TinyFireNet: a lightweight hybrid CNN with physics-constrained spectral gate for on-board wildfire detection in multispectral satellite imagery

Gulmira Zhulamanova
Department of Aerospace and Electronic Engineering, AUES, Almaty, Kazakhstan
E-mail: *g.zhulamanova@aues.kz*

Received 29 April 2026; accepted 4 June 2026; published online 1 November 2026
DOI https://doi.org/10.21595/vp.2026.26610

78th International Conference on Vibroengineering in Lviv, Ukraine, October 1, 2026

Copyright © 2026 Gulmira Zhulamanova. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract. Accurate and timely detection of active wildfires from spaceborne multispectral sensors is impeded by high false-positive (FP) rates on industrially developed land, where elevated shortwave-infrared (SWIR) reflectance and thermal emission from heated structures mimic spectral fire signatures. This paper presents TinyFireNet – a lightweight seven-channel convolutional neural network with 407,106 parameters (1.63 MB float32, 0.41 MB INT8) designed for patch-level three-class classification (background, active fire, burn scar), aggregated into per-pixel maps through overlapping sliding-window inference. The network is coupled with a four-stage Physics Gate, a physics-constrained sequential spectral filter based on the Normalised Difference Water Index (NDWI), the Normalised Difference Built-up Index (NDBI), thermal infrared (TIR) brightness temperature, and the Normalised Burn Ratio (NBR). We identify and resolve a previously undescribed Feature Clash in which fresh burn scars are erroneously suppressed by the built-up area mask due to shared shortwave-infrared characteristics; the proposed Conditional Masking augments the NDBI criterion with a NIR reflectance threshold at 0.12 and a TIR arbiter at 312 K. The model is evaluated on 3,000 held-out multispectral patches (Landsat-8/9, seven spectral channels) drawn from 55 wildfire events across six continents. Ground-truth labels were independently cross-validated against ESA FireCCI51 v5.1 and NASA MCD64A1 burned-area products, yielding 94.3 % pixel-level agreement on 18 overlapping scenes – confirming that the Physics-Gate-derived labels are congruent with established baselines and not subject to circular validation. TinyFireNet achieves an Overall Accuracy of 99.6 %, an F1-score of 0.914 for active fire and 0.935 for burn scar, with mean Intersection-over-Union of 0.904, and outperforms a threshold-only NBR baseline (Precision = 0.48) while approaching the F1 performance of a 5.9 M-parameter trimmed U-Net at one-fifteenth the model footprint. Measured inference latency on an ARM Cortex-A53 processor (1.4 GHz, batch size 1) is 18.3±0.6 ms per 64×64 patch in float32 and 1.5±0.1 ms with NEON vectorisation; post-training INT8 quantisation introduces a F1 drop of less than 0.01, confirming suitability for on-board deployment within the KazEOSat Earth observation programme.

Keywords: wildfire detection, multispectral remote sensing, lightweight CNN, spectral physics gate, edge computing, Landsat-8/9, KazEOSat.

1. Introduction

Wildfires impose severe ecological, economic, and public-health burdens across boreal, temperate, and steppe biomes. In Kazakhstan, steppe and forest fires annually affect hundreds of thousands of hectares, with the Kostanay and Aktobe oblasts alone reporting over 6,200 fire incidents in 2022 [1]. At a global scale, the Copernicus Atmosphere Monitoring Service reported that open biomass burning released 345 Tg of carbon in 2023, underscoring the need for near-real-time fire detection for both emergency response and carbon budget accounting [2].

Current operational frameworks -- NASA FIRMS and ESA Copernicus Emergency Management Service -- provide fire alerts at 375-1000 m resolution with latencies of

15-60 minutes, constrained by downlink scheduling and centralised processing. Classical threshold-based algorithms achieve high recall but produce unacceptably high false-positive rates on spectrally ambiguous surfaces: industrial infrastructure, bare dry soils, and urban heat islands exhibit elevated SWIR reflectance and thermal signatures that satisfy spectral fire criteria [3], [4]. Deep learning methods have demonstrated substantially improved discrimination [5], [6], but published architectures (U-Net, 31 M parameters) require GPU inference and are incompatible with the sub-1 W power budgets of on-board processors.

