IJCAI 2025 Workshop
on Deepfake Detection, Localization, and Interpretability

浙江大學
ZHEJIANG UNIVERSITY

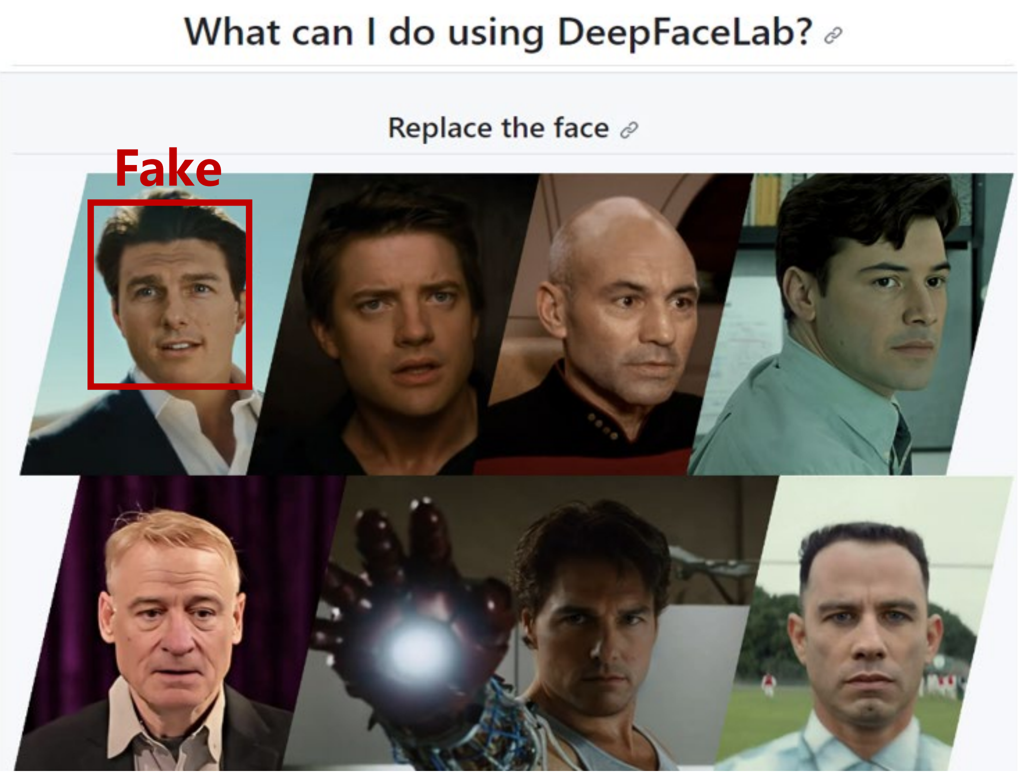
Morphology-optimized Multi-Scale Fusion: Combining Local Artifacts and Mesoscopic Semantics for Deepfake Detection and Localization

Chao Shuai , Gaojian Wang, Kun Pan, Tong Wu, Fanli Jin, HaohanTan, Mengxiang Li, Zhenguang Liu, Feng Lin and Kui Ren

Speaker: Qing Wen   2025.08.29

Background

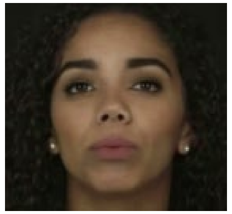
➤ Deepfake has given rise to concerns about the misuse of fake videos fabricating people's words and actions.



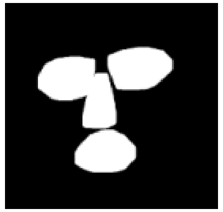
Challenge

Main Task:

- ✓ Detection
- ✓ Localization



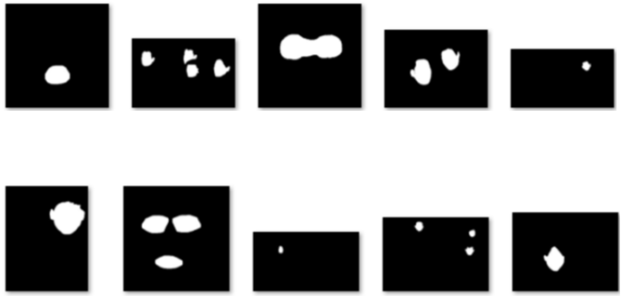
Real
or
Fake?



Where?

Main Challenges:

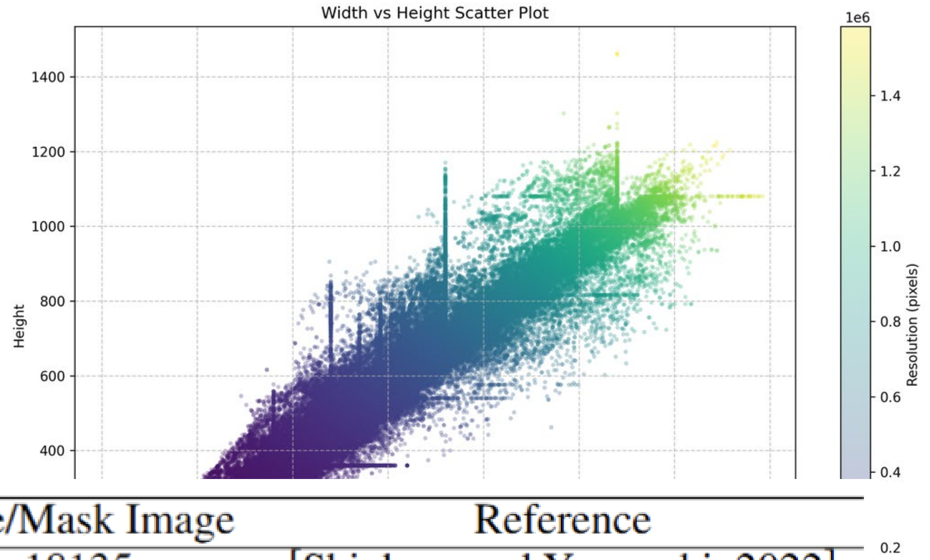
- ✓ Diverse generative model
- ✓ Various Image Degradation
- ✓ Multi-Scale Faces
- ✓ Precisely Localization



Method

➢ Image Analysis：

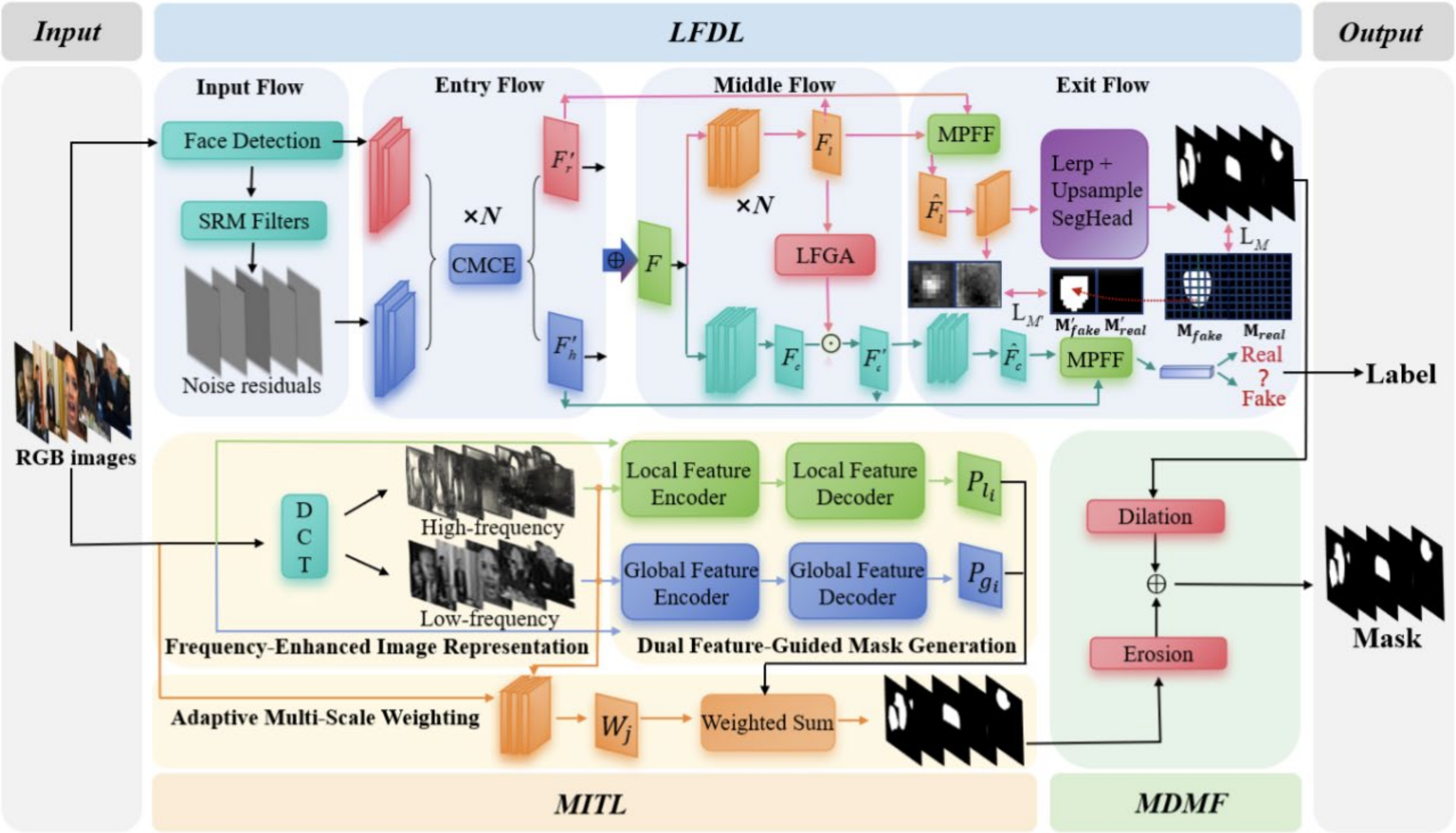
- ✓ Multi-faces ➡ End-to-end Detection
- ✓ Multi-scale ➡ Multi-scale features
- ✓ Strong degradation ➡ Data augmentation
- ✓ Multi-source ➡ Self-built Forgery Dataset



