



OPEN

Automated forest fire detection in ecological monitoring using enhanced deep learning networks

Akhyar Akhyar¹, Jaesung Lee², Norazman Shahar¹, Adhi Harmoko Saputro³ & Mohd Asyraf Zulkifley¹✉

The incidents of forest fires are escalating in frequency and intensity because of climate change and land use pressures, thus posing serious risks to ecosystems, biodiversity, and human safety. Detecting and segmenting forest fire regions accurately from aerial imagery remains a major challenge due to variations in fire size, smoke density, terrain complexity, and illumination. Traditional convolutional models often struggle with these heterogeneous conditions, leading to inaccurate or delayed responses. This study introduces DeepFire S3GA-Net, a novel deep learning-based segmentation framework designed for forest fire segmentation using unmanned aerial vehicle imagery. The architecture is built on a fully convolutional encoder-decoder network, enhanced with atrous spatial pyramid pooling in the encoder for capturing multiscale contextual features and group convolution modules in the decoder to improve spatial refinement and feature diversity. These components allow the model to distinguish both fine-grained fire edges and broad burned regions effectively. Extensive evaluation using standard metrics of mean intersection over union, accuracy, precision, recall, and Dice coefficient show that DeepFire S3GA-Net consistently outperforms existing models. Its robustness across complex visual environments highlights its potential as a scalable and reliable tool for real-time forest fire monitoring and ecological risk management.

Keywords Forest fire detection, UAV imagery, Deep learning, Multiscale features, Semantic segmentation

Forests play a critical role in maintaining ecological balance by supporting biodiversity, regulating climate, and delivering essential ecosystem services such as carbon sequestration and water regulation^{1,2}. However, these ecosystems face escalating threats due to the increasing frequency and intensity of wildfire, which have become one of the most destructive natural disturbances globally³. Forest fire causes immediate destruction of vegetation and wildlife and leads to long-term ecological degradation, including soil erosion, altered hydrological cycles, and disrupted forest regeneration dynamics^{4,5}. Climate change manifested through rising temperatures, extended droughts, and shifting weather patterns has been closely linked to the proliferation of wildfire events⁶. In addition, anthropogenic factors such as deforestation, land use changes, and poor management practices have further heightened fire susceptibility, even in ecosystems previously considered low risk. These trends underscore the urgent need for proactive and adaptive fire monitoring strategies^{6,7}.

Conventional forest fire monitoring often relies on satellite remote sensing due to its wide spatial coverage and temporal reach. However, its limitations in spatial resolution, revisit time, and sensitivity to smoke and cloud cover often hinder accurate real-time detection^{8,9}. Unmanned aerial vehicles (UAVs) or drones have emerged as valuable complements, offering high-resolution imagery and flexibility for localized fire surveillance, yet they remain constrained by short flight durations, limited area coverage, and regulatory challenges^{10,11}. In this context, deep learning networks, especially convolutional neural networks (CNNs), have shown remarkable promise for automated forest fire segmentation. CNNs can extract hierarchical features from UAV and satellite images, enabling accurate detection under challenging conditions, such as dense smoke, uneven illumination, and complex landscapes^{12,13}. Architectures such as U-Net, SegNet, and FCDenseNet have shown strong performance in segmenting fire-affected regions. However, existing CNN-based methods often struggle to preserve fine-grained boundary details, especially under occlusion caused by smoke or shadows, and may suffer from over-smoothing in large-scale fire regions. Moreover, many prior works lack mechanisms to fully capture multiscale context or to encourage sufficient diversity in learned features. Recent innovations further enhance

¹Department of Electrical, Electronic and Systems Engineering, Universiti Kebangsaan Malaysia, UKM, 43600 Bangi, Selangor, Malaysia. ²Department of Artificial Intelligence, Chung-Ang University, Heukseok-Dong, Dongjak-Gu, Seoul 06974, Republic of Korea. ³Department of Physics, Faculty Mathematics and Natural Science, Universitas Indonesia, Depok 16424, Indonesia. ✉email: asyraf.zulkifley@ukm.edu.my

CNN segmentation performance. Multiscale feature extraction allows models to identify small fire edges and large burned areas, while group shuffle mechanisms improve feature diversity with low computational cost^{14–16}. These strategies are crucial for developing scalable and accurate forest fire detection systems. Figure 1 illustrates the segmentation challenge using UAV imagery, highlighting the need for robust models capable of coping with occlusion, terrain variation, and smoke interference.

To address these challenges, this study introduces a novel segmentation framework that enhances conventional CNN architecture for forest fire detection. The new model enhances conventional CNN structures by using multiscale atrous spatial pyramid pooling (ASPP) to collect information at various scales and adding a group convolution (GC) module to increase feature diversity. The ASPP component allows the model to capture spatial information at multiple receptive fields, thereby enhancing segmentation performance across various fire geometries and scales. Meanwhile, the GC module promotes richer feature learning by enabling more effective cross-group information exchange, helping the model to build more robust and discriminative representations from UAV imagery, which often exhibits diverse landscape textures and is affected by occlusions such as smoke and shadows. The major contributions of this study are as follows:

1. Novel architecture design. This study proposes a CNN-based segmentation model that integrates multiscale ASPP and GC modules for forest fire detection using UAV imagery. ASPP enhances the model's ability to capture fire features at different spatial scales, while GC improves feature diversity and computational efficiency.
2. Benchmark experimental evaluation. Controlled experiments on UAV datasets are conducted to assess the effects of ASPP and GC. ASPP improves segmentation accuracy through improved multiscale context extraction, while GC enhances generalization and model compactness. Combined, both modules considerably outperform the baseline.
3. Comparative performance analysis. A comparative evaluation is conducted between the proposed model and several established deep learning architectures, demonstrating the model's effectiveness in accurately segmenting forest fire-affected areas.

The remainder of this paper is structured as follows: The next section presents the literature review, followed by the methodology detailing the proposed segmentation model. The subsequent section outlines the experimental results and provides a discussion of the findings. The paper concludes with a summary of the key outcomes derived from the study.

Literature review

Forest fire segmentation has become a critical area of research, driven by the increasing availability of remote sensing data and advances in computer vision. In recent years, deep learning, particularly CNNs, has shown strong potential in improving the accuracy and responsiveness of fire monitoring systems. Numerous studies have utilized CNNs for fire detection and segmentation using satellite and UAV imagery. The integration of UAVs has further enhanced capabilities by offering high-resolution, real-time views of fire affected areas. For instance, Qin et al. (2021)¹⁷ developed a lightweight fire detection system that combines depthwise separable CNN with YOLOv3, achieving a mean average precision (mAP) of 89.65% and 24.32 FPS. Its efficiency and compactness make it suitable for real-time UAV-based deployment. Similarly, Zhang et al. (2023)¹⁸ introduced MS-FRCNN, a multiscale Faster R-CNN model for detecting small-scale fires, which achieved a detection accuracy of 91.3% and an mAP of 87.6%. Meanwhile, Huot et al. (2020)¹⁹ proposed a deep learning model for wildfire prediction using a decade of remote sensing data, achieving an AUC of 83%, confirming its effectiveness for large-scale risk assessment. However, while these detection-oriented models have demonstrated high accuracy, they typically lack fine-grained segmentation capabilities necessary for delineating fire boundaries in UAV imagery.

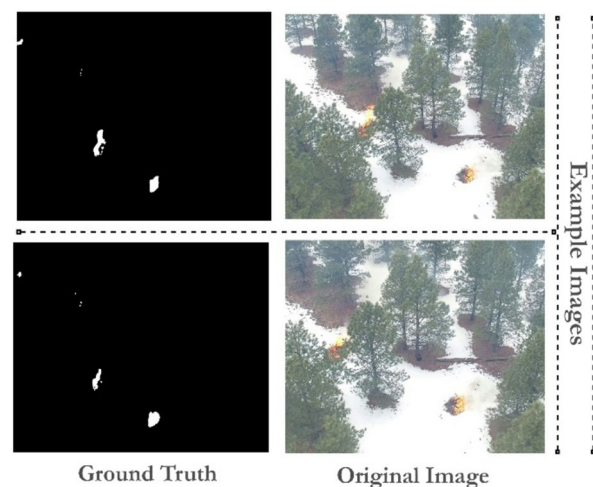


Fig. 1. Example of UAV forest fire image with corresponding ground truth segmentation.