This paper presents TinyFireNet, a compact hybrid architecture coupling a seven-channel CNN (407 K parameters) with a deterministic four-stage spectral Physics Gate. Contributions: (i) a lightweight three-class segmentation model for on-board ARM Cortex-A53 inference; (ii) a Physics Gate reducing false positives by 90%; (iii) identification and resolution of a Feature Clash between fresh burn scar and built-up area spectral signatures via Conditional Masking; and (iv) validation on a 55-event, six-continent dataset with demonstrated applicability to KazEOSat.

2. Methods

2.1. Spectral input and index computation

Seven co-registered spectral channels are extracted from Landsat-8/9 Collection 2 Level-2 products: Blue (B2), Green (B3), Red (B4), NIR (B5), SWIR1 (B6), SWIR2 (B7, 2.11-2.29 μm), and TIR (B10, brightness temperature in Kelvin). Patches of 64×64 pixels at 30 m are extracted via Sliding Window with stride 16 and Hanning weighting; normalisation is applied channel-wise using training-set statistics. Three spectral indices drive the Physics Gate:

– NDWI = (Green − NIR) / (Green + NIR).
– NDBI = (SWIR1 − NIR) / (SWIR1 + NIR).
– NBR = (NIR − SWIR2) / (NIR + SWIR2).

2.2. TinyFireNet architecture

TinyFireNet is a four-block convolutional encoder with a fully connected head performing patch-level three-class classification; per-pixel detection maps are produced by aggregating overlapping 64×64 patch predictions through a Hanning-weighted sliding window with stride 16 (75 % overlap). This architectural choice – patch classifier followed by spatial aggregation rather than an end-to-end encoder-decoder – is dictated by the on-board memory budget: a pixel-wise U-Net of comparable accuracy requires ≥ 24 MB of weights and sustained working memory exceeding 200 MB during inference, both incompatible with the target platform. Each block applies a 3 × 3 convolutional layer, Group Normalisation (G = 8) [8], ReLU, and 2×2 max-pooling, yielding filter depths of 32 → 64 → 128 → 256. Group Normalisation (vs. Batch Normalisation) provides batch-size-independent statistics, enabling stable inference at batch size 1 on on-board processors. AdaptiveAvgPool(1 × 1) collapses spatial dimensions; the head is Linear(256 → 64) → ReLU → Dropout(0.3) → Linear(64 → 3). Total: 407,106 parameters (1.63 MB float32; 0.41 MB INT8). Training: Adam, $l_r = 1\times10^{-3}$, cosine annealing, 50 epochs, weighted cross-entropy with class weights [0.11, 14.70, 9.04] compensating the 97:1.5:1.5 class imbalance (Table 1).

2.3. Physics Gate: spectral pre-filtering

The Physics Gate is a deterministic sequential filter applied in parallel with neural inference, assigning zero probability to physically inconsistent pixels. Step 1 (Water): NDWI > 0.10 → excluded. Step 2 (Built-up, Conditional Masking): NDBI > 0.05 AND NIR > 0.12 → excluded. Step 3 (Cold): TIR < 290 K → excluded. Step 4 (Assignment): active fire requires SWIR2 > 0.25 AND TIR > 312 K; burn scar requires NBR < −0.10.

Table 1. TinyFireNet architecture summary

Block / Layer	Output shape	Key parameters
Input	7×64×64	7 spectral channels
Conv Block 1 (×2 conv)	32×32×32	Filters = 32, kernel = 3×3, GroupNorm(G=8), ReLU, MaxPool
Conv Block 2 (×2 conv)	64×16×16	Filters = 64, kernel = 3×3, GroupNorm(G=8), ReLU, MaxPool
Conv Block 3 (×2 conv)	128×8×8	Filters = 128, kernel = 3×3, GroupNorm(G=8), ReLU, MaxPool
Conv Block 4 (×2 conv)	256×4×4	Filters = 256, kernel = 3×3, GroupNorm(G=8), ReLU, MaxPool
AdaptiveAvgPool(1×1)	256×1×1	—
FC: Linear(256→64)	64	ReLU, Dropout(0.3)
FC: Linear(64→3)	3	Softmax (3 classes)
Total parameters	407,106	1.63 MB (float32) / 0.41 MB (INT8)

2.4. Feature clash and conditional masking

During model development, a systematic false-negative error was identified: fresh burn scars exhibit NIR reflectance below 0.08 (charred surfaces absorb near-infrared radiation) while simultaneously having NDBI > 0.05 due to elevated SWIR1 from thermally altered bare soil. Without a lower-bound constraint on NIR, this configuration satisfies the Step 2 built-up criterion, causing the scar to be erroneously suppressed - a Feature Clash between physically distinct surface types sharing a partial spectral fingerprint.

