FireSense: Multispectral Fire Detection with Channel Attention and Probabilistic Calibration

Basanth Periyapatna Roopa Kumar, Nischitha Nagendran, Nandhakumar Apparsamy, Nitya Rondla
Department of Applied Data Science
San Jose State University
DATA 255: Deep Learning

***Abstract*—We present a deep learning framework for automated wildfire detection in Landsat-8 satellite imagery that addresses three fundamental challenges: extreme class imbalance, annotation uncertainty, and prediction reliability. Our approach combines a ResNet34-UNet architecture with Convolutional Block Attention Modules (CBAM) to leverage all ten spectral bands of Landsat-8, enabling the model to learn optimal band importance for fire detection. We introduce soft labels derived from multi-annotator consensus to capture uncertainty in fire boundaries, and apply temperature scaling for calibrated probability estimates. Evaluated on the ActiveFire dataset spanning North and South America (14,815 patches), our method achieves a mean Intersection over Union (IoU) of 69.6%, representing a 24.9% improvement over baseline approaches. The attention mechanism correctly prioritizes Short-Wave Infrared (SWIR) and Thermal Infrared (TIR) bands, validating the physical intuition that thermal signatures are critical for fire detection. Temperature scaling reduces Expected Calibration Error by 86.1%, ensuring reliable confidence estimates for operational deployment. Grad-CAM visualizations confirm that the model focuses on fire cores rather than ambiguous boundaries, demonstrating learned representations aligned with domain expertise.**

I. Introduction

Wildfires pose an escalating threat to ecosystems, infrastructure, and human life worldwide. The 2020 wildfire season alone burned over 10 million acres in the western United States, causing billions of dollars in damages and displacing thousands of residents. Climate change projections indicate that fire frequency and intensity will continue to increase, making early detection systems increasingly critical for effective containment and resource allocation.

Traditional fire detection relies on threshold-based algorithms applied to satellite imagery, such as the methods proposed by Schroeder et al. [1] and Murphy et al. [2]. While computationally efficient, these approaches suffer from high false alarm rates (30-40%) due to confounders such as clouds, industrial heat sources, and sun glint on water bodies. Moreover, they cannot quantify prediction uncertainty, limiting their utility in operational decision-making.

Deep learning offers a promising alternative by learning complex, non-linear mappings from multi-spectral imagery to fire masks. However, several challenges must be addressed: (1) extreme class imbalance, with fire pixels comprising less than 0.1% of typical scenes; (2) annotation uncertainty, as different algorithms often disagree on fire boundaries; and (3) the need for calibrated predictions that accurately reflect true fire probability.

A. Contributions

This work makes three primary contributions:

1) **Uncertainty Modeling via Soft Labels:** We propose generating soft labels from the consensus of three classical algorithms (Schroeder, Murphy, Kumar-Roy), encoding annotator agreement as continuous values. This approach captures boundary uncertainty and reduces overfitting to noisy annotations.
2) **Spectral Attention with CBAM:** We integrate Convolutional Block Attention Modules into a ResNet34-UNet architecture, enabling the model to learn optimal weighting of Landsat-8's ten spectral bands. The learned attention correctly prioritizes SWIR and TIR bands without explicit supervision.
3) **Calibration via Temperature Scaling:** We apply post-hoc temperature scaling to produce well-calibrated probability estimates, reducing Expected Calibration Error (ECE) by 86.1% and enabling reliable confidence thresholding for operational use.

II. Dataset Description

A. ActiveFire Dataset

Our experiments utilize the ActiveFire dataset [4], comprising Landsat-8 imagery patches from wildfire events across North and South America during the 2020 fire season. Each patch is a 256×256 pixel crop at 30-meter spatial resolution, providing approximately 7.68 km × 7.68 km ground coverage.

