DilAtSE-Net: An Encoder Decoder Network for Burnt Area Delineation

Konthalapalli Laalenthika*, Sarvesh Kumar*, Advika Srivastava†, Ananya Das†, Ananya Shriram‡, Ishita Choudhary‡

Manipal Institute of Technology, Manipal, India
{sarvesh.mitmpl2023, advika.mitmpl2024, ananya6.mitmpl2024, ananya4.mitmpl2024, ishita2.mitmpl2024, konthalapalli.mitmpl2023}@learner.manipal.edu

Abstract

Burned area delineation following forest wildfires plays a critical part in quantification for disaster management, post-disaster assessment, and restoration planning. In recent years, advances in deep learning and computer vision, coupled with the growing availability of Earth observation and remote sensing datasets, have significantly contributed to progress in addressing this problem. However, burned area delineation from satellite imagery faces several interrelated challenges that dampen the effectiveness of current approaches. Severe class imbalance causes models to exhibit majority-class bias and poor recall. Additionally, region-based loss functions optimize overall overlap rather than edge accuracy, often overlooking precise boundary detection. This study introduces an encoder-decoder based architecture, inspired by the conventional U-Net, that incorporates dilated convolution blocks for increased field of view, squeeze-and-excitation layers for dynamic channel weighting, and a self-attention bottleneck, trained with a combination of losses. The proposed DilAtSE-Net model outperforms methods previously benchmarked by achieving higher mean Intersection over Union (mIoU) and comparable Dice scores on the Wildfire-CEMS dataset with **12.8M** parameters compared to existing benchmarks ranging from **31M** to **64.1M** parameters.

Introduction

Wildfires are a growing environmental problem with vast impact on the environment, posing a threat to various ecosystems. While natural elements like lightning and volcanic activities cause fires, recently, climate change, along with activities like rapid urbanization and deforestation have been increasing the frequency of fires (Farid et al. 2024) making their study relevant for assessing long term effects, patterns and to construct methods to mitigate and manage these incidents. Satellite data is a reliable and time-efficient resource for studying the locations, frequency and time periods of wildfires.

Despite recent advancements, burned area delineation from satellite imagery presents several fundamental challenges that limit the effectiveness of existing approaches. Severe class imbalance is inherent to this task, making standard loss functions susceptible to majority-class bias (Rege Cambrin et al. 2023) (Hu, Zhang, and Ban 2023)

Precise fire perimeter delineation is essential for operational damage assessment, but methods relying solely on conventional region-based loss functions (such as Dice loss) tend to prioritize overall spatial overlap while neglecting explicit boundary alignment, leading to imprecise edge localization. (Shi et al. 2024) (Yin et al. 2023).

Finally, capturing both fine-grained local patterns and broad spatial context is necessary to distinguish burned areas from spectrally similar unburned regions, a task that standard convolutional architectures with fixed receptive fields struggle to aggregate multiscale information effectively. (Liu et al. 2025b)

In this work, we propose the **Dil**ation with **At**tention Bottleneck, **S**queeze-**E**xcitation blocks **Net**work (DilAtSE-Net), a lightweight encoder-decoder network specifically tailored for burnt area segmentation on the Wildfire-CEMS dataset. This architecture replaces standard convolutional layers with blocks incorporating dilation and channel reweighting, and integrates an attention-included bottleneck layer. Our network maintains significantly low parameter counts while simultaneously demonstrating a notable improvement in segmentation quality with around 1/5th of the parameters of benchmarks reported in (Arnaudo et al. 2023) and less than half the parameters as a U-Net while achieving an increase in mIoU and maintaining a comparable Dice score.

Related Works

Significant progress has been made in the area of burned area delineation in the past few years. Traditional approaches to burned area mapping rely on spectral indices such as the Normalized Burn Ratio (NBR) or differenced NBR (dNBR) derived from satellite imagery (Murphy, Reynolds, and Koltun 2008) (Escuin, Navarro, and Fernández 2008) (Giddey, Baard, and Kraaij 2022). However, these methods are sensitive to atmospheric conditions, suffer from threshold selection challenges, and struggle with spectral confusion between burned areas and similar land cover types such as bare soil, water bodies, and shadows (Seydi 2025). Regional datasets like the CaBuAr dataset, which focus on California wildfires, and the Indonesian tropical fires centric

*Joint first authors.
†Joint second authors.
‡Joint third authors.

