation models, and loss functions below. These results are divided into 1) benchmarking results for binary fire segmentation, and 2) benchmarking results for multiclass contextual feature segmentation, which includes a summary of the performance metrics across forest, water, and settlement segmentation.

A. Active Fire Segmentation

Benchmarking results for active fire segmentation, including precision, recall, and IoU for fire segmentation across all training configurations of band combinations, segmentation models, and loss functions are shown in Table 2. There are several points of interest in the fire segmentation results which suggest that the proposed methodology for deep learning-based fire segmentation of AMS imagery was successful. Firstly, the highest reported IoU across all training configurations was found in the training configuration using the fire emphasizing band combination as input, and the U-Net segmentation model with Dice Loss. This is perhaps unsurprising, as the particular band combination is intended to emphasize infrared-based hotspots where fires may be located, and the U-Net model configuration has been shown in prior literature to perform well in satellite and aerial image segmentation [38]. However, while this model configuration produced the highest IoU, other model configurations may be preferred if maximizing precision or recall are desired. The model configuration using color infrared bands with an FPN model and IoU loss produced the highest precision, at 1. However, while precision is maximized under this model configuration, it may not be a desired metric in active fire spotting, as the trade off with recall may result in missed fire hotspots. The model configuration using fire emphasizing bands with a U-Net model and IoU loss, alternatively, provided the highest recall and the highest overall precision-recall average. In this case, the loss function may be particularly relevant for maximizing recall, as IoU loss has been shown in prior literature to be particularly effective in deep learning-based object detection [39].

Table 2 Fire segmentation results across all training configurations.

Band Combination	Segmentation Model	Loss	IoU	Precision	Recall
Visible	U-Net	IoU	0.47	0.86	0.52
Visible	U-Net	Dice	0.51	0.94	0.52
Visible	FPN	IoU	0.5	0.97	0.51
Visible	FPN	Dice	0.5	0.97	0.51
Visible	Deeplab	IoU	0.5	0.9	0.52
Visible	Deeplab	Dice	0.51	0.9	0.53
IR	U-Net	IoU	0.55	0.87	0.61
IR	U-Net	Dice	0.59	0.92	0.59
IR	FPN	IoU	0.59	1	0.59
IR	FPN	Dice	0.58	0.92	0.59
IR	Deeplab	IoU	0.58	0.88	0.61
IR	Deeplab	Dice	0.59	0.92	0.6
Fire Emph.	U-Net	IoU	0.52	0.71	0.7
Fire Emph.	U-Net	Dice	0.62	0.87	0.65
Fire Emph.	FPN	IoU	0.59	0.81	0.66
Fire Emph.	FPN	Dice	0.58	0.82	0.65
Fire Emph.	Deeplab	IoU	0.57	0.85	0.62
Fire Emph.	Deeplab	Dice	0.56	0.83	0.62

Four examples of fire segmentation across test images for the model configuration with the best IoU (fire emphasizing bands with a U-Net model and Dice loss) are shown in Figure 5. The input images to the model are shown in the left column, the ground truth masks are shown in the middle column, and the model-generated prediction masks are shown in the right column. Here, the generated masks serve purely to visualize the performance of the models, and may help better manually identify the strengths and weaknesses of the segmentation. A comparison of the segmentation predictions with the ground truth masks shows that the model performs, generally, very well in identifying large areas of fire, and discerning exact boundaries where they exist. Additionally, the results indicate that the model has some ability

to segment through smoke, and generally segment the areas of active fire well without the need for prior data on terrain appearance. Overall, while further work is needed to determine the accuracy of the proposed method in various settings and types of fires, the conducted test of deep learning-based segmentation is broadly successful in identifying areas of active fire.