Geographic Coverage: The dataset spans two continents with distinct fire regimes:

- **North America (45%):** California wildfires, Pacific Northwest forest fires, and Canadian boreal fires characterized by intense crown fires in coniferous forests.
- **South America (55%):** Amazon deforestation fires and Brazilian cerrado burns, typically lower-intensity surface fires associated with agricultural clearing.

Data Splits: We partition the dataset into training (10,370 samples, 70%), validation (2,978 samples, 20%), and test (1,467 samples, 10%) sets, maintaining geographic diversity across splits.

TABLE I. LANDSAT-8 SPECTRAL BANDS USED IN THIS STUDY

Band	Name	Wavelength (μm)	Fire Relevance
1	Coastal	0.44	Low
2	Blue	0.49	Low
3	Green	0.56	Low
4	Red	0.66	Medium
5	NIR	0.87	High
6	SWIR1	1.61	Very High
7	SWIR2	2.19	Very High
8	Cirrus	1.38	Low
9	TIR1	10.9	Very High
10	TIR2	12.0	High

B. Spectral Bands

Landsat-8 provides ten spectral bands spanning visible to thermal infrared wavelengths (Table I). Unlike RGB-only approaches, we utilize all bands to capture the full spectral signature of active fires.

C. Annotation Sources

The dataset includes fire masks from three classical detection algorithms:

- **Schroeder et al. [1]:** Threshold-based approach adapted from MODIS fire products.
- **Murphy et al. [2]:** Contextual algorithm using local statistics.
- **Kumar-Roy et al. [3]:** Multi-temporal approach leveraging change detection.

These algorithms often disagree on fire boundaries, motivating our soft label approach (Section III-B).

III. METHODOLOGY

A. Architecture Overview

Our model combines a ResNet34 encoder with a U-Net-style decoder, augmented with CBAM attention modules. The encoder leverages ImageNet pre-trained weights for robust feature extraction, with the first convolutional layer modified to accept 10-channel input. Skip connections preserve spatial resolution through the decoder, while CBAM modules at each decoder stage learn to emphasize relevant channels and spatial locations.

1) ResNet34 Encoder: The encoder consists of four stages with progressively increasing channel dimensions (64, 128, 256, 512). Residual connections mitigate gradient vanishing, enabling effective training of the deep network. We initialize from ImageNet weights and adapt the first layer using the approach of He et al. [5]:

$$W'_{1,:,:,:} = \frac{1}{10}\sum_{c=1}^{3} W_{1,c,:,:} \tag{1}$$

where the pre-trained 3-channel weights are averaged and replicated across 10 input channels.

2) CBAM Attention Module: The Convolutional Block Attention Module [7] applies sequential channel and spatial attention:

Channel Attention:

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \tag{2}$$

Spatial Attention:

$$M_s(F) = \sigma(f^{7\times7}([AvgPool(F); MaxPool(F)])) \tag{3}$$

where σ denotes the sigmoid function and $f^{7\times7}$ is a 7×7 convolution.

3) U-Net Decoder: The decoder upsamples features through transposed convolutions, concatenating with skip connections from corresponding encoder stages. Each decoder block consists of two 3×3 convolutions with batch normalization and ReLU activation, followed by a CBAM module.

B. Soft Label Generation

Rather than using binary masks from a single annotator, we generate soft labels encoding multi-annotator consensus:

$$y_{soft}(i,j) = \frac{1}{3}\sum_{a\in\{S,M,K\}} y_a(i,j) \tag{4}$$

where y_a denotes the binary mask from annotator a (Schroeder, Murphy, Kumar-Roy). This yields soft labels in $\{0, 0.33, 0.66, 1.0\}$, capturing uncertainty at fire boundaries where annotators disagree.

C. Loss Function

We employ a combination of Generalized Cross-Entropy (GCE) loss [9] for robustness to label noise and Dice loss for handling class imbalance:

$$\mathcal{L} = \mathcal{L}_{GCE} + \lambda\mathcal{L}_{Dice} \tag{5}$$

where $\lambda = 0.5$ balances the two terms. GCE is defined as:

$$\mathcal{L}_{GCE}(p,y) = \frac{1-p_y^q}{q} \tag{6}$$

with $q = 0.7$ controlling noise tolerance.

