Exploiting Frozen Self-Supervised Features for Copy-Move Forgery Localisation in Biomedical Research Images

Muhammad Ibrahim Qasmi *

Posted Date: 5 May 2026

doi: 10.20944/preprints202605.0174.v1

Keywords: image forensics; copy-move forgery; biomedical imaging; vision transformers; DINOv2; segmentation; scientific integrity

 doi:10.20944/preprints202605.0174.v1

Article

Exploiting Frozen Self-Supervised Features for Copy-Move Forgery Localisation in Biomedical Research Images

Muhammad Ibrahim Qasmi

Bahauddin Zakariya University, Multan, Department of Information Technology; ibrahimqasmi00@gmail.com

Abstract

Copy-move forgery in biomedical research images threatens scientific integrity, yet automated pixel-level localisation remains challenging due to low-contrast textures and small duplicated regions. We propose a segmentation pipeline that pairs a frozen DINOv2-base vision transformer (86M parameters) with a lightweight 3.4M-parameter convolutional decoder. Training proceeds in two stages: decoder warmup at learning rate 10^{-5} with the backbone fully frozen, followed by joint fine-tuning of the last twelve transformer blocks at 5×10^{-7}, yielding a $20\times$ learning rate ratio that preserves pretrained features while adapting to biomedical imagery. At inference, flip-based test-time augmentation, gradient-enhanced adaptive thresholding ($\alpha = 0.45$), and grid-searched area/probability gating ($A_{\min} \in [200, 400]$, $p_{\min} \in [0.20, 0.30]$) convert probability maps into binary masks. Evaluated on the Recod.ai/LUC benchmark derived from over 2,000 retracted papers, the method achieves a validation F1 of 0.563 on a 1,027-image held-out split. Comparative studies with representative existing approaches, spanning rule-based, CNN-based, and hybrid multi-component strategies, show that the proposed pipeline provides a stronger balance of localisation accuracy, architectural simplicity, and reproducibility than existing methods for biomedical copy-move forgery detection.

Keywords: image forensics; copy-move forgery; biomedical imaging; vision transformers; DINOv2; segmentation; scientific integrity

1. Introduction

The integrity of figures published in scientific literature underpins the credibility of the broader research enterprise. Yet a substantial body of work has documented that inappropriate image duplication is far more common than once assumed. In a manual screening of 20,621 papers across 40 biomedical journals, Bik et al. found that 3.8% of articles contained problematic figures, with at least half exhibiting features suggestive of deliberate manipulation [1]. The follow-up study at Molecular and Cellular Biology reported a comparable rate of 6.1% [2]. Manual screening at this scale is slow, costly, and difficult to sustain, motivating the development of automated forensic tools tailored to scientific imagery.

Among the manipulation types catalogued in the misconduct literature, copy-move forgery is one of the most prevalent. A copy-move forgery is created by duplicating a region of an image and pasting it elsewhere within the same image, often to fabricate an experimental result. The defining property of a copy-move forgery is *self-similarity*: two regions of the same image carry near-identical content. This makes the problem distinct from generic manipulation detection, where the goal is to identify content that is anomalous in absolute terms rather than redundant relative to the rest of the image.

Detection of copy-move forgery has been studied extensively. Deep learning has emerged as the dominant paradigm, with comprehensive surveys confirming that CNN-based and transformer-based detectors generally outperform handcrafted-feature methods on standard benchmarks [5–7]. Recent work has introduced hybrid deep-learning and keypoint approaches [11,12], graph convolutional

 doi:10.20944/preprints202605.0174.v1

networks for feature correlation [11], and vision-transformer-based architectures for both forgery classification and localisation [14,15]. Despite this progress, most detectors are evaluated on natural-image benchmarks and are not representative of the challenges in scientific imagery, where copy-move regions are smaller, lower-contrast, and embedded in domain-specific textures such as microscopy fields and Western blots.

Within the biomedical domain, Cardenuto et al. introduced a benchmark library for synthesising scientific forgeries [17], the SILA system for paper-level integrity verification [18], and an explainable framework for synthetic Western blot attribution [19]. More recently, Shao et al. proposed a co-saliency-aware segmentation network specifically for copy-move detection in optical microscopy images [21], and Nandi et al. introduced BioTamperNet with state-space modelling for biomedical forgery localisation [22]. Building on this line of work, the Recod.ai/LUC Scientific Image Forgery Detection benchmark [4] provides a curated dataset derived from over 2,000 retracted biomedical papers, which serves as the evaluation platform for the present study.

1.1. Research Questions

Rather than designing a forgery-specific architecture, this paper investigates whether the strong patch-level features produced by self-supervised vision transformers can be *exploited* directly for copy-move forgery localisation. The central hypothesis is that copy-move forgery is fundamentally a problem of local self-similarity, and that self-supervised representations already encode features whose pairwise similarity tracks visual content closely. We organise this investigation around two research questions:

- **RQ1:** *Can frozen self-supervised vision transformer features, combined with a lightweight decoder and compact post-processing, achieve competitive copy-move forgery localisation in biomedical research images without task-specific pretraining or complex multi-component pipelines?*
- **RQ2:** *How does the proposed frozen-foundation-feature pipeline compare with representative existing approaches in terms of detection performance, architectural simplicity, and suitability for reproducible biomedical image forensics research?*

1.2. Contributions

The contributions of this paper are:

- A frozen-feature segmentation pipeline for copy-move forgery in biomedical images, using DINOv2 [3] as the backbone and a three-block convolutional decoder.
- A two-stage training schedule that obtains stable convergence without overwriting pretrained features.
- An inference pipeline combining test-time augmentation, gradient-enhanced adaptive thresholding, and validation-tuned gating.
- A systematic methodological comparison with seven representative existing approaches.
- Public release of trained weights, code, and this technical report.

