Constructing Deepfake Detectors with Different Modalities Based on Feature Encoders

IJCAI 2025 Workshop
On Deepfake Detection, Localization, and Interpretability

Speaker: Jiaming Chu

Authors: Jian Zhao[1], Jiaming Chu[1,3], Xin Zhang[1], Yuchu Jiang[1], Yuer Li[1], Xinru Wang[1],Mingxing Yuan[1], Xu Yang[2], Lei Jin[3†] , Chi Zhang[1], Xuelong Li[1†]

1 Institute of Artificial Intelligence (TeleAI), China Telecom
2 Southeast University, China;
3 Beijing University of Posts and Telecommunications

Backgrounds

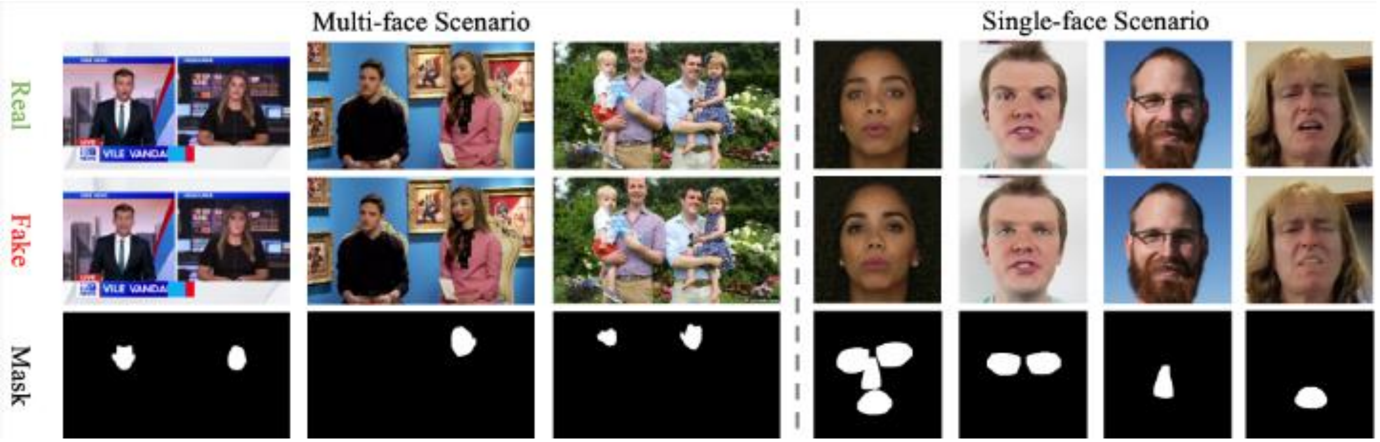
DeepFake

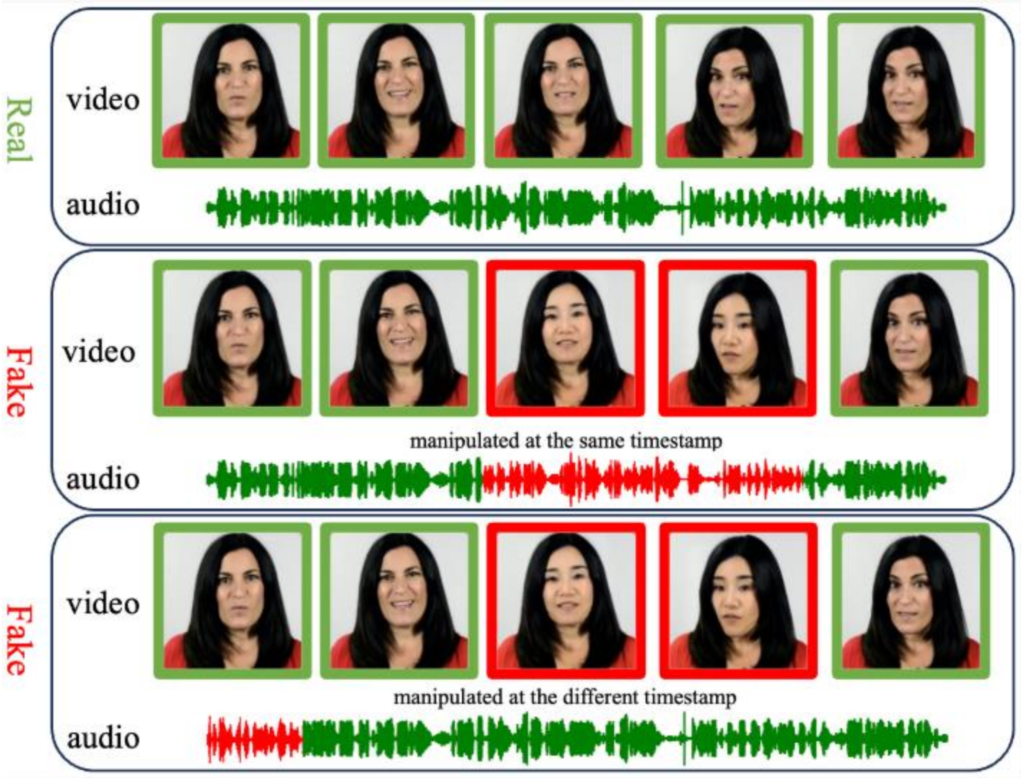
Challenging

Reality

Backgrounds

- Different Modality
- Concealment From partial modifications
- Diverse Generative Method





Method

Good Models are all Effective but Simple.