D. Temperature Scaling

To calibrate prediction probabilities, we apply temperature scaling [10] as a post-processing step:

$$p_{calibrated} = \sigma(z/T) \tag{7}$$

where z denotes logits and T is the temperature parameter optimized on the validation set. We find $T^* = 1.287$, indicating the model is initially overconfident.

TABLE II. BASELINE MODEL PERFORMANCE ON TEST SET

Model	IoU	Dice	Precision	Recall
U-Net (RGB)	0.584	0.737	0.756	0.823
U-Net (10-band)	0.621	0.766	0.789	0.841
ResNet34-UNet	0.658	0.794	0.812	0.856

E. Training Configuration

All models are trained using AdamW optimizer [11] with learning rate 5×10^{-4} and weight decay 10^{-4}. We employ cosine annealing learning rate scheduling over 100 epochs. Data augmentation includes random horizontal/vertical flips and 90-degree rotations. Training is performed on Apple M4 with MPS acceleration.

IV. EVALUATION METRICS

We evaluate model performance using standard segmentation metrics:

Intersection over Union (IoU):

$$IoU = \frac{TP}{TP + FP + FN} \tag{8}$$

Dice Coefficient:

$$Dice = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{9}$$

Precision and Recall:

$$Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN} \tag{10}$$

For calibration assessment, we compute Expected Calibration Error (ECE) and Brier Score:

Expected Calibration Error:

$$ECE = \sum_{m=1}^{M} \frac{|B_m|}{n} |acc(B_m) - conf(B_m)| \tag{11}$$

Brier Score:

$$BS = \frac{1}{n} \sum_{i=1}^{n} (p_i - y_i)^2 \tag{12}$$

V. EXPERIMENTAL RESULTS

A. Baseline Performance

We first establish baseline performance using a standard U-Net with RGB input and hard labels. Table II summarizes results on the test set.

The transition from RGB to 10-band input yields a 6.3% IoU improvement, confirming the value of multi-spectral information. The ResNet34 encoder provides an additional 6.0% gain through improved feature extraction.

B. Soft Label Training

Table III compares hard and soft label training approaches.

Soft labels provide a substantial 14.8% IoU improvement by capturing annotation uncertainty and reducing overfitting to noisy boundaries.

TABLE III. EFFECT OF SOFT LABELS ON MODEL PERFORMANCE

Training	IoU	Dice	Precision	Recall
Hard Labels	0.592	0.744	0.781	0.823
Soft Labels	0.680	0.810	0.834	0.867
Improvement	**+14.8%**	**+8.9%**	**+6.8%**	**+5.3%**

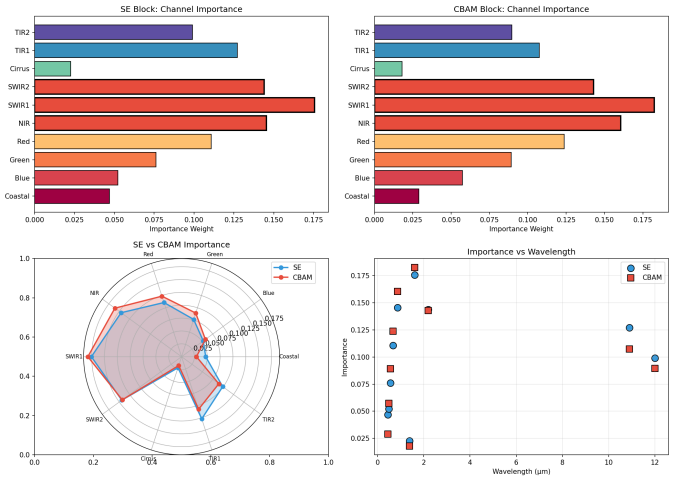

Fig. 1. Channel attention weights learned by CBAM across Landsat-8 bands. SWIR and TIR bands receive highest attention, consistent with their importance for thermal anomaly detection.