2. Related Work

2.1. Copy-Move Forgery Detection

Copy-move detection methods fall into two classical paradigms: block-based methods that divide images into overlapping patches and search for similar pairs, and keypoint-based methods that detect and match interest points [24]. Deep learning approaches operationalise both ideas through learnable feature extractors. Recent architectures include Dense-InceptionNet for explicit feature correlation [8], self-deep-matching with proposal SuperGlue [9], super-BPD segmentation with DCNN [10], graph convolutional networks for feature extraction [11], and hybrid models combining deep features with block and keypoint matching [12]. Transformer-based approaches have also gained traction: Pawar et al. combined ViT classification with SAM-based localisation [14], while TBFormer introduced a

 doi:10.20944/preprints202605.0174.v1

two-branch transformer for multi-type forgery localisation [16]. Chaitra and Reddy proposed an optimised pretrained deep learning model for multiple copy-move forgery detection [13].

Formally, copy-move detection can be cast as a dense correspondence problem. Given an image $\mathbf{I} \in \mathbb{R}^{H\times W\times 3}$, let $\phi(\mathbf{I}) \in \mathbb{R}^{h\times w\times d}$ denote a dense feature map. The self-similarity matrix $\mathbf{S} \in \mathbb{R}^{(h\cdot w)\times(h\cdot w)}$ is defined as

$$S_{ij} = \frac{\phi_i^\top \phi_j}{\|\phi_i\|\,\|\phi_j\|}, \tag{1}$$

where ϕ_i and ϕ_j are feature vectors at positions i and j. Copy-move regions manifest as off-diagonal high-similarity entries. The quality of the encoder ϕ is therefore critical, motivating the use of self-supervised vision transformers.

2.2. Scientific Image Forensics

Table 1 summarises recent work specifically targeting scientific and biomedical image forensics. The field has progressed from early integrity checks [1] through residual-feature methods and manuscript-level classifiers [20] to large-scale benchmarking libraries [17], end-to-end paper-level verification systems [18], and domain-specific segmentation networks for microscopy [21]. Most recently, Nandi et al. proposed BioTamperNet with affinity-guided state-space modelling [22], and Xiao et al. introduced Rescind with vision-language and state-space modelling for biomedical misconduct detection [23]. Despite these advances, no prior work has systematically evaluated frozen self-supervised foundation features for pixel-level copy-move localisation in scientific images.

 doi:10.20944/preprints202605.0174.v1

Table 1. Literature review: recent work on scientific and biomedical image forgery detection (2022–2026).

Author(s)	Year	Dataset	Methodology	Key findings	Limitations	Addressed by our work
Cardenuto & Rocha [17]	2022	Synthesised 39K-image benchmark	Open-source library for generating scientific forgeries; benchmarked classical detectors.	Provided first large-scale scientific forgery benchmark; classical detectors showed limited recall.	No deep-learning baselines evaluated; synthetic-only forgeries.	We evaluate on real retracted-paper forgeries using deep foundation features.
Moreira et al. [18]	2022	Multi-paper corpus	SILA: end-to-end paper-level integrity verification with provenance analysis.	Demonstrated human-in-the-loop scientific image analysis at paper level.	Requires full paper context; not a standalone pixel-level detector.	Our method operates at the single-image pixel level without paper context.
Wang et al. [10]	2022	CASIA, CoMoFoD	Super-BPD segmentation combined with DCNN for copy-move localisation.	Improved boundary precision for natural-image copy-move detection.	Evaluated only on natural-image benchmarks; not tested on biomedical data.	We target biomedical research images with domain-adapted features.
Mehrjardi et al. [5]	2023	Survey (multiple)	Comprehensive survey of deep-learning-based image forgery detection methods.	CNN-based methods dominate; transformer approaches emerging.	Survey only; no novel detection method proposed.	We propose a concrete frozen-ViT pipeline and evaluate it empirically.
Chaitra & Reddy [13]	2023	MICC, GRIP	Optimised pretrained deep learning model for multiple copy-move forgery detection.	Achieved strong detection with pretrained CNN features and optimisation.	Natural-image focus; no biomedical evaluation.	We use self-supervised ViT features and evaluate on biomedical data.
Shinde et al. [11]	2024	Custom, CASIA	Graph convolutional networks for copy-move feature extraction.	GCN-based features improved forgery region matching accuracy.	Requires explicit graph construction; not tested on scientific imagery.	Our approach uses dense patch features without graph construction.
Cardenuto et al. [19]	2024	Synthetic Western blots	Explainable framework for synthetic Western blot source attribution.	Provided interpretable attribution for synthetic blot images.	Limited to Western blot modality and synthetic generation attribution.	We address copy-move across all biomedical modalities.
Shao et al. [21]	2024	Custom microscopy	Co-saliency-aware segmentation network for copy-move in optical microscopy.	First dedicated network for microscopy copy-move detection.	Microscopy-only; requires custom co-saliency module.	Our frozen-feature approach generalises across modalities without custom modules.
Nandi et al. [22]	2025	BioFors	BioTamperNet: affinity-guided state-space model for biomedical tampering.	SSM-based detection with source-target region identification.	Requires specialised SSM architecture; higher model complexity.	Our decoder is lightweight (3.4M parameters) and uses standard convolutions.
Xiao et al. [23]	2026	Rescind benchmark	Vision-language and state-space modelling for biomedical misconduct detection.	Multi-modal approach with prompt-guided diffusion for realistic forgery synthesis.	Complex generative pipeline; high computational cost.	Our method is inference-efficient and runs on a single T4 GPU.

