IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, December 2025
Impact Factor: 7.67

Spectra-Temporal Attention Networks (STAN): A Dual-Stream Approach for Robust Deepfake Detection in Face Recognition Systems

Asst. Prof. Manisha Bharatram Bannagare[1], Mr. Vishal Ashok Ghuge[2], Mr. Rohit Vijay Shukla[3], Mr. Kaushik Rajvilas Moon[4], Mr. Om Avinash Jadhav[5], Mr. Sankalp Manohar Ganvir[6], Mr. Pratik Dnyaneshwar Shevane[7], Ms. Sakshi Shankar Kewat[8], Onkar Rajendra Dhanewar[9]

Guide, Department of Computer Science & Engineering[1]

Students, Final Year Department of Information Technology[2-9]

manisha180392@gmail.com

R.V. Parankar College of Engineering and Technology, Arvi, Maharashtra, India

ghugevishal25@gmail.com and rohitvijayshukla265@gmail.com

Abstract: *The rapid proliferation of deep learning-based synthetic media, commonly known as "deepfakes," poses a critical threat to the integrity of biometric security systems, particularly face recognition protocols. While early generation deepfakes were easily detectable by the human eye, modern auto-encoder and diffusion-based models can generate hyper-realistic artifacts that challenge even sophisticated detection algorithms. Traditional Convolutional Neural Networks (CNNs) often fail to generalize against these threats because they over-rely on spatial pixel patterns, which are easily masked by video compression algorithms used on social media platforms. To address this limitation, this paper introduces the* ***Dual-Stream Spectral-Temporal Attention Network (DS-STAN)****. This novel architecture moves beyond simple pixel analysis by exploiting two fundamental weaknesses in synthetic media: the frequency-level "fingerprints" left by upsampling operations and the subtle physiological inconsistencies inherent in generated video over time. By fusing a Frequency-based stream with a Video Vision Transformer (ViT) stream, DS-STAN achieves state-of-the-art performance. Experimental results on benchmark datasets demonstrate that our model not only detects known attack types with high accuracy but also generalizes significantly better to unseen deepfake methods compared to single-modality detectors.*

Keywords: *Deepfake Detection, Biometric Security, Vision Transformers, Frequency Analysis, Face Anti-Spoofing, Generative Adversarial Networks*

DOI: 10.48175/IJARSCT-30572