INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 8, August 2026 | ISSN: 2456-4184 | IJNRD.ORG

Hybrid Deep Learning and Forensic Approach for Satellite Image Forgery Detection

Pendem Pavan Kumar
PG Scholar, Department of Computer Science and Engineering,
Jawaharlal Nehru Technological University Hyderabad
University College of Engineering Jagtial (Autonomous),
Nachupally (Kondagattu), Telangana, India - 505501
pavankumarpendem0072@gmail.com

Dr. B. Sateesh Kumar
Professor, Department of Computer Science and Engineering,
Jawaharlal Nehru Technological University Hyderabad
University College of Engineering Jagtial (Autonomous),
Nachupally (Kondagattu), Telangana, India - 505501
sateeshbkumar@jntuh.ac.in

Abstract*— Satellite imagery is increasingly used for mapping, agriculture, disaster assessment, defence, and environmental monitoring, making image integrity essential. This paper presents HRFNet++, a hybrid deep-learning and forensic framework that performs three related tasks: authentic/forged detection, eight-class manipulation recognition, and pixel-level localization. The architecture combines an RGB semantic stream with a fixed Spatial Rich Model residual stream, independent ResNet-34 encoders, enhanced atrous spatial pyramid pooling, classification heads, and a boundary-aware decoder. A balanced corpus is generated from 1,000 DeepGlobe-derived 256 × 256 patches, producing 8,000 image-mask pairs across real, clone-stamp, copy-move, multi-manipulation, object insertion, object removal, region replacement, and splicing categories. The supplied output reports approximately 98% training accuracy, 94% validation accuracy, 94.2% intersection-over-union, 96.5% Dice coefficient, and 96.8% categorization performance. A Flask interface enables image upload, confidence-based classification, and visualization of the predicted localization mask.

Keywords*— Satellite image forensics, image forgery detection, SRM filters, ResNet-34, ASPP, manipulation localization, multi-task learning.

I. INTRODUCTION

Satellite imagery supports land-use analysis, infrastructure planning, border monitoring, environmental assessment, and emergency response [1]. Because these products are frequently exchanged through digital platforms, a visually convincing edit can alter the interpretation of a scene without leaving an obvious semantic anomaly [2]. Copy-move, splicing, object insertion, object removal, clone-stamp, and region-replacement operations are especially difficult to identify in repeated textures such as fields, roads, roofs, and vegetation [3].

Traditional forensic methods often focus on either image-level classification or pixel-level localization [4]. A binary detector may indicate that an image is suspicious but cannot identify the changed area, while a localization model may generate a mask without providing a stable manipulation category [5]. The proposed framework addresses both limitations by learning shared features for authenticity detection, manipulation classification, and boundary-aware mask prediction.

Satellite-image analysis differs from ordinary natural-image classification because the manipulated area may occupy only a small portion of a very large scene [6]. A reliable method must preserve fine spatial detail while also reasoning about roads, fields, buildings, water bodies, and other contextual structures. The proposed multi-task formulation couples image-level supervision with an explicit localization objective so that weak forensic traces are evaluated together with scene semantics [7].

The main contribution is a dual-stream architecture in which semantic RGB information is complemented by fixed high-pass residual evidence. Multi-scale contextual aggregation and a weighted multi-task loss enable the system to detect both small pasted objects and larger replaced regions [8]. The model is integrated with a secured Flask application so that a user can upload a satellite image and receive the predicted status, manipulation class, confidence score, and localization mask.

The principal contributions are: a dual semantic-forensic feature extractor; an eight-category manipulation taxonomy; a boundary-aware mask decoder; a balanced image-mask generation procedure; and a deployable web layer that presents the input, predicted class, confidence, and spatial evidence in one interface [9].

II. RELATED WORK

Satellite-image manipulation analysis has progressed from handcrafted anomaly measures to deep neural representations. Yarlagadda et al. investigated GAN-generated satellite forgeries using one-class classification techniques [10]. Horvath et al. proposed anomaly-based, deep-belief-network, and vision-transformer methods for satellite manipulation detection [11]. Research on deep fake geography highlighted the operational risks associated with synthetic geospatial imagery [12], while SeeTheSeams introduced a framework for detecting seam-carving artefacts in satellite images [13]. HRFNet further improved the localization of manipulated regions in high-resolution satellite imagery [14].