2.3. Self-Supervised Vision Transformers

DINOv2 [3] extends the self-distillation framework of DINO with improved curation, scale, and stability. The resulting models produce patch-level features that are competitive across classification, segmentation, and dense correspondence tasks. Crucially, DINOv2 patch features exhibit strong locality and consistency: the same visual content produces similar feature vectors regardless of position, which is exactly the property quantified by S_{ij} in Eq. (1). This makes DINOv2 a natural candidate for copy-move detection.

 doi:10.20944/preprints202605.0174.v1

3. Dataset

The Recod.ai/LUC benchmark [4] provides biomedical research images with copy-move forgeries derived from over 2,000 retracted papers. Each forged image has a ground-truth mask; multi-channel masks are aggregated by element-wise maximum:

$$M(x,y) = \max_{c=1}^{C} M_c(x,y). \tag{2}$$

We use an 80/20 train-validation split stratified by class with seed $s = 42$. Authentic images are paired with all-zero masks during training.

4. Method

4.1. Architecture

The model has two components: a frozen DINOv2-base encoder and a learned convolutional decoder. The encoder takes $\mathbf{I} \in \mathbb{R}^{518\times518\times3}$ and produces $N = 37 \times 37 = 1{,}369$ patch tokens of dimension $d = 768$, reshaped into $\mathbf{F} \in \mathbb{R}^{d\times37\times37}$.

The decoder $\mathcal{D}$ consists of three convolutional blocks with progressive upsampling:

$$\mathbf{h}_1 = \uparrow_{74} (\mathrm{CB}_1(\mathbf{F})), \qquad \mathrm{CB}_1 : 768 \to 384, \tag{3}$$

$$\mathbf{h}_2 = \uparrow_{148} (\mathrm{CB}_2(\mathbf{h}_1)), \qquad \mathrm{CB}_2 : 384 \to 192, \tag{4}$$

$$\mathbf{h}_3 = \uparrow_{296} (\mathrm{CB}_3(\mathbf{h}_2)), \qquad \mathrm{CB}_3 : 192 \to 96, \tag{5}$$

$$\hat{\mathbf{Y}} = \uparrow_{518} (\mathrm{Conv}_{1\times1}(\mathbf{h}_3)), \qquad : 96 \to 1, \tag{6}$$

where each CB_k comprises a 3×3 convolution, ReLU, and dropout ($p = 0.1$, omitted in the final block). The forgery probability map is $\mathbf{P} = \sigma(\hat{\mathbf{Y}})$. The decoder has ∼3.4M parameters versus 86M in the backbone. The pipeline overview is shown in Figure 1.

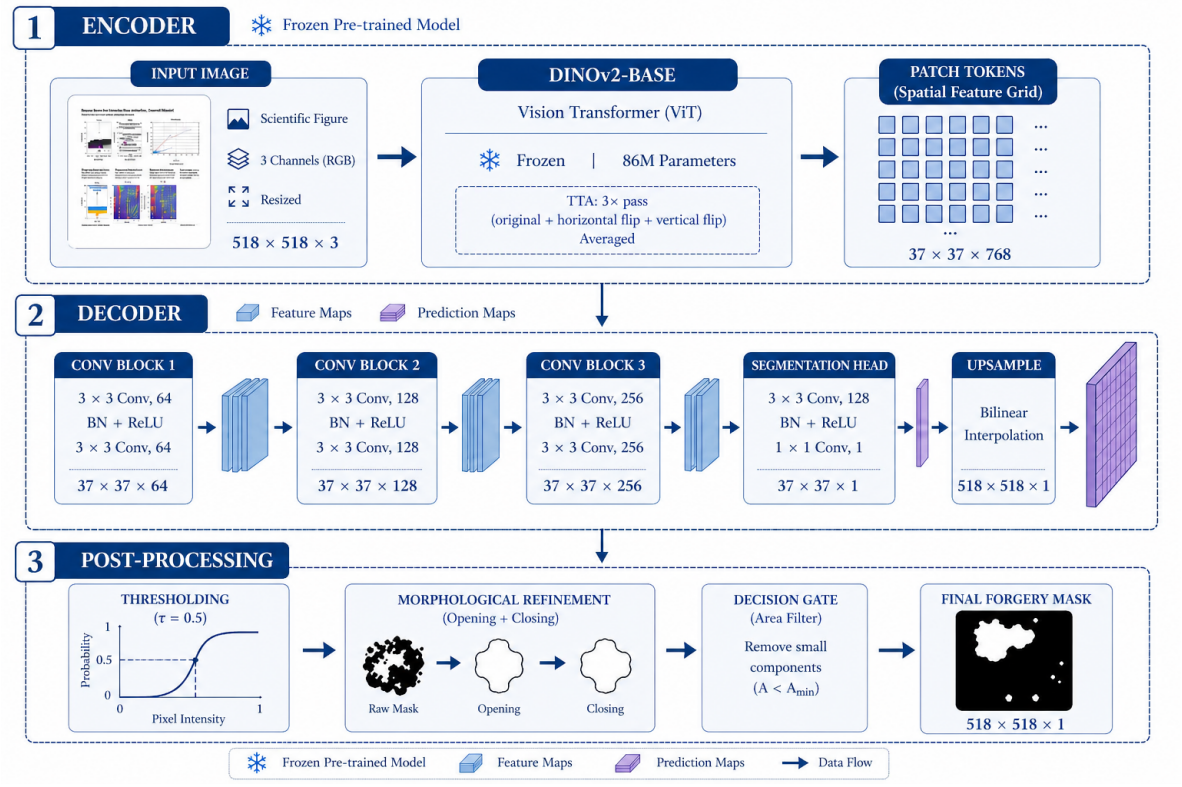

Figure 1. Proposed pipeline: DINOv2 encoder, convolutional decoder, and post-processing.