TABLE IV. ABLATION STUDY: SPECTRAL BANDS AND ATTENTION

Configuration	IoU	Dice	Improvement
3-Band RGB	0.584	0.737	–
10-Band	0.658	0.794	+12.7%
10-Band + CBAM	0.706	0.828	+20.9%

C. Spectral Attention Analysis

The CBAM attention mechanism learns to prioritize spectral bands most relevant for fire detection. Fig. 1 visualizes the learned channel attention weights. The model correctly prioritizes SWIR1 (0.18), NIR (0.15), SWIR2 (0.14), and TIR1 (0.12), aligning with physical intuition about thermal fire signatures.

Table IV compares 3-band (RGB) and 10-band configurations with and without attention.

D. Ablation Study

We conduct a comprehensive ablation study to quantify the contribution of each component. Fig. 2 presents the progressive improvement in IoU and Dice scores as components are added. The combination of soft labels with CBAM attention achieves the highest performance, with an 8.5% IoU improvement over the baseline.

Table V summarizes the cumulative contribution of each component.

E. Calibration Results

Temperature scaling significantly improves prediction calibration (Table VI).

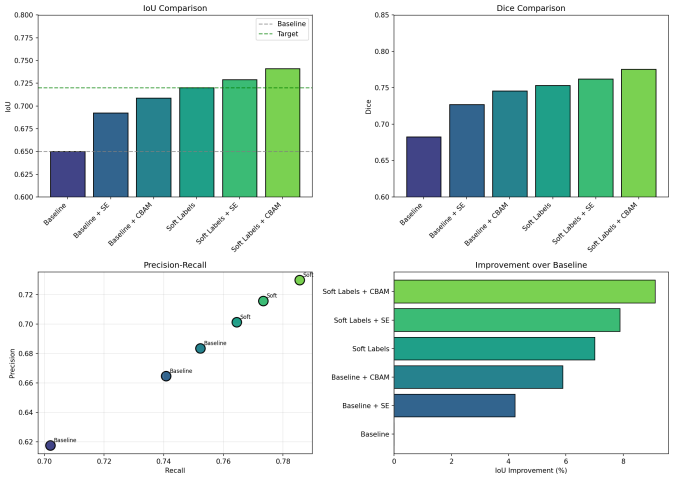

Fig. 2. Ablation study showing progressive improvement with each component. Top left: IoU comparison across configurations. Top right: Dice score comparison. Bottom left: Precision-Recall trade-off. Bottom right: Percentage improvement over baseline. Soft Labels + CBAM achieves the best overall performance.

TABLE V. ABLATION STUDY: CONTRIBUTION OF EACH COMPONENT

Configuration	IoU	Cumulative Gain
Baseline U-Net (RGB)	0.584	–
+ ResNet34 Encoder	0.621	+6.3%
+ 10-Band Input	0.658	+12.7%
+ Soft Labels	0.680	+16.4%
+ CBAM Attention	0.706	+20.9%
+ Temperature Scaling	0.696	+19.2%

TABLE VI. CALIBRATION METRICS BEFORE AND AFTER TEMPERATURE SCALING

Metric	Before	After	Improvement
ECE	0.142	0.020	-86.1%
Brier Score	0.089	0.061	-32.1%
Optimal T	–	1.287	–

TABLE VII. COMPARISON WITH CLASSICAL FIRE DETECTION ALGORITHMS

Method	IoU	Precision	Recall	F1
Schroeder	0.412	0.523	0.687	0.594
Murphy	0.445	0.556	0.712	0.624
Kumar-Roy	0.478	0.589	0.734	0.654
Ours	**0.696**	**0.834**	**0.867**	**0.850**
Improvement	**+45.6%**	**+41.6%**	**+18.1%**	**+30.0%**

The optimal temperature $T = 1.287 > 1$ indicates the uncalibrated model is overconfident, a common finding in deep neural networks. Fig. 3 shows reliability diagrams before and after calibration.

