Proceedings of the 11th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS'25)
Paris, France - August, 2025
Paper No. MVML 111
DOI: 10.11159/mvml25.111

A Comprehensive Analysis of Transfer Learning Algorithms for Image Segmentation of Irregular-Shaped Fire Object

Yoseob Heo[1], Jeong Kyu Kim[2], Tae-Eung Sung[3], and *Jongseok Kang[4]
[1]Division of Data Analysis, Busan-Ulsan-Gyeongnam Branch, Korea Institute of Science and Technology Information
41, Centum dong-ro, Haeundae-gu, Busan
joseph87@kisti.re.kr
[2, 3]Department of Computer Science, Yonsei University
1, Yeonsedae-gil, Wonju, Gangwon-do, Republic of Korea
[2]hanencia@yonsei.ac.kr, [3]tesung@yonsei.ac.kr
[4]Division of Data Analysis, Korea Institute of Science and Technology Information
66, Hoegi-ro, Dongdaemun-gu, Seoul, Republic of Korea
*Corresponding author: kangjs@kisti.re.kr

Abstract - This study investigates the application of transfer learning and attention mechanisms to improve image segmentation for fire detection, particularly for irregularly shaped fire regions. Five models, including U-Net and its variants with VGG16, ResNet50, DenseNet201, and EfficientNet-B7 backbones, were developed and evaluated with and without attention layers. A dataset of 5,000 images, segmented into training, validation, and test sets, was prepared, focusing on flame regions. Experimental results demonstrated that attention-based models consistently outperformed their non-attention counterparts, with the VGG16 U-Net attention model achieving the highest validation IoU score of 0.8220. By effectively capturing intricate fire boundaries, these models offer significant improvements in segmentation accuracy. The findings highlight the potential of combining attention mechanisms and transfer learning for real-time fire detection systems.

***Keywords*:** fire head, image segmentation, U-Net, attention model, transfer learning

1. Introduction

Fires rank among the most frequent and destructive disasters worldwide, causing substantial damage to both human life and property [1]. Early fire detection and rapid suppression are critically important for preventing the spread of flames and minimizing potential losses. Consequently, the development of technologies that enable automatic detection of fire and precise delineation of fire regions—often characterized by irregular shapes—has emerged as a vital area of research in the field of disaster safety [2].

Recently, deep learning methods have undergone significant development, prompting a growing body of work that applies deep neural networks to a variety of tasks in image analysis. In particular, semantic segmentation has garnered increasing attention for applications in which the target object—such as a fire—can change in shape, size, and appearance due to factors like smoke and ever-shifting boundaries. While deep learning-based object recognition and segmentation techniques often exhibit remarkable performance gains over traditional approaches, they generally require large, well-annotated datasets to avoid overfitting and to ensure stable training. In the context of fire detection, constructing such datasets is difficult and costly, in part due to the inherent danger and logistical hurdles involved in recording real fire scenarios. Consequently, transfer learning has emerged as a valuable method: models pretrained on large-scale datasets can be fine-tuned for more specialized tasks, thereby reducing data requirements and avoiding protracted training cycles [3].

In addition, various structural improvements to deep neural networks have been explored in an effort to more accurately capture complex fire boundaries. One such method involves the use of attention mechanisms, which guide the model to assign greater weight to salient fire features. Studies have shown that this approach can enhance segmentation of irregular or fine-grained contours [4]. Attention-based strategies have thus found wide application not only in fire detection but also in the segmentation of other objects with irregular shapes.

Against this backdrop, the present study aims to improve the semantic segmentation performance for fire regions by integrating transfer learning algorithms with diverse segmentation model architectures. Specifically, we employ pretrained weights from large-scale datasets to develop models specialized for identifying irregularly shaped objects such as fire. We then quantitatively and qualitatively evaluate the impact of incorporating attention layers on segmentation accuracy. Through these investigations, we seek to propose a practical framework that can be adopted in real-time fire detection and response systems.

2. Related Works

Because fire exhibits inherently irregular contours and fluctuating intensities, semantic segmentation of such objects is uniquely challenging; however, recent progress in deep learning has considerably advanced segmentation accuracy—particularly by refining feature extraction, boundary delineation, and context interpretation—thereby enabling more robust methods in complex fire scenarios.