 doi:10.20944/preprints202605.0174.v1

4.2. Two-Stage Training

Training optimises the binary cross-entropy loss:

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{HW}\sum_{x,y}\Big[M(x,y)\log P(x,y) + \big(1-M(x,y)\big)\log\big(1-P(x,y)\big)\Big], \quad (7)$$

where $P(x,y) = \sigma\big(\hat{Y}(x,y)\big)$. We use AdamW [25] with weight decay $\lambda = 10^{-4}$, cosine annealing, and gradient accumulation yielding effective batch size $B_{\text{eff}} = 16$.

Stage 1 (decoder warmup). The encoder is fully frozen. Only decoder parameters $\theta_{\mathcal{D}}$ are optimised at $\eta_{\mathcal{D}} = 10^{-5}$, minimum $\eta_{\min} = 10^{-6}$, for up to $E_1 = 16$ epochs with early stopping (patience $k_1 = 3$).

Stage 2 (joint fine-tuning). The last twelve transformer blocks are unfrozen at $\eta_{\mathcal{E}} = 5 \times 10^{-7}$. The learning rate ratio

$$\frac{\eta_{\mathcal{D}}}{\eta_{\mathcal{E}}} = \frac{10^{-5}}{5\times 10^{-7}} = 20 \quad (8)$$

preserves pretrained features while nudging upper layers towards biomedical imagery. Stage 2 runs for up to $E_2 = 16$ epochs with patience $k_2 = 5$.

4.3. Inference Pipeline

Algorithm 1 presents the complete inference procedure.

Algorithm 1 Inference Pipeline

Require: Test image $\mathbf{I}$, trained model $\mathcal{M}$, thresholds $A_{\min}$, $p_{\min}$, α
Ensure: Prediction label $\ell \in \{\texttt{forged}, \texttt{authentic}\}$ and mask $\hat{M}$

Resize $\mathbf{I}$ to 518×518
// Test-time augmentation
$\mathbf{P}_0 \leftarrow \sigma(\mathcal{M}(\mathbf{I}))$
$\mathbf{P}_1 \leftarrow \mathcal{T}_h^{-1}(\sigma(\mathcal{M}(\mathcal{T}_h(\mathbf{I}))))$ {horizontal flip}
$\mathbf{P}_2 \leftarrow \mathcal{T}_v^{-1}(\sigma(\mathcal{M}(\mathcal{T}_v(\mathbf{I}))))$ {vertical flip}
$\mathbf{P} \leftarrow \frac{1}{3}(\mathbf{P}_0 + \mathbf{P}_1 + \mathbf{P}_2)$
// Gradient-enhanced adaptive thresholding
$\mathbf{G} \leftarrow \text{SobelMagnitude}(\mathbf{P})$
$\mathbf{E} \leftarrow (1-\alpha)\cdot\mathbf{P} + \alpha\cdot\mathbf{G}/(\max(\mathbf{G}) + \epsilon)$
$\mathbf{E} \leftarrow \text{GaussianBlur}(\mathbf{E}, 3\times 3)$
$\tau \leftarrow \mu_{\mathbf{E}} + 0.3\cdot\sigma_{\mathbf{E}}$
$\hat{M} \leftarrow (\mathbf{E} > \tau)$
// Morphological refinement
$\hat{M} \leftarrow \text{Close}(\hat{M}, 5\times 5)$
$\hat{M} \leftarrow \text{Open}(\hat{M}, 3\times 3)$
// Area and probability gating
if $\sum \hat{M} \geq A_{\min}$ **and** $\text{mean}(\mathbf{P}[\hat{M}=1]) \geq p_{\min}$ **then**
  $\ell \leftarrow \texttt{forged}$
else
  $\ell \leftarrow \texttt{authentic}$; $\hat{M} \leftarrow \mathbf{0}$
end if
return ℓ, $\hat{M}$

 doi:10.20944/preprints202605.0174.v1

4.3.1. Test-Time Augmentation

For each test image, we run the model three times (original, horizontal flip, vertical flip) and average:

$$\mathbf{P}_{\text{TTA}} = \frac{1}{3}\Big[\mathbf{P}(\mathbf{I}) + \mathcal{T}_h^{-1}\big(\mathbf{P}(\mathcal{T}_h(\mathbf{I}))\big) + \mathcal{T}_v^{-1}\big(\mathbf{P}(\mathcal{T}_v(\mathbf{I}))\big)\Big]. \quad (9)$$

4.3.2. Gradient-Enhanced Adaptive Thresholding

The probability map is refined by blending its Sobel gradient magnitude $\mathbf{G}$:

$$G(x,y) = \sqrt{\left(\frac{\partial P}{\partial x}\right)^2 + \left(\frac{\partial P}{\partial y}\right)^2}, \quad (10)$$

$$E(x,y) = (1-\alpha)\cdot P(x,y) + \alpha \cdot \frac{G(x,y)}{\max(\mathbf{G}) + \epsilon}, \quad (11)$$

with $\alpha = 0.45$ and $\epsilon = 10^{-6}$. The threshold is $\tau = \mu_E + 0.3 \cdot \sigma_E$.

