ForgEgg-ViT: A Forgery-Aware Adaptive Transformer for Copy-Move Detection in Parasitic Egg Microscopy

Chouari Hala

Université de Sousse, Institut Supérieur des Sciences Appliquées et de Technologie de Sousse, LATIS- Laboratory of Advanced Technology and Intelligent Systems, 4023, Sousse, Tunisia; chouarihala@gmail.com

Lamia Rzouga Haddada

Université de Sousse, Institut Supérieur des Sciences Appliquées et de Technologie de Sousse, LATIS- Laboratory of Advanced Technology and Intelligent Systems, 4023, Sousse, Tunisia; lamia.rzouga@issatso.u-sousse.tn

Mohamed Ben Hammouda

Université de Sousse, Institut Supérieur des Sciences Appliquées et de Technologie de Sousse, LATIS- Laboratory of Advanced Technology and Intelligent Systems, 4023, Sousse, Tunisia; mohamed.ben.hammouda.pro@gmail.com

Najoua Essoukri Ben Amara

Université de Sousse, École Nationale d'Ingénieurs de Sousse, LATIS- Laboratory of Advanced Technology and Intelligent Systems, 4023, Sousse, Tunisia; najoua.benamara@eniso.rnu.tn

***Abstract*—The proliferation of advanced image editing tools poses an escalating threat to the integrity of biomedical imaging, where accurate diagnosis relies on authentic microscopic evidence. Existing detection approaches, whether Convolutional Neural Network (CNN)-based or generic transformer models, fail to simultaneously capture subtle local manipulation artifacts and global contextual inconsistencies in the visually homogeneous domain of parasitic egg microscopy. To address this gap, we propose ForgEgg-ViT, a domain-adapted FatFormer, a forgery-aware Vision Transformer originally developed for natural images, tailored for copy-move forgery detection in parasitic egg microscopy. The model integrates a Forgery-Aware Adapter (FAA) for dual spatial-frequency manipulation trace detection and a Language-Guided Alignment (LGA) module to localize semantic inconsistencies between duplicated and authentic regions. Domain adaptation is achieved through selective layer unfreezing, Focal Loss with label smoothing, and a Weighted Random Sampler. ForgEgg-ViT achieves 98.40% accuracy, 99.33% recall, and an F1-score of 0.984, outperforming both an adapted BusterNet baseline and CMSeg-Net without any pixel-level supervision, establishing a scalable, annotation-efficient framework for securing biomedical images against sophisticated manipulation.**

***Index Terms*—Copy-move forgery detection, parasitic egg microscopy, FatFormer, Vision Transformer, domain adaptation, biomedical image forensics.**

I. Introduction

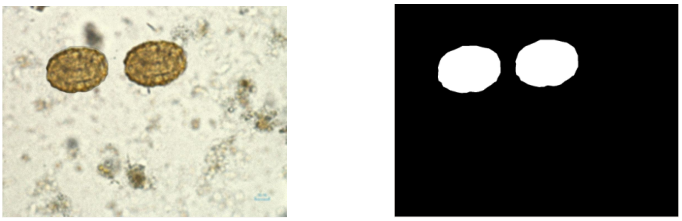
Scientific progress rests on a foundational assumption: that experimental images faithfully represent observed biological reality. This assumption currently faces increasing opposition. The past decade has revealed through high-profile misconduct cases which included systematic image duplication in cancer research and falsified Alzheimer's imagery and integrity-related academic resignations that image fraud constitutes a major problem in biomedical publications. Large-scale analyses of biomedical literature indicate that a non-negligible fraction of published articles contain inappropriate image duplications [1] which include copy-move manipulation [2] as one of the most common methods.

Beyond technical considerations, this issue is driven by structural incentives within the scientific ecosystem. The pressure to publish statistically significant and visually compelling results, combined with career and funding constraints, creates implicit motivations for subtle image manipulation. Rather than fabricating entire experiments, manipulations often involve localized operations such as Copy-Move Forgery (CMF), where regions are duplicated and repositioned within the same image to enhance visual evidence.

The implications of such falsifications extend beyond academic credibility. Manipulated microscopic images can distort clinical decision-making, contaminate medical AI training datasets, and ultimately impact patient safety [3], [4]. These challenges are particularly critical in parasitology, where microscopic identification of parasitic eggs directly informs diagnosis and treatment for hundreds of millions of patients worldwide.