Building on earlier work in wildfire prediction, several recent studies have advanced the application of deep learning for active fire detection and segmentation. De Almeida et al. (2021)²⁰ introduced a large-scale dataset for fire detection using Landsat-8 imagery, achieving a precision of 87.2% and recall of 92.4% on annotated datasets. Dewangan et al. (2022)²¹ developed the Fire Ignition Library (FIgLib), comprising approximately 25,000 labeled wildfire smoke images, and proposed SmokeyNet, a deep learning model that utilizes spatiotemporal features from static camera imagery. SmokeyNet achieved 92.1% accuracy, outperforming ConvLSTM and 3D-CNN baselines and demonstrating performance comparable to human annotators. Kang et al. (2022)²² presented Fire-Net, a multiscale CNN that integrates spectral, spatial, and temporal features from Himawari-8 satellite imagery. Fire-Net achieved high detection accuracy and reduced average detection time by 12 min compared with conventional satellite methods. However, most of these models emphasize detection rather than fine-grained segmentation, limiting their use in detailed fire mapping. Recently, Muksimova et al. (2024)²³ proposed a modified Miti-DETR architecture for UAV-based wildfire detection, integrating a redesigned AlexNet backbone with residual depthwise separable convolution and a residual self-attention mechanism. The model demonstrated enhanced detection accuracy and inference efficiency on the FLAME dataset, achieving a precision of 0.86 and recall of 0.84, particularly under complex aerial and low-visibility conditions, making it well suited for real-time wildfire monitoring using UAV imagery. However, the model cannot effectively delineate fine-grained fire boundaries, which are crucial for precise fire mapping and emergency response.

This growing body of research underscores how the integration of UAV platforms with advanced deep learning frameworks is reshaping fire detection systems, enabling more accurate, real-time analysis across diverse environmental conditions. In line with this progress, Zhan et al. (2021)²⁴ proposed PDAM-STPNNet, a deep learning framework for detecting wildland fire smoke using remote sensing data from fixed wing and multirotor UAVs. The model incorporates spatiotemporal feature learning and attention mechanisms, enhancing its ability to recognize smoke patterns in complex visual environments. PDAM-STPNNet achieved a detection accuracy of 94.3% and reduced the false alarm rate by 21% compared with baseline models, demonstrating the value of combining UAV flexibility with intelligent image analysis for real-time monitoring. Building on these developments, Gong et al. (2023)²⁵ introduced a lightweight model for detecting forest fire smoke (LMDFS) based on the YOLOv7 architecture and optimized it for UAV deployment. LMDFS attained an mAP of 91.7% while maintaining a lightweight design suitable for edge computing. Compared with YOLOv5 and YOLOv4, LMDFS improved inference speed by 32.6%, offering a practical solution for fast, accurate smoke detection in constrained environments. Further advancing the field, Niu et al. (2023)²⁶ developed YOLOv5s-Seg, a segmentation model designed for UAV-based infrared imagery. The model integrates object detection with semantic segmentation to address challenges such as smoke occlusion and low illumination. Achieving a precision of 92.4%, recall of 90.8%, and intersection over union (IoU) of 88.1%, YOLOv5s-Seg outperformed conventional infrared detection methods. These findings emphasize the growing importance of deep learning-enhanced UAV systems for precise, responsive wildfire detection and segmentation. Nonetheless, many of these models focus primarily on detection tasks and provide limited segmentation granularity for precise fire boundary mapping.

In parallel with these segmentation-focused developments, other researchers have concentrated on improving the speed, efficiency, and scalability of real-time fire detection using lightweight deep learning models. Wang et al. (2021)²⁷ developed a real-time detection system using a MobileNetV3-YOLOv4 framework, achieving 94.6% accuracy while reducing inference time to 37.2 ms per image and memory usage by 42%, making it ideal for resource-constrained UAV platforms. Building on this, Yun et al. (2025)²⁸ proposed an enhanced YOLOv8-SPDConv model optimized for detecting small-scale and blurred fire targets in UAV imagery. Their architecture achieved an mAP of 89.5% and improved small fire detection by 12.6% over baseline YOLOv8, showing strength in early-stage wildfire conditions with low visibility. Complementing these efforts, Marques et al. (2024)²⁹ introduced a modular CNN-based system that integrates UAV imagery with multisensor data, including visual and thermal inputs for forest fire detection and monitoring. Their model achieved an F1-score of 0.93, precision of 0.91, and recall of 0.94 on a diverse, multienvironment dataset. This sensor fusion approach enabled distributed and reliable fire monitoring across broad ecological landscapes, underscoring the value of combining deep learning with multisource UAV data for improved situational awareness and fire response.

While multisensor integration enhances detection reliability, recent advancements in CNN architecture have also focused on improving segmentation accuracy from UAV imagery. Feng et al. (2025)³⁰ introduced U3UNet, a nested U-shaped network designed for precise wildfire segmentation using UAV imagery. Published in neurocomputing, the model achieved excellent performance with a Dice coefficient of 0.923, IoU of 0.887, and pixel accuracy of 98.6% on the UAV-FF dataset. Its strength lies in retaining fine spatial features through enhanced multilevel feature fusion, making it suitable for generating detailed fire maps vital for situational awareness. Fralenko (2024)³¹ conducted an analytical review of neural network approaches for early forest fire detection using UAV imagery and video. The study highlighted CNN modifications that improved F1-scores by up to 20%, particularly in scenarios requiring fast response and adaptability in forest environments. Collectively, these studies highlight how well CNN-based segmentation and detection models, especially when integrated with UAV platforms, help advance the precision, speed, and operational scalability of forest fire monitoring systems. Building responsive early warning systems in areas prone for wildfires depends on these developments.

Methodology

This study proposes a deep learning-based segmentation framework designed to enhance forest fire detection using UAV imagery. The model adopts a fully convolutional encoder–decoder architecture with targeted enhancements to improve feature extraction and reconstruction of the feature maps. ASPP modules are integrated at key stages of the encoder to support contextual understanding, enabling multiscale feature extraction without

adding considerable computational complexity. These modules help the model identify fine fire contours and broad burned areas across diverse environments. In the decoder, a GC module is employed to enrich feature diversity and spatial consistency by efficiently redistributing grouped feature maps across channels. This design maintains lightweight performance while enhancing representation quality. Combined, these components establish a cohesive architectural strategy for delivering accurate and adaptive fire segmentation in complex scenes. The next subsection presents the detailed design of the proposed framework.

Architectural design of the proposed framework

The proposed framework builds its foundational architecture on a conventional CNN-based image segmentation approach, utilizing a symmetrical encoder–decoder structure. This design allows for progressive feature extraction in the encoder and structured spatial reconstruction in the decoder, forming a robust backbone for forest fire segmentation. The encoder has three parts, each made up of several convolutional blocks, and each part reduces the size of the image while capturing more complex features. Specifically, the first block uses two convolutional layers with 64 filters each, the second with 128 filters, and the third comprises three convolutional layers with 256 filters each. All layers are followed by ReLU activations, enabling the model to learn intricate patterns such as the texture and edge contours commonly associated with forest fire imagery³².

The decoder is carefully designed to reconstruct spatial resolution while preserving critical feature information and thus complement the encoding process. It mirrors the encoder's structure but utilizes 2D transposed convolutions for upsampling. These layers progressively reconstruct image dimensions while refining features critical for accurate segmentation. Skip connections are integrated between corresponding encoder and decoder layers to preserve spatial information lost during down sampling, enhancing the model's ability to capture fine fire boundaries and subtle spread variations^{33,34}.

Although effective, this baseline architecture has limits in complex scenes with high variability in fire intensity and background texture. These obstacles emphasize the necessity of a more diverse feature learning approach and improved multiscale feature representation. To overcome these challenges, this study introduces DeepFire S3GA-Net, a novel fully convolutional encoder–decoder model designed to enhance forest fire segmentation performance. The architecture is augmented with ASPP and GC modules. These enhancements augment the model's capacity to discern coarse and fine characteristics, hence improving its robustness in practical forest fire monitoring contexts, as illustrated in Fig. 2.

