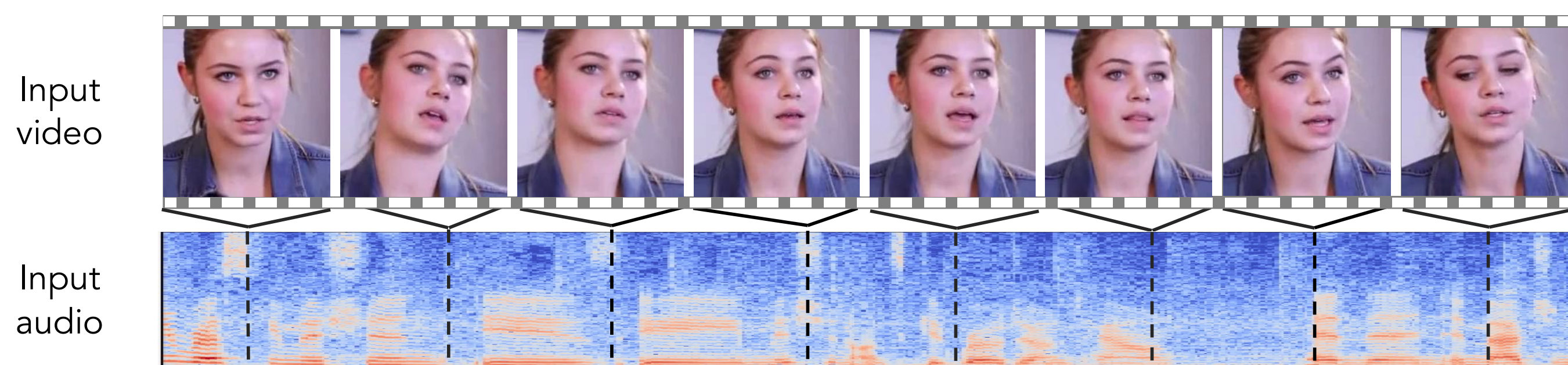


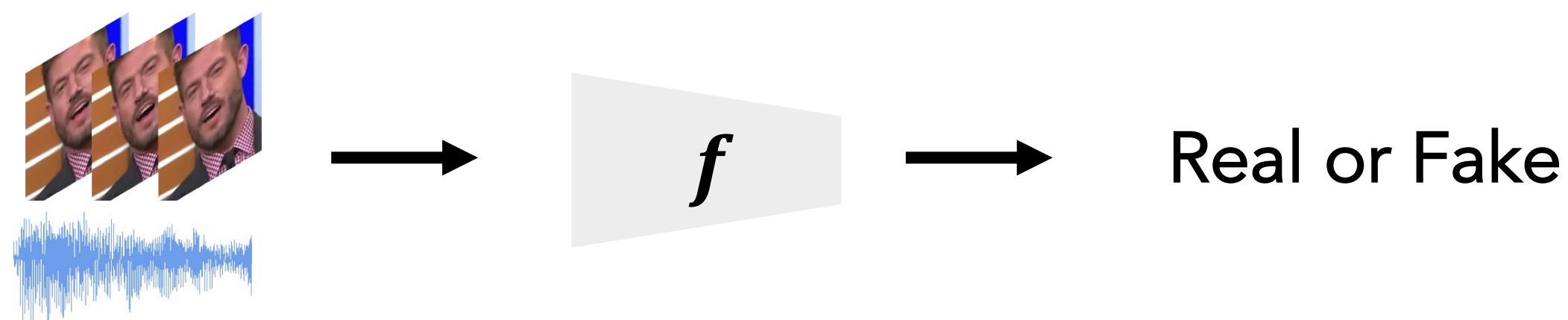
Motivation

Goal: Learning to detect manipulations through self-supervision without fake videos. Building likelihood function on real data and leveraging anomaly detection to flag manipulated videos.