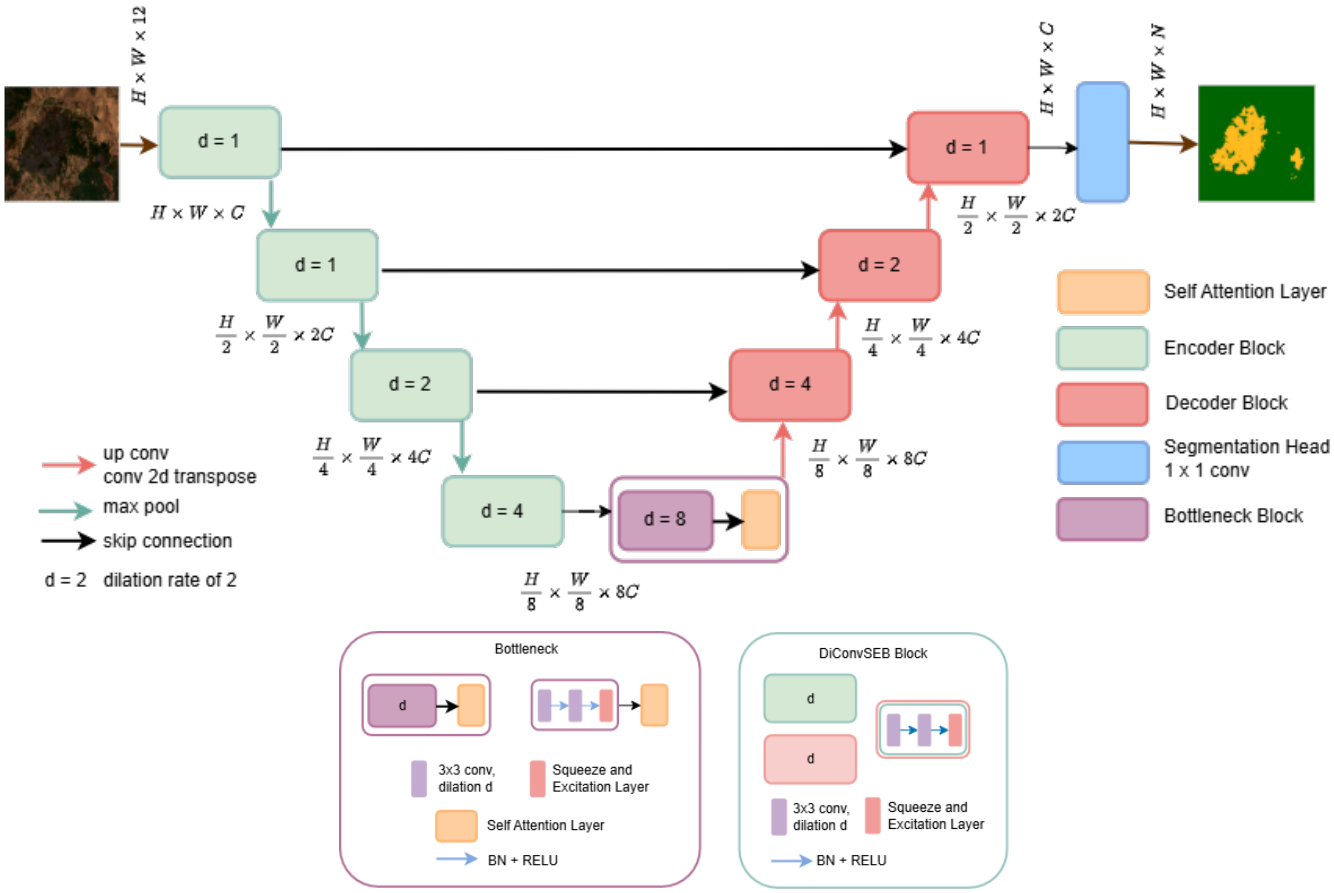

Figure 1: Model architecture for DilAtSE-Net

dataset developed by (Prabowo et al. 2022), offer valuable geographic context but lack diversity across different environmental and climatic conditions. (Colomba et al. 2022) introduced the first CEMS-based dataset consisting of 73 European fire events from 2017-2019, combining Copernicus Emergency Management Service's (EMS) expert annotations with Sentinel-2 imagery for burned area delineation and severity estimation. Building on this foundation, (Arnaudo et al. 2023). released an extended Wildfire-CEMS dataset with 171 fire events spanning 2017-2023, adding multitask annotations for both burned area delineation and land cover classification using ESA WorldCover labels which we utilise to benchmark the effectiveness of our network.