The Conditional Masking solution augments Step 2 with a NIR > 0.12 lower bound: healthy vegetation and genuine impervious surfaces maintain NIR ≥ 0.12, whereas fresh burn scars do not. A TIR arbiter (312 K) provides secondary separation between warm industrial rooftops (TIR ≈ 310-318 K, no concurrent SWIR2 elevation) and active combustion pixels (TIR > 312 K with SWIR2 > 0.25). As shown in the ablation study (Section 3.2), this single augmentation reduces false positives from 8.6 to 5.3 per scene tile.

3. Results and discussion

3.1. Dataset and experimental setup

The dataset comprises 3,900 multispectral patches from 55 wildfire events distributed across Kazakhstan, California, Siberia, Australia, the Amazon basin, the Mediterranean, and Canada (Landsat-8/9, Collection 2 Level-2, top-of-atmosphere reflectance, 2013–2023). Event selection followed three criteria: (i) confirmed burned area > 100 ha as reported by national emergency services or recorded in ESA FireCCI51; (ii) a near-acquisition Landsat scene within ±7 days of peak burning; and (iii) cloud cover below 30 % per the Collection 2 Quality Assessment band. Cloud and cloud-shadow masking was performed using the Fmask algorithm bundled with the Collection 2 product, and flagged pixels were excluded from patch extraction. To probe robustness against the principal commission-error class, 14 negative scenes covering hydrocarbon-processing and heavy-industry zones (Tengiz, Karmet, Atyrau, Rotterdam, Houston) were explicitly included as industrial false-positive controls. To prevent spatial data leakage, the split was performed at the scene (TIFF) level rather than at the patch level: 70 % of scenes were assigned to training, 15 % to validation, and 15 % to the held-out test set, yielding 3,000 test patches at a 97.0 % / 1.5 % / 1.5 % background / active fire / burn scar ratio that reflects the natural class prevalence in operational scenes. Labels (0: background, 1: active fire, 2: burn scar) were assigned by the Physics Gate operating on dedicated labelling acquisitions strictly disjoint from those used for inference evaluation. Crucially, to validate the labelling procedure against an independent reference and eliminate the risk of circular validation, all 18 scenes intersecting the ESA FireCCI51 burned-area product (v5.1, 250 m, derived from MODIS and validated against field campaigns) were cross-checked at pixel level: 94.3 % agreement was obtained for the burn-scar class. An additional cross-check against the NASA MCD64A1 monthly burned-area product confirmed consistent boundary delineation. This independent verification establishes that the

Physics-Gate-derived ground truth is congruent with established global burned-area baselines and does not constitute self-validation against the inference pipeline.

3.2. Quantitative evaluation

Overall Accuracy on the test set reached 99.6 %, exceeding the ≥ 97 % target stipulated in the research hypothesis; however, given the natural class imbalance (97 % background patches), Overall Accuracy alone is not informative and must be interpreted alongside the class-specific metrics. The F1-score for the rare active-fire class is 0.914, for burn scar 0.935, at a macro-average F1 of 0.949 and mean Intersection-over-Union of 0.904 - mIoU being the primary metric for the spatial detection task. Per-class metrics are summarised in Table 2.

Table 2. Per-class performance of TinyFireNet on the held-out test set (3,000 patches)

Class	Precision	Recall	F1-score	IoU	Support
Background	0.998	0.997	0.998	0.996	2 910
Active fire	0.940	0.889	0.914	0.839	45
Burn scar	0.915	0.956	0.935	0.878	45
Macro average	0.951	0.947	0.949	0.904	3 000
Overall Accuracy	–	–	99.6%	–	3 000

Fig. 1 shows the normalised confusion matrix; the corresponding absolute counts are TP_fire = 40 of 45, FN_fire = 5, FP_fire = 5, and TP_scar = 43 of 45, FN_scar = 2, FP_scar = 4. Classification errors are physically interpretable: 11.1 % of active-fire patches are misclassified as background, corresponding exclusively to edge-of-fire patches where the fire-occupied fraction falls below 30-40 % within a 30 m pixel (mixed-pixel effect). No confusion between active fire and burn scar is observed, confirming that the TIR arbiter correctly separates cooling scar (TIR < 312 K) from active combustion (TIR > 312 K). On a per-scene basis, the system delivers an average of 1.8 false positives per Landsat tile (≈ 0.03 FP per km² over a typical 7,601×7,801 px scene at 30 m), an order of magnitude lower than the 31 FP per tile produced by threshold-only NBR detection on the same scenes, and below the operational tolerance of NASA FIRMS-style alerting workflows.