One of the key innovations in this architecture is the integration of ASPP modules within the encoder to enable more effective multiscale feature extraction. Specifically, two ASPP modules are strategically embedded into the encoder: one at the end of the convolutional block with 64 filters and another at the end of the 128-filter block, both positioned immediately before their respective max pooling layers. These modules are designed to enhance the model's ability to extract multiscale contextual features without degrading spatial resolution. By applying parallel atrous convolutions with varying dilation rates, the ASPP modules allow the model to capture fine and coarse fire-related patterns early in the encoding process.

The rationale for positioning the ASPP modules in early encoder stages lies in the need to capture multiscale contextual features before significant spatial down-sampling occurs. This placement enables the model to retain fine-grained boundary details while simultaneously encoding broader contextual information, a key factor in distinguishing flame regions from surrounding vegetation or smoke. Preliminary ablation results showed that inserting ASPP after the first and second convolutional block improved mean IoU compared with placement in deeper encoder layers.

Simultaneously, the decoder incorporates GC modules at the final stages of the 256-filter and 128-filter decoder blocks. These modules promote efficient feature recombination by distributing grouped feature maps across channels, thereby enhancing spatial diversity and minimizing redundancy within the learned representations. This addition plays a critical role in strengthening feature refinement during upsampling processes, especially in spatially complex fire scenes. The GC module was specifically selected for its ability to perform lightweight channel reorganization with minimal parameter overhead, effectively balancing accuracy and computational efficiency.

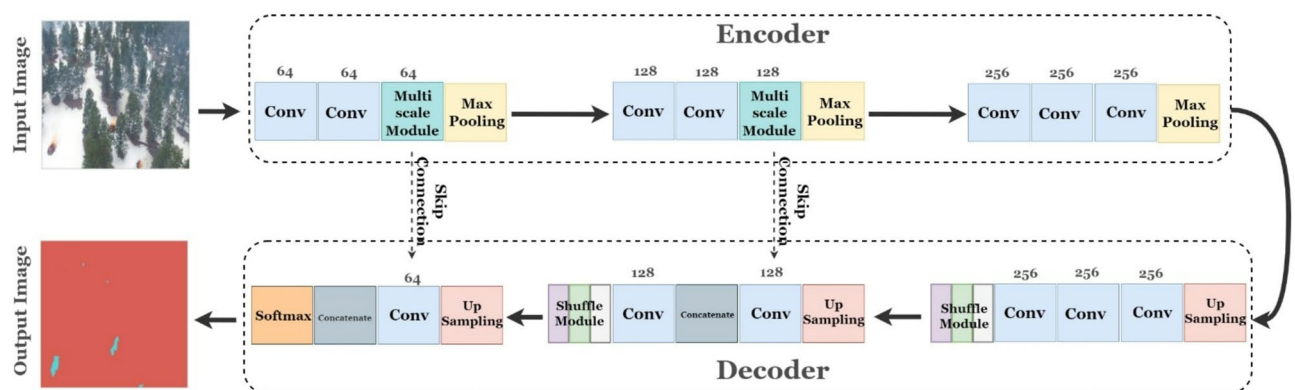


Fig. 2. Proposed model: DeepFire S3GA-Net architecture.

Skip connections are applied between the ASPP outputs in the encoder and their corresponding decoder stages after upsampling operations to support further the integration of contextual and spatial information. These concatenated connections help merge multiscale contextual features with decoder representations, allowing the network to better retain subtle fire boundaries throughout reconstruction. This design harmonizes local spatial precision from the encoder with global contextual cues derived from the ASPP–GC fusion, ensuring balanced segmentation performance across varying fire intensities. By harmoniously integrating ASPP, GC modules, and targeted skip connections, DeepFire S3GA-Net achieves a balanced design capable of accurately segmenting fire-affected regions in visually diverse and challenging environments. The next section presents the detailed architectural configuration of these components.

Atrous spatial pyramid pooling (ASPP)

ASPP is a widely adopted semantic segmentation technique that remarkably enhances a model's ability to capture contextual information at multiple spatial scales. By applying parallel convolutions with varying dilation rates, ASPP expands the receptive field without increasing computational cost, making it particularly suitable for detecting objects of varying shapes and sizes such as smoke and flames in complex aerial or satellite imagery³⁵. In the proposed framework, ASPP is strategically embedded within the encoder to address the scale variability commonly observed in forest fire imagery. Given that fire regions often appear in different intensities, contours, and sizes, fixed receptive fields of traditional convolution layers are insufficient. Hence, two ASPP modules are integrated after the 64-filter and 128-filter encoder blocks, just before their respective max pooling operations. These placements ensure that multiscale features are captured early in the encoding process, improving the model's ability to adapt to heterogeneous fire patterns³⁶.

Each ASPP module is composed of specialized convolutional branches designed to extract local and global features. Specifically, the modules include parallel convolutions with kernel sizes of 1×1 , 2×2 , and 3×3 , combined with dilation rates of 2, 4, and 6. This configuration enables the model to gather spatial features from various receptive fields while preserving resolution, as shown in Fig. 3. A global average pooling branch is also included to capture long-range dependencies, allowing the model to differentiate true fire zones from visually similar artifacts such as sunlit terrain or haze.

Although the ASPP modules remarkably improve multiscale feature extraction and contextual awareness, additional refinement is required to enhance spatial detail preservation and increase feature diversity during reconstruction. The ASPP outputs, processed at different stages of the network, help retain fine-grained fire

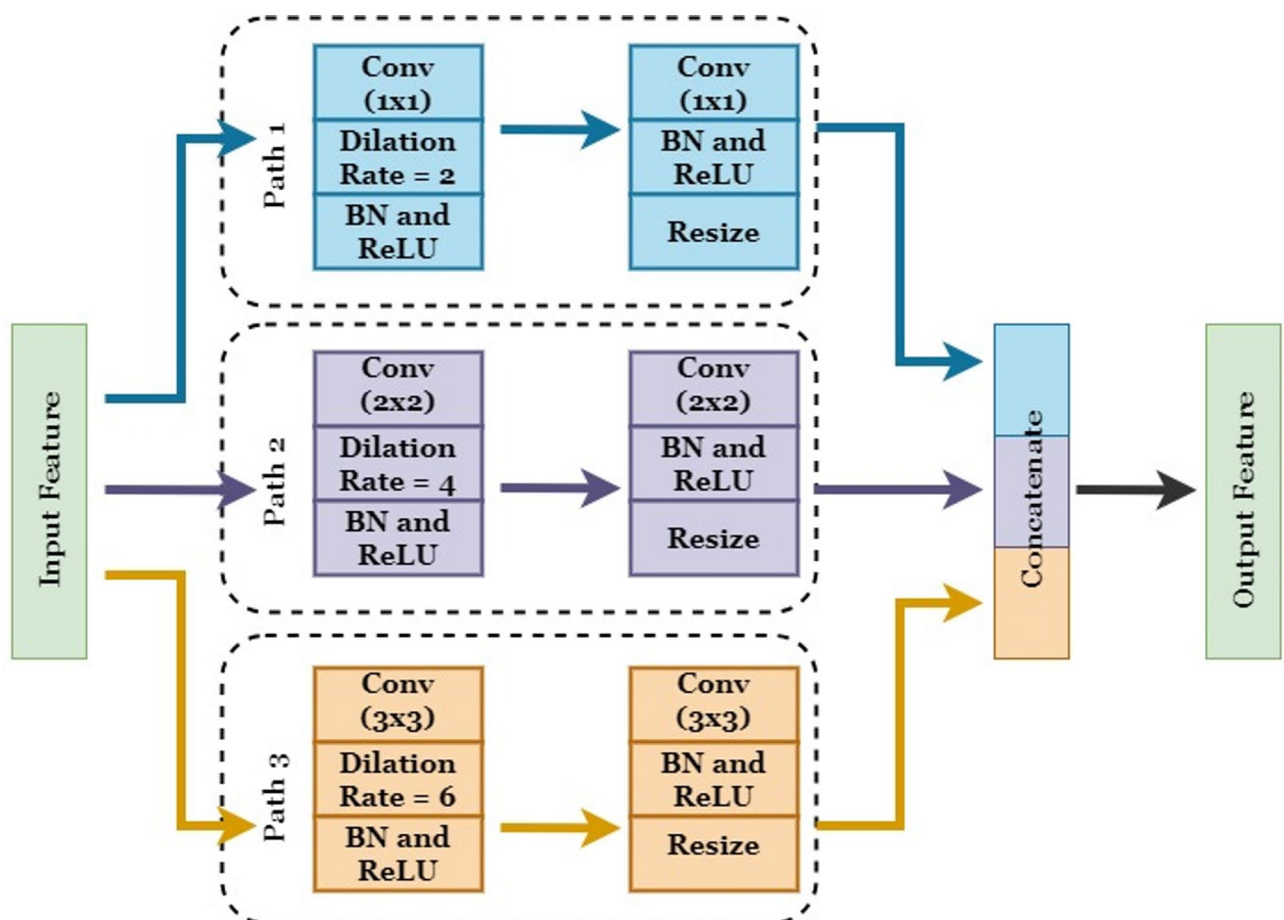


Fig. 3. Multiscale ASPP module.