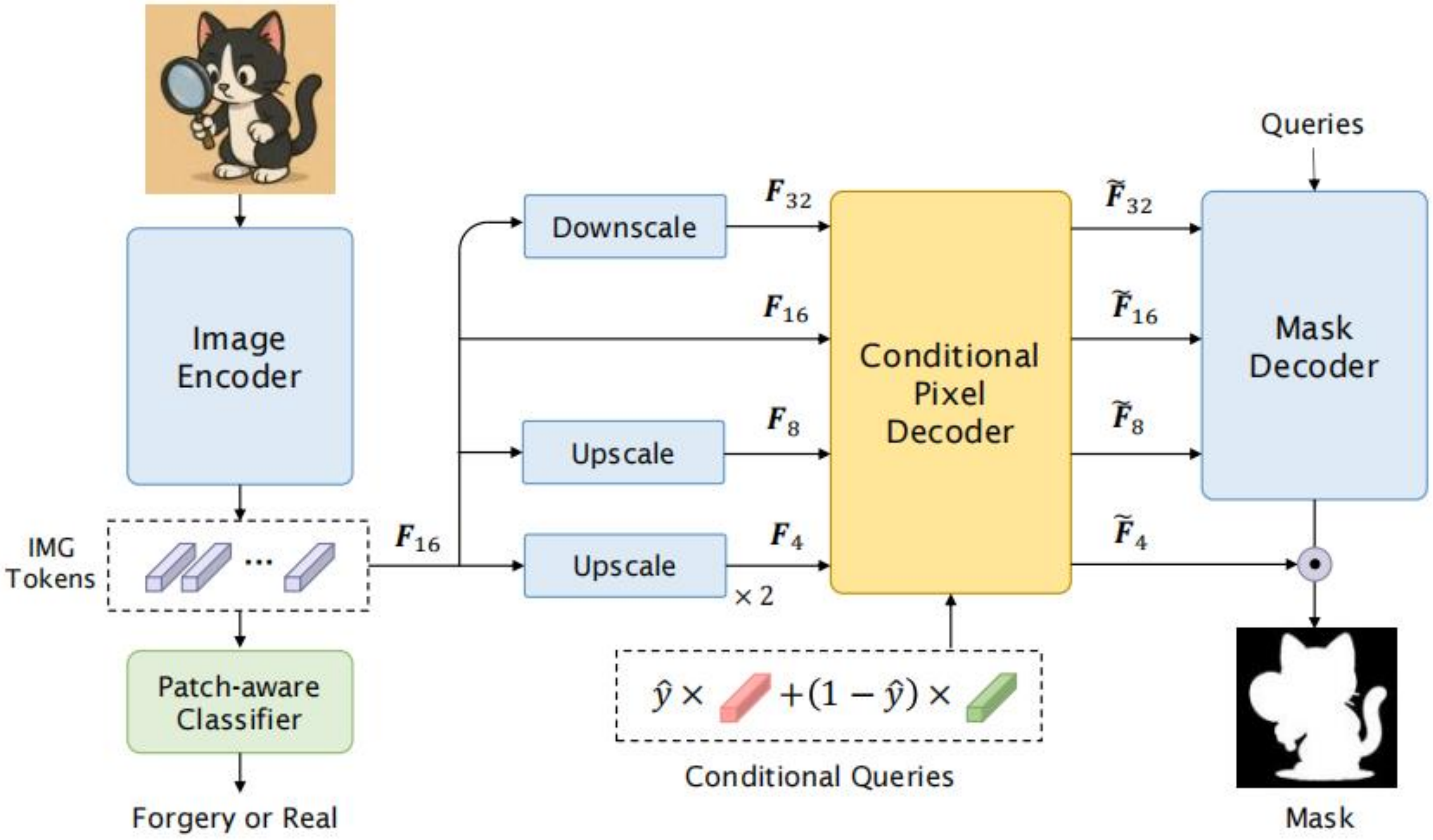
- Good Feature Extractor
- A Complete Locator and Classifier
- Good Loss Function Design