Parasitic egg micrographs present unique challenges for Copy Move Forgery Detection (CMFD). Eggs from different species exhibit highly similar shapes, colors, and sizes, allowing duplicated regions to blend seamlessly into the image. Complex backgrounds containing biological artifacts closely resemble foreground structures, making manipulated regions difficult to distinguish. These forgeries are often performed with deliberate adversarial intent, employing smoothing and boundary blending to evade detection [5], [6].

Figure 1 illustrates a typical CMF encountered in this domain. The forged image (a) shows a parasitic egg duplicated to simulate a higher parasite load; the ground truth mask (b) highlights the manipulated region.



(a) Forged image (b) Ground truth mask

Fig. 1: Example of CMF in parasitic egg microscopy

Existing detection methods struggle to address these compounded challenges. Traditional feature-based approaches rely on strong texture patterns absent in homogeneous egg interiors. Convolutional Neural Networks (CNN) are inherently limited by fixed receptive fields, restricting their capacity to capture global contextual inconsistencies. Most critically, the absence of dedicated biomedical CMFD datasets has systematically

hindered the development of robust detectors adapted to this domain.

To address these limitations, we propose ForgEgg-ViT (A Forgery-Aware Adaptive Transformer for CMFD in Parasitic Egg Microscopy) and make the following contributions:

- Adaptation of forgery-aware adaptive transformers to biomedical microscopy. We propose ForgEgg-ViT, a domain-adapted forgery-aware adaptive transformer for CMFD in parasitic egg images, where homogeneous textures, small objects, and complex backgrounds present unique forensic challenges.
- A specialized training pipeline for biomedical forgery detection. We introduce a training methodology combining selective fine-tuning, Focal Loss with label smoothing, and a weighted random sampler, each component designed to address the unique forensic challenges of parasitic egg microscopy.

This paper is arranged as follows. Related work is discussed in Section II. Section III presents the proposed methodology. Experimental results are reported in Section IV. Finally, Section V concludes the study and suggests future research avenues.

II. Related Work

A. Traditional CMFD Methods

Traditional approaches to CMFD depend on handcrafted feature extraction [7] with similarity-based matching. Block-based strategies partition images into overlapping or non-overlapping blocks, then compare descriptors from transforms including Discrete Cosine Transform (DCT), Principal Component Analysis (PCA), and Local Binary Patterns (LBP) [8], [9]. Keypoint-oriented methods identify local features using Scale-Invariant Feature Transform (SIFT) or Speeded-Up Robust Features (SURF), then match features to locate duplicated regions [10], [11]. While effective under controlled conditions, they show limited resilience to geometric distortions, noise, and post-processing. More critically, their reliance on low-level texture fails in parasitic egg microscopy, where egg interiors are uniform and species differences subtle. Consequently, traditional CMFD approaches generally fail to generalize in complex real-world scenarios, especially biomedical imagery [12].

B. Deep Learning-Based Methods

To overcome the limitations of handcrafted features, deep learning-based approaches have been introduced. CNNs leverage hierarchical feature extraction to capture multi-scale patterns and are widely used for CMFD. BusterNet [13] introduced a dual-branch design combining manipulation feature extraction with self-correlation computation. Dense-InceptionNet [14] integrated multi-scale dense features for improved correlation matching. Recent works incorporate attention mechanisms [15] and generative models [16] to enhance detection. Lightweight architectures have also been explored for efficient CMFD [17]–[19].

Despite these advances, CNN-based methods remain limited by their local receptive fields, restricting their ability to capture long-range dependencies needed for detecting subtle copy-move manipulations. This limitation is particularly critical in CMFD, where duplicated regions may appear at different points in the same image. Consequently, CNN-based approaches often fail to detect global contextual differences, especially in visually similar biomedical images.

C. Transformer-Based Methods

Transformer architectures have recently emerged as powerful alternatives due to their ability to model global contextual relationships via self-attention. Vision Transformer (ViT) [20] and its variants show outstanding results in various computer vision tasks. Hybrid CNN-transformer architectures like ObjectFormer [21] and CNN-Transformer GAN models [22] integrate local feature extraction with global reasoning. More recently, FatFormer [23], a forgery-aware transformer, introduced specialized modules to capture manipulation traces in spatial and frequency domains while leveraging semantic alignment. Though these methods outperform CNN-based approaches via global reasoning, most are developed and evaluated on natural image datasets. Consequently, their effectiveness in biomedical imaging remains uncertain due to the substantial domain gap, characterized by low texture diversity, small object structures, and complex biological backgrounds. Therefore, directly applying these models to biomedical microscopy without domain adaptation may lead to suboptimal performance. In particular, adapting forgery-aware transformer architectures to biomedical data remains largely unexplored.