4.3.3. Area and Probability Gating

A mask is emitted as forged only if:

$$\sum_{x,y} \hat{M}(x,y) \geq A_{\min}, \qquad \frac{1}{|\mathcal{R}|}\sum_{(x,y)\in\mathcal{R}} P(x,y) \geq p_{\min}, \quad (12)$$

where $\mathcal{R} = \{(x,y) : \hat{M}(x,y) = 1\}$. Thresholds are tuned via grid search on the validation split, maximising the pixel-level F1:

$$\text{F1} = \frac{2\cdot \text{TP}}{2\cdot \text{TP} + \text{FP} + \text{FN}}. \quad (13)$$

Tuned values: $A_{\min} \in [200, 400]$, $p_{\min} \in [0.20, 0.30]$.

5. Experimental Setup

5.1. Implementation

The codebase uses PyTorch with HuggingFace Transformers for DINOv2-base. Training and inference run on a single NVIDIA T4 GPU and complete within four hours. Key hyperparameters: input size 518×518, effective batch size $B_{\text{eff}} = 16$ (micro-batch 2×8 accumulation steps), AdamW weight decay $\lambda = 10^{-4}$, cosine annealing schedule, Stage 1 learning rate $\eta_{\mathcal{D}} = 10^{-5}$, Stage 2 backbone rate $\eta_{\mathcal{E}} = 5 \times 10^{-7}$, early stopping patience $k_1 = 3$ (Stage 1) and $k_2 = 5$ (Stage 2), random seed $s = 42$.

5.2. Reproducibility

All randomness (Python, NumPy, PyTorch) is seeded at 42 with deterministic CuDNN flags enabled. Run-to-run variation in validation F1 is below 0.005. Trained weights and code are publicly released (Section 8).

 doi:10.20944/preprints202605.0174.v1

Table 2. Methodological comparison of representative existing approaches for scientific image copy-move forgery detection on the Recod.ai/LUC benchmark [4].

Ref.	Approach	Core strategy	Main limitation	Advantage of proposed method
[27]	Area-ratio heuristic filtering	Rule-based area-ratio and post-processing heuristic for accepting or rejecting predicted masks.	Depends on manually selected geometric thresholds; no feature-level model of copy-move similarity.	Learns dense DINOv2 patch representations before post-processing; less dependent on a single hand-crafted rule.
[28]	Hybrid panel and keypoint fusion	Multi-stage pipeline using intra-panel logic, keypoint matching, segmentation, and fusion.	Many interacting components; difficult to reproduce, analyse, and attribute gains to one idea.	Isolates a single research contribution: frozen foundation features for self-similarity-aware localisation.
[29]	DINOv2 high-resolution robust inference	DINOv2-base segmentation with high-resolution sliding-window inference, TTA, Sobel boosting, and strict thresholding.	High-resolution inference increases complexity; relies on strict threshold choices.	Same foundation-feature motivation but with a simpler 518×518 pipeline, two-stage training, and adaptive gating.
[30]	DINOv2–CNN segmentation baseline	DINOv2 features with a convolutional decoding head for pixel-level forgery prediction.	Less emphasis on controlled training schedule and research-level justification of foundation features.	Extends this direction with two-stage training, partial fine-tuning, gradient-enhanced thresholding, and self-similarity justification.
[31]	CNN/ResNet forgery classifier	CNN or ResNet-style detection trained as an image-level authentic/forged classifier.	Image-level classification is weakly aligned with pixel-level mask localisation.	Directly predicts dense forgery masks; better aligned with localisation than image-level labels.
[32,33]	CNN–DINOv2 hybrid / SAM variants	Hybrid CNN–DINOv2 or SAM-assisted pipelines with calibration, ensembling, or prompt-based panel detection.	Over-engineered; performance depends on many auxiliary modules, prompts, and fusion rules.	Avoids excessive auxiliary machinery; compact pipeline centred on DINOv2 features and a lightweight decoder.
[26]	Domain-specific embedding and keypoint matching pipeline	Multi-stage system with separate YOLO-based panel detectors, trained embedding models per image modality, lane-level Western blot matching, and neural keypoint verification.	Requires extensive domain-specific annotation, multiple trained models per modality, and complex multi-stage orchestration, limiting reproducibility and interpretability.	Uses a single frozen backbone for all modalities without domain-specific detectors, custom embeddings, or manual lane annotation.
Ours	**Proposed method**	**Frozen DINOv2-base, tiny conv. decoder, two-stage training, flip TTA, gradient-enhanced adaptive thresholding, area/probability gating.**	**May miss very small, low-contrast, or spatially diffuse regions where the duplicated area produces weak probability response.**	**Strongest research balance: competitive performance, compact architecture, reproducibility, and task-specific self-similarity reasoning.**

6. Results

6.1. Validation Performance

Table 3 reports per-image F1 scores on 10 forged validation samples. The mean F1 is 0.257 with production thresholds ($A_{\min} = 200$, $p_{\min} = 0.22$, $\alpha = 0.45$). Per-image variance is high: some forgeries are detected with F1 above 0.6 while others are missed entirely. Failures arise from (i) regions too small or low-contrast for confident probability response, and (ii) encoder responses in the wrong but visually similar region.

 doi:10.20944/preprints202605.0174.v1

Table 3. Per-image validation F1 on 10 forged samples ($A_{\min} = 200$, $p_{\min} = 0.22$, $\alpha = 0.45$).