↓

- Good Detector and Classifier

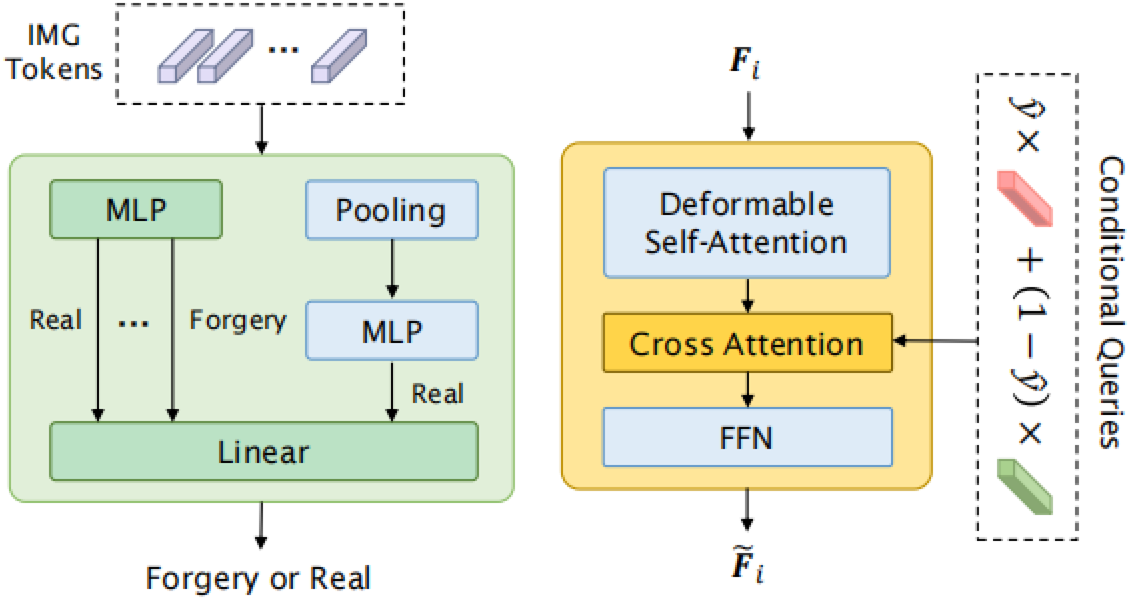
Method

Track1 - Loupe



Method

Track1 - Loupe



(a) Patch-aware classifier (b) Conditional pixel decoder layer

$$\mathcal{L}_{\text{patch}} = \frac{1}{N}\sum_{i=1}^{N}\left[-\alpha(1-p_i)^{\gamma}\log(p_i) + \epsilon(1-p_i)^{\gamma+1}\right]. \tag{1}$$

$$\mathcal{L}_{\text{cls}} = \mathcal{L}_{\text{patch}} + \mathcal{L}_{\text{global}}. \tag{2}$$

$$\mathcal{L}_{\text{tversky}} = 1 - \frac{\text{TP}}{\text{TP} + \alpha\cdot\text{FP} + \beta\cdot\text{FN}}, \tag{3}$$

$$\mathcal{L}_{\text{seg}} = \lambda_1\mathcal{L}_{\text{mask}} + \lambda_2\mathcal{L}_{\text{tversky}} + \lambda_3\mathcal{L}_{\text{box}}, \tag{4}$$

Experiment

Track1 - Loupe

Table 1: **Leaderboard of the IJCAI 2025 Deepfake Detection and Localization Challenge**. The *overall* score is computed as the average of AUC, F1, and IoU.

Rank	**AUC**	**F1**	**IoU**	**Overall**
1 (ours)	0.963	0.756	0.819	**0.846**
2	-	-	-	0.8161
3	-	-	-	0.8151
4	-	-	-	0.815
5	-	-	-	0.815

Table 2: Ablation on patch prediction.

	AUC
Loupe (ours)	**0.946**
– patch prediction	0.920

Table 3: Ablation on conditional queries of our modified pixel decoder and training objectives.