D. Biomedical Image Forgery Detection

Forgery detection in biomedical imaging remains relatively underexplored. Recent efforts introduced domain-specific datasets such as FakeParaEgg [24], highlighting the unique challenges of detecting copy-move manipulations in parasitic egg microscopy. CMSeg-Net [25], the only previous research dedicated to this dataset, solves the problem via a pixel-based co-saliency segmentation method and achieves strong localization performance.

However, segmentation-based approaches require pixel-level annotations, which are costly and time-consuming to obtain, limiting their scalability in real-world forensic applications. Moreover, biomedical images present distinct challenges compared to natural images: homogeneous textures, subtle manipulation artifacts, high inter-class similarity, and small object structures, making CMFD significantly more difficult.

These limitations highlight a critical gap: the absence of an image-level, forgery-aware framework specifically adapted to biomedical microscopy that can effectively capture both local inconsistencies and global contextual relationships without relying on pixel-level annotations. This gap motivates the proposed approach.

III. Methodology

This approach aims to detect CMF in parasitic egg microscopy images using image classification to determine whether each image is authentic or forged. By adapting FatFormer [23], a forgery-aware adaptive transformer pre-trained on natural images, through a specialized fine-tuning pipeline, this methodology effectively transfers manipulation-sensitive representations to the biomedical domain.

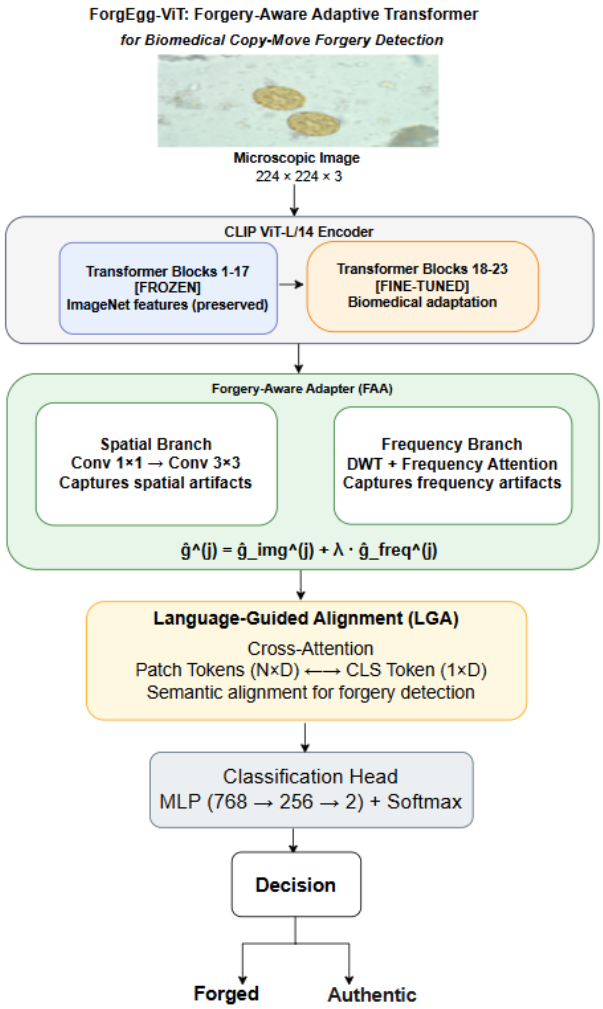


Fig. 2: Overall framework of the proposed ForgEgg-ViT for biomedical CMFD

As shown in Fig. 2, the proposed framework integrates spatial and frequency-aware modules within a transformer backbone to effectively capture both local inconsistencies and global contextual relationships, a critical capability for CMFD in parasitic egg microscopy.

A. Backbone: FatFormer

FatFormer [23] is built upon the Contrastive Language-Image Pre-training (CLIP) ViT-L/14 image encoder [26] and extends it with two core components specifically designed for forgery detection.