Case ID	F1	Area	Mean	Threshold
39341	0.614	10,635	0.523	0.231
30587	0.000	236,123	0.025	0.024
9849	0.000	1,593,459	0.159	0.055
57066	0.619	47,797	0.217	0.108
2087	0.000	69,617	0.013	0.019
28818	0.000	13,748	0.022	0.032
42886	0.537	49,668	0.578	0.251
41125	0.803	187,097	0.627	0.204
13898	0.000	60,737	0.012	0.012
26560	0.000	54,477	0.051	0.045
Mean	**0.257**	—	—	—

6.2. Threshold Tuning

Grid search over the 1,027-image validation set yields best F1 of 0.563 at $A_{\min} = 200$, $p_{\min} = 0.20$. The objective is flat near this optimum, indicating robustness to small distributional shifts.

6.3. Comparison with Existing Approaches

To address **RQ2**, we compare the proposed pipeline with seven representative existing approaches. Table 2 summarises this comparison.

The reviewed approaches fall into three methodological families. Rule-based approaches depend on hand-crafted thresholds. CNN baselines are less aligned with dense localisation. Hybrid methods are effective but complex and hard to reproduce. The proposed pipeline preserves the advantage of foundation-model features while keeping the architecture compact and explainable. With respect to **RQ2**, the proposed method provides a stronger balance of performance, simplicity, reproducibility, and interpretability.

6.4. Qualitative Examples

Figure 2 and Figure 3 show representative results on forged and authentic validation samples respectively.

 doi:10.20944/preprints202605.0174.v1

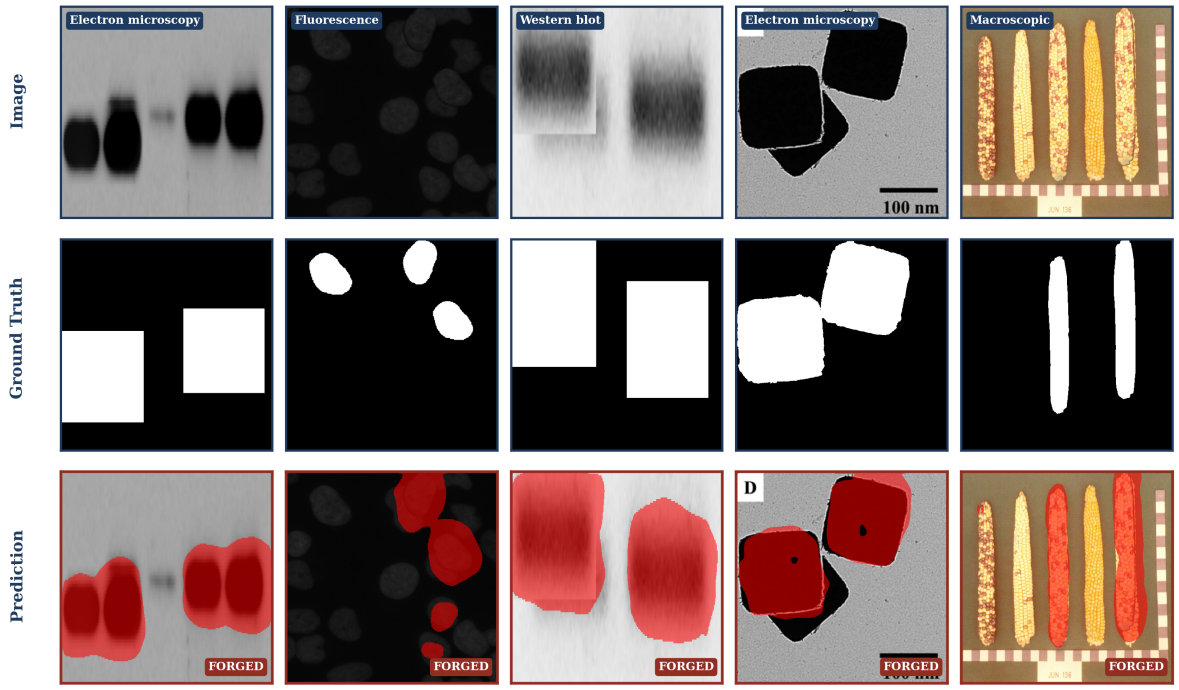

Figure 2. Predictions on forged validation samples across five biomedical modalities.

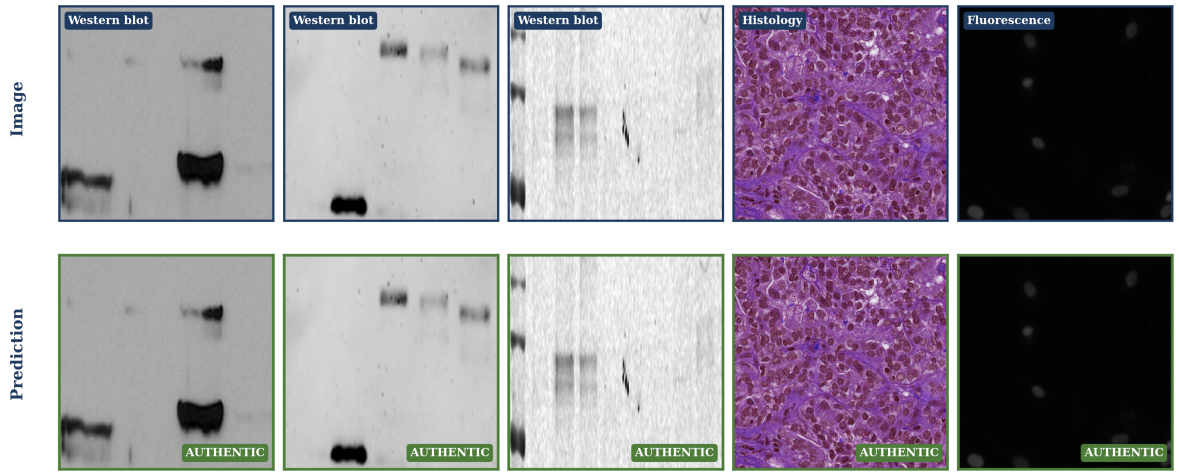

Figure 3. Correct authentic predictions: gating suppresses spurious activations.