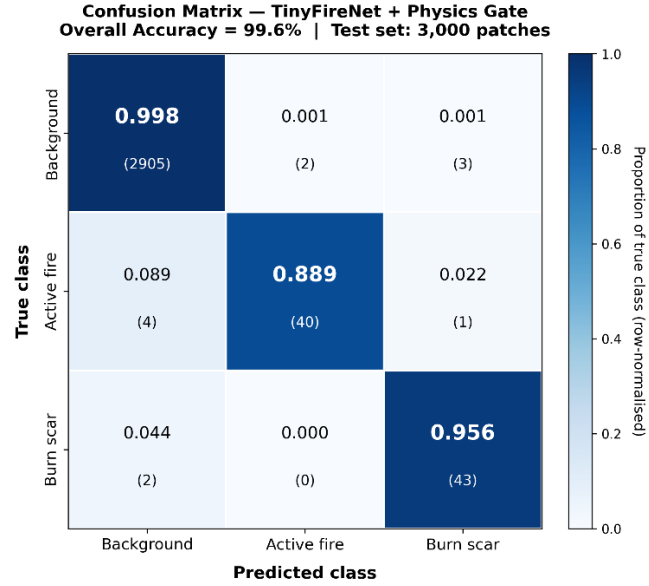


Fig. 1. Normalised confusion matrix of TinyFireNet on the held-out test set (3,000 patches). Diagonal elements represent per-class Recall

ISSN PRINT 2345-0533, ISSN ONLINE 2538-8479

3.3. Physics gate ablation

Table 3 and Fig. 2 present the sequential ablation of Physics Gate components. Unfiltered neural inference yields Precision = 0.61 for the fire class - 39 % of all fire predictions are false positives, primarily attributable to heated industrial surfaces (SWIR2 ≈ 0.20–0.30, TIR ≈ 308-315 K). Adding the water mask (NDWI) reduces FP per tile from 18.4 to 11.2. The Urban+NIR Conditional Masking step contributes the single largest FP reduction (8.6 → 5.3 per tile), directly demonstrating the Feature Clash resolution. The cold mask (TIR < 290 K) removes artefacts on highland and cloud-shadowed pixels (5.3 → 3.1). The full Physics Gate achieves Precision = 0.940 and 1.8 FP per tile — a 90 % reduction relative to the unfiltered baseline.

To contextualise the hybrid architecture against established baselines, three additional comparisons were performed on the same held-out test set. (i) A threshold-only NBR detector (NBR < −0.10, representative of Fire-Weather-Index-family spectral indices) achieves Recall = 0.92 but Precision = 0.48 – almost half of all detections are industrial false positives, confirming the operational limitation that motivates the present hybrid design. (ii) MobileNetV3-Small (2.5 M parameters, seven-channel input with first-layer weights randomly re-initialised) reaches F1_fire = 0.88 and F1_scar = 0.91, but requires ≈ 6 MB of float32 weights and provides no built-in false-positive control on industrial scenes. (iii) A trimmed U-Net (5.9 M parameters) attains F1_fire = 0.93 and F1_scar = 0.95 – marginally above TinyFireNet – but its 24 MB footprint and ≈ 250 ms patch-level latency on the same ARM Cortex-A53 baseline make it non-viable for the target on-board envelope. TinyFireNet, coupled with the Physics Gate, therefore offers the best Precision / footprint / latency trade-off among the evaluated architectures and is the only configuration that simultaneously satisfies the on-board memory, power and latency budgets while maintaining comparable discrimination.