2.1. Multi-Scale and Attention-Based Approaches

The use of multi-scale semantic segmentation techniques, such as the combination of U-Net with multi-scale residual group attention (MRGA), has been shown to effectively enhance the perception of small-scale smoke and fire features. This approach leverages global information to improve accuracy, particularly for thin smoke at image edges, achieving a mean Intersection over Union (mIoU) of 91.83% [5]. Similarly, models leveraging dual attention mechanisms, such as the DAMN architecture, mitigate issues including slow fitting and edge blurring by processing feature maps in parallel through atrous spatial pyramid pooling (ASPP), thereby enhancing segmentation performance and achieving an average IoU of 85.77% on the Fire-Smoke dataset [6].

2.2. Deep Learning Architectures

Deep learning architectures such as Deep-RegSeg have been developed to segment fire pixels and detect precise fire shapes in complex environments. This model has demonstrated high performance in segmenting small fire areas under various conditions, outperforming recent state-of-the-art techniques [7].

The Global Position Guidance (GPG) and Multi-path explicit Edge information Interaction (MEI) modules have been proposed to refine fire segmentation by restraining local segmentation errors and utilizing edge information, achieving superior IoU scores across multiple test sets [8].

2.3. UAV and Remote Sensing Applications

Unmanned aerial vehicle(UAV)-based models like FBC-ANet focus on forest fire monitoring, where fires often have small areas and irregular contours. This model combines boundary enhancement and context-aware modules to improve segmentation accuracy, achieving an IoU of 83.08% on the FLAME dataset [9].

Wang et al. proposes a novel network model, "Smoke-Unet," which integrates an improved U-Net architecture with attention mechanisms and residual blocks to enhance early forest fire smoke detection in remote sensing imagery. Experiments show that the proposed Smoke-Unet outperforms the standard U-Net by 3.1% in smoke pixel segmentation accuracy, effectively distinguishing early smoke from confounding factors like clouds or fog [10].

3. Methodology

In this study, 600 fire images were collected from the internet, news articles, YouTube videos, and publicly available fire video databases. For image segmentation, only the regions containing flames were labeled to generate the corresponding mask data. Each image was then divided into 256×256 patches, retaining only those patches that contained fire; the same procedure was applied to the mask data. Consequently, a total of 5,000 fire images and their associated mask data were collected and preprocessed.

To evaluate segmentation performance on fire images for each model, this study employed the basic image segmentation model (U-Net), four transfer-learning models that incorporate various convolutional neural network backbones pre-trained on ImageNet (VGG16 U-Net, ResNet50 U-Net, DenseNet201 U-Net, and EfficientNet-B7 U-Net), and five attention-based variants in which an Attention layer was added to each model (U-Net Attention, VGG16 U-Net Attention, ResNet50 U-Net

Attention, DenseNet201 U-Net Attention, and EfficientNet-B7 U-Net Attention). Fire image segmentation tests were then conducted using these models to compare their performance (Table 1). In this study, out of a total of 5,000 fire image–mask pairs, 70% were allocated to the training dataset, 20% to the test dataset, and 10% to the validation dataset. The Adam optimizer was employed with a learning rate of 0.001. Each training phase was conducted for 200 epochs with a batch size of 16, with employing data augmentation techniques.

Table 1: Types of models employed for fire image segmentation test.

Backbone	U-Net model	U-Net attention model
-	Simple U-Net	Simple U-Net attention
VGG16	VGG16 U-Net	VGG16 U-Net attention
ResNet50	ResNet50 U-Net	ResNet50 U-Net attention
DenseNet201	DenseNet201 U-Net	DenseNet201 U-Net attention
EfficientNet-B7	EfficientNet-B7 U-Net	EfficientNet-B7 U-Net attention

Both training and testing were conducted on a computing system equipped with an AMD Ryzen 9 7950X 16-core processor and an NVIDIA GeForce RTX 4080 GPU with 16 GB of memory. For all models, binary cross-entropy (BCE) was utilized as the loss function, enabling fast and stable training in binary segmentation while achieving high accuracy. The formulation of this function is presented in Equation (1), where y denotes the true labels, p represents the predicted outcomes, and N signifies the set of all pixels.

$$BCE = -\frac{1}{N}\sum_{i=1}^{N} y_i \log p_i + (1 - y_i) \log (1 - p_i) \tag{1}$$

To perform an effective and comprehensive evaluation of each image segmentation model, this study employs four primary evaluation metrics: Intersection of Union (IoU), Precision, Recall and F1 score. These metrics facilitate the assessment of model performance from diverse perspectives [11].