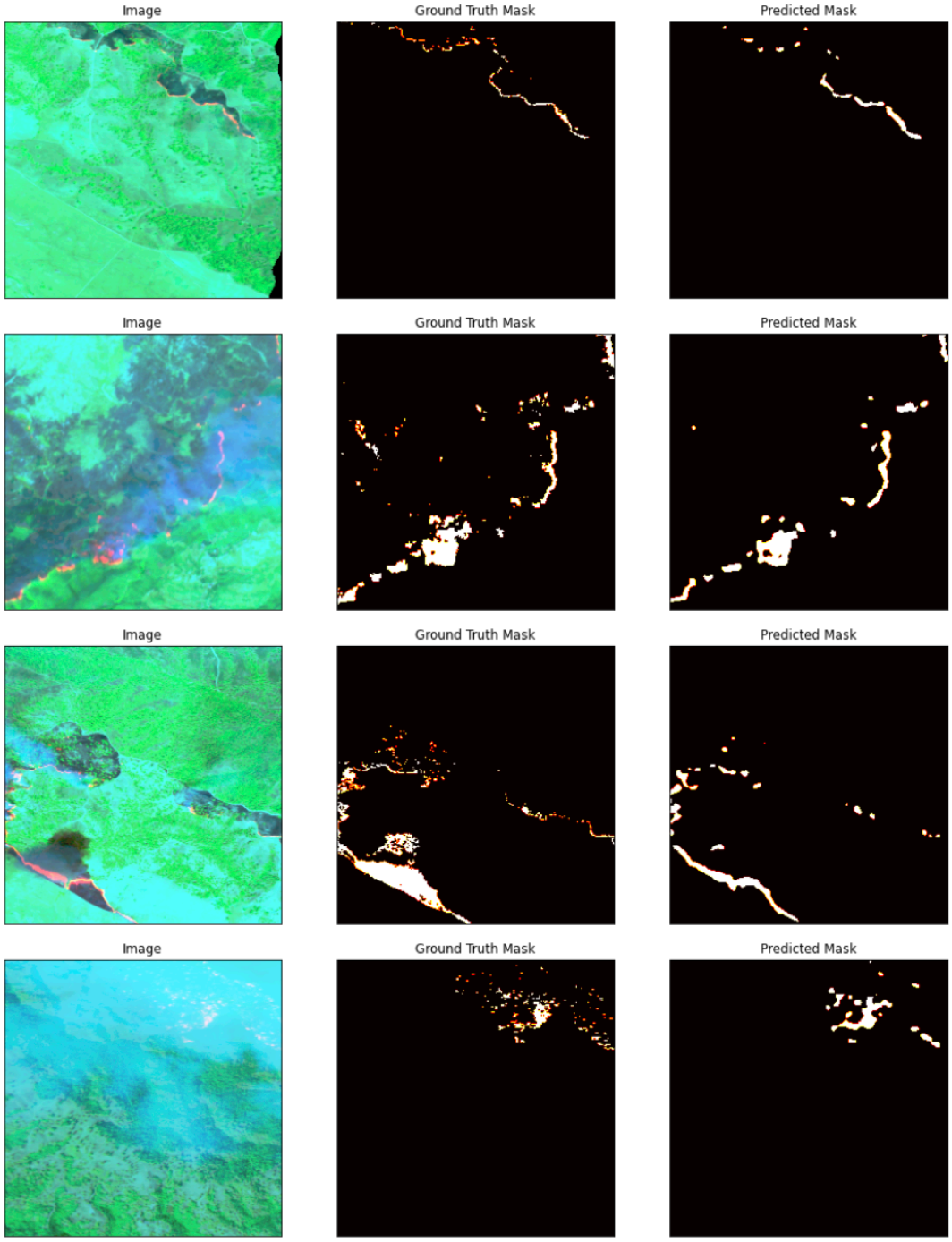

Fig. 5 Results from segmentation of active fire in AMS imagery. Left: Original AMS imagery visualized using the fire emphasizing band combination; Middle: Ground truth masks of active fire determined using dNBR ratio; Right: Predicted active fire regions.