characteristics and broad spatial context, enabling more accurate segmentation in complex environments. However, segmentation performance in visually challenging scenarios can still be limited without proper feature reorganization in the decoding process. To address this, the following subsection introduces the GC module, designed to enrich the representational capacity of the decoder by improving channel-wise diversity and spatial consistency while maintaining computational efficiency suitable for real-time forest fire monitoring applications.

Group convolution module

GC is a computationally efficient variant of standard convolutional operations in CNNs, designed to reduce the number of parameters and computational cost while potentially enhancing feature diversity. In standard convolution, each filter connects to all input channels. By contrast, GC divides the input channels into g distinct groups. Convolution filters are also divided into the same number of groups, and each filter group is only convolved with its corresponding input channel group^{37,38}. Mathematically, for an input feature map $X \in \mathbb{R}^{H \times W \times C_{in}}$ and output feature map $Y \in \mathbb{R}^{H' \times W' \times C_{out}}$, where C_{in} is divided into g groups of size C_{in}/g , and similarly for C_{out} into C_{out}/g filters per group, the convolution operation for the i^{th} group can be represented as $Y_i = X_i * K_i$, where X_i is the i^{th} group of input channels, K_i is the i^{th} group of filters, and $*$ denotes the convolution operation. The final output Y is then a concatenation of all group outputs $Y = [Y_1, Y_2, \dots, Y_g]$. This approach effectively reduces the parameter count by a factor of g compared with standard convolution with the same number of filters³⁹.

However, GC can limit information flow between different channel groups, potentially hindering representational power. Thus, channel shuffle is introduced as a complementary operation, particularly in efficient CNN architectures such as ShuffleNet. Channel shuffle aims to enable cross group information exchange, enhancing feature integration across all input channels. Following GC, channel shuffle rearranges the feature channels. If we consider the output of GC with dimensions $(N, H', W', G \times C')$, where G is the number of groups and C' is the number of channels per group, channel shuffle first reshapes the output to (N, H', W', G, C') , then transposes the last two dimensions to (N, H', W', C', G) , and finally flattens it back to $(N, H', W', G \times C')$. This sequence of operations effectively shuffles the channel order, allowing features learned within different channel groups in the GC to be mixed and combined in subsequent layers, thus improving overall network performance^{40,41}.

The combined use of GC and channel shuffle provides a powerful strategy for designing efficient and effective CNN architectures, especially for resource-constrained environments. GC reduces computational cost, while channel shuffle mitigates the information isolation problem. These techniques are pivotal in models aimed at deployment on mobile and embedded devices, where computational resources are limited⁴². As shown in Fig. 4, the combination of GC and channel shuffle in DeepFire S3GA-Net enhances segmentation by improving spatial awareness and channel-wise feature interaction, ultimately leading to superior fire region detection and classification performance.

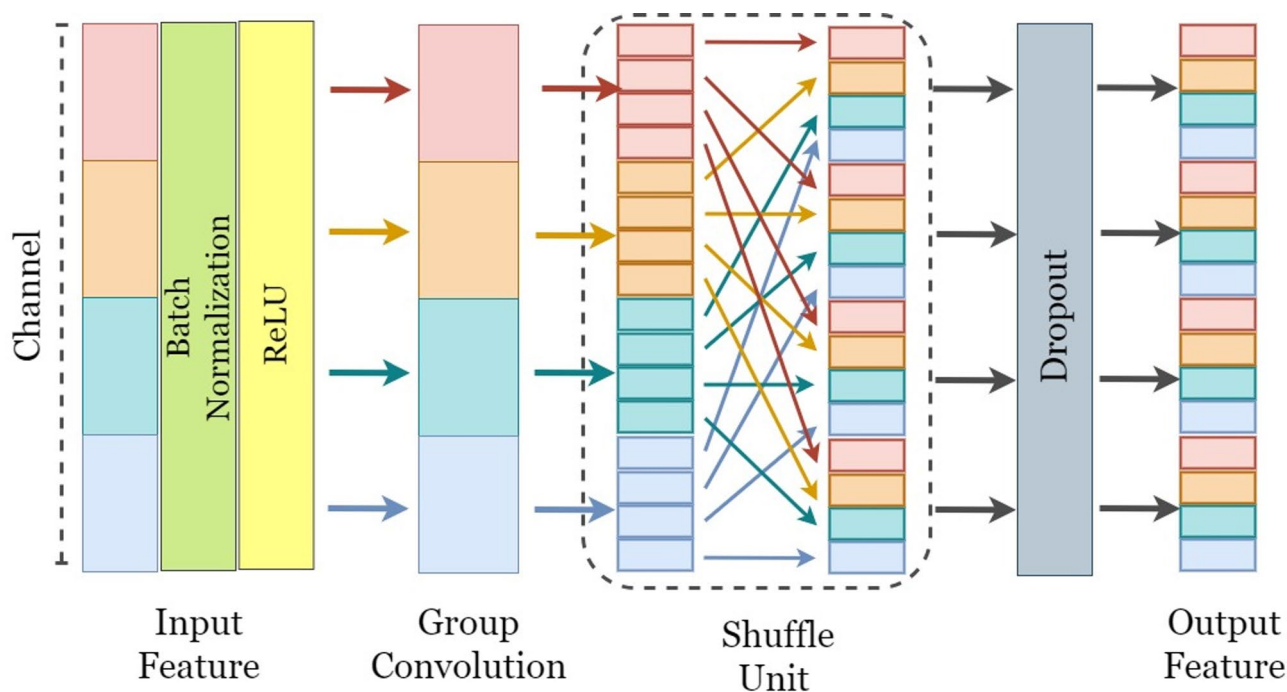


Fig. 4. Group convolution module.

Results

This section presents the simulation results obtained from the evaluation of DeepFire S3GA-Net for forest fire segmentation. The performance of the proposed model is analyzed using standard segmentation metrics, including IoU, accuracy, precision, recall, and Dice coefficient, to assess its effectiveness in distinguishing fire-affected regions from the background. Comparative evaluations are conducted against state-of-the-art deep learning architectures to highlight the improvements achieved by DeepFire S3GA-Net.

Dataset

The dataset used in this work comes from the work of Shamsoshoara et al. (2021)⁴³, which provides a complete set of forest fire photos specifically meant for segmentation needs. Given its vast coverage of a wide range of fire situations, including differences in weather conditions, fire intensity, and backdrop complexity, this dataset is fit for evaluating the effectiveness of deep learning models in practical applications. The dataset has 2,000 images split into 1,600 images for training and 400 images for testing. Every image is available in two forms: ground truth masks, which offer pixel wise annotations of fire areas, and original images, including raw forest fire sequences. For supervised learning, these annotations are essential because they let the model efficiently distinguish between fire and nonfire zones. The dataset ensures the model's generalizability and robustness by including photos taken in a range of environments, such as those with different lighting, smoke levels, and background complexity.