1) Forgery-Aware Adapter (FAA)

The FAA captures manipulation traces in both spatial and frequency domains using a lightweight convolutional module and a Discrete Wavelet Transform (DWT)-based attention mechanism. This dual-domain approach is particularly effective for parasitic egg microscopy because CMF often leaves subtle fingerprints in both domains: spatially, duplicated regions may exhibit imperceptible boundary artifacts or texture discontinuities; in the frequency domain, manipulation introduces high-frequency inconsistencies that are invisible to the human eye but detectable through wavelet analysis. The outputs from spatial and frequency branches are fused through a learnable scale factor λ that adaptively balances their contributions based on the input image characteristics as described in Equation (1):

$$\hat{g}^{(j)} = \hat{g}^{(j)}_{\text{img}} + \lambda \cdot \hat{g}^{(j)}_{\text{freq}} \tag{1}$$

2) Language-Guided Alignment (LGA)

The LGA module improves forgery detection through its ability to connect local image patch tokens with global image embeddings using cross-attention mechanism. This mechanism directs the encoder to concentrate on areas which contain forgery evidence through semantic matching. The intuition is that manipulated regions, while visually similar to authentic areas, often exhibit subtle semantic inconsistencies when compared to the global context of the image. For example, a forged egg may be perfectly duplicated in terms of texture, but its spatial relationship with surrounding biological structures may be semantically inconsistent. The LGA module leverages CLIP's pre-trained language-image alignment capabilities to learn these semantic relationships, effectively directing the model's attention to regions where local-global semantic coherence is disrupted. The final prediction probability is given by Equation (2):

$$\hat{P}^{(i)} = \frac{\exp\big((S^{(i)} + S'^{(i)})/\tau\big)}{\sum_k \exp\big((S^{(k)} + S'^{(k)})/\tau\big)} \tag{2}$$

where $S^{(i)}$ is the cosine similarity between the image CLS token and the i-th class embedding, $S'^{(i)}$ is the average patch-level similarity, and τ is the temperature parameter. The sum in the denominator runs over all classes k. The model is initialized from the publicly available checkpoint pre-trained on 4-class Progressive Growing of GANs (ProGAN) data.

B. Selective Fine-Tuning Strategy for Biomedical Domain Adaptation

To transfer forgery-sensitive representations from natural images to biomedical microscopy while avoiding catastrophic forgetting, all parameters of the ViT-L/14 encoder are frozen except for the last six Transformer blocks (residual blocks 18–23), which are selectively unfrozen to enable high-level semantic adaptation. This selective approach is motivated by the observation that shallow layers capture generic visual features (edges, textures, simple patterns) that transfer well across domains, while deeper layers encode task-specific representations that require adaptation to the unique characteristics of parasitic egg images: homogeneous egg interiors, subtle texture variations, and complex biological backgrounds. All FAA and LGA modules remain fully trainable. A differential learning rate is applied: 5×10^{-7} for the unfrozen encoder blocks and 5×10^{-5} for the FAA, LGA, and remaining adapter modules, ensuring that pre-trained representations are only minimally adjusted while meaningful domain adaptation is achieved.

C. Training Configuration for Biomedical Forgery Detection

To address the inherent challenges of biomedical forgery detection, we implement a specialized training configuration comprising three key components.

1) Weighted Random Sampler

The Weighted Random Sampler gives authentic-class samples a sampling weight of 1.5, which increases their likelihood of being selected. The dataset maintains an equal 50/50 distribution, yet authentic images remain simpler to identify than expert forgeries. The sampler controls which authentic samples the

model will see during training, and this process helps the model learn to identify real images while reducing the risk of authentic images being misidentified as forged, which is vital for preserving trust in forensic work.

2) Focal Loss with Label Smoothing

Focal Loss [27] with label smoothing is adopted as the training objective to focus learning on hard-to-classify examples as expressed in Equation (3):

$$\mathcal{L}_{\text{focal}} = -\frac{1}{N}\sum_{i=1}^{N}\Big[y_i\alpha_{\text{forged}}(1-p_i)^{\gamma}\log p_i + (1-y_i)\alpha_{\text{authentic}}p_i^{\gamma}\log(1-p_i)\Big] \tag{3}$$

where $\gamma = 2.0$ controls the focusing strength, $\alpha_{\text{forged}} = 1.0$ for the forged class, $\alpha_{\text{authentic}} = 1.5$ for the authentic class, and label smoothing $\varepsilon = 0.05$ prevents overconfidence.

3) Optimization and Regularization

Optimization is performed using Adam with weight decay (AdamW) [28] with weight decay 10^{-4} and a CosineAnnealingWarmRestarts scheduler. Gradient clipping with max norm 1.0 stabilizes training. The batch size is set to 8, with a maximum of 15 epochs and early stopping after 5 consecutive epochs without improvement.

