Exposing DeepFakes via Hyperspectral Domain Mapping (Student Abstract)

Aditya Mehta, Swarnim Chaudhary, Pratik Narang, Jagat Sesh Challa

Department of Computer Science and Information Systems,
Birla Institute of Technology and Science, Pilani, Pilani Campus, Vidya Vihar, Pilani, Rajasthan 333031, India
{p20230303, f20231040, pratik.narang, jagatsesh}@pilani.bits-pilani.ac.in

Abstract

Modern generative and diffusion models produce highly realistic images that can mislead human perception and even sophisticated automated detection systems. Most detection methods operate in RGB space and thus analyze only three spectral channels. We propose **HSI-Detect**, a two-stage pipeline that reconstructs a 31-channel hyperspectral image from a standard RGB input and performs detection in the hyperspectral domain. Expanding the input representation into denser spectral bands amplifies manipulation artifacts that are often weak or invisible in the RGB domain, particularly in specific frequency bands. We evaluate **HSI-Detect** across FaceForensics++ dataset and show the consistent improvements over RGB-only baselines, illustrating the promise of spectral-domain mapping for Deepfake detection.

Introduction

Rapid progress in generative adversarial networks (GANs) and diffusion models has made it increasingly easy to synthesize highly realistic human faces, voices, and videos. These so-called *deepfakes* are no longer confined to entertainment and creative media; they pose risks of misinformation, impersonation, and even legal or political manipulation. Consequently, robust and generalizable detection methods have become an urgent research need.

A key limitation of current deepfake detectors is their reliance on RGB images, which capture only three broad spectral channels. While suitable for visualization, RGB compresses much of the fine spectral information in natural images, causing subtle generative artifacts to be averaged out. These artifacts often lie in narrow spectral bands or specific frequency ranges, making RGB-based detectors vulnerable to missed detections and poor cross-dataset generalization.

Research has shown that spectral expansion can reveal inconsistencies invisible in RGB space, but most frequency-aware methods still operate on RGB inputs and remain limited by their three-channel representation. In contrast, hyperspectral imaging (HSI) captures tens or hundreds of narrow bands, enabling finer analysis of subtle variations. HSI has proven to be valuable in domains such as remote sensing (Peyghambari and Zhang 2021) and environmental monitoring (Wright, Levermore, and Kelly 2019), where fine spectral cues are critical. Motivated by these findings, we propose that deepfake detection can similarly benefit from hyperspectral representations, as illustrated in Figure 1.

In this work, we introduce **HSI-Detect**, a hyperspectral-guided deepfake detection framework. Instead of training detectors on RGB images alone, we reconstruct a 31-channel hyperspectral image from RGB inputs. These expanded spectral bands capture latent artifacts that generative models unintentionally introduce, particularly in low- and high-frequency regions (Dong et al. 2022). A dedicated classification network then analyzes the hyperspectral representation to detect whether the input is real or fake. By moving beyond the three-channel bottleneck, **HSI-Detect** seeks to improve robustness and generalization in the battle against deepfakes by providing a richer input to the detector.

Proposed Method

To propose HSI-Detect, we hypothesize that multi-band information in the reconstructed hyperspectral images can improve the separability between real and fake content by exposing spectral artifacts that remain hidden in RGB space. In particular, while RGB compresses the visible spectrum into three broad channels, hyperspectral reconstruction expands this into 31 narrow bands, allowing the detector to analyze subtle frequency irregularities and inter-band inconsistencies that generative models inadvertently introduce. Our method therefore consists of two main stages:

1. **Hyperspectral Reconstruction (HSR).** To reconstruct hyperspectral images from RGB inputs, we use the MST++ model (Cai et al. 2022), a transformer-based framework tailored for spectral reconstruction. MST++ employs *spectral-wise self-attention* to capture inter-band correlations often missed by CNNs focused on spatial features. Its multi-stage U-shaped encoder–decoder progressively refines outputs, enabling high-fidelity recovery of subtle spectral signatures. By emphasizing spectral self-similarity alongside local details, MST++ provides 31-channel hyperspectral estimates that form the input to our detection module.

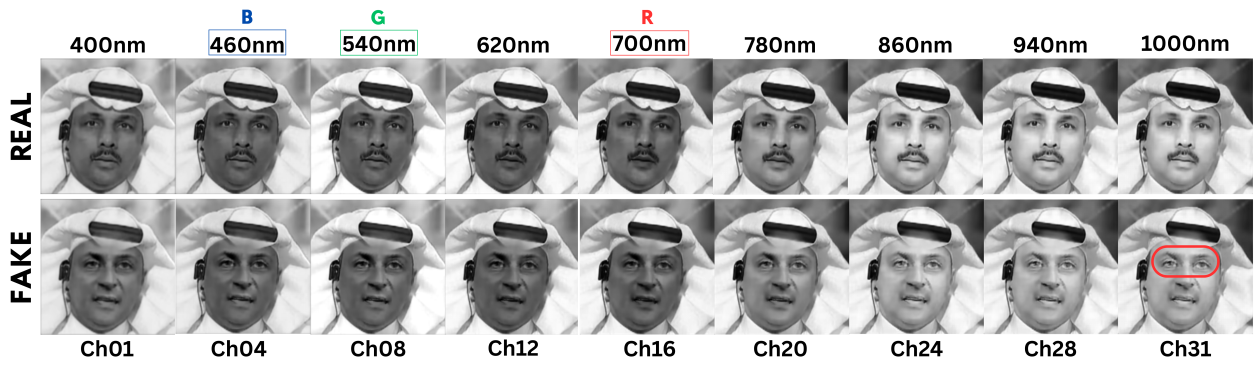

Figure 1: Motivation: Hyperspectral imaging offers richer spectral information than RGB, revealing deepfake artifacts—such as distortions near the eyes—that remain hidden in RGB but become prominent at lower frequencies.