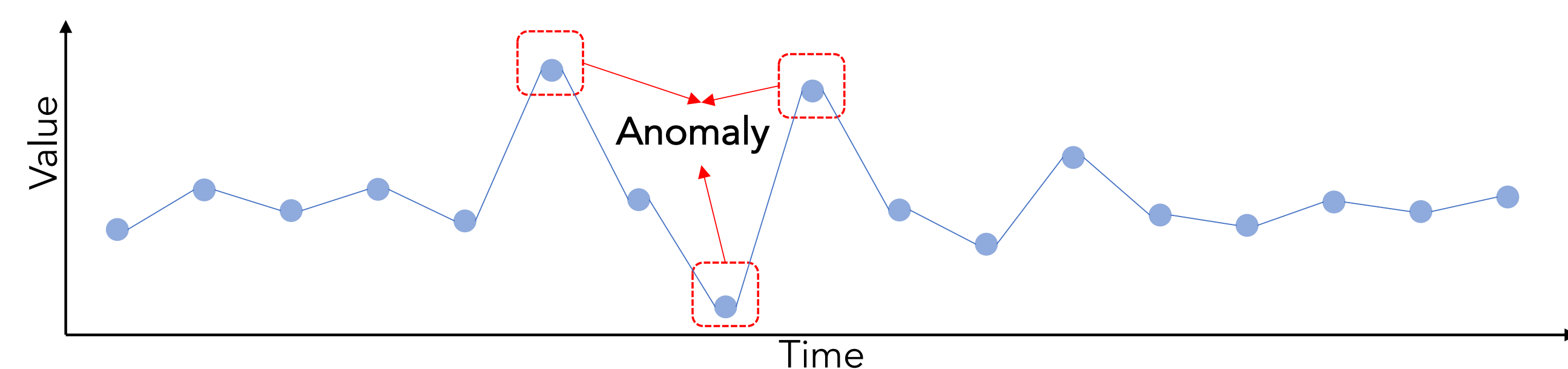


Background

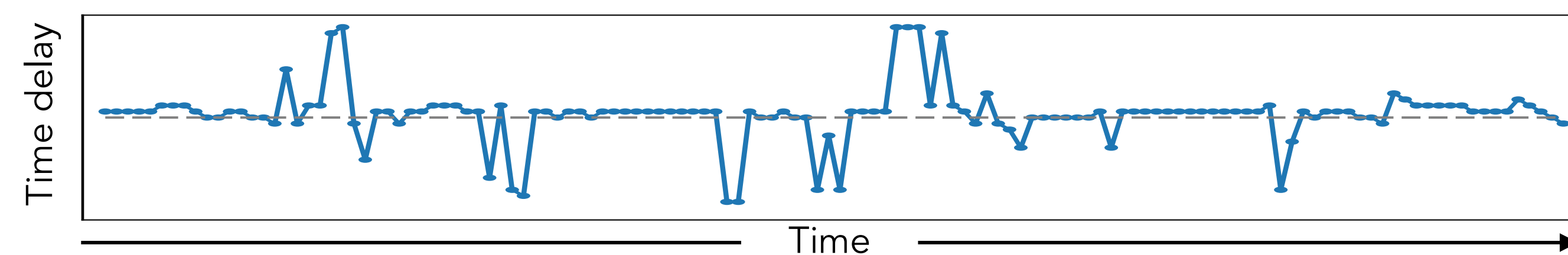
Video forensics: usually done in a supervised manner but it has generalization issue.