IV. Experiments and Results

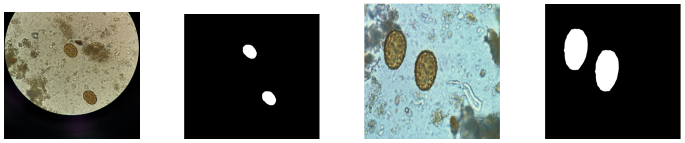
A. Dataset

The assessment uses a balanced dataset which combines two data sources that are available to the public. Authentic images are drawn from the Chula-ParasiteEgg-11 dataset [29], which contains 13,200 microscopic images of parasitic eggs across 11 species. Forged images are sourced from FakeParaEgg [24], comprising 1,500 copy-move forged microscopic images augmented through data augmentation techniques that include rotation, affine transformations, color jittering, and grayscale conversion to 7,500 samples. The resulting balanced dataset of 15,000 images is divided into three splits: training (9,000 images), validation (3,000 images), and testing (3,000 images), each maintaining a strict 50/50 class balance. Table I summarizes the dataset distribution.

TABLE I: Dataset Distribution

Split	Authentic	Forged	Total
Train	4,500	4,500	9,000
Validation	1,500	1,500	3,000
Test	1,500	1,500	3,000
Total	**7,500**	**7,500**	**15,000**

Fig. 3 shows two examples from the used FakeParaEgg dataset.



(a) Forged image (b) Ground truth mask (c) Forged image (d) Ground truth mask

Fig. 3: Examples from the FakeParaEgg dataset

B. Evaluation Metrics

We use multiple metrics [30] to evaluate the proposed ForgEgg-ViT model, including Precision, Recall, F1-Score, and Accuracy. These metrics are derived from the confusion matrix, which defines True Positives (TP, forged correctly detected), True Negatives (TN, authentic correctly identified), False Positives (FP, authentic misclassified as forged), and False Negatives (FN, forged missed).

Precision, which is given by Equation (4), measures the proportion of correctly detected forgeries among all images predicted as forged. Recall (Equation (5)) quantifies the model's ability to identify all actual forged images. The F1-Score (Equation (6)) provides a harmonic mean of precision and recall. Accuracy (Equation (7)) represents the overall correct classification rate.

$$\text{Precision} = \frac{TP}{TP+FP} \tag{4}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{5}$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{7}$$

C. Implementation Details

All experiments are implemented in PyTorch and run on an NVIDIA A6000 GPU. The model is initialized from the official FatFormer checkpoint pre-trained on 4-class ProGAN data. Input images are resized to 256×256 pixels using bicubic interpolation and normalized using CLIP pre-training statistics ($\mu = [0.481, 0.458, 0.408]$, $\sigma = [0.269, 0.261, 0.276]$). A center crop to 224×224 is applied at test time.

The key hyperparameters used in this work are summarized in Table II.

TABLE II: Summary of Training Hyperparameters

Parameter	Value
Input size	224×224
Optimizer	AdamW
Weight decay	1×10^{-4}
lr (encoder blocks 18–23)	5×10^{-7}
lr (FAA, LGA, adapters)	5×10^{-5}
Scheduler	CosineAnnealingWarmRestarts
T_0	5
T_{mult}	2
$\eta_{\min}$	1×10^{-8}
Batch size	8
Max epochs	15
Early stopping patience	5
Gradient clipping	1.0
Label smoothing ε	0.05
Focal loss γ	2.0
Class weight (forged) α_{forged}	1.0
Class weight (authentic) $\alpha_{\text{authentic}}$	1.5

D. Results

Table III presents the image-level classification performance of ForgEgg-ViT on the test set comprising 3,000 images. The model achieves 98.40% accuracy, 97.51% precision, 99.33% recall, and an F1-score of 0.984.

TABLE III: Detection Performance of the Proposed ForgEgg-ViT Model

Metric	Accuracy	Precision	Recall	F1-Score
Value	98.40%	97.51%	99.33%	0.984

Figure 4 shows the confusion matrix of the proposed ForgEgg-ViT. The results demonstrate that 1,490 out of 1,500 authentic images and 1,462 out of 1,500 forged images are correctly classified, with only 10 false positives and 38 false negatives.