Performance metrics

Several standard performance metrics are utilized to evaluate comprehensively the performance of DeepFire S3GA-Net in forest fire segmentation, including IoU and accuracy. These metrics provide a detailed assessment of the model's accuracy, robustness, and ability to distinguish fire regions from nonfire areas. IoU, also known as the Jaccard index, measures the ratio of correctly predicted fire pixels to the total number of pixels in the union of the predicted and ground-truth regions⁴⁴. It can be calculated using the following equation:

$$IoU(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

where A is the predicted segmentation mask, and B is the ground-truth mask. A higher IoU value indicates better segmentation accuracy.

Accuracy in segmentation tasks is a more general metric that measures the proportion of correctly classified pixels over the entire image. It can be calculated using the following equation:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

In forest fire segmentation tasks, model evaluation commonly relies on the classification of pixels into four categories: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). TP refers to fire pixels correctly identified as fire, indicating the model's success in detecting affected regions. TN represents nonfire pixels accurately labeled as nonfire, reflecting the model's ability to distinguish background or unaffected areas. Misclassifications occur when FP labels nonfire pixels as fire, potentially triggering false alarms, or when FN classifies actual fire pixels as nonfire, leading to missed detections^{45–47}. These components serve as the foundation for several standard performance metrics in deep learning-based segmentation.

Key metrics include precision, recall, and Dice coefficient, all of which provide insight into model effectiveness. Precision measures how many predicted fire pixels are truly fire, with higher values indicating fewer false alarms. It can be calculated using the following equation:

$$Prec = \frac{TP}{TP + FP} \quad (3)$$

Recall assesses the model's ability to detect all actual fire regions, where higher scores signify fewer missed detections, and it can be mathematically expressed as follows:

$$Rcl = \frac{TP}{TP + FN} \quad (4)$$

The Dice coefficient measures the degree of overlap between the predicted and ground-truth masks, making it particularly useful for imbalanced datasets where fire pixels are sparse. The corresponding equation is expressed as follows:

$$Dice = \frac{2TP}{2TP + FP + FN} \quad (5)$$

Together, these metrics guide performance tuning and help assess the practical reliability of segmentation models in real-world wildfire scenarios^{48,49}.

Benchmarking performance evaluation

The effectiveness of DeepFire S3GA-Net in forest fire segmentation is validated through rigorous performance evaluation and comparative analysis against several benchmarking deep learning models. This evaluation is designed to assess the model's ability to identify fire regions across diverse environmental conditions accurately, including varying smoke densities, fire intensities, and complex background textures. Standard segmentation metrics such as mean IoU, pixel accuracy, precision, Dice coefficient, and recall are employed to provide a comprehensive performance assessment.

As shown in Table 1, DeepFire S3GA-Net achieves the highest performance across all evaluated metrics. It records a mean IoU of 0.8723, which reflects a strong agreement between the predicted segmentation masks and the ground-truth fire regions. This value considerably surpasses that of the next best model, SegNet, which achieves an IoU of 0.7942, indicating a more accurate boundary delineation of fire-affected regions. In terms of pixel classification accuracy, DeepFire S3GA-Net achieves 99.84%, the highest among all models tested, demonstrating its precision in distinguishing fire from non-fire pixels.

Moreover, the model yields a precision score of 0.9347, surpassing DABNet's 0.8569, the next best-performing model in this metric. This high precision indicates DeepFire S3GA-Net's effectiveness in minimizing false positives, a crucial requirement in real-time monitoring scenarios. The Dice coefficient of 0.9269 further underscores the model's robustness in capturing spatial overlap between predicted and actual fire regions, exceeding SegNet's Dice score of 0.8808. Additionally, DeepFire S3GA-Net attains a recall score of 0.9194, which demonstrates its superior sensitivity in detecting true fire regions, outperforming FCDenseNet (0.8987) and U-Net (0.8873).

In terms of computational performance, DeepFire S3GA-Net achieves an inference time of 13.01 ± 0.05 ms per image, translating to a frame rate of 76.88 FPS, which confirms its ability to deliver real-time segmentation without compromising accuracy. This performance is comparable to SegNet (12.86 ms, 77.78 FPS) and U-Net (12.75 ms, 78.41 FPS), but with significantly better accuracy across all metrics, demonstrating a well-balanced trade-off between speed and precision.

This quantitative superiority is further supported by the qualitative results illustrated in Fig. 5. Compared with other models, the output image generated by DeepFire S3GA-Net more closely matches the ground truth, demonstrating clearer segmentation boundaries and better localization of fire regions. Models such as FCN8 and U-Net show evident undersegmentation or fragmented detection, failing to capture certain fire-affected areas. SegNet and DABNet perform moderately well but still exhibit inconsistencies in identifying small or dispersed fire regions. Meanwhile, DeepFire S3GA-Net produces a segmentation map that preserves the shape and spatial extent of the fire zones with minimal noise or false detections. This visual evidence reinforces the numerical results, confirming the model's ability to combine high precision, recall, and Dice similarity into more accurate and consistent segmentation outputs. These findings highlight DeepFire S3GA-Net's capability to deliver statistically and visually reliable results in complex forest fire detection scenarios.

Collectively, these results illustrate that DeepFire S3GA-Net not only achieves strong performance in general classification and segmentation but also maintains high accuracy and consistency across all critical evaluation metrics. Its outstanding performance highlights its potential for reliable deployment in real world forest fire detection and ecological monitoring systems.

Ablation study

An ablation study was conducted to quantify the impact of the proposed ASPP and GC modules and thus further validate the contributions of the individual architectural components within DeepFire S3GA-Net. This analysis isolates and evaluates the effect of each enhancement by incrementally integrating them into a common baseline encoder decoder architecture. The objective is to assess how each component contributes to segmentation accuracy, spatial coherence, and feature representation fidelity. As shown in Table 2, the baseline model (BL) achieves a strong starting performance with a mean IoU of 0.8562, accuracy of 99.82%, and recall of 0.9217, reflecting a well-optimized conventional CNN segmentation framework. Adding one ASPP module (BL + ASPP) results in a slight improvement, with mean IoU increasing to 0.8590, Dice to 0.9181, and recall to 0.9239, suggesting a modest enhancement from a single multiscale context module. However, adding a second ASPP module (BL + 2 ASPP) leads to a minor drop in mean IoU to 0.8533, while Dice remains relatively high at 0.9142, indicating that increasing encoder context alone may saturate gains without parallel improvements in the decoder.

Model	Performance metrics							
	Mean IoU	Accuracy (%)	Precision	Dice	Recall	Flop (G)	Inferences time (ms/image)	FPS (image/s)
FCN8	0.5290	99.44	0.6734	0.5584	0.5356	30.713	4.77 ± 0.07 ms	209.71
PSPNet	0.7022	99.62	0.8642	0.7889	0.7397	53.358	17.26 ± 0.20 ms	57.94
U-Net	0.7178	99.48	0.7514	0.8047	0.8873	83.768	12.75 ± 0.07 ms	78.41
DABNet	0.7313	99.64	0.8569	0.8170	0.7852	3.611	11.72 ± 1.04 ms	85.34
FCDenseNet	0.7686	99.64	0.8121	0.8500	0.8987	58.046	41.03 ± 0.21 ms	24.37
SegNet	0.7942	99.70	0.8602	0.8808	0.8702	49.611	12.86 ± 0.25 ms	77.78
DeepFire S3GA-Net	0.8723	99.84	0.9347	0.9269	0.9194	40.510	13.01 ± 0.05 ms	76.88

Table 1. Comparative performance of deepfire S3GA-Net and benchmark segmentation Models.