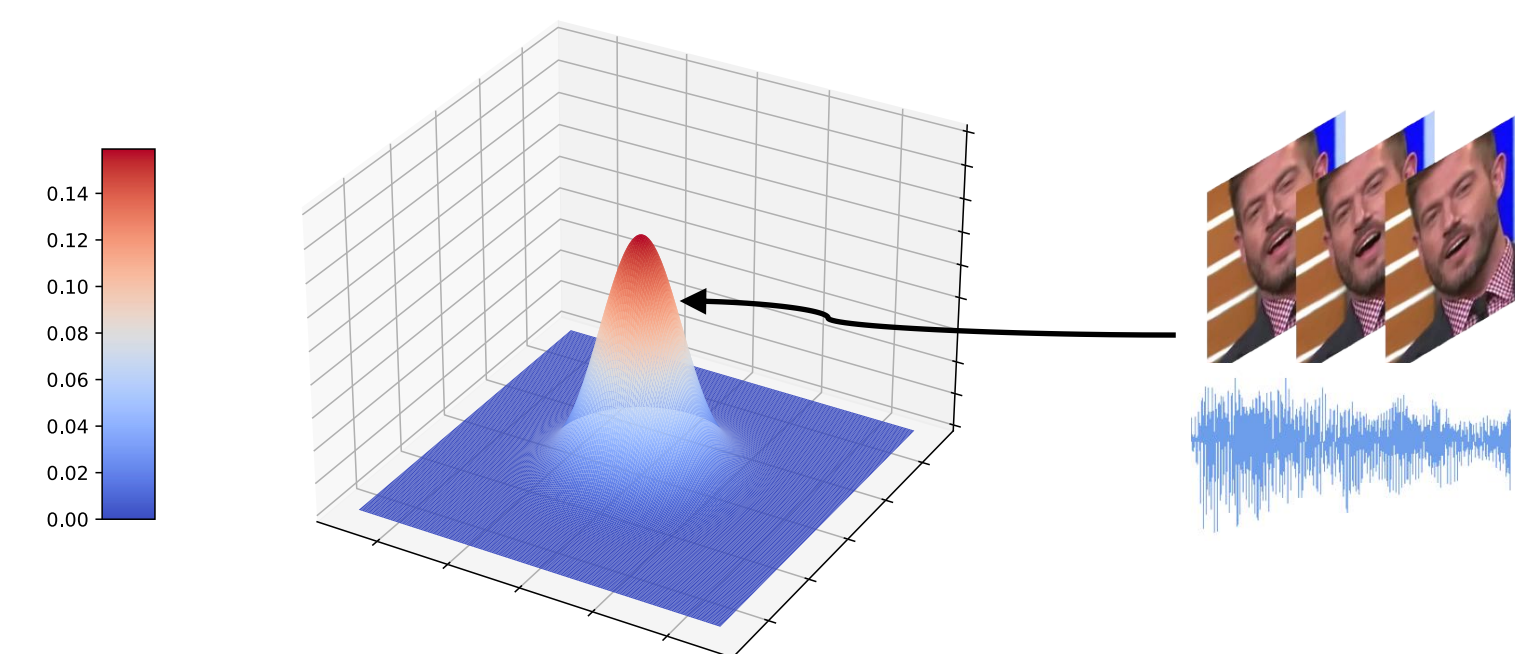
Anomaly detection: identification of rare items, events, or observations which deviate significantly from normal patterns.



Audio-visual misalignment: alignment degree between lip motion and sound (time delay) can be used as a cue for anomaly detection.