B. Land Feature Segmentation

Results for segmentation benchmarking of contextual land features are shown in Table 3. Benchmarking results are combined across land features for reported indices. The results at a high level suggest that, although less successful than fire segmentation, the deep learning architectures are able to learn, with reasonable accuracy, to segment contextual land features. The best performing model is the U-Net architecture with IoU loss on the visible spectrum band combination. There are several possible reasons for this result; for example, in addition to the U-Net architecture being conducive to remote sensing image segmentation, the pretrained Imagenet weights used in the model architectures are better suited to segmenting features in the visible spectrum. Because the input images were normalized using these weights, images in the visible spectrum may provide more tuned information than images in other band combinations. Generally, the visible spectrum band combination resulted in a higher training performance benchmarks, suggesting that the band composition itself is the likely reason for the positive training results.

Table 3 Land feature segmentation results across all training configurations.

Band Combination	Segmentation Model	Loss	IoU	Precision	Recall
Visible	U-Net	IoU	0.4632	0.6077	0.6689
Visible	U-Net	Dice	0.4624	0.5779	0.6798
Visible	FPN	IoU	0.4504	0.5988	0.6275
Visible	FPN	Dice	0.4489	0.5533	0.6647
Visible	Deeplab	IoU	0.4514	0.5953	0.6195
Visible	Deeplab	Dice	0.4353	0.6173	0.5963
IR	U-Net	IoU	0.3900	0.5520	0.5287
IR	U-Net	Dice	0.3921	0.5635	0.5420
IR	FPN	IoU	0.3596	0.4838	0.5114
IR	FPN	Dice	0.3632	0.5581	0.4816
IR	Deeplab	IoU	0.3718	0.4907	0.5331
IR	Deeplab	Dice	0.3697	0.5630	0.5165
Fire Emph.	U-Net	IoU	0.4248	0.5343	0.6376
Fire Emph.	U-Net	Dice	0.4312	0.5609	0.5846
Fire Emph.	FPN	IoU	0.4040	0.5392	0.5461
Fire Emph.	FPN	Dice	0.4051	0.6066	0.5467
Fire Emph.	Deeplab	IoU	0.3947	0.5477	0.5536
Fire Emph.	Deeplab	Dice	0.3979	0.6103	0.5231

Examples of image segmentation for forests, water, and human settlements using the best performing training configuration are shown in Figures 6, 7, and 8 respectively. Forest segmentation, upon visual inspection, appears to have reasonable accurate results. A comparison of the ground truth and predicted mask shows that visibly contrasting, dense areas of forests are well detected, even if there is variability in their appearance across and within images. However, the more detailed delineation between forest and bare terrain is lost in the segmentation, which may be improved in the future with higher resolution imagery and more detailed training labels. Water, similarly, showed good segmentation results, despite relatively poor training labels from the USGS data. Bodies of water, unlike more static features like forests, can vary significantly by time of year and location, making concrete classifications of pixels from external datasets difficult. As a result, the accuracy of the segmentation was artificially bounded by the limitations of training labels. This may be improved in the future with more accurate and time-specific training labels. Finally, segmentation of human settlements showed poor performance, as seen in Figure 8. This is likely due to the insufficient number of human settlement instances in the training data, since areas of large fires were more likely to occur in more remote, forest regions rather than densely populated areas. As a result, the model architectures were unable to learn features of human settlements needed for smaller detection areas. This may be improved in the future by introducing additional examples of urban areas, possibly through transfer learning if no direct examples of settlements in fire regions exist.