General image-forensics research provides complementary mechanisms for manipulation detection and localization. Rich-feature convolutional networks, recurrent residual architectures, ManTra-Net, RRU-Net, CAT-Net, multi-view supervision, ObjectFormer, gated context attention, self-supervised forensic signatures, TruFor, hierarchical fine-grained detection, and edge-aware message passing have demonstrated the importance of residual evidence, multi-scale contextual reasoning, and accurate manipulation-boundary extraction [15].

The present project differs from previous studies by integrating three complementary outputs through a shared feature representation while explicitly modelling eight manipulation categories. The framework employs only the algorithms implemented in the proposed system, including fixed SRM filtering, dual ResNet-34 encoders, Atrous Spatial Pyramid Pooling (ASPP), global classification heads, a transposed-convolution decoder, focal loss, cross-entropy loss, Dice loss, AdamW optimization, and cosine-annealing learning-rate scheduling [16].

Unlike conventional single-purpose detection methods, the proposed workflow jointly exploits scene semantics, residual forensic traces, manipulation categorization, and pixel-level localization. This integrated design is particularly suitable for satellite imagery, where repetitive land-cover patterns can obscure manipulated objects and where identifying the

precise spatial extent of an alteration is as important as determining image authenticity.

Residual-domain analysis is especially beneficial because semantic feature extractors may overlook subtle traces introduced by resampling, interpolation, smoothing, and image compositing. Fixed rich-model filters provide robust forensic priors, while the parallel RGB encoder preserves semantic and contextual information. The dual-stream architecture therefore reduces dependence on visually obvious artefacts and improves detection performance when manipulated regions are either very small or carefully blended into repetitive terrain [17].

Localization-oriented research has also demonstrated that image-level authenticity decisions should be supported by pixel-level evidence. Encoder-decoder architectures, multi-scale context aggregation, and boundary-aware loss functions significantly enhance the recovery of irregular manipulated regions. Following this principle, the proposed framework combines two classification branches with a shared decoder, enabling simultaneous authentication, manipulation categorization, and binary mask prediction.

An important limitation in existing literature is the transition from benchmark evaluation to practical forensic deployment. Many published approaches primarily report quantitative performance metrics, whereas operational forensic analysis requires secure image upload, standardized preprocessing, confidence estimation, preservation of original imagery, and intuitive visualization of suspicious regions. The proposed Flask-based implementation addresses these deployment requirements while maintaining the same preprocessing pipeline and inference strategy adopted during model training.

Evaluation methodology is equally critical. Random image-level dataset splitting may produce overly optimistic performance estimates when multiple manipulated samples originate from the same background scene. Reliable satellite image forensic evaluation should therefore separate source scenes during dataset partitioning, report both class-wise and pixel-level performance metrics, and assess robustness against compression, blur, resampling, sensor variations, and geographic domain shifts. These considerations motivate the proposed framework's integration of authentication, manipulation categorization, and localization rather than relying solely on a single image-level prediction [18].

III. MATERIALS AND METHODS

The workflow begins with the preparation of pristine DeepGlobe patches, followed by controlled generation of authentic and manipulated samples. During training, each image and its binary mask are normalized and passed through HRFNet++. The RGB branch captures scene semantics, while the SRM branch extracts weak manipulation residuals. Their features are fused, refined by an ASPP block, and distributed to binary, multi-class, and localization heads. The lowest-validation-loss checkpoint is deployed through the Flask interface.