Burned area delineation has evolved from time-series-based methods (MARINHO et al. 1999) to supervised methods such as SVMs and Random Forest (Chandel et al. 2022). Convolutional neural networks (CNNs) based approaches to burned area mapping primarily use encoder-decoder architectures, such as U-Net (Brand and Manandhar 2021), for semantic segmentation in remote sensing (Zhong et al. 2022) (Sun et al. 2022) (Dimitrovski et al. 2024) due to their ability to capture both spatial details and contextual information through skip connections. However, these approaches lack specific mechanisms to perform channel reweighting, which is crucial for multi-channel input data like that in CEMS. Further upgrades include ASPP (Atrous Spatial Pyramid Pooling) for multi-scale context, CBAM (Convolutional Block Attention Module) extensions for channel and spatial reweighting (Liu et al. 2025a), and dual-path attention residual U-Nets that fuse Sentinel images for robustness (Khankeshizadeh et al. 2024).

Dilated convolutions introduced by (Yu and Koltun 2016) expand the receptive field and capture multi-scale context without losing resolution. (Chen et al. 2017) later popularized this idea in the DeepLab framework, referring to it as atrous convolution for semantic segmentation. DRUNET (2018) incorporated dilated convolutions within residual U-Net blocks, progressively increasing dilation with depth, resulting in improved segmentation outcomes by leveraging broader context while preserving fine details (Devalla et al. 2018). Subsequent variants, such as SDU-Net (2022) introduced in (Wang et al. 2022), further explored stacked parallel multi-rate dilations to produce multi-scale features in parallel.

(Hu, Shen, and Sun 2018) introduced the Squeeze-and-Excitation (SE) block to adjust channel-wise feature calibration by boosting important feature maps and suppress-

ing less useful ones at minimal cost. SE modules prove effective in remote sensing segmentation with their ability to address multi-band data. (Prasanna et al. 2022) introduced the Squeeze-Excitation Embedded Attention U-Net, which combines SE modules with attention gates, jointly leveraging channel recalibration and spatial attention to achieve improved Jaccard scores and reduced losses with minimal additional parameters. (Tang et al. 2021) introduced HBA-U-Net, which integrates hierarchical bottleneck attention capture both local details and global structure in fundus images, making it possible to model long-range dependencies with fewer parameters and lower computational cost than at full resolution. Building on these, SEDARU-Net (Lafraxo et al. 2025) puts forward a comprehensive framework combining dilated convolutions, SE-based residual blocks, and attention gates within a U-Net backbone for melanoma lesion segmentation, highlighting the strengths of these complementary mechanisms when brought together.

Class imbalance is one of the major challenges faced in burned area delineation. Focal loss, first introduced by (Lin et al. 2017) aids in minimizing class imbalance. (Wei et al. 2022) demonstrated that CNNs with focal loss outperform sampling-based approaches for seismic fault detection in remote sensing. The unified focal loss framework proposed by (Yeung et al. 2021) provides a hierarchical generalization of dice and focal-based losses, demonstrating consistent improvements across five class-imbalanced medical imaging datasets. Their asymmetric formulation combines focal loss for handling class imbalance with Tversky-Dice losses for region optimization. (Kervadec et al. 2019) introduced boundary loss for highly unbalanced segmentation. Building on this foundation, (Yin et al. 2023) proposed focal boundary cice loss for breast tumor segmentation, which simultaneously addresses class imbalance, region-level accuracy, and boundary precision through a triple-component formulation. Given the demonstrated effectiveness of these composite loss structures, our method adopts a similar weighted combination of focal, dice and boundary losses, for comprehensive loss optimization.