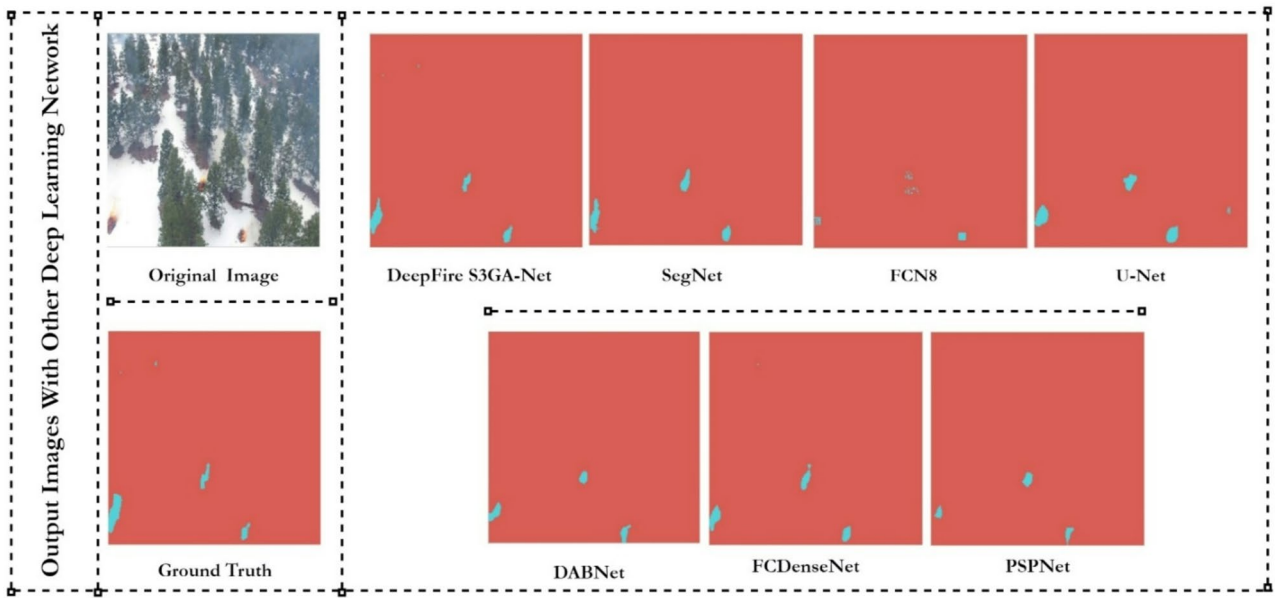


Fig. 5. Output Image DeepFire S3GA-Net with other deep learning networks.

Model	Performance metrics				
	Mean IoU	Accuracy (%)	Precision	Dice	Recall
BL	0.8562	99.82	0.9108	0.9162	0.9217
BL + ASPP	0.8590	99.82	0.9124	0.9181	0.9239
BL + 2 ASPP	0.8533	99.81	0.9113	0.9142	0.9172
BL + GC	0.5442	99.49	0.8278	0.5842	0.5490
BL + 2 GC	0.5083	99.48	0.9857	0.5199	0.5109
BL + ASPP + GC	0.6280	99.56	0.8602	0.7055	0.6442
BL + ASPP + 2GC	0.8180	99.77	0.9005	0.8890	0.8782
BL + 2 ASPP + GC	0.5899	99.54	0.8957	0.6545	0.5966
DeepFire S3GA-Net (BL + 2 ASPP + 2 GC)	0.8723	99.84	0.9347	0.9269	0.9194

Table 2. Impact of ASPP and group convolution modules on baseline performances.

In contrast, when a GC module is added without ASPP (BL + GC or BL + 2 GC), segmentation performance declines sharply mean IoU falls to 0.5442 and 0.5083, respectively, despite relatively high precision values (0.8278 and 0.9857). This suggests that decoder-level transformation without supporting contextual cues from the encoder can impair spatial learning and cause overfitting or poor generalization. A more balanced configuration (BL + ASPP + GC) yields moderate performance (mean IoU: 0.6280, Dice: 0.7055), but still falls short due to underutilization of decoder refinement. However, when two GC modules are integrated alongside one ASPP (BL + ASPP + 2 GC), performance increases significantly, achieving a mean IoU of 0.8180 and Dice of 0.8890, confirming that deeper decoder grouping improves reconstruction consistency and feature refinement.

The configuration combining two ASPP modules and only one GC (BL + 2 ASPP + GC) shows suboptimal results (mean IoU: 0.5899, Dice: 0.6545), suggesting that a single GC is insufficient to process the denser multiscale features generated in the encoder. Finally, the complete DeepFire S3GA-Net configuration (BL + 2 ASPP + 2 GC) delivers the best overall performance, with mean IoU of 0.8723, accuracy of 99.84%, precision of 0.9347, Dice of 0.9269, and recall of 0.9194. These results affirm that synergistic integration of encoder multiscale context (ASPP) and decoder grouping (GC) is crucial for achieving superior segmentation fidelity, especially in complex fire environments.

The qualitative output results further support these quantitative findings, as visualized in Fig. 6. Compared with other configurations, the final model (BL + 2 ASPP + 2 GC) produces the most accurate segmentation maps, closely aligning with the ground truth in both shape and spatial extent. Simpler variants such as BL + ASPP or BL + 2 ASPP provide moderate improvement but struggle to delineate finer fire boundaries. Meanwhile, configurations using only one GC module, such as BL + ASPP + GC, suffer from blurry and incomplete segmentations. In contrast, DeepFire S3GA-Net demonstrates sharp, well-preserved fire zones, effectively capturing small and dispersed fire patterns that are commonly missed. This validates that multiscale encoder

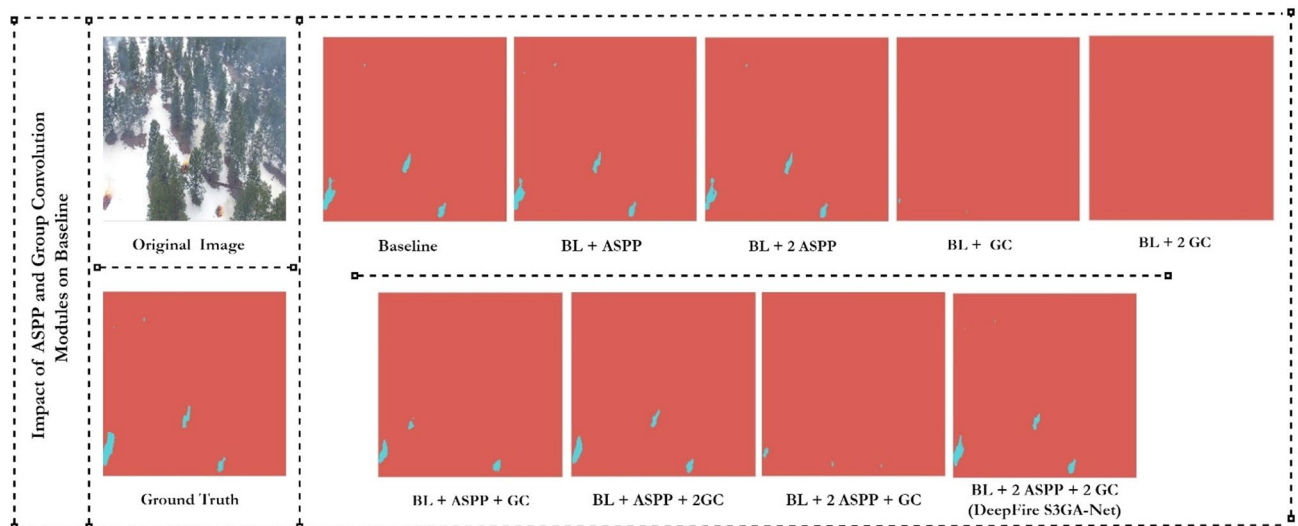


Fig. 6. Output image for impact ASPP and group convolution module on baseline.

context combined with grouped decoder refinement substantially enhances the model's ability to handle variable fire shapes, densities, and textures in real-world forest fire scenarios.

Discussion

The evaluation of DeepFire S3GA-Net against several state-of-the-art segmentation models clearly demonstrates its superior performance in forest fire segmentation tasks. As reported in Table 1, DeepFire S3GA-Net achieves the highest scores across all key performance metrics, including mean IoU (0.8723), accuracy (99.84%), precision (0.9347), Dice coefficient (0.9269), and recall (0.9194). These results reflect the model's ability to distinguish between fire and nonfire regions effectively with a high degree of spatial and semantic precision, which is critical in ecological monitoring where oversegmentation and undersegmentation carry operational risks.

Compared with classic models such as SegNet (IoU=0.7942) and U-Net (IoU=0.7178), DeepFire S3GA-Net shows considerable improvement, particularly in detecting fine-grained fire boundaries. Although these architectures provide a strong foundation through encoder-decoder designs, they lack explicit multiscale context integration and advanced spatial refinement mechanisms. Similarly, more recent models such as FCDenseNet and DABNet exhibit competitive Dice and recall scores, but they still fall short in overall IoU and precision compared with DeepFire S3GA-Net. This result indicates that while dense connectivity and attention mechanisms contribute to feature richness, they are not sufficient without enhanced spatial decoding strategies.