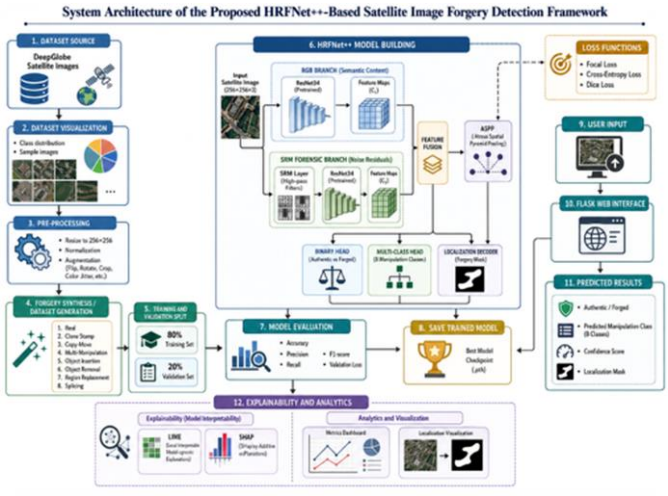

Fig. 1. System Architecture

The architecture retains the spatial feature maps of two ImageNet-pretrained ResNet-34 backbones. Their concatenated 1,024-channel tensor is processed using atrous convolutions with dilation rates 6, 12, and 18. Global average pooling supports authenticity and class prediction, while transposed-convolution stages reconstruct the localization mask.

A) Dataset Collection

The pristine image source is the DeepGlobe satellite-image dataset [19]. The preprocessing script searches the downloaded archive for satellite photographs and randomly crops 1,000 high-resolution patches of 256 × 256 pixels. Each pristine patch is saved as a base image for controlled manipulation synthesis.

For every base patch, the generator stores one authentic sample and seven forged variants: clone-stamp, copy-move, multi-manipulation, object insertion, object removal, region replacement, and splicing. The corresponding ground-truth mask is saved with the same filename, where white pixels identify the manipulated region and black pixels represent authentic content. The intended complete corpus therefore contains 8,000 RGB images and 8,000 paired masks, with 1,000 samples in each class [20].

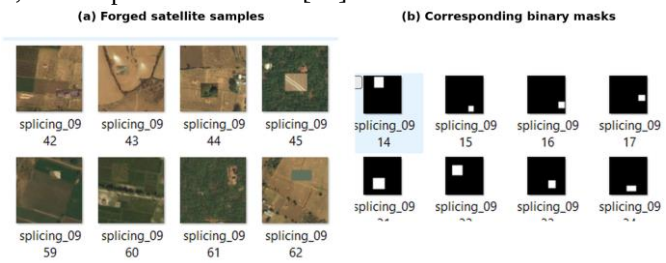

Fig. 2. Representative forged satellite images and their corresponding binary localization masks.

The controlled synthesis provides exact pixel-level annotations and a balanced class distribution. Since related variants originate from the same pristine patch, a strict benchmark should separate data by source patch before generating class variants to prevent background leakage between training and evaluation partitions.

B) Visualization

Dataset visualization was performed to verify image resolution, class balance, and image-mask alignment before training. The representative samples in Fig. 2 show pasted structures, duplicated land parcels, and their exact binary masks. This review helps identify synthesis artefacts that could otherwise become shortcuts during model learning.

C) Preprocessing

All images are converted to RGB and resized to 256 × 256 pixels. During training, horizontal flipping is applied with a

probability of 0.5. Each colour channel is normalized using ImageNet mean values (0.485, 0.456, 0.406) and standard deviations (0.229, 0.224, 0.225), after which the image is converted to a PyTorch tensor using Albumentations [21]. Validation images receive normalization without random augmentation.

Masks are converted to grayscale, thresholded at 127, and represented as binary tensors of size 1 × 256 × 256. The same geometric transform is applied jointly to image and mask to preserve pixel correspondence. The training script samples 300 records in sprint mode and randomly assigns 80% to training and 20% to validation, producing 240 and 60 images, respectively. Batch size is eight, and the code-configured experiment uses five epochs; the application dashboard displays a longer 30-epoch convergence visualization.

D) Algorithms

SRM Forensic Filtering: The Spatial Rich Model layer applies three fixed 5 × 5 high-pass kernels to each RGB channel. These non-trainable filters suppress slowly varying semantic content and emphasize local prediction errors associated with resampling, smoothing, cloning, and compositing [22]. For input image X and fixed kernel Kk, the residual response is

$$R_k = K_k * X \tag{1}$$

Dual ResNet-34 Encoders: Independent pretrained ResNet-34 networks process the RGB image and SRM residual maps [23]. Their pooling and fully connected layers are removed, preserving 512-channel spatial features from each branch. Feature-level concatenation retains semantic and forensic evidence without averaging away weak residual patterns.