Dataset

In this work, we utilize the Wildfire CEMS dataset (Arnaudo et al. 2023), consisting of Sentinel-2 Level-2A satellite imagery covering the burned areas across Europe spanning from 2017 to 2023, which was developed under the Copernicus Emergency Management Service and provides paired data for burned area delineation, damage grading, cloud cover and land cover classification.

The dataset comprises 281 samples in the train set, 53 samples in the validation set and 99 samples in the test set. The satellite imagery from Sentinel-2 (S2L2A) and the corresponding delineation (DEL) masks were used for the experiments.

Methodology

Our proposed architecture, the DilAtSE-Net, is an encoder-decoder network inspired by the conventional U-Net (Ronneberger, Fischer, and Brox 2015), composed of our DiConvSEB blocks, and specifically designed for the task of burnt area segmentation.

DiConvSEB

The DiConvSEB block (**Di**lation **Conv**olution + **S**queeze-and-**E**xcitation **B**lock) acts as the fundamental building block across both encoder and decoder paths, combining dilated convolutions and a Squeeze-and-Excitation (SE) block for enhanced feature representation. Each block consists of two sequential 3×3 dilated convolutional layers, each followed by Batch Normalization and ReLU activation, concluding with an SE block that adaptively recalibrates channel-wise responses. By using dilated convolutions, the network effectively enlarges its receptive field without increasing kernel size or computational cost, enabling a broader contextual understanding crucial for segmenting large and irregular burnt areas, while the SE mechanism emphasizes the most informative channels throughout the process.

Encoder-Decoder Structure

The network is designed to handle 12 input channels, corresponding to Sentinel-2 and CEMS data, and follows a standard feature progression. The initial convolutional block expands the feature dimension to 64 channels, which then doubles at each subsequent down-sampling stage, reaching 512 channels in the bottleneck. To mitigate the significant increase in parameters inherent in deeper U-Net architectures, we designed the bottleneck to cap at 512 channels and incorporated a specialized Self-Attention layer to efficiently model long-range dependencies.

The encoder's feature maps are preserved as skip connections and concatenated with the corresponding upsampled features in the decoder. This ensures that fine-grained spatial information from the early layers is combined with the high-level semantic context from the deeper layers. The down-sampling in the encoder is achieved via a 2×2 max pooling operation between DiConvSEB blocks.

At each decoder stage, the upsampled feature map is concatenated with the corresponding skip connection from the encoder, the resulting concatenated is input to the subsequent DiConvSEB processing block. The DiConvSEB block then processes these combined features,and halves the number of channels, while progressively decreasing dilation rates to prioritize boundary localization

The decoder path concludes with a segmentation head consisting of a 1×1 convolutional layer applied to the final feature map to produce the output segmentation map.

Self Attention Bottleneck

The deepest layer, the bottleneck, incorporates a DiConvSEB block retaining spatial dimensions followed by a self-attention block similar to the module used in (Zhang et al. 2019) to model long-range spatial dependencies(Tang et al. 2021) across the entire compressed feature map. Within the attention mechanism, the input features are projected into query, key, and value representations via separate 1×1 convolutions. The attention map is formed by taking the dot

