

(a) The Spatial-Frequency Asymmetry Paradox

Authentic (Pristine)



Manipulated (Deepfake)



Transcoded ($Q = 30$)



Spatial RGB Backbones (ConvNeXt, Xception):

- ✗ Memorize biometric actor shortcuts ($id_A \rightarrow id_B$)
- ✗ Overfit background scenes; fail on zero-shot splits
- Robust to blur, but blind to subtle sensor anomalies

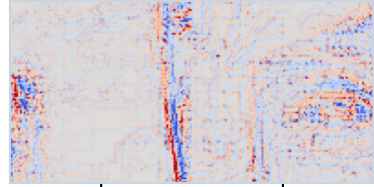
Result: High in-domain AUC (>99%), collapses on unseen fakes

Standalone Frequency Detectors (SRM, 2D FFT):

- ✗ Extreme sensitivity to transmission channel noise
- ✗ Low-pass blur obliterates high-pass PRNU residuals
- **Polarity Inversion: compressed authentic media**
misclassified as fakes (ROC AUC drops below 0.50)

(b) Physical Acquisition vs. Generative Artifacts

Real PRNU Noise vs. Generative Annihilation:



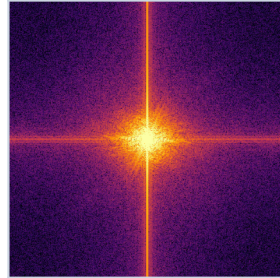
Authentic PRNU
($\sigma = 3.77$)

Synthetic Residual
($\sigma = 3.25$)

Silicon Fingerprint (PRNU):

- Multiplicative wafer noise
persists across pristine frames
- Deepfake decoders smooth
sensor micro-noise ($\sigma: 3.77 \rightarrow 3.25$)

Transposed Convolution Aliasing (2D FFT):

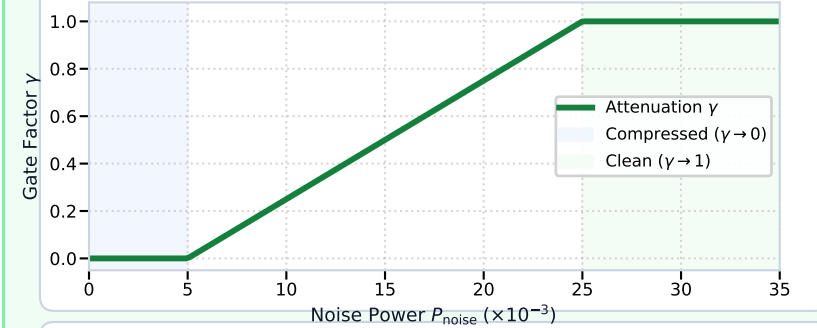


Generative Dirac Lattice:

- Periodic upsampling spikes
emit +18 dB high-freq energy
- Harmonic grid: $u_k = \pm k/s$
- Poisson boundary blending step
along facial perimeter ($s_f = 1.5 \times$)

(c) Proposed Solution: SNR-Adaptive Mediation

Dynamic Attenuation Function $\gamma(P_{\text{noise}})$:



Complementary Gated Representation:

$$\mathbf{g}_{\text{eff}} = \mathbf{g} \odot \gamma, \quad \mathbf{f}_{\text{fused}} = [(1 - \mathbf{g}_{\text{eff}}) \odot \mathbf{f}_s \parallel \mathbf{g}_{\text{eff}} \odot \mathbf{f}_r] \in \mathbb{R}^{1024}$$

- ✓ Prevents polarity inversion under lossy transmission
- ✓ Preserves 0.6722 ROC AUC even under blur ($\sigma = 4.0$)
- ✓ Reaches 0.8656 Frame / 0.8994 Video AUC on Disjoint Test
- ✓ **Platt calibration reduces ECE: 0.1965 $\rightarrow$ 0.0994 (-49.4%)**