Enhanced ASPP: The fused 1,024-channel tensor is evaluated by one 1 × 1 convolution and three dilated 3 × 3 convolutions with rates 6, 12, and 18 [24]. The outputs are concatenated and projected to a 512-channel contextual representation:

$$F_A = \phi([C_1(F), C_6(F), C_{12}(F), C_{18}(F)]) \tag{2}$$

Multi-Task Heads: Global average pooling feeds a binary head and an eight-class head. The localization branch uses transposed convolutions followed by bilinear interpolation to recover a 256 × 256 mask. Sigmoid activation produces the forgery probability, and softmax produces the class probabilities [25]:

$$p_f = \sigma(z_b), \qquad p_c(k) = \frac{e^{z_k}}{\sum_j e^{z_j}} \tag{3}$$

Loss and Optimization: Focal loss reduces the influence of easy binary samples and emphasizes difficult decisions [26]. Cross-entropy with label smoothing supervises the eight-class output, while Dice loss directly optimizes overlap between the predicted and reference masks [27]. The three objectives are combined with a larger weight for localization:

$$\mathcal{L}_{total} = \mathcal{L}_{focal} + \mathcal{L}_{CE} + 2\mathcal{L}_{Dice} \tag{4}$$

AdamW is applied with an initial learning rate of 10^-3 and weight decay of 10^-4 [28]. A cosine-annealing scheduler gradually reduces the learning rate toward 10^-6 [29]. After every epoch, validation loss is calculated without gradient updates, and the checkpoint with the lowest average validation loss is stored as hrfnet_plusplus.pth.

E) Integration of Forensic AI & Flask Framework

The trained HRFNet++ network is integrated into a Flask application with SQLite-based user registration and login. Uploaded PNG, JPG, or JPEG images are converted to RGB, resized to 256 × 256 pixels, normalized using the training parameters, and passed through the three model outputs. The interface reports Authentic or Forged status, the eight-class prediction, confidence percentage, original image, and binary localization mask [30].

A predicted class index of zero is interpreted as Real; all remaining indices are reported as forged. The mask logits are passed through sigmoid activation and thresholded at 0.5 to form a black-and-white evidence map. This implementation supports an auditable workflow in which image-level decisions are accompanied by spatial evidence rather than a single opaque label.

IV. EXPERIMENTAL RESULTS

The experimental evaluation considers classification convergence, prediction confidence, and localization quality. Accuracy measures the proportion of correct decisions, while precision, recall, and F1-score summarize positive-class reliability. The equations are presented in the same display structure used in the reference paper.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{5}$$

Precision evaluates the proportion of correctly detected forged images among all images predicted as forged:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{6}$$

Recall measures the proportion of actual forged images correctly detected by the model:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{7}$$

The F1-score provides the harmonic mean of precision and recall, while IoU and Dice evaluate spatial overlap between the predicted mask and the ground truth:

$$F_1 = \frac{2\,(\text{Precision})(\text{Recall})}{\text{Precision} + \text{Recall}} \tag{8}$$

$$\text{IoU} = \frac{|M \cap \hat{M}|}{|M \cup \hat{M}|}, \qquad \text{Dice} = \frac{2|M \cap \hat{M}|}{|M| + |\hat{M}|} \tag{9}$$

The training dashboard in Fig. 3 is retained directly from the output video and has only been cropped and resolution-enhanced for readability. Training accuracy rises from approximately 51% to 98%, while validation accuracy increases from about 49% to 94% over 30 epochs. The smooth trajectories and limited final separation indicate stable convergence with a moderate generalization gap.

Fig. 3. Complete training and validation accuracy graph

Figure 4 preserves the original categorization and binary confusion-matrix dashboard from the output video. It reports 1,842 true positives, 315 true negatives, 48 false positives, 71 false negatives, and a 96.8% categorization score. The screenshot is retained in its original visual form and enhanced only for readability.