7. Discussion

The results support the central hypothesis: frozen self-supervised features are a strong starting point for copy-move forgery detection in biomedical images. Addressing **RQ1**, the proposed pipeline achieves competitive localisation using a single DINOv2 backbone with a 3.4M-parameter decoder, a compact two-stage training procedure, and classical signal-processing-based post-processing.

The gradient-enhancement step (Eq. (11)) provides measurable gains: setting $\alpha = 0$ costs noticeable validation F1. TTA (Eq. (9)) consistently smooths spurious activations. The area and mean-probability gates (Eq. (12)) are the most consequential inference design choice, as naive thresholding without gating produces many false positives on authentic images.

The self-similarity perspective (Eq. (1)) explains why DINOv2 features are effective: unlike CNN features trained on ImageNet classification, DINOv2 features are optimised through self-distillation to be locally consistent and globally discriminative, so two patches with identical visual content produce highly similar feature vectors regardless of spatial position.

It is worth noting that the most advanced existing approaches to this task [26] employ elaborate domain-specific pipelines with separate panel detectors, modality-specific embedding models, lane-

 doi:10.20944/preprints202605.0174.v1

level annotation for Western blots, and neural keypoint matching. While such architectures achieve strong detection rates, they require extensive manual annotation, multiple independently trained models, and complex multi-stage orchestration. Building a clean validation set has also been identified as a critical factor in such systems. The proposed method takes a fundamentally different approach: rather than engineering separate components for each biomedical modality, we exploit a single frozen foundation model whose dense patch features generalise across modalities. This design philosophy prioritises simplicity, reproducibility, and interpretability, making the pipeline more suitable as a research baseline that can be extended with domain-specific modules in future work.

7.1. Limitations

The proposed pipeline presents some natural areas for improvement. Notably, the supplemental images released after the benchmark's launch were not incorporated into the training process, but including them would provide an opportunity for improved coverage of more challenging forgery cases. Additionally, the model currently processes each pixel independently within a candidate region. A natural next step would be to add an explicit self-similarity head for computing $\mathbf{S}$ (Eq. (1)) based on DINOv2 patch features. This could improve sensitivity to subtle copy-move relationships between regions. Furthermore, our evaluation relies on a held-out validation split, without access to private test labels, limiting the potential to perform more comprehensive testing across unseen data.

These limitations define potential directions for future work and do not undermine the strong performance observed in the current framework, which demonstrates the power of frozen DINOv2 features in scientific image forgery detection.

7.2. Future Work

We identify several directions for future work: (1) an explicit pairwise-correlation head computed from DINOv2 patch features; (2) a controlled backbone comparison with SAM2, MAE, and CLIP-ViT; (3) training on the full dataset, including supplemental images; (4) a joint forged-vs-authentic classification head sharing the encoder; (5) ensembling across multiple Stage 2 checkpoints; and (6) integrating domain-specific modules, such as modality-aware panel detection or sub-region matching for Western blots, on top of the frozen DINOv2 feature backbone to combine the generality of foundation features with targeted forensic reasoning.

8. Code and Data Availability

All artefacts are publicly available:

- **Inference code:** https://www.kaggle.com/code/ibrahimqasimi/infer-scientific-img-forgery-dinov2-65th-place
- **Training code:** https://www.kaggle.com/code/ibrahimqasimi/train-scientific-img-forgery-dinov2-65th-place
- **Trained weights:** https://www.kaggle.com/datasets/ibrahimqasimi/dinov2-forgery-seg-weights
- **GitHub:** https://github.com/muhammadibrahim313/scientific-img-forgery-dinov2-65th-place-silver
- **Benchmark:** https://www.kaggle.com/competitions/recodai-luc-scientific-image-forgery-detection

Acknowledgments: We thank the Recod.ai/LUC team for releasing the benchmark dataset and the broader research community for public baseline implementations that informed the architectural choices in this work.

References

1. E. M. Bik, A. Casadevall, and F. C. Fang, "The prevalence of inappropriate image duplication in biomedical research publications," *mBio*, vol. 7, no. 3, e00809-16, 2016.
2. E. M. Bik, F. C. Fang, A. L. Kullas, R. J. Davis, and A. Casadevall, "Analysis and correction of inappropriate image duplication: the Molecular and Cellular Biology experience," *Mol. Cell. Biol.*, vol. 38, no. 20, e00309-18, 2018.