Intersection over Union (IoU) for the fire class quantifies the extent of overlap between the predicted segmentation and the ground truth segmentation labels. It is calculated using the following formula:

$$IoU = \frac{TP_i}{TP_i + FP_i + FN_i} \tag{2}$$

In this equation, TP_i denotes True Positives (correctly identified flame pixels), FP_i represents False Positives (pixels incorrectly labeled as fire), and FN_i stands for False Negatives (actual fire pixels that were not detected).

Precision indicates the proportion of correctly identified fire pixels out of all pixels predicted as fire. It is defined by the following equation:

$$Precision = \frac{TP_i}{TP_i + FP_i} \tag{3}$$

Recall measures the ratio of correctly identified fire pixels to the total number of actual fire pixels. It is calculated as:

$$Recall = \frac{TP_i}{TP_i + FN_i} \tag{3}$$

The F1 Score is particularly valuable for evaluating the balance between Precision and Recall. It is the harmonic mean of these two metrics, providing a single value that accounts for both false positives and false negatives. In the context of fire segmentation, a higher F1 Score signifies a better balance between detecting a large number of fire pixels (Recall) and ensuring that the detected pixels are accurately identified as fire (Precision). The F1 Score is computed using the following formula:

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{3}$$

Including the F1 Score as an evaluation metric offers a more holistic view of the model's performance, especially in situations where Precision and Recall are equally important.

These evaluation metrics enable a thorough and objective assessment of the model's performance in fire segmentation tasks. They allow for a precise understanding and comparison of different methodologies. By examining these metrics, we can identify the model's strengths and weaknesses and pinpoint areas for improvement. Additionally, comparing these metrics across various models facilitates informed decisions regarding which models are most suitable for deployment in real-world fire detection scenarios. Given the critical need for timely and accurate detection to prevent the spread and damage of fires, reliable evaluation metrics are essential for validating the model's effectiveness and ensuring its operational reliability.

4. Results

The training results for the various models previously described, including the loss, accuracy, and IoU metrics for both the training and validation datasets, are presented in Table 2.

Table 2: Training and validation results of various fire image segmentation models.

Model	Train loss	Train accuracy	Train IoU	Validation loss	Validation accuracy	Validation IoU
Simple U-Net	0.1601	0.9305	0.8001	0.1666	0.9207	0.7847
Simple U-Net attention	0.1498	0.9339	0.8356	0.1814	0.9320	0.8009
VGG16 U-Net	0.1810	0.9308	0.8143	0.1978	0.9289	0.8015
VGG16 U-Net attention	0.1739	0.9277	0.8128	0.1759	0.9316	0.8220
ResNet50 U-Net	0.1405	0.9376	0.8658	0.1799	0.9333	0.7901
ResNet50 U-Net attention	0.1200	0.9418	0.8497	0.1841	0.9316	0.8021
DenseNet201 U-Net	0.1346	0.9397	0.8553	0.1735	0.9340	0.7910
DenseNet201 U-Net attention	0.1342	0.9398	0.8517	0.1816	0.9323	0.8001
EfficientNet-B7 U-Net	0.1286	0.9360	0.8497	0.1761	0.9330	0.7857
EfficientNet-B7 U-Net attention	0.1271	0.9405	0.8625	0.1696	0.9345	0.7995

The VGG16 U-Net with attention mechanism exhibited the highest performance on the validation dataset, achieving a validation IoU score of 0.8220. Additionally, the experimental results consistently indicated that models incorporating attention layers outperformed those without them in terms of performance metrics.

Table 3 displays the evaluation metrics for the test dataset. The results observed in the training and validation datasets were consistent with those obtained when evaluating the model on the test dataset.