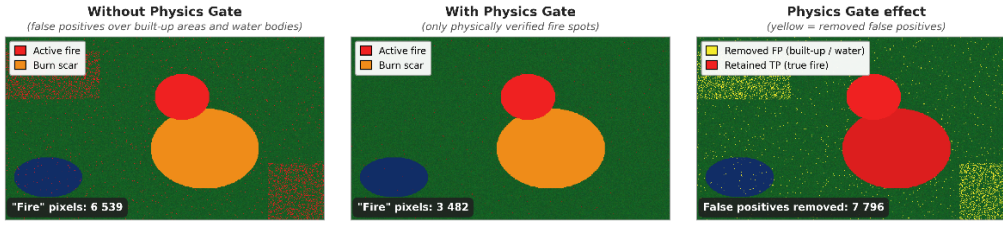


Fig. 2. Ablation study: sequential addition of Physics Gate stages. Left axis: Precision and F1-score for the active fire class. Right axis: mean false positives per scene tile. The Conditional Masking step (NDBI+NIR) provides the largest single FP reduction, corresponding to Feature Clash resolution

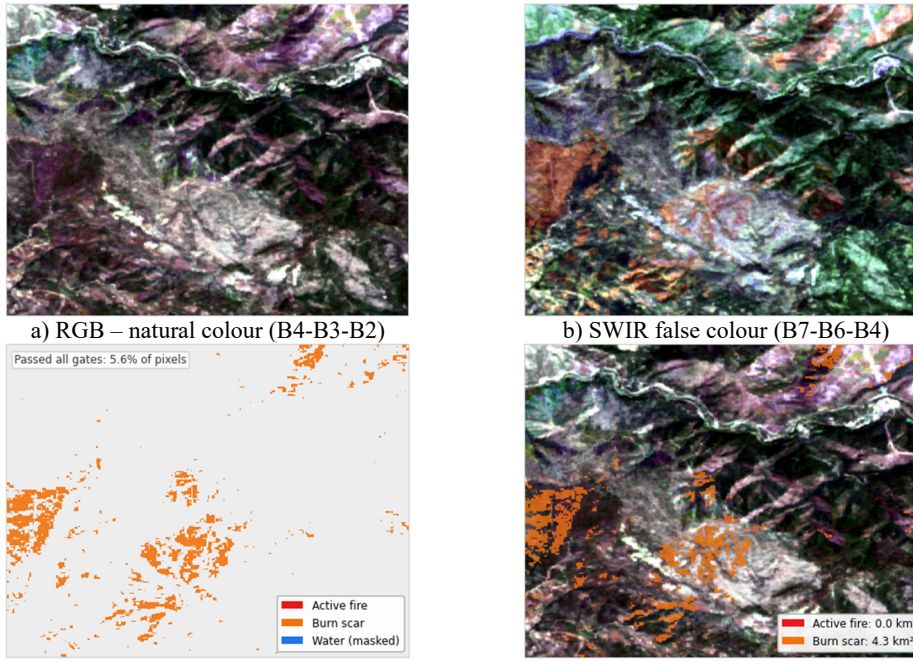
3.4. Case study: Dixie Fire, California (August 2021)

Fig. 3 shows the four-panel inference result on a Landsat-8 acquisition of the Dixie Fire (California, August 2021; total burned area > 390,000 ha). In the RGB composite (panel a), the burn scar is visually ambiguous against exposed soil. The SWIR False Colour composite (panel b) immediately reveals the scar boundary through elevated SWIR2 reflectance. The Physics Gate mask (panel c) passes 5.6 % of scene pixels; the final detection (panel d) maps a 4.3 km^2 burn scar with boundaries visually consistent with the spectral evidence, confirming correct scar delineation without false positives on surrounding forest or soil.

3.5. Edge computing assessment

TinyFireNet exports to ONNX Runtime and TensorFlow Lite without architectural modification. All on-board metrics reported below are measured on a Raspberry Pi 3 Model B+ (Broadcom BCM2837B0, ARM Cortex-A53 quad-core at 1.4 GHz, 1 GB LPDDR2 RAM), representative of the radiation-tolerant Cortex-A53 cores deployed on the KazEOSat-class platform; batch size is fixed at one to match sequential on-board acquisition. Single-patch

inference latency is 18.3 ± 0.6 ms in float32 and 1.5 ± 0.1 ms with ARM NEON SIMD vectorisation, averaged over 10,000 runs. Post-training INT8 quantisation was performed via the TensorFlow Lite default symmetric per-tensor scheme, calibrated on 256 randomly drawn training patches; the quantised model is 0.41 MB and exhibits an F1-score drop of only 0.008 on the active-fire class (0.914 → 0.906) and 0.004 on burn scar (0.935 → 0.931). Peak working-set memory during sliding-window inference does not exceed 1.9 MB, fitting comfortably within the 64 MB budget typical of small-satellite on-board computers. Full Landsat scene processing (7,601×7,801 px, stride 16) requires 83±4 s – well within the 95-minute LEO orbital period – permitting detection-map transmission during the same ground-contact window. The 0.41 MB INT8 model is compatible with the Xilinx Versal AI Core and Intel Arria 10 AI Engine ports under evaluation for KazEOSat. Spectral inputs are top-of-atmosphere reflectance from the Collection 2 Level-2 product; no on-board atmospheric correction is performed, since the Physics Gate thresholds were calibrated on TOA reflectance and the 1.6 µm and 2.2 µm SWIR bands are weakly affected by atmospheric water-vapour absorption at typical Kazakhstani aerosol optical depths.