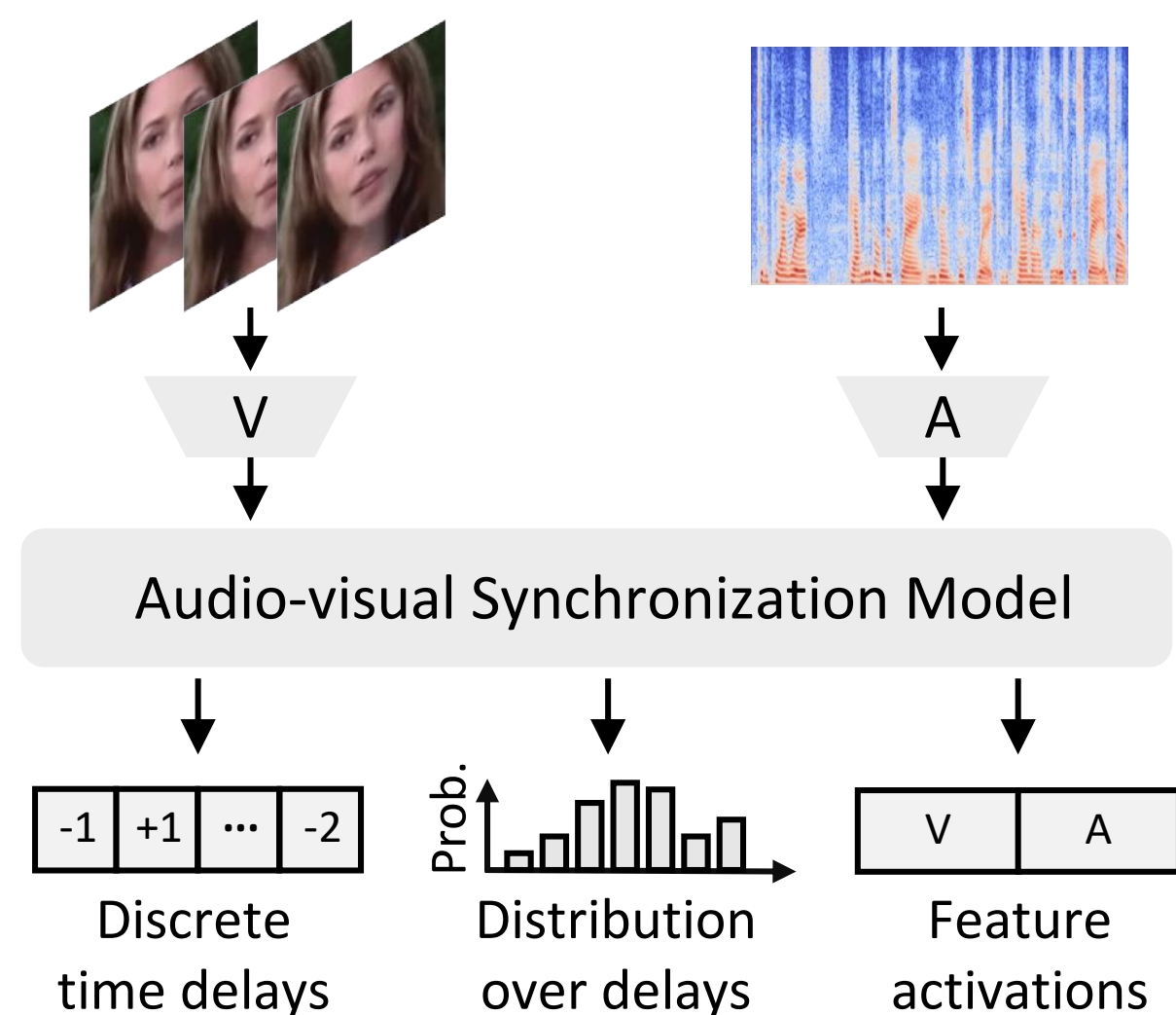
Audio-Visual Anomaly Detection



Idea: we use audio-visual anomaly detection to treat low-probability examples as fake.