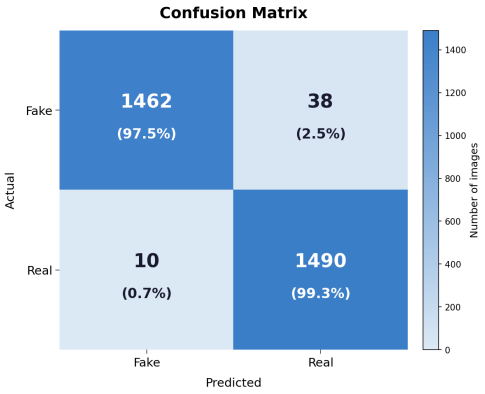


Fig. 4: Confusion matrix of the proposed ForgEgg-ViT model

To situate ForgEgg-ViT within the existing literature, two baselines are considered. CMSeg-Net [25] is the only prior work designed and evaluated on the FakeParaEgg dataset. It frames copy-move detection as a pixel-level co-saliency segmentation task and requires dense ground-truth masks for training. To provide a strong CNN-based image-level reference point, BusterNet [13] is adapted for binary classification: its dual-branch VGG-based architecture, which jointly models manipulation cues and self-correlation across image regions, is preserved in full, while its pixel-level localization outputs are replaced by a global average pooling layer followed by a binary classification head. This architectural adaptation shifts the model from segmentation to image-level forgery decision without altering its core feature learning capacity. To ensure a strictly fair comparison, the adapted BusterNet is trained on the same dataset, the same splits, and under the same optimization strategy as ForgEgg-ViT.

Figure 5 benchmarks ForgEgg-ViT against CMSeg-Net [25] and the adapted BusterNet baseline across the three metrics shared between pixel-level and image-level methods (F1, Precision, Recall), providing a cross-paradigm assessment of detection capability on the FakeParaEgg dataset.

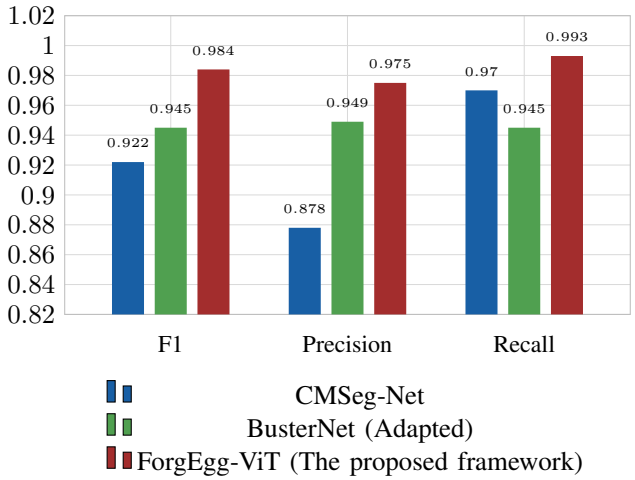


Fig. 5: Comparison with state-of-the-art methods on FakeParaEgg dataset

ForgEgg-ViT outperforms both baselines across all shared metrics, achieving a +3.9% F1 and +4.8% recall gain over BusterNet, and a +6.2% F1 and +9.7% precision gain over CMSeg-Net. Remarkably, while CMSeg-Net is a pixel-level segmentation method trained with dense ground-truth masks, ForgEgg-ViT exceeds it on every shared detection metric operating solely at the image level, with no pixel-level annotations, which are costly, time-consuming, and rarely available at scale in the medical domain. Unlike CMSeg-Net, which relies on CNN-based segmentation with limited receptive fields, and BusterNet, whose dual-branch correlation architecture was designed for natural-image localization and struggles with the homogeneous textures and subtle inter-class differences of parasitic egg microscopy, ForgEgg-ViT leverages the Vision Transformer's self-attention mechanism to capture long-range dependencies between spatially distant duplicated regions, a capability that is precisely what copy-move detection demands. These results demonstrate that adapting forgery-aware Transformers to biomedical imaging is not only feasible but also highly effective, offering a robust and scalable framework for securing biomedical images against advanced tampering attempts.

Table IV compares the confusion matrix metrics of proposed ForgEgg-ViT and adapted BusterNet. Adapted BusterNet misclassifies 153 forged images as authentic (FN) and 12 authentic images as forged (FP), revealing the inherent difficulty CNN-based architectures face in this domain. In contrast, proposed ForgEgg-ViT produces only 38 false negatives and 10 false positives on the same test set, confirming its substantially stronger sensitivity to copy-move manipulations while maintaining a very low false alarm rate, a critical requirement in biomedical forensic applications.

TABLE IV: Comparison of confusion matrix metrics on the test set (1500 authentic, 1500 forged).

Model	TP	TN	FP	FN
Proposed ForgEgg-ViT	1462	1490	10	38
Adapted BusterNet	1347	1488	12	153

E. Discussion

ForgEgg-ViT demonstrates strong performance in CMFD for parasitic egg microscopy, showing the effectiveness of transformer-based architectures in biomedical image forensics. The model's success is driven by its ability to jointly model local manipulation artifacts and global contextual relationships. The FAA detects subtle spatial and frequency inconsistencies, while the LGA enhances sensitivity to semantic discrepancies. The selective fine-tuning strategy effectively transfers forgery-aware representations from natural images to the biomedical domain. The image-level classification approach enables the model to avoid costly pixel-level annotations while achieving better system scalability. Nevertheless, the proposed approach presents certain limitations. The model is evaluated on a single dataset, which may limit generalization to other biomedical imaging modalities. Furthermore, the computational complexity of the ViT-L/14 backbone raises concerns regarding model size, inference time, and memory footprint in resource-constrained clinical environments.