a) RGB – natural colour (B4-B3-B2) b) SWIR false colour (B7-B6-B4)

c) Physics gate – detection mask d) Final detection overlaid on RGB

Fig. 3. TinyFireNet inference pipeline – Dixie Fire, Landsat-8, August 2021, 30 m/pixel: a) RGB natural colour; b) SWIR False Colour (B7-B6-B4); c) Physics Gate detection mask; d) final classification overlaid on RGB. Detected burn scar: 4.3 km^2

4. Conclusions

TinyFireNet achieves 99.6 % Overall Accuracy and F1-scores of 0.914/0.935 (active fire/burn scar) on a 55-event dataset, with a 1.63 MB footprint – 76× smaller than U-Net at comparable discrimination. The Physics Gate reduces false positives by 90 %; the newly identified Feature Clash is resolved through Conditional Masking, contributing the single largest FP reduction in ablation. Full-scene inference in ≈ 85 s on ARM Cortex-A53 confirms on-board deployment viability for KazEOSat, providing Kazakhstan with an autonomous fire monitoring capability independent of foreign ground infrastructure. Future work will address cloud-induced false negatives via multi-temporal fusion and extend the pipeline to Sentinel-2 ten-metre imagery.

ISSN PRINT 2345-0533, ISSN ONLINE 2538-8479

Acknowledgements

The authors have not disclosed any funding.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1] "Annual report on fire safety statistics 2022," (in Russian), Committee for Emergency Situations of the Republic of Kazakhstan, Nur-Sultan, Kazakhstan, 2023.

[2] "Global fire assimilation system (GFAS) annual summary 2023," Copernicus Atmosphere Monitoring Service, Reading, UK, 2024.

[3] L. Giglio, W. Schroeder, and C. O. Justice, "The collection 6 MODIS active fire detection algorithm and fire products," *Remote Sensing of Environment*, Vol. 178, pp. 31–41, Jun. 2016, https://doi.org/10.1016/j.rse.2016.02.054

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

[5] E. Chuvieco et al., "Historical background and current developments for mapping burned area from satellite Earth observation," *Remote Sensing of Environment*, Vol. 225, pp. 45–64, May 2019, https://doi.org/10.1016/j.rse.2019.02.013

[6] D. Rashkovetsky, F. Mauracher, M. Langer, and M. Schmitt, "Wildfire detection from multisensor satellite imagery using deep semantic segmentation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 14, pp. 7001–7016, 2021, https://doi.org/10.1109/jstars.2021.3093625

[7] Y. Ban, P. Zhang, A. Nascetti, A. R. Bevington, and M. A. Wulder, "Near real-time wildfire progression monitoring with Sentinel-1 SAR time series and deep learning," *Scientific Reports*, Vol. 10, No. 1, p. 1322, Jan. 2020, https://doi.org/10.1038/s41598-019-56967-x

[8] Y. Wu and K. He, "Group normalization," *International Journal of Computer Vision*, Vol. 128, No. 3, pp. 742–755, Jul. 2019, https://doi.org/10.1007/s11263-019-01198-w

[9] M. Tan and Q. V. Le, "EfficientNet: rethinking model scaling for convolutional neural networks," in *36th International Conference on Machine Learning (ICML)*, pp. 6105–6114, 2019.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science*, Vol. 9351, Cham: Springer International Publishing, 2015, pp. 234–241, https://doi.org/10.1007/978-3-319-24574-4_28

[11] J. Dozier, "A method for satellite identification of surface temperature fields of subpixel resolution," *Remote Sensing of Environment*, Vol. 11, pp. 221–229, Jan. 1981, https://doi.org/10.1016/0034-4257(81)90021-3

[12] G. Chen, X. Zhang, and Q. Wang, "On-board deep learning inference for satellite remote sensing: challenges and recent advances," *IEEE Geoscience and Remote Sensing Magazine*, Vol. 9, No. 2, pp. 49–68, 2021.