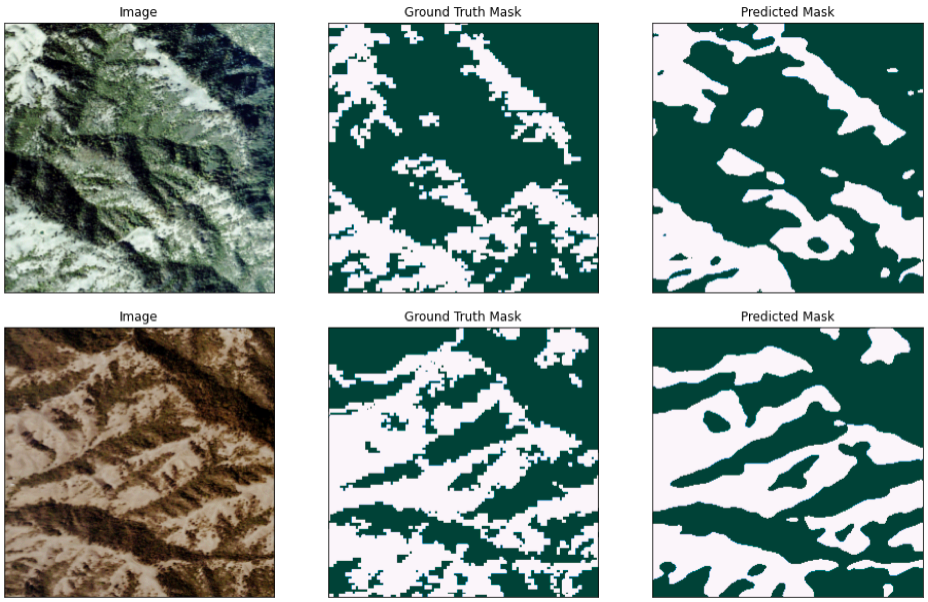

Fig. 6 Results from segmentation of forests in AMS imagery. Segmentation showed good performance, but highly variable depending on the type and size of forest. Generally, there is signficinat overlap across large, dense areas of forest, but further work is needed to improve the training data and output accuracy.

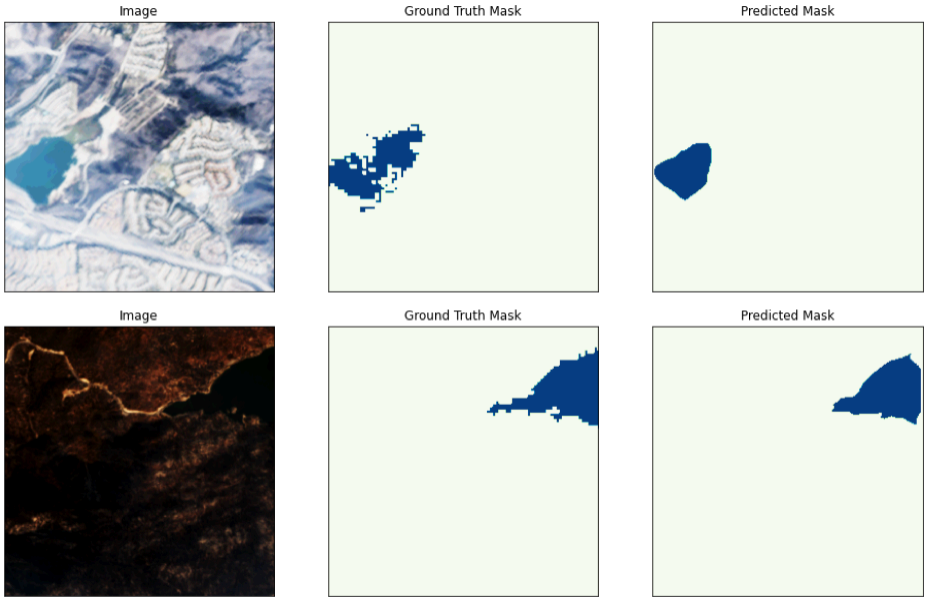

Fig. 7 Results from segmentation of water in AMS imagery. Segmentation showed moderate to good performance, but missed smaller bodies of water. This is likely due to the insufficient number of training instances, as the AMS imagery was taken in California where naturally occurring bodies of water are rare. Training labels may also be improved in the future to accurately represent the location of water given seasonal changes in water levels.