	F1	**IoU**
Loupe (ours)	**0.880**	**0.886**
– conditional queries	0.870	0.874

IJCAI 2025 MONTREAL

EVOL Lab@TeleAI
Evolutionary Vision+x Oriented Learning

Method

Track2 - ERF-BA-TFD+

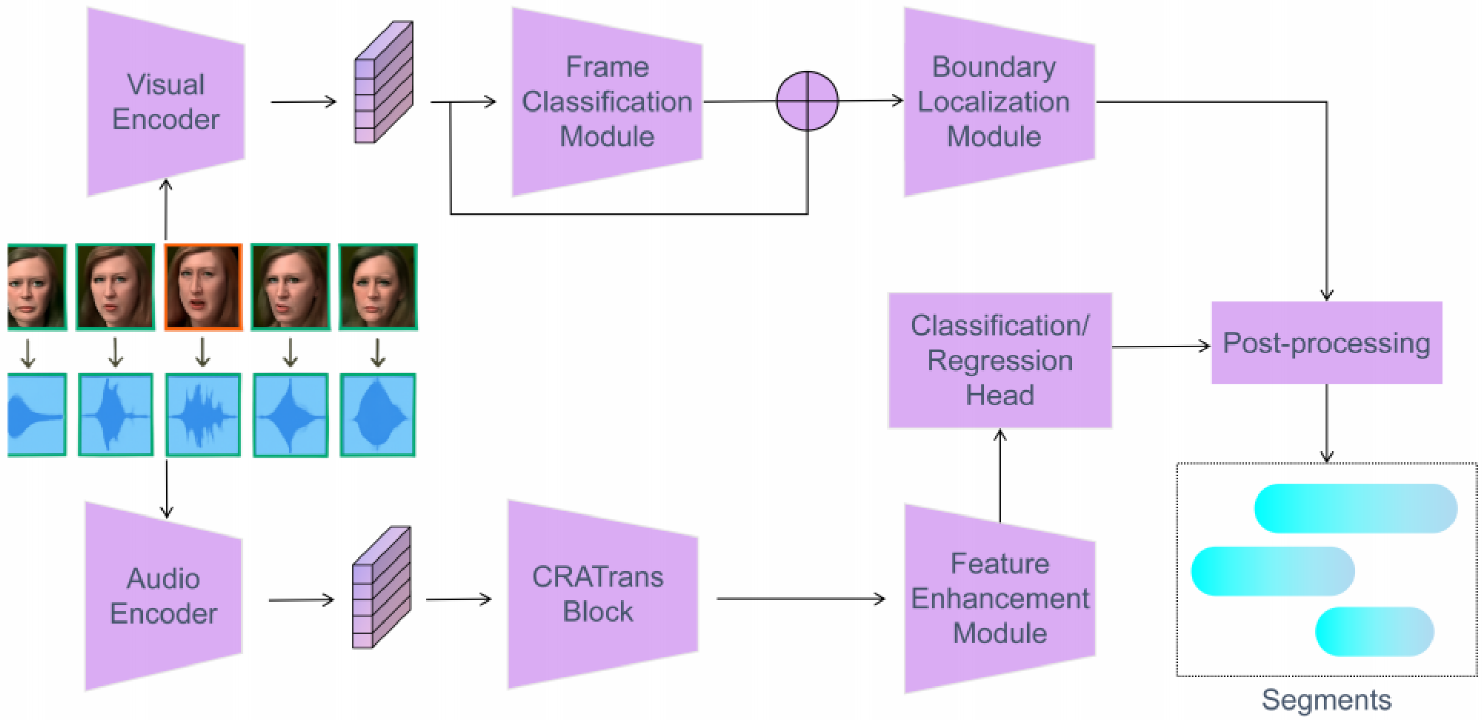


Figure 1: ERF-BA-TFD+ Model Architecture

Experiment Track2 - ERF-BA-TFD+

Table 3: Performance Metrics After UMMA Integration on DDL-AV Dataset (Fusion Modality)

Metric	Score
AP@0.5	0.9243
AP@0.75	0.8050
AP@0.95	0.0451
AR@90	0.8246
AR@50	0.8121
AR@20	0.8039
AR@10	0.7952

Table 1: Comparison of Performanc
Dataset vs Trained on DDL-AV Data

Metric	LAV-DF Sco	
AP@0.5	0.9630	
AP@0.75	0.8498	
AP@0.95	0.0446	
AR@100	0.8160	
AR@50	0.8048	
AR@20	0.7940	
AR@10	0.7876	0.4130

e Metrics for Fusion Modality (Base-
set

	Score
.5	0.0163
.75	0.0117
.95	0.0014
00	0.2290
0	0.1681
AR@20	0.1182

IJCAI 2025 MONTREAL

EVOL Lab@TeleAI

Evolutionary Vision+x Oriented Learning

Experiment Track2 - ERF-BA-TFD+

Table 3: Performance Metrics After UMMA Integration on DDL-AV Dataset (Fusion Modality)

Metric	Score
AP@0.5	0.9243
AP@0.75	0.8050
AP@0.95	0.0451
AR@90	0.8246
AR@50	0.8121
AR@20	0.8039
AR@10	0.7952

Table 4: Performance Comparison on Sampled Validation Set (Before and After ERF Integration)

Metric	Before	ERF Integration
AP@0.5	0.6472	0.8214
AP@0.75	0.5431	0.7287
AP@0.95	0.0704	0.0951
AR@100	0.6513	0.7886
AR@50	0.6342	0.7732
AR@20	0.6012	0.7464
AR@10	0.5836	0.7397

Ending

Thank you for your listening.

Code

Paper1

Paper2