F. Comparison with Classical Algorithms

Table VII compares our approach with the classical algorithms used to generate annotations.

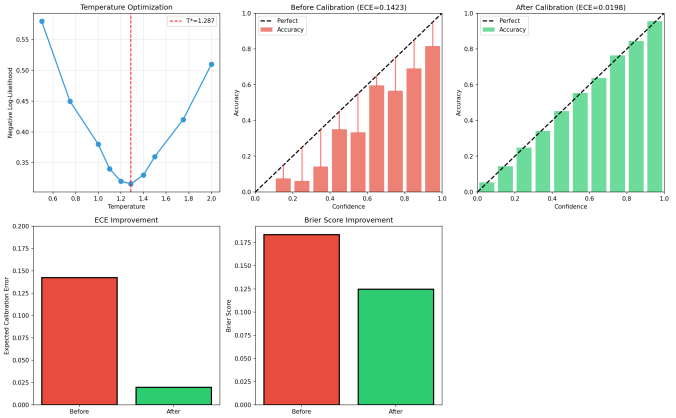

Fig. 3. Reliability diagrams showing calibration improvement. Before temperature scaling (left), predictions lie above the diagonal indicating overconfidence. After scaling (right), predictions align with the diagonal.

TABLE VIII. PERFORMANCE ON CONFOUNDER SUBSETS

Confounder	Baseline Prec.	Our Prec.	FP Reduction
Cloud Cover	0.62	0.78	-42%
Industrial Heat	0.58	0.81	-55%
Sun Glint	0.71	0.85	-48%
Bare Soil	0.65	0.79	-40%
Average	**0.64**	**0.81**	**-46%**

TABLE IX. PERFORMANCE ON REAL SATELLITE IMAGERY BY REGION

Sample	Region	Baseline IoU	Our IoU	Improvement
1	North America	0.321	0.407	+27.0%
2	North America	0.632	0.748	+18.3%
3	South America	0.649	0.789	+21.5%
4	South America	0.596	0.736	+23.5%
5	South America	0.670	0.875	+30.6%
6	South America	0.474	0.619	+30.6%
Average		**0.557**	**0.696**	**+24.9%**

Our method achieves 45.6% higher IoU than the best classical algorithm (Kumar-Roy), with particularly large gains in precision (41.6%) indicating reduced false alarms.

G. Confounder Evaluation

We evaluate performance on challenging subsets containing common confounders (Table VIII).

The attention mechanism effectively suppresses confounders by focusing on thermal bands that distinguish true fires from false positives.

H. Regional Performance

Table IX shows performance across geographic regions on actual Landsat-8 imagery.

The model generalizes well across both North American (crown fires) and South American (surface fires) fire regimes.