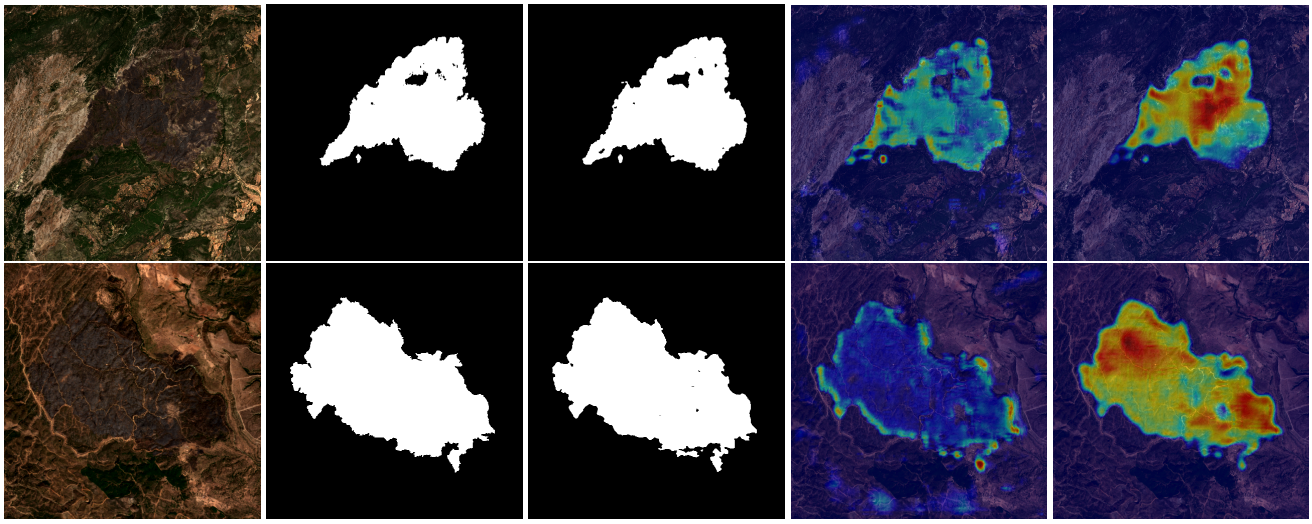

Figure 2: Visualization of RGB channels, ground truth, predictions and GradCAM maps from encoder and decoder blocks before and after bottleneck.

product between queries and keys followed by normalization with a softmax. This attention map is applied to the value projection followed by scaling with γ, a learnable parameter before being added to the initial input through a residual connection.

Loss Functions

We utilize a combination of dice loss, focal loss, and boundary loss in our final objective function.

Focal loss(Lin et al. 2017), given by equation (1) is an extension of the cross-entropy loss designed to solve the class imbalance problem by down-weighting the loss contributed by easy-to-classify examples (Lin et al. 2017) like large background areas. This forces the network to focus its learning effort on the hard, misclassified pixels or those near the boundary.

$$\mathcal{L}_{\text{focal}} = -\sum_{c=1}^{C} \alpha_c \, (1 - p_c)^{\gamma} \, y_c \, \log(p_c) \tag{1}$$

where α_c is the class weighting factor that balances the importance of each class, which is set to 1 for this binary segmentation task, p_c is the model's predicted probability for class c and $\log(p_c)$ is the log-likelihood term that penalizes incorrect or low-confidence predictions. γ is set to 2 in our experimentation.

Dice loss (Milletari, Navab, and Ahmadi 2016) as in equation (2) is designed to directly optimize the spatial overlap between the predicted segmentation and the ground truth. Instead of evaluating predictions on a per-pixel basis, dice loss measures the overall similarity between the predicted and true regions by maximizing their intersection over union.

$$\mathcal{L}_{\text{dice}} = 1 - \frac{2 \sum_{i=1}^{N} p_i g_i + \epsilon}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} g_i + \epsilon} \tag{2}$$

where p_i is the predicted probability for pixel i, g_i is the ground truth label, ϵ is a small constant for numerical stability.

We also incorporate a boundary loss inspired by other multi-task networks (Kervadec et al. 2019) (Yin et al. 2023) and remote sensing literature to explicitly promote precise boundary localization.

The boundary loss is calculated as the L1 distance between the edge map of the predicted probabilities and the edge map of the one-hot encoded ground truth. The edge maps are generated by convolving the probability and target tensors with a standard edge-detection kernel. This loss directly penalizes discrepancies in the shape and location of the boundary lines (Colomba et al. 2022), which results in sharper and more accurate borders for the segmented burnt area. We experiment with 3 common kernels - Sobel, Scharr and Laplacian of Gaussian.

Our final loss function is as follows,

$$\mathcal{L} = w_1 \mathcal{L}_{\text{focal}} + w_2 \mathcal{L}_{\text{dice}} + w_3 \mathcal{L}_{\text{boundary}} \tag{3}$$

where w_i indicates the weight given to each loss. A comparative study of the weights given to these losses is mentioned in the Results section.

Training Strategy

All experimentations were trained for 50 epochs, using the Adam optimizer (Kingma and Ba 2015) with a learning rate of 1e-4 and a batch size of 4 on an NVIDIA P100 GPU. As mentioned in (Arnaudo et al. 2023), we resize images to a resolution of 512×512 and use the training, validation and testing split provided.