In addition to the quantitative performance metrics, the visual comparison of segmentation outputs further validates the effectiveness of the proposed model. As shown in Fig. 6, DeepFire S3GA-Net demonstrates superior capability in accurately delineating fire regions with minimal misclassification, even in visually complex environments. In contrast to models such as FCN8, which exhibits oversimplified segmentations, or U-Net, which tends to miss fragmented fire zones, DeepFire S3GA-Net produces output that closely aligns with the ground truth in terms of shape and spatial coverage. This consistency reflects the model's enhanced multiscale contextual understanding and robust spatial decoding, driven by the integration of ASPP and GC modules. The visual clarity and completeness of DeepFire S3GA-Net's output also suggest improved generalization in handling noise, background heterogeneity, and subtle fire boundaries, which are common challenges in UAV-based forest fire imagery. Overall, these visual results reinforce the reliability of DeepFire S3GA-Net for practical implementation in real-time wildfire monitoring systems.

The results from Table 2 further highlight the incremental impact of the architectural components embedded within DeepFire S3GA-Net. Starting from the baseline model, which achieves a strong mean IoU of 0.8562, the addition of a single ASPP module (BL + ASPP) yields a slight improvement in IoU (0.8590) and Dice (0.9181). However, adding a second ASPP (BL + 2 ASPP) slightly reduces the IoU to 0.8533, indicating that increasing contextual paths in the encoder without appropriate decoder support may lead to diminishing returns. This is likely due to oversmoothing and feature redundancy caused by excessive context aggregation without adequate spatial differentiation, leading to blurred boundaries and reduced segmentation sharpness.

The role of the GC module is particularly notable. When a single GC is added to the baseline (BL + GC), performance significantly drops to IoU = 0.5442 and Dice = 0.5842, suggesting that decoder refinement without adequate encoder support disrupts spatial consistency. Adding two GC modules (BL + 2 GC) further reduces the IoU to 0.5083, albeit with a very high precision (0.9857), but at the cost of poor Dice and recall.

Interestingly, a model configuration that includes two ASPP modules but only one GC module (BL + 2 ASPP + GC) performs even worse (IoU = 0.5899), reinforcing the importance of architectural balance. In this case, the rich multiscale features generated by ASPP in the encoder are not effectively utilized because of inadequate spatial refinement in the decoder. This outcome indicates that without proper decoding support, the

encoder's representational gains are not translated into improved segmentation output. It highlights the need for well-paired enhancements in the encoder and decoder to maximize performance.

The segmentation results underscore the critical role of the GC module, particularly the embedded shuffle unit, in enhancing decoder performance. In the DeepFire S3GA-Net configuration (BL+2 ASPP+2 GC), fire boundaries are not only more complete but also spatially coherent compared with other variants. The shuffle mechanism enhances intergroup feature interaction by redistributing channel-wise information, reducing redundancy and increasing feature diversity. This results in richer representations that better capture small, scattered, or partially occluded fire areas an essential capability for real-world forest fire monitoring. By contrast, configurations such as BL+2 ASPP+GC, which incorporate only a single GC, show limited decoder refinement, leading to visibly degraded segmentation despite sufficient contextual encoding from the ASPP modules. These observations suggest that the decoder's ability to integrate features spatially and maintain detail relies heavily on the presence and completeness of group-based refinement.

Figure 6 further supports these findings through a qualitative comparison of segmentation outputs across different architectural variants. While the baseline model detects primary fire regions, it fails to capture fine edges and smaller fire patches. The inclusion of one or two ASPP modules improves contour accuracy modestly, aligning with their slight metric gains in Table 2. However, performance drops noticeably when decoder refinement is incomplete, as demonstrated in outputs with only one GC module. Notably, DeepFire S3GA-Net, with two ASPP and two GC modules, delivers segmentation maps that most closely resemble the ground truth, accurately capturing dominant and subtle fire patterns. These results confirm the synergistic effect of combining multiscale context extraction in the encoder with enhanced spatial refinement in the decoder. In summary, the integrated ASPP and GC modules enable DeepFire S3GA-Net to deliver high precision, recall, and boundary delineation consistently, making it well suited for robust, real-time forest fire segmentation using UAV imagery under complex environmental conditions.

Conclusion

This study introduced DeepFire S3GA-Net, a fully convolutional encoder–decoder architecture designed for precise forest fire segmentation using UAV imagery. By integrating ASPP modules in the encoder and GC modules in the decoder, the model effectively captures multiscale contextual features and fine spatial details. Experimental results across multiple performance metrics demonstrate that DeepFire S3GA-Net consistently outperforms several state-of-the-art segmentation models in terms of accuracy, precision, and boundary delineation. The inclusion of multiscale processing and feature reorganization mechanisms allows the model to generalize well across varying fire intensities and environmental complexities. Visual assessments further confirm its ability to segment fire-affected areas with great clarity and structural consistency. Overall, the proposed framework offers a robust and efficient solution for UAV-based forest fire monitoring and segmentation in complex and dynamic wildfire scenarios.

Future work

Future work will explore model compression for real-time deployment, integration of temporal cues from video data, and the use of multimodal inputs, such as thermal and infrared imagery, to improve robustness, accuracy, and adaptability in diverse fire scenarios.

Data availability

The data that support the findings of this study are available on request from the corresponding author.

Received: 23 June 2025; Accepted: 4 December 2025

Published online: 11 December 2025

References

- Hernández-Morcillo, M. et al. Scanning the solutions for the sustainable supply of forest ecosystem services in Europe. *Sustain. Sci.* **17**, 2013–2029 (2022).
- Barausse, A. et al. A methodological proposal for the climate change risk assessment of coastal habitats based on the evaluation of ecosystem services: lessons learnt from the INTERREG Project ECO-SMART. *Sustainability* **14**(13), 7567 (2022).
- Sun, Z. & Zhou, Y. Trade-offs between forest carbon sinks and multidimensional forest values: from assessment to management strategies. *J. Phys. Conf. Ser.* **2665**, 12020 (2023).
- Dalias, P. et al. Wildfire effects on the soil respiration and bacterial microbiota composition in mediterranean-type ecosystems. *Fire* **7**, 213 (2024).
- George, M. B. & Peter-Anyebe, A. C. The role of U.S. Environmental diplomacy in international wildfire management and sustainable grassland burning practices. *Int. J. Sci. Res. Mod. Technol.* **3**, 21–38 (2024).
- Heidari, H., Arabi, M. & Warziniack, T. Effects of climate change on natural-caused fire activity in western U.S. national forests. *Atmosphere* <https://doi.org/10.3390/atmos12080981> (2021).
- Kala, C. P. Environmental and socioeconomic impacts of forest fires: A call for multilateral Cooperation and management interventions. *Nat. Hazards Res.* **3**, 286–294 (2023).
- Xu, H., Zhang, G., Zhou, Z., Zhou, X. & Zhou, C. Forest fire monitoring and positioning improvement at subpixel level: application to himawari-8 fire products. *Remote Sensing* **14**(10), 2460 (2022).
- Saleh, A. et al. Forest fire surveillance systems: A review of deep learning methods. *Heliyon* **10**, e23127 (2024).
- Wang, Y., Ning, W., Wang, X., Zhang, S. & Yang, D. A novel method for analyzing infrared images taken by unmanned aerial vehicles for forest fire monitoring. *Trait Du Signal.* **40**, 1219–1226 (2023).
- Liu, H., Hu, H., Zhou, F. & Yuan, H. Forest flame detection in unmanned aerial vehicle imagery based on YOLOv5. *Fire* **6**(7), 279 (2023).
- Lee, S., Zulkifley, M. A., Lee, J. K. & Lee, J. Neural Network-Based pipeline with High-Resolution feature enhancement and Low-Resolution feature preservation for automated treatment decision of graves' orbitopathy patients. *IEEE Access.* **12**, 98426–98435 (2024).