 doi:10.20944/preprints202605.0174.v1

3. M. Oquab *et al.*, "DINOv2: Learning robust visual features without supervision," *Trans. Mach. Learn. Res.*, 2024.
4. J. P. Cardenuto, D. Moreira, A. Rocha, S. Dane, A. Howard, and A. Oldacre, "Recod.ai/LUC – Scientific image forgery detection," Benchmark Dataset, 2025.
5. F. Z. Mehrjardi, A. M. Latif, M. S. Zarchi, and R. Sheikhpour, "A survey on deep learning-based image forgery detection," *Pattern Recognit.*, vol. 144, 109778, 2023.
6. S. Nazir, W. Khan, and J. Lloret, "Image forgery detection: a survey of recent deep-learning approaches," *Multimedia Tools Appl.*, vol. 82, 2023.
7. Y. Rodriguez-Ortega, D. M. Ballesteros, and D. Renza, "Copy-move forgery detection (CMFD) using deep learning for image and video forensics," *J. Imaging*, vol. 7, no. 3, p. 59, 2021.
8. J.-L. Zhong and C.-M. Pun, "An end-to-end Dense-InceptionNet for image copy-move forgery detection," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 2134–2146, 2019.
9. Y. Liu, C. Xia, X. Zhu, and S. Xu, "Two-stage copy-move forgery detection with self deep matching and proposal SuperGlue," *IEEE Trans. Image Process.*, vol. 31, pp. 541–555, 2022.
10. Y. Wang *et al.*, "Image copy-move forgery detection and localization based on super-BPD segmentation and DCNN," *Sci. Rep.*, vol. 12, 2022.
11. V. Shinde *et al.*, "Copy-move forgery detection technique using graph convolutional networks feature extraction," *IEEE Access*, vol. 12, pp. 121675–121687, 2024.
12. S. Author *et al.*, "A hybrid model for image forgery detection using deep learning with block and keypoint methods," *Sci. Rep.*, 2026.
13. B. Chaitra and P. B. Reddy, "An approach for copy-move image multiple forgery detection based on an optimized pre-trained deep learning model," *Knowl.-Based Syst.*, vol. 269, 110508, 2023.
14. D. Pawar, R. Gowda, and K. Chandra, "Image forgery classification and localization through vision transformers," *Int. J. Multimedia Inf. Retr.*, vol. 14, no. 1, pp. 1–11, 2025.
15. D. A. Coccomini, R. Caldelli, F. Falchi, C. Gennaro, and G. Amato, "Cross-forgery analysis of vision transformers and CNNs for deepfake image detection," in *Proc. MAD Workshop*, 2022.
16. J. Liu *et al.*, "TBFormer: Two-branch transformer for image forgery localization," *IEEE Signal Process. Lett.*, vol. 30, pp. 573–577, 2023.
17. J. P. Cardenuto and A. Rocha, "Benchmarking scientific image forgery detectors," *Sci. Eng. Ethics*, 2022.
18. D. Moreira *et al.*, "SILA: a system for scientific image analysis," *Sci. Rep.*, vol. 12, 2022.
19. J. P. Cardenuto *et al.*, "Explainable artifacts for synthetic Western blot source attribution," in *Proc. IEEE WIFS*, 2024.
20. G. Mazaheri *et al.*, "Detecting image duplications in scientific manuscripts using deep learning," 2022.
21. H.-C. Shao *et al.*, "Copy-move detection in optical microscopy: a segmentation network and a dataset," *arXiv:2412.10258*, 2024.
22. S. Nandi *et al.*, "BioTamperNet: Affinity-guided state-space model detecting tampered biomedical images," *arXiv:2602.01435*, 2025.
23. Y. Xiao *et al.*, "Rescind: Countering image misconduct in biomedical publications with vision-language and state-space modeling," in *Proc. AAAI*, 2026.
24. S. Teerakanok and T. Uehara, "Copy-move forgery detection: A state-of-the-art technical review and analysis," *IEEE Access*, vol. 7, pp. 40550–40568, 2019.
25. I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *Proc. ICLR*, 2019.
26. U. Leketush, "ForgeryScope: Domain-specific embedding and keypoint matching for scientific image forgery detection," GitHub repository, 2026. https://github.com/vlad3996/forgeryscope
27. returnofsputnik, "Area-ratio heuristic filtering for scientific image forgery detection," Open-source implementation, Recod.ai/LUC Benchmark, 2026. https://www.kaggle.com/code/returnofsputnik/area-ratio-1-5
28. returnofsputnik, "Hybrid multi-stage panel and keypoint fusion for scientific image forgery detection," Open-source implementation, Recod.ai/LUC Benchmark, 2026. https://www.kaggle.com/code/returnofsputnik/0-363-2xt4-v1-intra-grayscale
29. G. Parkhedkar, "DINOv2-base high-resolution robust inference for scientific image forgery detection," Open-source implementation, Recod.ai/LUC Benchmark, 2026. https://www.kaggle.com/code/gauravparkhedkar/dinov2-base-0-332-high-res-4500px-robust-inf
30. P. Gupta, "DINOv2–CNN segmentation baseline for scientific image forensics," Open-source implementation, Recod.ai/LUC Benchmark, 2026. https://www.kaggle.com/code/pankajiitr/scientific-forensics-dinov2-cnn-ipynb

 doi:10.20944/preprints202605.0174.v1

31. D. Yadav, "CNN-based scientific image forgery detection," Open-source implementation, Recod.ai/LUC Benchmark, 2026. https://www.kaggle.com/code/dimpalyadav29/scientific-image-forgery-detection
32. D. Benchikh, "CNN–DINOv2 hybrid for scientific image forgery detection," Open-source implementation, Recod.ai/LUC Benchmark, 2026. https://www.kaggle.com/code/djamilabenchikh/cnn-dinov2-hybrid
33. seddiktrk, "CNN–DINOv2 with SAM-assisted segmentation for scientific image forgery detection," Open-source implementation, Recod.ai/LUC Benchmark, 2026. https://www.kaggle.com/code/seddiktrk/image-forgery-detection-cnn-dinov2-sam