V. Conclusion

This study presented ForgEgg-ViT, a domain-adapted forgery-aware adaptive Transformer for CMFD in parasitic egg microscopy. By combining transformer-based global modeling with a tailored adaptation and training strategy, the model achieves effective detection of subtle manipulation artifacts in biomedical images. Experimental results demonstrate strong performance, achieving 98.40% accuracy and an F1-score of 0.984, outperforming both CMSeg-Net and an adapted BusterNet baseline without pixel-level annotations, demonstrating the capacity of transformer-based models for scalable biomedical image forensics.

This performance implies that automated forensic screening can reliably flag tampered microscopic images in clinical diagnostics and scholarly peer review, reducing misdiagnosis risks and upholding research integrity at scale.

Future work will focus on extending the method to more complex forgery types such as retouching and splicing, improving generalization across diverse datasets, and investigating model compression and distillation strategies to reduce model size, inference time, and memory footprint for real-time clinical deployment.

References

[1] E. Sabir, S. Nandi, W. AbdAlmageed, and P. Natarajan, "MONet: Multi-scale overlap network for duplication detection in biomedical images," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Bordeaux, France, Oct. 2022, pp. 3793–3797.

[2] I. Shallal, L. Rzouga Haddada, and N. Essoukri Ben Amara, "Improved image forgery detection based on VGG16, cosine similarity, and support vector machines," in *Proc. 11th Int. Conf. Control, Decision Inf. Technol. (CoDIT)*, Split, Croatia, 2025, pp. 888–893.

[3] P. Capasso, G. Cattaneo, and M. De Marsico, "A comprehensive survey on methods for image integrity," *ACM Transactions on Multimedia Computing, Communications & Applications*, vol. 20, no. 11, pp. 1-34, Nov. 2024.

[4] X. Lin, W. Tang, H. Wang, Y. Liu, Y. Ju, S. Wang, and Z. Yu, "Exposing image splicing traces in scientific publications via uncertainty-guided refinement," *Patterns*, vol. 5, p. 101038, Nov. 2024.

[5] G. Zingarini, D. Cozzolino, G. Poggi, and L. Verdoliva, "M3Dsynth: A dataset of medical 3D images with AI-generated local manipulations," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Seoul, South Korea, Apr. 2024, pp. 1-5.

[6] S. Li, Z. Xing, H. Wang, P. Hao, X. Li, Z. Liu, and L. Zhu, "Toward medical deepfake detection: A comprehensive dataset and novel method," in *Proc. Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Daejeon, Korea, Sep. 2025, pp. 626-637.

[7] I. Shallal, L. Rzouga Haddada, and N. Essoukri Ben Amara, "Enhanced detection of copy-move forgery by fusing scores from handcrafted and deep learning-based detection systems," in *Proc. 17th Int. Conf. Develop. eSystems Eng. (DeSE)*, 2024, pp. 113–118.

[8] Koritala, S.P.; Chimata, M.; Polavarapu, S.N.; Vangapandu, B.S.; Manikandan, V.M., "An Efficient Copy-Move Forgery Detection using Discrete Cosine Transform with Block-wise Peak-Pixel-based Block Clustering," in *Proc. 2024 15th Int. Conf. Computing Communication and Networking Technologies (ICCCNT)*, Mandi, India, June 2024, pp. 1–6.

[9] Hamidon, N.R.; Salleh, S.F.M.; Warif, N.B.A., "An Analysis of Copy-Move Forgery Detection based on Discrete Cosine Transform against JPEG Compression and Brightness Changes Attack," in *Proc. 2024 1st Int. Conf. Cyber Security and Computing (CyberComp2024)*, Melaka, Malaysia, Nov. 2024, pp. 45–50.

[10] Fu, G.; Zhang, Y.; Wang, Y., "Image copy-move forgery detection based on fused features and density clustering," *Appl. Sci.*, vol. 13, no. 13, p. 7528, 2023.

[11] D. Tralic, I. Dunaj, and S. Grgic, "Copy-move forgery detection using SIFT and spatially constrained matching," *Multimed. Tools Appl.*, vol. 81, pp. 27407-27428, 2022.