Method

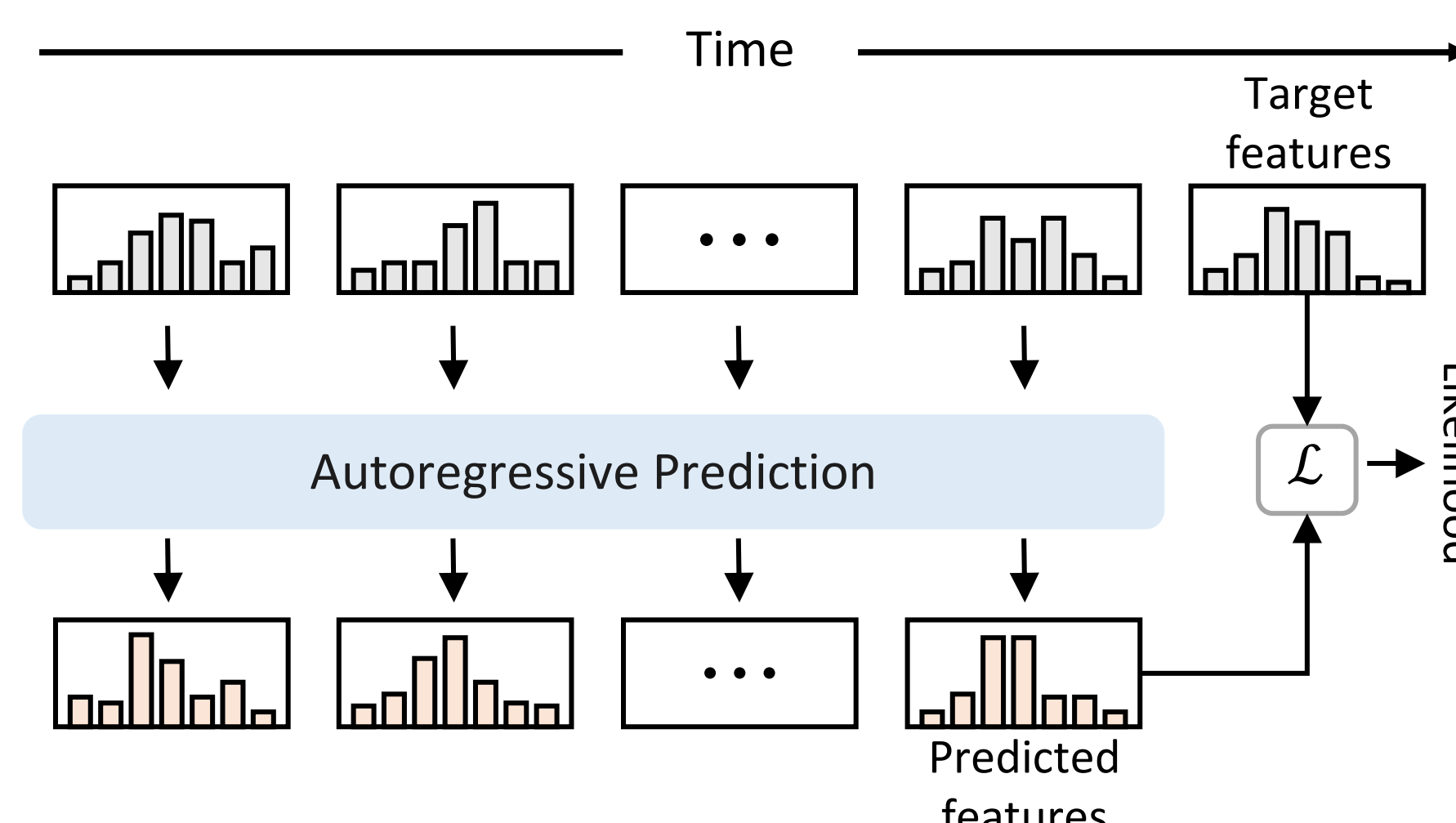
Self-supervised feature sets



Stage 1 Learning audio-visual synchronization feature sets:

$$S(i, j) = \frac{\exp(\phi(V_i, A_j))}{\sum_{k=i-\tau}^{i+\tau} \exp(\phi(V_i, A_k))}$$