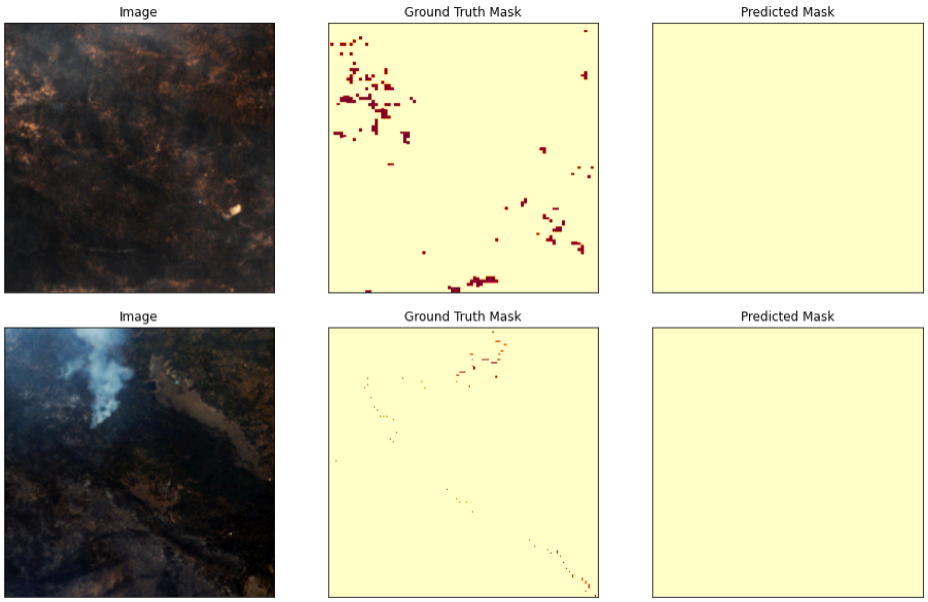

Fig. 8 Results from segmentation of settlements in AMS imagery. Segmentation showed poor performance, missing large roads and clusters of houses. This is likely due to insufficient number training instances, as the AMS imagery was largely taken in forest, desert, or mountainous terrain. Future work will include augmenting the models with existing remote sensing data on settlements to improve segmentation.

VI. Conclusion

In this work, we present a benchmarking of modern, deep learning-based segmentation models in classifying areas of AMS imagery, fusing existing sources of aerial and remote sensing imagery. We observe that the proposed system performs well in segmenting areas of active fire, and has reasonable performance segmenting other land features such as forests and bodies of water. It also can be easily adopted using publicly available data, and provides an additional verification technique conducive to active wildfire spotting scenarios. This system has several strengths which can help support research in deep learning for fire spotting. Firstly, the system fuses remote sensing-derived data with aerial imagery of active fire regions, allowing for improved contextual understanding and accuracy of a given region. Second, the proposed system helps provide fire spotting experts with a method to augmenting existing manual labelling efforts, which also helps identify nearby land features of interest in wildfire response. Third, the proposed system helps aid in improved discriminative classification; that is, the deep learning models can help understand features which are explicitly not active fire, improving downstream classification performance. Finally, the proposed system provides a useful baseline for future studies in fire region segmentation using deep learning models.

VII. Future Work

Further research includes testing the proposed system on a wider range of aerial imagery, and improving segmentation of images through other commonly used vision techniques such as transfer learning and active learning. Robustness tests across a wide variety of environments are critical to ensure that vision systems are not constrained to a particular region or natural environment. Wildfires, in particular, may occur in many different visual and ecological settings, so the proposed system must be tested across these environments to ensure broad applicability. Furthermore, the system may incorporate machine learning techniques which can facilitate the deployment of the system in limited-resource environments. For example, transfer learning may help achieve model convergence with fewer training examples by using images of previous fires to assist in learning; active learning, on the other hand, can help incorporate human-annotated labelling feedback into the training loop, and reduce the total number of images needed for training.

Acknowledgments

The authors would like to thank the NASA Ames Scalable Traffic Management for Emergency Response Operations (STEReO) project and the NASA Langley Research Center for their support in this work.

References

[1] Hoover, K., and Hanson, L. A., "Wildfire Statistics," *Frontiers in Ecology and the Environment*, 2021.

[2] Running, S. W., "Is global warming causing more, larger wildfires?" *Science*, 2006.

[3] Westerling, A. L., Hidalgo, H. G., Cayan, D. R., and Swetnam, T. W., "Warming and earlier spring increase western US forest wildfire activity," *science*, Vol. 313, No. 5789, 2006, pp. 940–943.

[4] Morton, D. C., Roessing, M. E., Camp, A. E., and Tyrrell, M. L., "Assessing the environmental, social, and economic impacts of wildfire," *GISF Research Paper*, Vol. 1, 2003.