Table 3: Evaluation metrics for image segmentation models on the test set.

Model	Test IoU	Precision	Recall	F1 Score
Simple U-Net	0.7143	0.8537	0.8140	0.8333
Simple U-Net attention	0.7254	0.8333	0.8192	0.8408
VGG16 U-Net	0.7887	0.9036	0.8611	0.8818
VGG16 U-Net attention	0.7938	0.9102	0.8613	0.8851
ResNet50 U-Net	0.7764	0.9025	0.8475	0.8741
ResNet50 U-Net attention	0.7845	0.9114	0.8493	0.8793
DenseNet201 U-Net	0.7618	0.8995	0.8326	0.8648
DenseNet201 U-Net attention	0.7708	0.9091	0.8351	0.8706
EfficientNet-B7 U-Net	0.7500	0.9036	0.8152	0.8572
EfficientNet-B7 U-Net attention	0.7513	0.9042	0.8162	0.8580

The actual image segmentation results are presented in Fig. 1. Although the segmentation outputs do not vary significantly across different models, it is evident that models incorporating attention layers more effectively capture the intricate structures of fire images compared to those without such layers. Specifically, the mask images corresponding to the ground truth become simplified during the patchifying process, which impairs the accurate delineation of fire boundaries. In contrast, the segmentation results for irregularly shaped fire images obtained using various models demonstrate a better ability to define fire boundaries than the original masks.

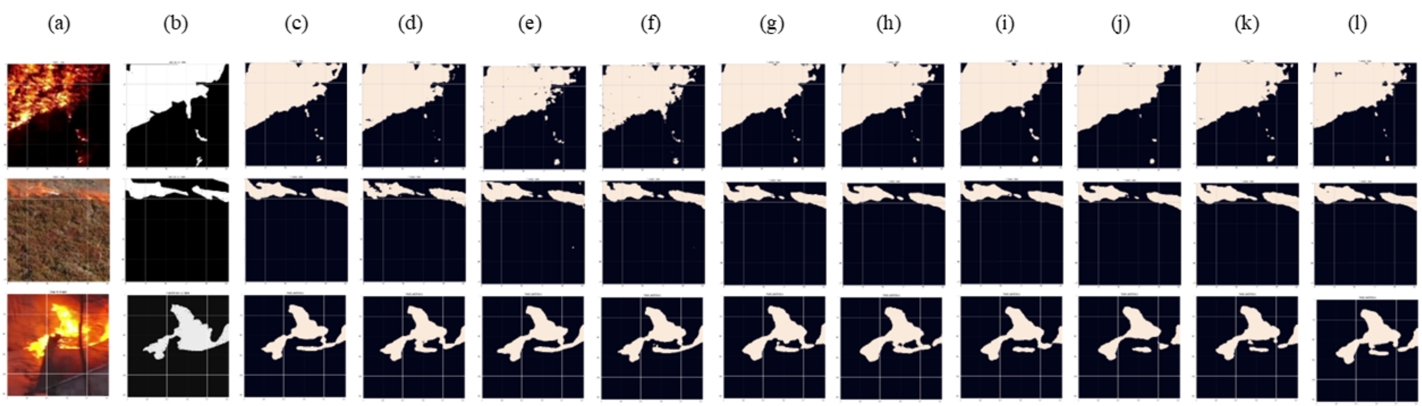

Fig. 1: Fire image segmentation results using various image segmentation models : (a) Image (b) Mask (c) Simple U-Net (d) Simple U-Net attention (e) VGG16 U-Net (f) VGG16 U-Net attention (g) ResNet50 U-Net (h) ResNet50 U-Net attention (i) DenseNet201 U-Net (j) DenseNet201 U-Net attention (k) EfficientNet-B7 U-Net (l) EfficientNet-B7 U-Net attention

4. Conclusion

This study demonstrates the significant potential of integrating transfer learning and attention mechanisms to address the challenging task of fire image segmentation, particularly for irregularly shaped fire regions. By employing various U-Net-based architectures with and without attention layers, we provided a comparative analysis that highlights the superiority of attention-based models in capturing complex and fine-grained fire boundaries. Among the tested models, the VGG16 U-Net with attention achieved the highest validation IoU score of 0.8220, underscoring its effectiveness in balancing precision and recall for fire segmentation tasks.