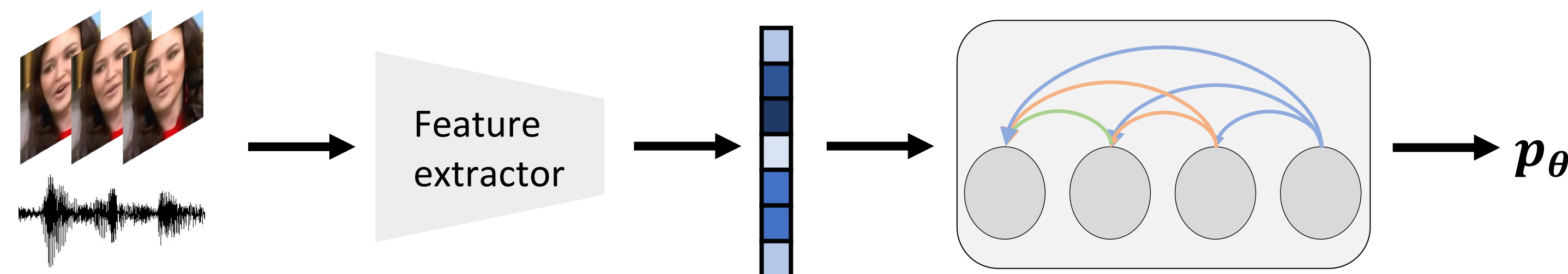
Generative anomaly detector



Stage 2 Learning autoregressive model on self-supervised audio-visual feature sets:

$$p_{\theta}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) = \prod_{i=0}^{N-1} p_{\theta}(\mathbf{x}_{i+1} | \mathbf{x}_1, \dots, \mathbf{x}_i)$$

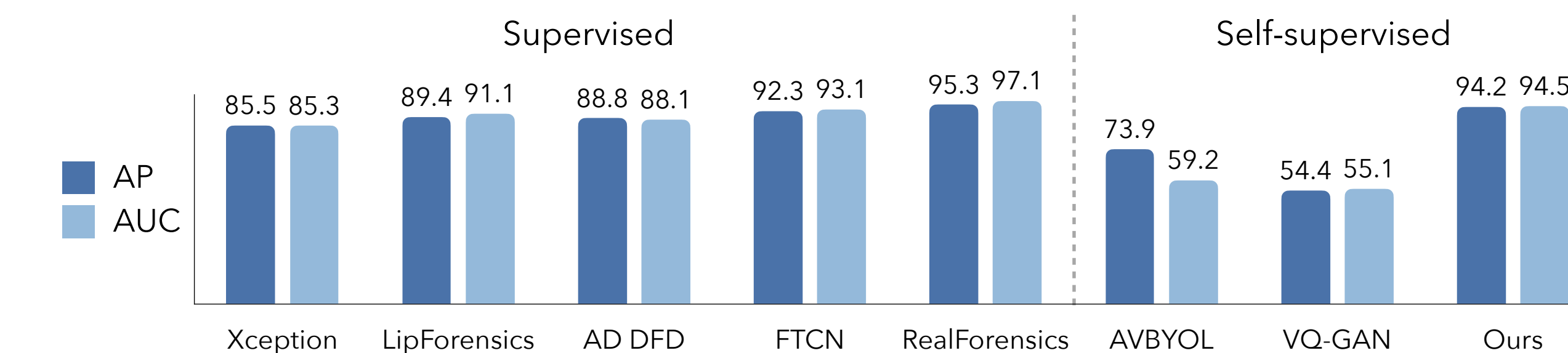
Test-time: we flag fake videos with low log-probability.