2. **Spectral Detection Network.** We employ an enhanced version of UCF (Yan et al. 2023a) to mitigate overfitting to specific forgery cues. The architecture uses a Disentanglement Framework that exploits spectral artefacts across 31 hyperspectral channels. It consists of an encoder, a decoder, and two classification heads. The encoder has a content encoder and a fingerprint encoder to extract content and forgery features, while the decoder applies Adaptive Instance Normalization (AdaIN) (Huang and Belongie 2017) to reconstruct images from these features:

$$\mathrm{AdaIN}(x, y) = \sigma(y)\left(\frac{x - \mu(x)}{\sigma(x)}\right) + \mu(y), \quad (1)$$

where $\mathbf{x}$ is the content vector, $\mathbf{y}$ the style vector, and μ, σ denote mean and standard deviation.

To support detection, we introduce three loss functions: (1) Multi-task Classification loss for learning forgery-specific and shared features, (2) Contrastive Regularization loss to enhance discrimination between real and fake, and (3) Reconstruction loss to ensure consistency between original and reconstructed images.

Experimental Evaluation

Our model has been trained on Neural Textures, one of the four manipulation techniques in FaceForensics++ (Rossler et al. 2019) dataset. All preprocessing and training codebases utilized in this study follow DeepfakeBench (Yan et al. 2023b) to ensure alignment with standardized settings. We utilize the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) as our evaluation metric. AUC quantifies the area beneath the ROC curve. A higher AUC score indicates better detection performance. For comparison with State-of-the-Art methods, we incorporate three baseline detectors: ViT (Dosovitskiy et al. 2020), RECCE (Cao et al. 2022) and MoE-FFD (Kong et al. 2025).

As shown in Table 1, HSI-Detect consistently outperforms prior methods across multiple unseen manipulation types, achieving the best overall average AUC. The gains are especially clear on DeepFakes and FaceSwap, where hyperspectral cues provide strong discriminative power. These results highlight that our approach not only improves over RGB-only detectors but also advances beyond recent high-quality conference benchmarks, establishing HSI-Detect as a promising step forward for more robust and generalizable deepfake detection.

Method	DF	FF	FS	AVG
ViT (ICLR'21)	78.46	68.31	45.07	63.95
RECCE (CVPR'22)	72.37	64.69	51.61	62.89
MoE-FFD (TDSC'25)	80.02	**73.02**	51.94	68.33
HSI-Detect (Ours)	**85.31**	67.31	**54.15**	**68.92**

Table 1: HSI-Detect achieves the highest average AUC for cross-manipulation detection, demonstrating the advantage of hyperspectral features over RGB-only baselines.

Conclusion and Future Work

We introduce HSI-Detect, a framework that leverages reconstructed hyperspectral images for deepfake detection. Our experiments show that extending beyond RGB enables multi-band spectral cues to reveal artifacts often hidden in standard visual space, confirming the promise of hyperspectral analysis for forensics.

Future progress hinges on two directions: improving hyperspectral reconstruction, ideally with face-focused training to capture subtle details, and designing detection architectures tailored to hyperspectral inputs rather than adapted from RGB models.

Overall, HSI-Detect establishes a strong proof-of-concept and highlights the potential of hyperspectral imaging to deliver more robust and generalizable deepfake detection systems.

References

Cai, Y.; Lin, J.; Lin, Z.; Wang, H.; Zhang, Y.; Pfister, H.; Timofte, R.; and Van Gool, L. 2022. Mst++: Multi-stage spectral-wise transformer for efficient spectral reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 745–755.

Cao, J.; Ma, C.; Yao, T.; Chen, S.; Ding, S.; and Yang, X. 2022. End-to-end reconstruction-classification learning for face forgery detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 4113–4122.

Dong, C.; Wang, R.; Zhang, L.; Shi, B.; and Yang, J. 2022. Think Twice Before Detecting GAN-Generated Fake Images from their Spectral Artifacts. In *CVPR*.

Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Huang, X.; and Belongie, S. 2017. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE international conference on computer vision*, 1501–1510.

Kong, C.; Luo, A.; Bao, P.; Yu, Y.; Li, H.; Zheng, Z.; Wang, S.; and Kot, A. C. 2025. Moe-ffd: Mixture of experts for generalized and parameter-efficient face forgery detection. *IEEE Transactions on Dependable and Secure Computing*.

Peyghambari, S.; and Zhang, Y. 2021. Hyperspectral remote sensing in lithological mapping, mineral exploration, and environmental geology: an updated review. *Journal of Applied Remote Sensing*, 15(3): 031501–031501.

Rossler, A.; Cozzolino, D.; Verdoliva, L.; Riess, C.; Thies, J.; and Nießner, M. 2019. Faceforensics++: Learning to detect manipulated facial images. In *Proceedings of the IEEE/CVF international conference on computer vision*, 1–11.

Wright, S. L.; Levermore, J. M.; and Kelly, F. J. 2019. Raman spectral imaging for the detection of inhalable microplastics in ambient particulate matter samples. *Environmental science & technology*, 53(15): 8947–8956.

Yan, Z.; Zhang, Y.; Fan, Y.; and Wu, B. 2023a. Ucf: Uncovering common features for generalizable deepfake detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, 22412–22423.

Yan, Z.; Zhang, Y.; Yuan, X.; Lyu, S.; and Wu, B. 2023b. Deepfakebench: A comprehensive benchmark of deepfake detection. *arXiv preprint arXiv:2307.01426*.