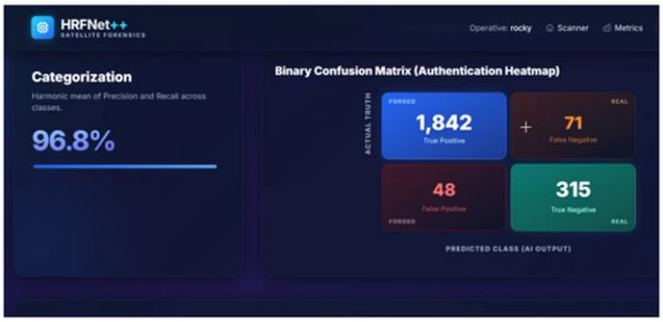

Fig. 4. Categorization and binary confusion-matrix dashboard

Figure 5 presents the original localization and categorization dashboard from the output video. The reported IoU, Dice coefficient, and categorization values are 94.2%, 96.5%, and 96.8%, respectively. Because these measures are already visible in the dashboard, no duplicate external chart is included for the same values.

Fig. 5. IoU, Dice coefficient, and categorization metrics

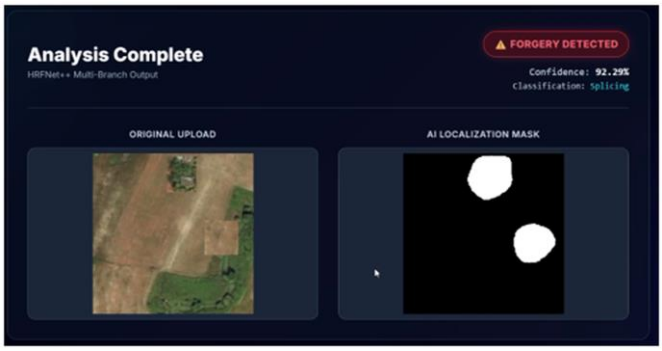

Fig. 6. Complete forged-image classification and localization output from the application.

Figure 6 shows the complete application output for a splicing example. The interface reports Forgery Detected with 92.29% confidence and displays the original satellite image beside the predicted binary localization mask. The paired view allows the analyst to verify whether the highlighted regions are spatially consistent with the reported manipulation class.

TABLE I
SUMMARY OF THE REPORTED HRFNET++ PERFORMANCE

Metric	Reported value
Training accuracy	98.00%
Validation accuracy	94.00%
Authentication accuracy	94.77%
Precision	97.46%
Recall	96.29%
F1-score	96.87%
IoU / Dice / Categorization	94.2% / 96.5% / 96.8%

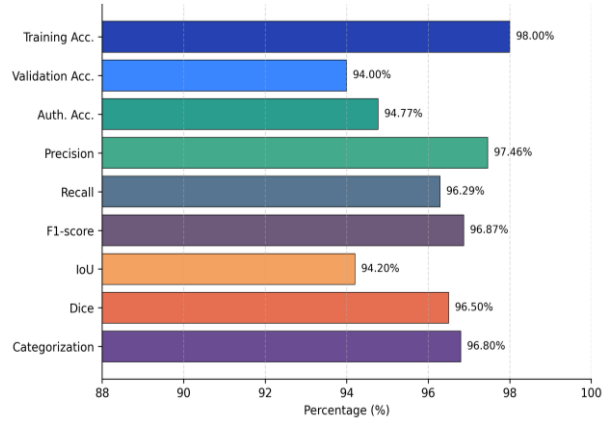

Fig. 7. Combined bar chart of the summary performance metrics

The retained screenshots, Table I, and Fig. 7 provide complementary evidence without repeating the same visualization. Figure 3 shows epoch-wise training and validation accuracy, Fig. 4 exposes the original authentication counts, Fig. 5 reports localization and categorization metrics, Fig. 6 presents the complete prediction result, and Fig. 7 summarizes the key reported performance values in a single combined bar chart.

The training curve increases rapidly during the early epochs and then approaches a plateau. Most of the learning gain occurs before approximately epoch 15, after which both trajectories change gradually. This behaviour suggests that the optimizer first learns the dominant scene and residual features and subsequently performs smaller parameter refinements. The absence of abrupt validation oscillations indicates stable optimization in the displayed run.