[5] Chan, H., Tran-Thanh, L., and Viswanathan, V., "Fighting Wildfires under Uncertainty: A Sequential Resource Allocation Approach," 2020.

[6] Branson-Potts, H., "L.A. Fire Department used drones for the first time during Skirball fire," `https://www.latimes.com/local/lanow/la-me-ln-lafd-drone-skirball-fire-20171214-story.html`, 2017. Accessed: 2021-05-30.

[7] Bernknopf, R., Kuwayama, Y., Gibson, R., Blakely, J., Mabee, B., Clifford, T., Quayle, B., Epting, J., Hardy, T., and Goodrich, D., "The cost-effectiveness of satellite Earth observations to inform a post-wildfire response," *Resources for the Future*, 2019.

[8] Harris, S., Veraverbeke, S., and Hook, S., "Evaluating spectral indices for assessing fire severity in chaparral ecosystems (Southern California) using MODIS/ASTER (MASTER) airborne simulator data," *Remote sensing*, Vol. 3, No. 11, 2011, pp. 2403–2419.

[9] Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., and Garcia-Rodriguez, J., "A review on deep learning techniques applied to semantic segmentation," *arXiv preprint arXiv:1704.06857*, 2017.

[10] Ambrosia, V., Myers, J., Hildum, E., Field, M., and Field, M., "NASA's Autonomous Modular Scanner (AMS)–Wildfire Sensor: Improving wildfire observations from airborne platforms," *Proceedings of the 34th International Symposium for Remote Sensing of the Environment. Sidney, Australia*, 2011.

[11] Jain, P., Coogan, S. C., Subramanian, S. G., Crowley, M., Taylor, S., and Flannigan, M. D., "A review of machine learning applications in wildfire science and management," *Environmental Reviews*, Vol. 28, No. 4, 2020, pp. 478–505.

[12] Twidwell, D., Allen, C. R., Detweiler, C., Higgins, J., Laney, C., and Elbaum, S., "Smokey comes of age: unmanned aerial systems for fire management," *Frontiers in Ecology and the Environment*, Vol. 14, No. 6, 2016, pp. 333–339.

[13] Akhloufi, M. A., Couturier, A., and Castro, N. A., "Unmanned aerial vehicles for wildland fires: Sensing, perception, cooperation and assistance," *Drones*, Vol. 5, No. 1, 2021, p. 15.

[14] Martínez-de Dios, J. R., Merino, L., and Ollero, A., "Fire detection using autonomous aerial vehicles with infrared and visual cameras," *IFAC Proceedings Volumes*, Vol. 38, No. 1, 2005, pp. 660–665.

[15] Yuan, C., Liu, Z., and Zhang, Y., "Fire detection using infrared images for UAV-based forest fire surveillance," *2017 International Conference on Unmanned Aircraft Systems (ICUAS)*, IEEE, 2017, pp. 567–572.

[16] Merino, L., Caballero, F., Martinez-de Dios, J., and Ollero, A., "Cooperative fire detection using unmanned aerial vehicles," *Proceedings of the 2005 IEEE international conference on robotics and automation*, IEEE, 2005, pp. 1884–1889.

[17] Aspragathos, N., Dogkas, E., Konstantinidis, P., Koutmos, P., Lamprinou, N., Moulianitis, V. C., Paterakis, G., Psarakis, E. Z., Sartinas, E., Souflas, K., et al., "From pillars to AI technology-based Forest fire protection systems," *Intelligent System and Computing*, IntechOpen, 2019.

[18] Akhloufi, M. A., Tokime, R. B., and Elassady, H., "Wildland fires detection and segmentation using deep learning," *Pattern recognition and tracking xxix*, Vol. 10649, International Society for Optics and Photonics, 2018, p. 106490B.

[19] Yuan, F., Zhang, L., Xia, X., Wan, B., Huang, Q., and Li, X., "Deep smoke segmentation," *Neurocomputing*, Vol. 357, 2019, pp. 248–260.