Results

Our comprehensive evaluation demonstrates that the proposed model achieves comparable performance with signif-

Loss Functions (Weights)			Kernels					
Focal	Dice	Boundary	Sobel		Scharr		Laplacian	
			Dice	mIoU	Dice	mIoU	Dice	mIoU
0.4	0.4	0.2	0.9284	0.8807	0.9284	0.8810	0.9287	0.8807
0.3	0.3	0.3	0.9179	0.8687	**0.9297**	**0.8816**	0.9265	0.8787
0.5	0.3	0.2	0.9258	0.8781	0.9269	0.8787	0.9250	0.8770

Table 1: Proposed model trained for burnt area delineation with different kernels for boundary losses and different weights to each loss.

icantly lesser computational requirements compared to existing benchmarks. This efficiency is achieved with massive gains in size: our model utilizes only 12 million parameters, representing a $5.4\times$ reduction in size compared to the 65 million parameters required by the UPerNet (ResNet-50) benchmark, reported in (Arnaudo et al. 2023).

This substantial improvement in efficiency stems from the integration of custom architectural components within the encoder–decoder framework. By combining the DiConvSEB block with a self-attention bottleneck inspired by SAGAN (Zhang et al. 2019), the network effectively captures high-level semantic and global contextual information, eliminating the need for the computationally intensive 1024-channel bottleneck typically used in conventional U-Net architectures. Instead, our design employs a shallower 512-channel bottleneck, reducing the parameter count to less than half while achieving slightly superior segmentation performance.

We present a quantitative comparison of our model with existing benchmarks on this dataset in Table (2), a study of the effect of different weights for each loss with different kernels in Table (1). Additionally, we visualize qualitative results of our model with corresponding attention maps from the encoder and decoder blocks prior and after the bottleneck layer in Figure (2).

Model	Params (M)	mIoU	Dice
U-Net	31.0	0.8676	0.9173
DeepLabv3	42.0	0.8655	0.9188
SegFormer (MiT-B3) [1]	44.6	0.8338	0.9094
UPerNet (ResNet-50)[1]	64.1	0.8494	0.9186
UPerNet (ViT-S)[1]	57.9	0.8298	0.9069
DilAtSE-N	**12.8**	**0.8816**	**0.9297**

Table 2: Comparative evaluation of DilAtSE-Net against existing architectures.

Conclusion

Our proposed model DilAtSE-Net, a lightweight and efficient architecture customized for burned area delineation from multi-spectral satellite imagery, combines squeeze-and-excitation blocks as well as dilated convolutions with progressive receptive field expansion across down-sampling and up-sampling layers, and self-attention mechanism at the bottleneck stage, showing the potential to capture both localized and global contextual cues for accurate wildfire delineation. Our network is trained using a compound loss function that collectively handles class imbalance, region-level accuracy, and boundary precision, allowing accurate segmentation of challenging fire-affected regions. Evaluated on the Wildfire-CEMS dataset, DilAtSE-Net demonstrates competitive segmentation performance when compared to heavier transformer-based baselines, with significant parameter reduction. These results showcase the ability of DiConvSEB block and attention bottleneck layer to yield high-resolution burned area maps. Future work will explore further experimentation for more environmental applications.

[1]Best results between Single-Task-Learning (STL) and Multi-Task-Learning (MTL) reported from (Arnaudo et al. 2023)

References

Arnaudo, E.; Barco, L.; Merlo, M.; and Rossi, C. 2023. Robust Burned Area Delineation through Multitask Learning. arXiv:2309.08368.

Brand, A.; and Manandhar, A. 2021. Semantic segmentation of burned areas in satellite images using a U-net-based convolutional neural network. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43: 47–53.

Chandel, A.; Sarwat, W.; Najah, A.; Dhanagare, S.; and Agarwala, M. 2022. Evaluating methods to map burned area at 30-meter resolution in forests and agricultural areas of Central India. *Frontiers in Forests and Global Change*, Volume 5 - 2022.

Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; and Yuille, A. L. 2017. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4): 834–848.