[12] S. Gazzah, L. R. Haddada, I. Shallal, and N. Essoukri Ben Amara, "Digital Image Forgery Detection with Focus on a Copy-Move Forgery Detection: A Survey," in *Proc. 2023 Int. Conf. Cyberworlds (CW)*, Sousse, Tunisia, 2023, pp. 240–247.

[13] J. Li, P. Wei, and Y. Zhang, "BusterNet 2.0: Enhanced copy-move forgery detection with improved source/target localization," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, 2023, pp. 1234-1243.

[14] J. L. Zhong and C. M. Pun, "An end-to-end dense-inceptionnet for image copy-move forgery detection," *IEEE Trans. Inf. Forensics Secur.*, vol. 15, pp. 2134–2146, 2020.

[15] J. L. Zhong, J. X. Yang, Y. F. Gan, L. Huang, and H. Zeng, "Coarse-to-fine spatial-channel-boundary attention network for image copy-move forgery detection," *Soft Comput.*, vol. 26, no. 21, pp. 11461–11478, 2022.

[16] A. Islam, C. Long, A. Basharat, and A. Hoogs, "DOA-GAN: Dual-order attentive generative adversarial network for image copy-move forgery detection and localization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 4676–4685.

[17] I. Shallal, L. Rzouga Haddada, and N. Essoukri Ben Amara, "Lightweight hybrid model combining MobileNetV2 and PCA for copy-move forgery detection," in *Proc. 17th Int. Conf. Develop. eSystems Eng. (DeSE)*, 2024, pp. 107–112.

[18] M. N. Abbas, M. S. Ansari, M. N. Asghar, N. Kanwal, T. O'Neill, and B. Lee, "Lightweight deep learning model for detection of copy-move image forgery with post-processed attacks," in *Proc. IEEE World Symp. Appl. Mach. Intell. Informat. (SAMI)*, 2021, pp. 1–6.

[19] O. Kuznetsov, E. Frontoni, L. Romeo, and R. Rosati, "Enhancing copy-move forgery detection through a novel CNN architecture and comprehensive dataset analysis," *Multimed. Tools Appl.*, vol. 83, no. 21, pp. 59783–59817, 2024.

[20] A. Dosovitskiy *et al.*, "An image is worth 16×16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2021.

[21] J. Wang *et al.*, "ObjectFormer for image manipulation detection and localization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022, pp. 2364–2373.

[22] Y. Zhang *et al.*, "CNN-transformer based generative adversarial network for copy-move source/target distinguishment," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 5, pp. 2019–2032, 2023.

[23] H. Liu, Z. Tan, C. Tan, Y. Wei, Y. Zhao, and J. Wang, "Forgery-aware adaptive transformer for generalizable synthetic image detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2024.

[24] "FakeParaEgg dataset," https://github.com/YoursEver/FakeParaEgg, 2023.

[25] H.-C. Shao, Y.-R. Liao, T.-Y. Tseng, Y.-L. Chuo, and F.-Y. Lin, "Copy-move detection in optical microscopy: A segmentation network and a dataset," *arXiv preprint arXiv:2412.10258*, 2024.

[26] M. Cherti, R. Beaumont, R. Wightman, M. Wortsman, G. Ilharco, C. Gordon, C. Schuhmann, L. Schmidt, and J. Jitsev, "Reproducible scaling laws for contrastive language-image learning," *arXiv preprint arXiv:2212.07143*, 2022.

[27] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020.

[28] P. Zhou, X. Xie, Z. Lin, and S. Yan, "Towards understanding convergence and generalization of AdamW," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 9, pp. 6486–6493, Sep. 2024.

[29] N. Anantrasirichai, T. H. Chalidabhongse, D. Palasuwan, K. Naruenatthanaset, T. Kobchaisawat, N. Nunthanasup, K. Boonpeng, X. Ma, and A. Achim. *ICIP 2022 Challenge on Parasitic Egg Detection and Classification in Microscopic Images: Dataset, Methods and Results*. In 2022 IEEE International Conference on Image Processing (ICIP), IEEE, 2022. DOI: 10.1109/ICIP46576.2022.9897267. Dataset available on IEEE DataPort: https://ieee-dataport.org/competitions/parasitic-egg-detection-and-classification-microscopic-images.

[30] I. Shallal, L. Rzouga Haddada, and N. Essoukri Ben Amara, "Image forgery detection with focus on copy-move: An overview, real world challenges and future directions," *Appl. Sci.*, vol. 15, no. 21, p. 11774, 2025.