The importance of this research lies in its practical implications for real-world fire detection systems. The proposed models offer not only improved accuracy but also operational robustness, which is critical for timely and reliable fire detection. This advancement has the potential to significantly reduce the response time and mitigate the damage caused by fires, particularly in scenarios where early detection is crucial.

Furthermore, this work contributes to the growing body of knowledge on attention mechanisms in semantic segmentation, demonstrating their ability to enhance model performance by emphasizing salient features. By utilizing transfer learning, the study also addresses the challenge of limited fire datasets, enabling efficient model training without requiring extensive annotated data.

Future research will focus on scaling these models to larger datasets, exploring additional architectural innovations, and integrating real-time processing capabilities. The insights from this study serve as a foundational step towards developing intelligent, automated fire detection systems that can be deployed in diverse and challenging environments, ultimately improving disaster management and safety.

Acknowledgements

This work was supported by the Korea Planning & Evaluation Institute of Industrial Technology and funded by the Korean government (Ministry of Trade, Industry, and Energy, Ministry of Science and ICT, Ministry of the Interior and Safety, National Fire Agency, Project Title: Development of gas hydrate extinguishing agent and fire suppression technology for inaccessible fires, Project Number: 1761002860).

References

[1] Y. Kim, Y. Heo, B. Jin, Y. Bae, "Real-time fire classification models based on deep learning for building an intelligent multi-sensor system," *Fire*, vol. 7, no. 9, 329, 2024.

[2] P. Barmpoutis, P. Papaioannou, K. Dimitropoulos, and N. Grammalidis, "A review on early forest fire detection systems using optical remote sensing," *Sensors*, vol. 20, no. 22, 6442, 2020.

[3] M. Yandouzi, M. Grari, I. Idrissi, M. Boukabous, O. Mous-saoui, M. Azizi, K. Ghoumid and A. K. Elmiad, "Forest fires detection using deep transfer learning," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no.8, pp. 268-275, 2022.

[4] M. Niknejad and A. Bernardino, "Attention on classification for fire segmentation," in *Proceedings of 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Pasadena, CA, 2021, pp. 616-621.

[5] Y. Zheng, Z. Wang, B. Xu and Y. Niu, "Multi-scale semantic segmentation for fire smoke image based on global information and U-Net," *Electronics*, vol. 11, no.17, 2718, 2022.

[6] Y. Wang, Z. Luo, D. Chen, and Y. Li, "Semantic segmentation of fire and smoke images based on dual attention mechanism," in *Proceedings of 2022 4th International Conference on Frontiers Technology of Information and Computer (ICFTIC),* Qingdao, China, 2022, pp. 185-190.

[7] R. Ghali, M. A. Akhloufi, W. Souidene Mseddi, and M. Jmal, "Wildfire segmentation using deep-regseg semantic segmentation architecture," in *Proceedings of the 19th International Conference on Content-Based Multimedia Indexing*, Graz, Austria, 2022, pp. 149-154.

[8] S. Li, P. Liu, Q. Yan and R. Qian, "Optimized deep learning model for fire semantic segmentation," *Comput. Mater. Continua*, vol. 72, no. 3, pp. 4999-5013, 2022.

[9] L. Zhang, M. Wang, Y. Ding, T. Wan, B. Qi, and Y. Pang, "FBC-ANet: A semantic segmentation model for UAV forest fire images combining boundary enhancement and context awareness," *Drones*, vol. 7, no. 7, 456, 2023.

[10] Z. Wang, P. Yang, H. Liang, C. Zheng, J. Yin, Y. Tian and W. Cui, "Semantic segmentation and analysis on sensitive parameters of forest fire smoke using smoke-unet and landsat-8 imagery," *Remote Sens.*, vol. 14, no. 1, 45, 2021.

[11] Z. Guan, X. Miao, Y. Mu, Q. Sun, Q. Ye, and D. Gao, "Forest fire segmentation from aerial imagery data using an improved instance segmentation model," *Remote Sens.*, vol. 14, no. 13, 3159, 2022.