13. Min, K. S., Zulkifley, M. A. & Kamari, N. A. M. A review of automated Micro-expression analysis. *J. Kejuruter*. **34**, 763–775 (2022).
14. Lim, S. V., Zulkifley, M. A., Saleh, A., Saputro, A. H. & Abdani, S. R. Attention-based semantic segmentation networks for forest applications. *Forests* **14**(12), 2437 (2023).
15. Chen, Y. et al. California wildfire spread derived using VIIRS satellite observations and an object-based tracking system. *Sci. Data* **9**(1), 249 (2022).
16. Zulkifley, M. A., Aziz, N. A. A. & Stofa, M. M. Burned Forest Areas Mapping Using Semantic Segmentation Networks with Group and Shuffle Embeddings. in *21st IEEE International Colloquium on Signal Processing & Its Applications (CSPA)* 200–204 (2025). 200–204 (2025). <https://doi.org/10.1109/CSPA64953.2025.10933135>
17. Qin, Y. Y., Cao, J. T. & Ji, X. F. Fire detection method based on depthwise separable Convolution and YOLOv3. *Int. J. Autom. Comput.* **18**, 300–310 (2021).
18. Zhang, L., Wang, M., Ding, Y. & Bu, X. A multi-scale faster RCNN model for small target forest fire detection. *Forests* <https://doi.org/10.3390/f14030616> (2023).
19. Huot, F. et al. Deep learning models for predicting wildfires from historical remote-sensing data. *arXiv preprint arXiv:1602.07360* at (2020). <https://doi.org/10.48550/ARXIV.2010.07445>
20. de Almeida Pereira, G. H., Fusioka, A. M., Nassu, B. T. & Minetto, R. Active fire detection in Landsat-8 imagery: A large-scale dataset and a deep-learning study. *ISPRS J. Photogramm Remote Sens.* **178**, 171–186 (2021).
21. Dewangan, A. et al. Dataset and deep learning model for real-time wildland fire smoke detection. *Remote Sensing* **14**(4), 1007 (2022).
22. Kang, Y., Jang, E., Im, J. & Kwon, C. A deep learning model using geostationary satellite data for forest fire detection with reduced detection latency. *GIScience Remote Sens.* **59**, 2019–2035 (2022).
23. Muksimova, S., Umirzakova, S., Mardieva, S., Abdullaev, M. & Cho, Y. I. Revolutionizing Wildfire Detection Through UAV-Driven Fire Monitoring with a Transformer-Based Approach. *Fire* <https://doi.org/10.3390/fire7120443> (2024).
24. Zhan, J., Hu, Y., Cai, W., Zhou, G. & Li, L. PDAM-STPNNet: A small target detection approach for wildland fire smoke through remote sensing images. *Symmetry (Basel)*. **13**, 2260 (2021).
25. Gong, C. et al. LMDFS: A lightweight model for detecting forest fire smoke in UAV images based on YOLOv7. *Remote Sens.* **15**, 3790 (2023).
26. Niu, K. et al. An improved YOLOv5s-Seg detection and segmentation model for the accurate identification of forest fires based on UAV infrared image. *Remote Sens.* **15**, 4694 (2023).
27. Wang, S. et al. A real-time deep learning forest fire monitoring algorithm based on an improved Pruned + KD model. *J. Real-Time Image Process.* **18**, 2319–2329 (2021).
28. Yun, B. et al. An improved unmanned aerial vehicle forest fire detection model based on YOLOv8. *Fire* **8**(4), 138 (2025).
29. Marques, T., Carreira, S., Miragaia, R., Ramos, J. & Pereira, A. Applying deep learning to real-time UAV-based forest monitoring: leveraging multi-sensor imagery for improved results. *Expert Syst. Appl.* **245**, 123107 (2024).
30. Feng, H. et al. U3UNet: an accurate and reliable segmentation model for forest fire monitoring based on UAV vision. *Neural Netw.* **185**, 107207 (2025).
31. Fralenko, V. P. Neural network methods for detecting wild forest fires. *Sci. Tech. Inf. Process.* **51**, 497–505 (2024).
32. Jonnalagadda, A. V., Hashim, H. A. & SegNet A segmented deep learning based convolutional neural network approach for drones wildfire detection. *Remote Sens. Appl. Soc. Environ.* **34**, 101181 (2024).
33. Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N. & Liang, J. UNet++: A nested U-Net architecture for medical image segmentation. 3–11 (2018). https://doi.org/10.1007/978-3-030-00889-5_1
34. Zulkifley, M. A., Saleh, A., Akhyar, A. & Saputro, A. H. Multiscale semantic segmentation networks for forest fire mapping. in *IEEE International Conference on Artificial Intelligence and Mechatronics Systems (AIMS)* 1–6 (2024). <https://doi.org/10.1109/AIMS618.12.2024.10512964>
35. Li, M. et al. A Real-time fire segmentation method based on a deep learning approach. *IFAC-PapersOnLine* **55**, 145–150 (2022).
36. Thanasutives, P., Fukui, K., Numao, M. & Kijirikul, B. Encoder-Decoder based convolutional neural networks with Multi-Scale-Aware modules for crowd counting. (2020). <https://doi.org/10.48550/arxiv.2003.05586>
37. Rizwan, I., Haque, I. & Neubert, J. Deep learning approaches to biomedical image segmentation. *Inf. Med. Unlocked*. **18**, 100297 (2020).
38. Guo, Q., Wu, X. J., Kittler, J. & Feng, Z. Self-grouping convolutional neural networks. *Neural Netw.* **132**, 491–505 (2020).
39. Zhang, X., Zhou, X., Lin, M., Sun, J. & Shufflenet An extremely efficient convolutional neural network for mobile devices. in *Proceedings of the IEEE conference on computer vision and pattern recognition* 6848–6856 (2018).
40. Wu, Y. & He, K. Group normalization. *Int. J. Comput. Vis.* **128**, 742–755 (2020).
41. Liu, D. & Zhou, Z. AdaptiveConv2d: A novel convolutional module for medical image segmentation. *Appl. Sci.* **15**(4), 2169 (2025).
42. Lee, Y., Park, J. H. & Lee, C. O. Balanced group convolution: an improved group convolution based on approximability estimates. *Pattern Analysis and Applications* **28**(4), 161 (2024).
43. Shamsoshoara, A. et al. Aerial imagery pile burn detection using deep learning: the FLAME dataset. *Comput. Networks*. **193**, 108001 (2021).
44. Shoaib, M. et al. Comparative studies of deep learning segmentation models for left ventricle segmentation. *Front Public. Heal* **10**, 981019 (2022).
45. Saponara, S., Elhanashi, A. & Gagliardi, A. Real-time video fire/smoke detection based on CNN in antifire surveillance systems. *J. Real-Time Image Process.* **18**, 889–900 (2021).
46. Frizzi, S., Bouchouicha, M., Ginoux, J., Moreau, É. & Sayadi, M. Convolutional neural network for smoke and fire semantic segmentation. *Iet Image Process.* **15**, 634–647 (2021).
47. Zafar, A. et al. An optimization approach for convolutional neural network using Non-Dominated sorted genetic Algorithm-II. *Comput. Mater. Contin.* **74**, 5641–5661 (2022).
48. Rashid, J. et al. Multi-level deep learning model for potato leaf disease recognition. *Electronics* **10**(17), 2064 (2021).
49. Neary, P. Automatic Hyperparameter Tuning in Deep Convolutional Neural Networks Using Asynchronous Reinforcement Learning. in *IEEE International Conference on Cognitive Computing (ICCC)* 73–77 (2018). 73–77 (2018). (2018). <https://doi.org/10.1109/ICCC.2018.00017>

Author contributions

A.A, J.L, N.S, A.H.S, and M.A.Z provided the conceptualization of the papers. A.A, J.L, N.S, A.H.S, and M.A.Z performed the analysis on the reviewed papers. A.A, J.L, N.S, A.H.S, and M.A.Z wrote the main text. All authors reviewed the initial and final manuscript.

Funding

The project is funded by the Ministry of Higher Education Malaysia through Fundamental Research Grant Scheme (FRGS/1/2022/TK07/UKM/02/4) and Asia Pacific Telecommunity Grant through EBC-K Contribution (KK-2022-026).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.A.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025