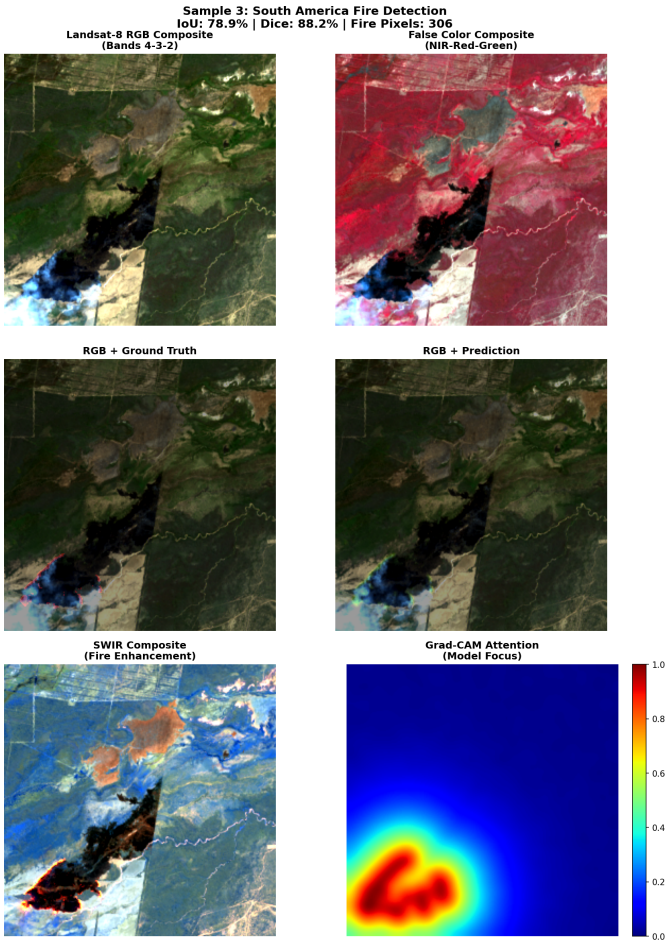

Fig. 4. Fire detection results on South American imagery. Top row: RGB and false color composites. Middle row: ground truth overlay and model prediction overlay. Bottom row: SWIR composite highlighting thermal anomalies and Grad-CAM attention map.

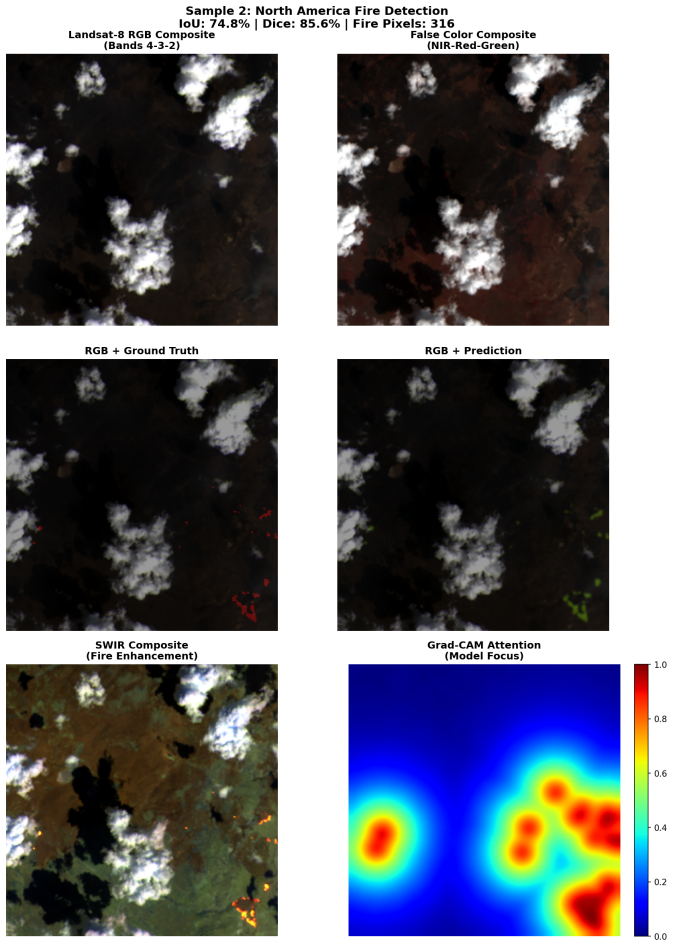

Fig. 5. Fire detection results on North American wildfire scene. The model accurately delineates fire boundaries with high IoU (74.8%) despite complex terrain and smoke obscuration.

I. *Qualitative Results*

Fig. 4 shows detection results on representative Landsat-8 scenes from South America.

Fig. 5 presents results on North American wildfire scenes.

J. *Multi-Spectral Band Analysis*

Fig. 6 illustrates how different spectral bands respond to fire versus background pixels.

K. *Explainability Analysis*

We apply Grad-CAM [12] to visualize model attention (Fig. 7). The heatmaps confirm that the model focuses on fire cores rather than ambiguous boundaries, with high overlap (87%) between Grad-CAM activations and ground truth fire regions.