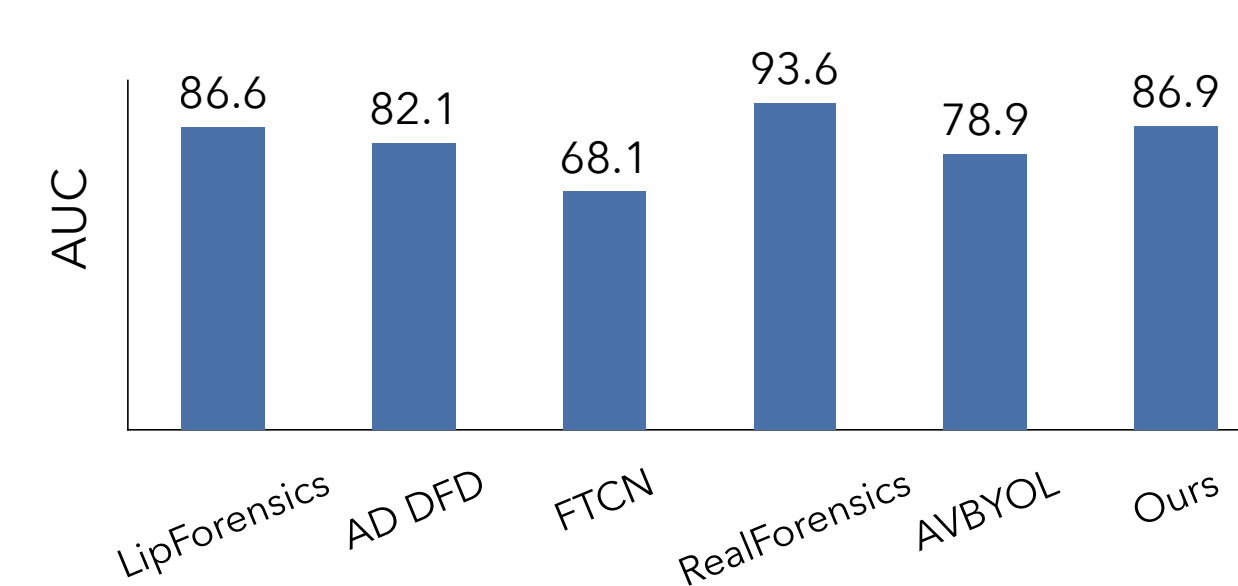
Experiments

Training datasets: Lip Reading Sentences (LRS) 2/3.

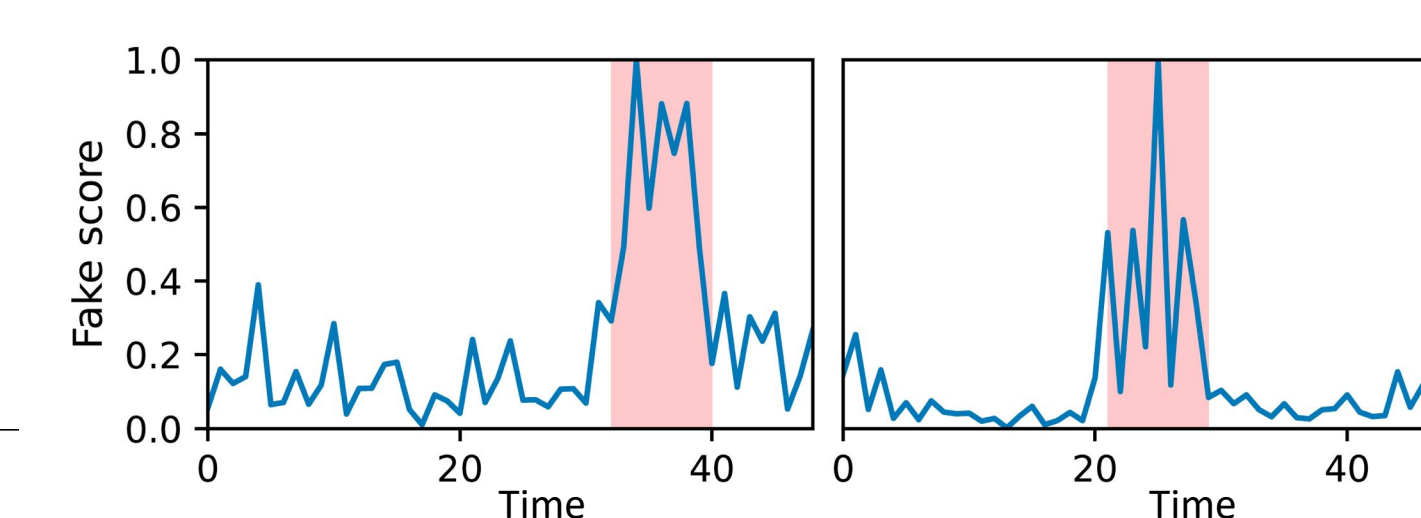
Cross-manipulation generalization:



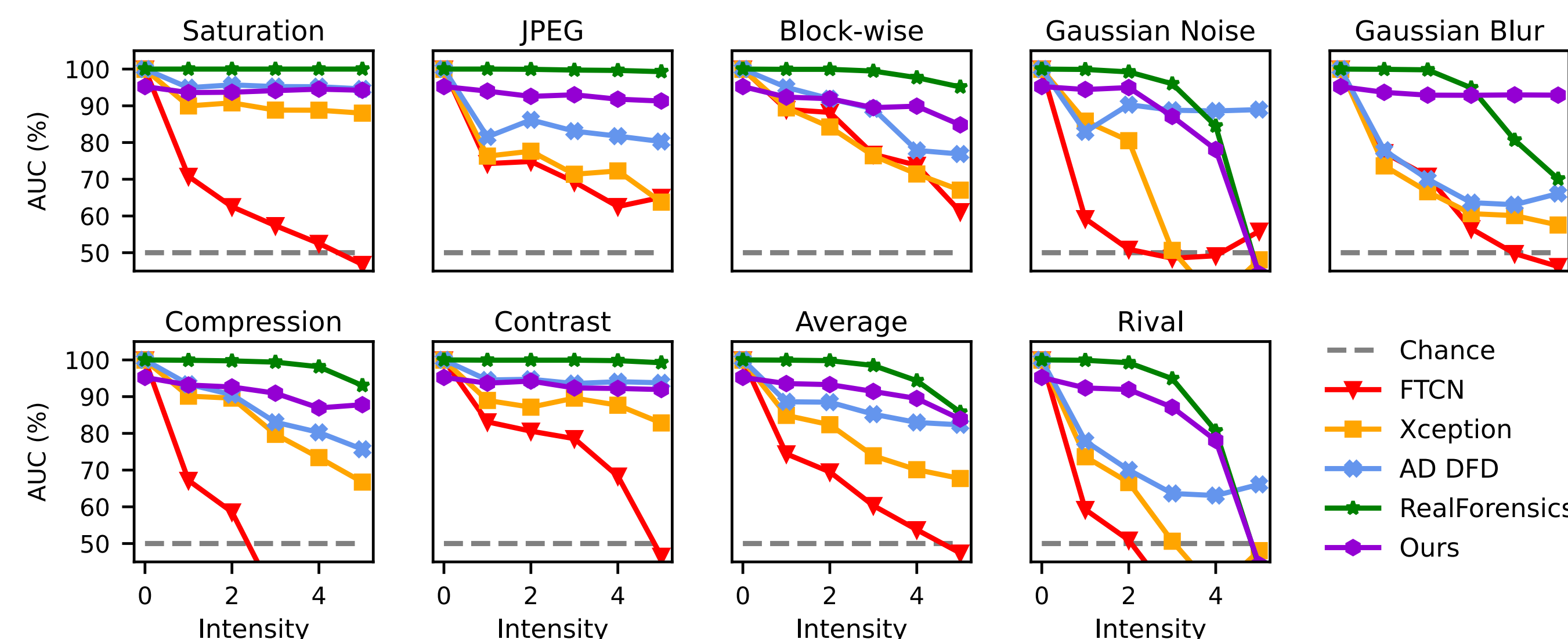
Cross-dataset generalization:



Temporal localization:



Robustness to unseen perturbation:



References

- A. Haliassos, R. Mira, S. Petridis, M. Pantic. Leveraging Real Talking Faces via Self-Supervision for Robust Forgery Detection, CVPR 2022.
- H. Khalid, S. Tariq, M. Kim, S. Woo. FakeAVCeleb: A Novel Audio-Video Multimodal Deepfake Dataset. NeurIPS Datasets and Benchmarks 2021.
- P. Kwon, J. You, G. Nam, S. Park, G. Chae. KoDF: A Large-scale Korean DeepFake Detection Dataset. ICCV 2021.
- Y. Zhou and S. Lim. Joint Audio-Visual Deepfake Detection. ICCV 2021.
- Y. Zheng, J. Bao, D. Chen, M. Zeng, F. Wen. Exploring Temporal Coherence for More General Video Face Forgery Detection. ICCV 2021.
- L. Jiang, R. Li, W. Wu, C. Qian, C. Loy. DeeperForensics-1.0: A Large-Scale Dataset for Real-World Face Forgery Detection. CVPR 2020