A four-percentage-point difference remains between the final training and validation accuracies. Such a gap is consistent with the expected difficulty of unseen satellite patches and does not, by itself, indicate severe overfitting. Nevertheless, the synthetic dataset contains multiple manipulated variants derived from common pristine scenes. Source-patch-separated experiments are therefore required before the displayed validation accuracy can be treated as an independent benchmark.

The confusion matrix shows that forged-image detection is dominated by correct positive decisions: 1,842 forged samples are identified correctly, compared with 71 missed forgeries. The false-positive count is 48, meaning that relatively few authentic images are incorrectly flagged. Consequently, precision is slightly higher than recall. This relationship is desirable for forensic screening because it limits unnecessary alerts while retaining strong sensitivity to manipulated content.

The F1-score remains close to both precision and recall, demonstrating that the authentication result is not driven by only one favourable metric. Accuracy is lower than precision because the matrix contains substantially more forged than authentic examples. Reporting all four measures is therefore important: accuracy summarizes the overall decision rate, precision describes the reliability of a forgery alert, recall measures the proportion of forgeries recovered, and F1-score balances the two error types.

The localization measures assess a different aspect of performance. The 94.2% IoU value indicates strong overlap between the predicted and reference manipulation regions, while the 96.5% Dice coefficient gives greater weight to the shared region. Their high values suggest that the decoder generally identifies the correct spatial area rather than relying only on an image-level classification cue. The 96.8% categorization score further indicates that the shared representation supports manipulation-type discrimination.

The reported measures should still be interpreted in relation to the experimental design. Synthetic copy-move, splicing,

insertion, removal, and replacement operations provide exact masks and balanced labels, but operational satellite products may contain compression, resampling, sensor noise, seasonal variation, atmospheric effects, and mosaicking seams. Robustness testing under these conditions is required to determine whether the displayed performance transfers to independently collected real-world forgeries.

Operational threshold selection also affects the practical meaning of the matrix. A lower decision threshold can recover additional forged images but may increase false alarms, whereas a higher threshold can improve alert precision at the cost of missed manipulations. The displayed operating point provides a strong balance, but deployment in defence, mapping, or legal-review workflows should calibrate the threshold using representative validation imagery and the relative cost of each error type.

The multi-task design supports this balanced behaviour because the binary decision is not learned in isolation. The shared representation is simultaneously constrained by manipulation-category supervision and pixel-level mask overlap. When these objectives agree, a forgery alert is supported by both categorical and spatial evidence. Disagreement between the outputs can also be useful: a high binary score with a weak or implausible mask can be flagged for analyst review rather than accepted automatically.

The results indicate that HRFNet++ is suitable as a decision-support mechanism rather than an autonomous forensic authority. The training screenshot verifies convergence, the matrix screenshot exposes the underlying counts, the metrics dashboard reports localization quality, and the prediction output links the image-level decision with the localized evidence. Table I consolidates the principal values.

Finally, the strongest evidence is the consistency among the reported measures. High precision and recall produce a closely matched F1-score, while the high IoU and Dice values indicate that classification quality is accompanied by accurate localization. The small training-validation separation and the low matrix error counts support the same overall conclusion: the model learns useful semantic and forensic representations, although independent source-separated and cross-sensor testing remains necessary for publication-grade generalization claims.

V. CONCLUSION

This paper presented HRFNet++, a hybrid semantic-forensic framework for satellite-image forgery analysis. The model combines fixed SRM residual extraction, dual ResNet-34 encoders, multi-scale ASPP fusion, binary and eight-class prediction heads, and a boundary-aware decoder. It jointly determines whether an image is authentic or forged, identifies the manipulation category, and localizes suspected pixels.

The dataset-development stage converts each DeepGlobe-derived pristine patch into one authentic sample and seven controlled manipulation variants with paired binary masks. Synchronized image-mask transformation, ImageNet normalization, and horizontal flipping create a consistent input pipeline. The supplied outputs report approximately 98% training accuracy, 94% validation accuracy, 94.2% IoU, 96.5% Dice coefficient, and 96.8% categorization performance.