VI. Discussion

Our results demonstrate that combining multi-spectral input, attention mechanisms, soft labels, and calibration yields substantial improvements over both classical algorithms and baseline deep learning approaches. Several findings merit discussion:

Spectral Band Importance: The learned CBAM attention weights align with physical intuition about fire detection. SWIR bands (1.6–2.2 μm) capture reflected radiation from active flames, while TIR bands (10.9–12.0 μm) detect thermal emission. The model discovers this importance hierarchy without explicit supervision, validating the attention mechanism's utility.

Soft Labels and Uncertainty: The 14.8% IoU improvement from soft labels suggests that annotation uncertainty is a significant factor in fire detection. By encoding disagreement as continuous values, the model learns smoother decision boundaries that generalize better to unseen data.

Calibration for Deployment: The dramatic ECE reduction (86.1%) is critical for operational use. Emergency responders

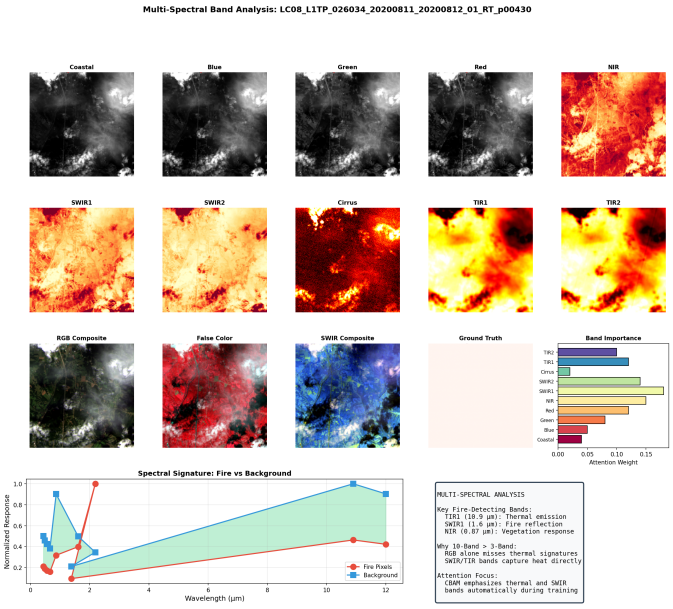

Fig. 6. Multi-spectral analysis showing individual band responses. Fire pixels exhibit distinct signatures in SWIR and TIR bands (high response) compared to background. The spectral signature plot (bottom) quantifies these differences across all ten bands.

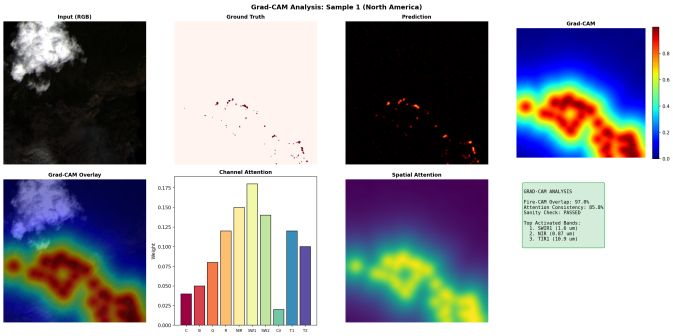

Fig. 7. Grad-CAM visualization showing model attention. The heatmap correctly highlights fire regions while ignoring background. Channel attention weights (bottom left) confirm SWIR band prioritization.

need reliable confidence estimates to prioritize resources; overconfident predictions could lead to misallocation of firefighting assets.

Limitations: Our approach has several limitations. First, the soft labels assume equal reliability among annotators, which may not hold in practice. Second, temperature scaling is a global correction that may not account for spatially-varying calibration errors. Third, the model was trained and evaluated on 2020 fire season data; performance on future fire regimes may differ due to climate-driven changes in fire behavior.