[20] Frizzi, S., Bouchouicha, M., Ginoux, J.-M., Moreau, E., and Sayadi, M., "Convolutional neural network for smoke and fire semantic segmentation," *IET Image Processing*, Vol. 15, No. 3, 2021, pp. 634–647.

[21] Zhao, Y., Ma, J., Li, X., and Zhang, J., "Saliency detection and deep learning-based wildfire identification in UAV imagery," *Sensors*, Vol. 18, No. 3, 2018, p. 712.

[22] Muruganandham, S., "Semantic segmentation of satellite images using deep learning," , 2016.

[23] Khryashchev, V., Ivanovsky, L., Pavlov, V., Ostrovskaya, A., and Rubtsov, A., "Comparison of different convolutional neural network architectures for satellite image segmentation," *2018 23rd Conference of Open Innovations Association (FRUCT)*, IEEE, 2018, pp. 172–179.

[24] Ambrosia, V. G., Wegener, S., Zajkowski, T., Sullivan, D., Buechel, S., Enomoto, F., Lobitz, B., Johan, S., Brass, J., and Hinkley, E., "The Ikhana unmanned airborne system (UAS) western states fire imaging missions: from concept to reality (2006–2010)," *Geocarto International*, Vol. 26, No. 2, 2011, pp. 85–101.

[25] Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Bender, S. M., Case, A., Costello, C., Dewitz, J., Fry, J., et al., "A new generation of the United States National Land Cover Database: Requirements, research priorities, design, and implementation strategies," *ISPRS journal of photogrammetry and remote sensing*, Vol. 146, 2018, pp. 108–123.

[26] Soverel, N. O., Perrakis, D. D., and Coops, N. C., "Estimating burn severity from Landsat dNBR and RdNBR indices across western Canada," *Remote Sensing of Environment*, Vol. 114, No. 9, 2010, pp. 1896–1909.

[27] Cocke, A. E., Fulé, P. Z., and Crouse, J. E., "Comparison of burn severity assessments using Differenced Normalized Burn Ratio and ground data," *International Journal of Wildland Fire*, Vol. 14, No. 2, 2005, pp. 189–198.

[28] Holden, Z., Smith, A., Morgan, P., Rollins, M., and Gessler, P., "Evaluation of novel thermally enhanced spectral indices for mapping fire perimeters and comparisons with fire atlas data," *International Journal of Remote Sensing*, Vol. 26, No. 21, 2005, pp. 4801–4808.

[29] Ronneberger, O., Fischer, P., and Brox, T., "U-net: Convolutional networks for biomedical image segmentation," *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, pp. 234–241.

[30] Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S., "Feature pyramid networks for object detection," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117–2125.

[31] Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L., "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, Vol. 40, No. 4, 2017, pp. 834–848.

[32] Minaee, S., Boykov, Y. Y., Porikli, F., Plaza, A. J., Kehtarnavaz, N., and Terzopoulos, D., "Image segmentation using deep learning: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[33] Hu, J., Li, L., Lin, Y., Wu, F., and Zhao, J., "A comparison and strategy of semantic segmentation on remote sensing images," *The International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery*, Springer, 2019, pp. 21–29.

[34] Guo, Y., Liu, Y., Georgiou, T., and Lew, M. S., "A review of semantic segmentation using deep neural networks," *International journal of multimedia information retrieval*, Vol. 7, No. 2, 2018, pp. 87–93.

[35] Jadon, S., "A survey of loss functions for semantic segmentation," *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, IEEE, 2020, pp. 1–7.

[36] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L., "Imagenet: A large-scale hierarchical image database," *2009 IEEE conference on computer vision and pattern recognition*, Ieee, 2009, pp. 248–255.

[37] Kingma, D. P., and Ba, J., "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[38] Neupane, B., Horanont, T., and Aryal, J., "Deep Learning-Based Semantic Segmentation of Urban Features in Satellite Images: A Review and Meta-Analysis," *Remote Sensing*, Vol. 13, No. 4, 2021, p. 808.

[39] Wang, Q., Ma, Y., Zhao, K., and Tian, Y., "A comprehensive survey of loss functions in machine learning," *Annals of Data Science*, 2020, pp. 1–26.