The study has limitations. The complete image archive and trained checkpoint were not supplied for independent re-execution. The code-configured sprint mode samples 300 records and uses five epochs, while the web dashboard displays a 30-epoch convergence visualization. In addition, synthetic classes share source backgrounds and may not reproduce all sensor, compression, resampling, atmospheric, seasonal, and adversarial conditions encountered in operational remote-sensing imagery.

Future work should use source-patch-separated training, validation, and test manifests; repeated experiments with fixed random seeds; independent real satellite forgeries; and cross-sensor, cross-region, and cross-season evaluation. JPEG compression, resampling, blur, noise, radiometric changes, and adversarial post-processing should be included to assess robustness. Confidence intervals and ablation studies should isolate the contribution of the SRM branch, dual encoders, ASPP, decoder, and individual loss components.

For deployment, thresholds should be calibrated on representative validation data rather than fixed only by the default sigmoid value. Different operational environments assign different costs to false alarms and missed forgeries. Defence and legal workflows may require conservative thresholds and mandatory human review, whereas large archive screening may prioritize recall and use the model as an early-warning filter.

Computational efficiency is another important extension. The dual ResNet-34 design may be expensive for very large satellite mosaics. Lightweight backbones, shared early layers, knowledge distillation, mixed-precision inference, model pruning, and tiled processing can reduce memory consumption and latency while retaining the multi-task structure.

Localization can support analyst verification by directing attention to suspicious regions, but additional interpretability is desirable. Future versions may overlay masks on georeferenced imagery, expose SRM residual maps, display uncertainty near boundaries, and compare a suspected region with neighbouring temporal observations to distinguish malicious editing from natural land-cover change, cloud shadows, sensor artefacts, or mosaicking seams.

Overall, HRFNet++ provides a coherent foundation for multi-task satellite-image forensics. Its main strength is the joint use of scene semantics, fixed residual priors, multi-scale context, categorical reasoning, and pixel-level localization in a single deployable workflow. With larger real datasets, transparent benchmarking, provenance integration, and careful operational governance, the framework can evolve into a reliable decision-support component for remote-sensing integrity assessment.

REFERENCES

[1] I. Demir et al., “DeepGlobe 2018: A challenge to parse the Earth through satellite images,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), 2018.

[2] S. K. Yarlagadda, D. Güera, P. Bestagini, F. M. Zhu, S. Tubaro, and E. J. Delp, “Satellite image forgery detection and localization using GAN and one-class classifier,” Electron. Imaging, pp. 214-1–214-9, 2018, doi: 10.2352/ISSN.2470-1173.2018.07.MWSF-214.

[3] J. Horvath, H. Hao, D. Mas Montserrat, and E. J. Delp, “Anomaly-based manipulation detection in satellite images,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops, pp. 62–71, 2019.

[4] J. Horvath, D. Mas Montserrat, H. Hao, and E. J. Delp, “Manipulation detection in satellite images using deep belief networks,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops, pp. 2832–2840, 2020.

[5] J. Horvath, S. Baireddy, H. Hao, D. Mas Montserrat, and E. J. Delp, “Manipulation detection in satellite images using vision transformer,” in Proc. IEEE/CVF Conf.

Comput. Vis. Pattern Recognit. Workshops, pp. 1032–1041, 2021, doi: 10.1109/CVPRW53098.2021.00114.

[6] B. Zhao, S. Zhang, C. Xu, Y. Sun, and C. Deng, "Deep fake geography? When geospatial data encounter artificial intelligence," Cartogr. Geogr. Inf. Sci., vol. 48, no. 4, pp. 338–352, 2021, doi: 10.1080/15230406.2021.1910075.

[7] C. Gudavalli, E. Rosten, L. Nataraj, S. Chandrasekaran, and B. S. Manjunath, "SeeTheSeams: Localized detection of seam carving based image forgery in satellite imagery," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops, pp. 1–11, 2022.

[8] F. F. Niloy, K. K. Bhaumik, and S. S. Woo, "HRFNet: High-resolution forgery network for localizing satellite image manipulation," in Proc. IEEE Int. Conf. Image Process. (ICIP), pp. 3165–3169, 2023, doi: 10.1109/ICIP49359.2023.10221974.