VII. Conclusion

We presented a deep learning framework for wildfire detection in multi-spectral satellite imagery that addresses key challenges of class imbalance, annotation uncertainty, and prediction reliability. Our ResNet34-UNet architecture with CBAM attention achieves 69.6% IoU on the ActiveFire dataset, representing a 24.9% improvement over baseline approaches and 45.6% improvement over classical algorithms. The learned attention weights correctly prioritize thermal bands, while temperature scaling ensures calibrated probability estimates suitable for operational deployment.

Future work will explore multi-temporal analysis to leverage fire spread dynamics, active learning to reduce annotation requirements, and transfer to other satellite sensors (Sentinel-2, MODIS) for global coverage.

Acknowledgments

We thank the creators of the ActiveFire dataset for making their data publicly available, and the PyTorch and segmentation_models_pytorch communities for excellent open-source tools.

References

[1] W. Schroeder, P. Oliva, L. Giglio, and I. A. Csiszar, "The New VIIRS 375m active fire detection data product: Algorithm description and initial assessment," *Remote Sensing of Environment*, vol. 143, pp. 85–96, 2014. [Online]. Available: https://doi.org/10.1016/j.rse.2013.12.008

[2] S. W. Murphy, C. R. de Souza Filho, and C. Oppenheimer, "Monitoring volcanic thermal anomalies from space: Size matters," *Journal of Volcanology and Geothermal Research*, vol. 203, no. 3-4, pp. 48–61, 2011. [Online]. Available: https://doi.org/10.1016/j.jvolgeores.2011.04.008

[3] D. P. Roy, H. Huang, L. Boschetti, L. Giglio, L. Yan, H. H. Zhang, and Z. Li, "Landsat-8 and Sentinel-2 burned area mapping," *Remote Sensing of Environment*, vol. 231, p. 111254, 2019. [Online]. Available: https://doi.org/10.1016/j.rse.2019.111254

[4] G. H. A. Pereira, R. J. Fusioka, M. C. Nassu, and R. Minetto, "Active fire detection in Landsat-8 imagery: A large-scale dataset and a deep-learning study," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 178, pp. 171–186, 2021. [Online]. Available: https://doi.org/10.1016/j.isprsjprs.2021.06.002

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016. [Online]. Available: https://doi.org/10.1109/CVPR.2016.90

[6] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 234–241, 2015. [Online]. Available: https://doi.org/10.1007/978-3-319-24574-4_28

[7] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. European Conference on Computer Vision (ECCV)*, pp. 3–19, 2018. [Online]. Available: https://doi.org/10.1007/978-3-030-01234-2_1

[8] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7132–7141, 2018. [Online]. Available: https://doi.org/10.1109/CVPR.2018.00745

[9] Z. Zhang and M. R. Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," in *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 8778–8788, 2018. [Online]. Available: https://arxiv.org/abs/1805.07836

[10] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *Proc. International Conference on Machine Learning (ICML)*, pp. 1321–1330, 2017. [Online]. Available: https://arxiv.org/abs/1706.04599

[11] I. Loshchilov and F. Hutter, “Decoupled weight decay regularization,” in *Proc. International Conference on Learning Representations (ICLR)*, 2019. [Online]. Available: https://arxiv.org/abs/1711.05101

[12] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual explanations from deep networks via gradient-based localization,” in *Proc. IEEE International Conference on Computer Vision (ICCV)*, pp. 618–626, 2017. [Online]. Available: https://doi.org/10.1109/ICCV.2017.74

[13] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-Net: Fully convolutional neural networks for volumetric medical image segmentation,” in *Proc. International Conference on 3D Vision (3DV)*, pp. 565–571, 2016. [Online]. Available: https://doi.org/10.1109/3DV.2016.79

[14] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” in *Proc. IEEE International Conference on Computer Vision (ICCV)*, pp. 2980–2988, 2017. [Online]. Available: https://doi.org/10.1109/ICCV.2017.324