Colomba, L.; Farasin, A.; Monaco, S.; Greco, S.; Garza, P.; Apiletti, D.; Baralis, E.; and Cerquitelli, T. 2022. A Dataset for Burned Area Delineation and Severity Estimation from Satellite Imagery. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, CIKM '22, 3893–3897. New York, NY, USA: Association for Computing Machinery. ISBN 9781450392365.

Devalla, S. K.; Renukanand, P. K.; Sreedhar, B. K.; Subramanian, G.; Zhang, L.; Perera, S.; Mari, J.-M.; Chin, K. S.;

Tun, T. A.; Strouthidis, N. G.; et al. 2018. DRUNET: a dilated-residual U-Net deep learning network to segment optic nerve head tissues in optical coherence tomography images. *Biomedical optics express*, 9(7): 3244–3265.

Dimitrovski, I.; Spasev, V.; Loshkovska, S.; and Kitanovski, I. 2024. U-net ensemble for enhanced semantic segmentation in remote sensing imagery. *Remote Sensing*, 16(12): 2077.

Escuin, S.; Navarro, R.; and Fernández, P. 2008. Fire severity assessment by using NBR (Normalized Burn Ratio) and NDVI (Normalized Difference Vegetation Index) derived from LANDSAT TM/ETM images. *International Journal of Remote Sensing*, 29(4): 1053–1073.

Farid, A.; Alam, M. K.; Goli, V. S. N. S.; Akin, I. D.; Akinleye, T.; Chen, X.; Cheng, Q.; Cleall, P.; Cuomo, S.; Foresta, V.; Ge, S.; Iervolino, L.; Iradukunda, P.; Luce, C. H.; Koda, E.; Mickovski, S. B.; O'Kelly, B. C.; Paleologos, E. K.; Peduto, D.; Ricketts, E. J.; Sadegh, M.; Sarris, T. S.; Singh, D. N.; Singh, P.; Tang, C.-S.; Tardio, G.; Vaverková, M. D.; Veneris, M.; and Winkler, J. 2024. A Review of the Occurrence and Causes for Wildfires and Their Impacts on the Geoenvironment. *Fire*, 7(8).

Giddey, B. L.; Baard, J. A.; and Kraaij, T. 2022. Verification of the differenced Normalised Burn Ratio (dNBR) as an index of fire severity in Afrotemperate Forest. *South African Journal of Botany*, 146: 348–353.

Hu, J.; Shen, L.; and Sun, G. 2018. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7132–7141.

Hu, X.; Zhang, P.; and Ban, Y. 2023. Large-scale burn severity mapping in multispectral imagery using deep semantic segmentation models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 196: 228–240.

Kervadec, H.; Bouchtiba, J.; Desrosiers, C.; Granger, E.; Dolz, J.; and Ayed, I. B. 2019. Boundary loss for highly unbalanced segmentation. In *International conference on medical imaging with deep learning*, 285–296. PMLR.

Khankeshizadeh, E.; Tahermanesh, S.; Mohsenifar, A.; Moghimi, A.; and Mohammadzadeh, A. 2024. FBA-DPAttResU-Net: Forest burned area detection using a novel end-to-end dual-path attention residual-based U-Net from post-fire Sentinel-1 and Sentinel-2 images. *Ecological Indicators*, 167: 112589.

Kingma, D. P.; and Ba, J. 2015. Adam: A Method for Stochastic Optimization. In Bengio, Y.; and LeCun, Y., eds., *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Lafraxo, S.; El Ansari, M.; Koutti, L.; Kerkaou, Z.; and Souaidi, M. 2025. SEDARU-net: a squeeze-excitation dilated based residual U-Net with attention mechanism for automatic melanoma lesion segmentation. *Multimedia Tools and Applications*, 84(21): 23935–23962.

Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; and Dollár, P. 2017. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, 2980–2988.

Liu, L.; Guo, Y.; Chen, E.; Li, Z.; Li, Y.; Liu, Y.; Zhang, Q.; and Wang, B. 2025a. A lightweight Deeplab V3+ network integrating deep transitive transfer learning and attention mechanism for burned area identification. *Scientific Reports*, 15(1): 15969.