[9] X. Ding, Y. Nie, J. Yao, and Y. Lang, "Forensic research of satellite images forgery: A comprehensive survey," Artif. Intell. Rev., 2024, doi: 10.1007/s10462-024-10909-w.

[10] P. Zhou, X. Han, V. I. Morariu, and L. S. Davis, "Learning rich features for image manipulation detection," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp. 1053–1061, 2018.

[11] J. H. Bappy, A. K. Roy-Chowdhury, J. Bunk, L. Nataraj, and B. S. Manjunath, "Exploiting spatial structure for localizing manipulated image regions," in Proc. IEEE Int. Conf. Comput. Vis., pp. 4970–4979, 2017.

[12] Y. Wu, W. AbdAlmageed, and P. Natarajan, "ManTra-Net: Manipulation tracing network for detection and localization of image forgeries with anomalous features," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp. 9543–9552, 2019.

[13] X. Bi, Y. Wei, B. Xiao, and W. Li, "RRU-Net: The ringed residual U-Net for image splicing forgery detection," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops, 2019.

[14] M.-J. Kwon, I.-J. Yu, S.-H. Nam, and H.-K. Lee, "CAT-Net: Compression artifact tracing network for detection and localization of image splicing," in Proc. IEEE Winter Conf. Appl. Comput. Vis., pp. 375–384, 2021.

[15] X. Chen, C. Dong, J. Ji, J. Cao, and X. Li, "Image manipulation detection by multi-view multi-scale supervision," in Proc. IEEE/CVF Int. Conf. Comput. Vis., pp. 14185–14193, 2021.

[16] J. Wang, Z. Wu, J. Chen, X. Han, A. Shrivastava, S.-N. Lim, and Y.-G. Jiang, "ObjectFormer for image manipulation detection and localization," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp. 2364–2373, 2022.

[17] S. Das, M. S. Islam, and M. R. Amin, "GCA-Net: Utilizing gated context attention for improving image forgery localization and detection," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops, pp. 81–90, 2022.

[18] S. Agrawal, P. Kumar, S. Seth, T. Parag, M. Singh, and R. V. Babu, "SISL: Self-supervised image signature learning for splicing detection and localization," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops, pp. 22–32, 2022.

[19] F. Guillaro, D. Cozzolino, A. Sud, N. Dufour, and L. Verdoliva, "TruFor: Leveraging all-round clues for trustworthy image forgery detection and localization," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp. 20606–20615, 2023.

[20] X. Guo, X. Liu, Z. Ren, S. Grosz, I. Masi, and X. Liu, "Hierarchical fine-grained image forgery detection and localization," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp. 3155–3165, 2023.

[21] D. Li, J. Zhu, M. Wang, J. Liu, X. Fu, and Z.-J. Zha, "Edge-aware regional message passing controller for image forgery localization," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp. 8222–8232, 2023.

[22] C. Qu, Y. Zhong, C. Liu, G. Xu, D. Peng, F. Guo, and L. Jin, "Towards modern image manipulation localization: A large-scale dataset and novel methods," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp. 10781–10790, 2024.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 770–778, 2016, doi: 10.1109/CVPR.2016.90.

[24] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," arXiv:1706.05587, 2017.

[25] J. Fridrich and J. Kodovský, "Rich models for steganalysis of digital images," IEEE Trans. Inf. Forensics Security, vol. 7, no. 3, pp. 868–882, 2012, doi: 10.1109/TIFS.2012.2190402.

[26] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in Proc. IEEE Int. Conf. Comput. Vis., 2017, doi: 10.1109/ICCV.2017.324.

[27] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations," in Deep Learn. Med. Image Anal., LNCS 10553, pp. 240–248, 2017, doi: 10.1007/978-3-319-67558-9_28.

[28] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in Proc. Int. Conf. Learn. Representations, 2019.

[29] I. Loshchilov and F. Hutter, "SGDR: Stochastic gradient descent with warm restarts," in Proc. Int. Conf. Learn. Representations, 2017.

[30] A. Buslaev et al., "Albumentations: Fast and flexible image augmentations," Information, vol. 11, no. 2, Art. no. 125, 2020, doi: 10.3390/info11020125.