Liu, Y.; Liu, X.; Zheng, Z.; Luo, S.; Xiao, Y.; Wang, S.; Liu, X.; and Sun, Z. 2025b. Power Transmission Corridors Wildfire Detection for Multi-Scale Fusion and Adaptive Texture Learning Based on Transformers. *IEEE Access*, 13: 140973–140983.

MARINHO, B. P.; GREGOIRE, J.-M.; CARDOSO, P. J.; et al. 1999. An Algorithm for Extracting Burned Areas from Time Series of AVHRR GAC Data Applied at a Continental Scale. *Remote Sensing of Environment*.

Milletari, F.; Navab, N.; and Ahmadi, S.-A. 2016. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. arXiv:1606.04797.

Murphy, K. A.; Reynolds, J. H.; and Koltun, J. M. 2008. Evaluating the ability of the differenced Normalized Burn Ratio (dNBR) to predict ecologically significant burn severity in Alaskan boreal forests. *International Journal of Wildland Fire*, 17(4): 490–499.

Prabowo, Y.; Sakti, A. D.; Pradono, K. A.; Amriyah, Q.; Rasyidy, F. H.; Bengkulah, I.; Ulfa, K.; Candra, D. S.; Imdad, M. T.; and Ali, S. 2022. Deep Learning Dataset for Estimating Burned Areas: Case Study, Indonesia. *Data*, 7(6).

Prasanna, G.; Ernest, J. R.; Lalitha, G.; and Narayanan, S. 2022. Squeeze excitation embedded attention u-net for brain tumor segmentation. In *International Conference on Emerging Electronics and Automation*, 107–117. Springer.

Rege Cambrin, D.; Colomba, L.; Garza, P.; et al. 2023. Vision Transformers for Burned Area Delineation. In *CEUR Workshop Proceedings*, volume 3343. CEUR-WS.

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

Seydi, S. T. 2025. Deep Learning-Based Burned Area Mapping Using Bi-Temporal Siamese Networks and AlphaEarth Foundation Datasets. *arXiv preprint arXiv:2509.07852*.

Shi, P.; Hu, J.; Yang, Y.; Gao, Z.; Liu, W.; and Ma, T. 2024. Centerline Boundary Dice Loss for Vascular Segmentation. arXiv:2407.01517.

Sun, Y.; Bi, F.; Gao, Y.; Chen, L.; and Feng, S. 2022. A multi-attention UNet for semantic segmentation in remote sensing images. *Symmetry*, 14(5): 906.

Tang, S.; Qi, Z.; Granley, J.; and Beyeler, M. 2021. U-net with hierarchical bottleneck attention for landmark detection in fundus images of the degenerated retina. In *International Workshop on Ophthalmic Medical Image Analysis*, 62–71. Springer.

Wang, S.; Singh, V. K.; Cheah, E.; Wang, X.; Li, Q.; Chou, S.-H.; Lehman, C. D.; Kumar, V.; and Samir, A. E. 2022. Stacked dilated convolutions and asymmetric architecture for U-Net-based medical image segmentation. *Computers in biology and medicine*, 148: 105891.

Wei, X.-L.; Zhang, C.-X.; Kim, S.-W.; Jing, K.-L.; Wang, Y.-J.; Xu, S.; and Xie, Z.-Z. 2022. Seismic fault detection using convolutional neural networks with focal loss. *Computers & Geosciences*, 158: 104968.

Yeung, M.; Sala, E.; Schönlieb, C.-B.; and Rundo, L. 2021. Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation. arXiv:2102.04525.

Yin, X.-X.; Jian, Y.; Shen, J.; Wu, J.; Zhang, Y.; and Wang, W. 2023. Focal boundary dice: improved breast tumor segmentation from MRI scan. *Journal of Cancer*, 14(5): 717.

Yu, F.; and Koltun, V. 2016. Multi-Scale Context Aggregation by Dilated Convolutions. arXiv:1511.07122.

Zhang, H.; Goodfellow, I.; Metaxas, D.; and Odena, A. 2019. Self-Attention Generative Adversarial Networks. arXiv:1805.08318.

Zhong, L.; Lin, Y.; Su, Y.; and Fang, X. 2022. Improved U-Net network segmentation method for remote sensing image. In *2022 IEEE 6th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, 1034–1039. IEEE.