Model Type	Method	Forgery Types	Fake/Mask Image	Reference
Image Edit	SBIs	FaceSwap	18135	[Shiohara and Yamasaki, 2022]
	Random combination	FaceSwap	17728	-
GAN	Simswap	FaceSwap	14999	[Chen *et al.*, 2020]
	MaskFaceGAN	Face Attribute Editing	14999	[Pernuš *et al.*, 2023]
	Facedancer	FaceSwap	20000	[Rosberg *et al.*, 2023]
Diffusion Model	BELM	Diffusion Inversion	14674	[Wang *et al.*, 2024]
	SD-inpanting	Inpanting	18347	[Podell *et al.*, 2023]

Method

Overview of our proposed framework

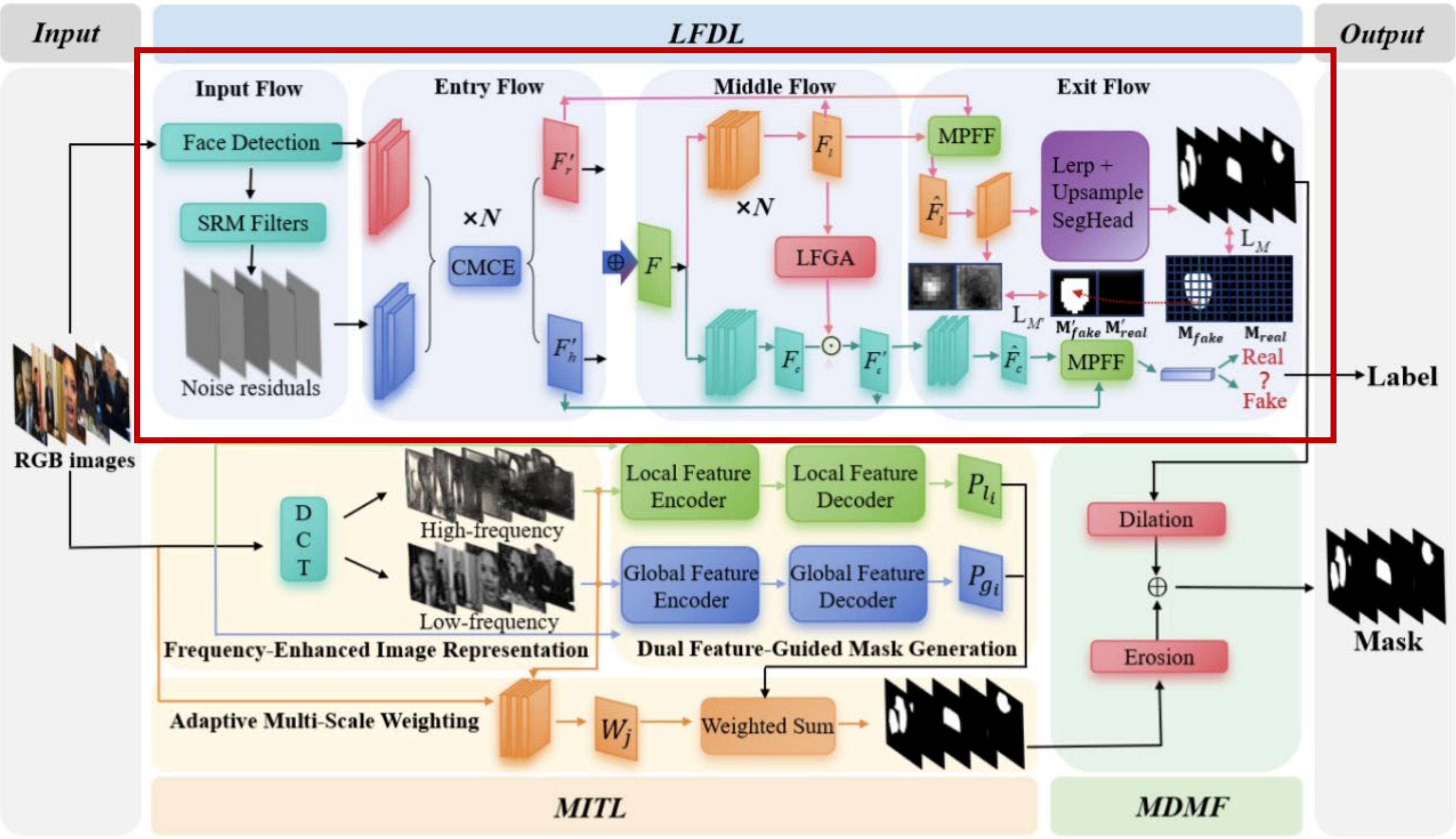


Three key components:

- LFDL: Local Facial Forgery Detection and Location
- MITL: Mesoscopic Image Tampering Localization
- MDMF: Morphology-Driven Mask Fusion for Comprehensive Forgery Localization

Method

✓ LFDL: Local Facial Forgery Detection and Location



Key points:

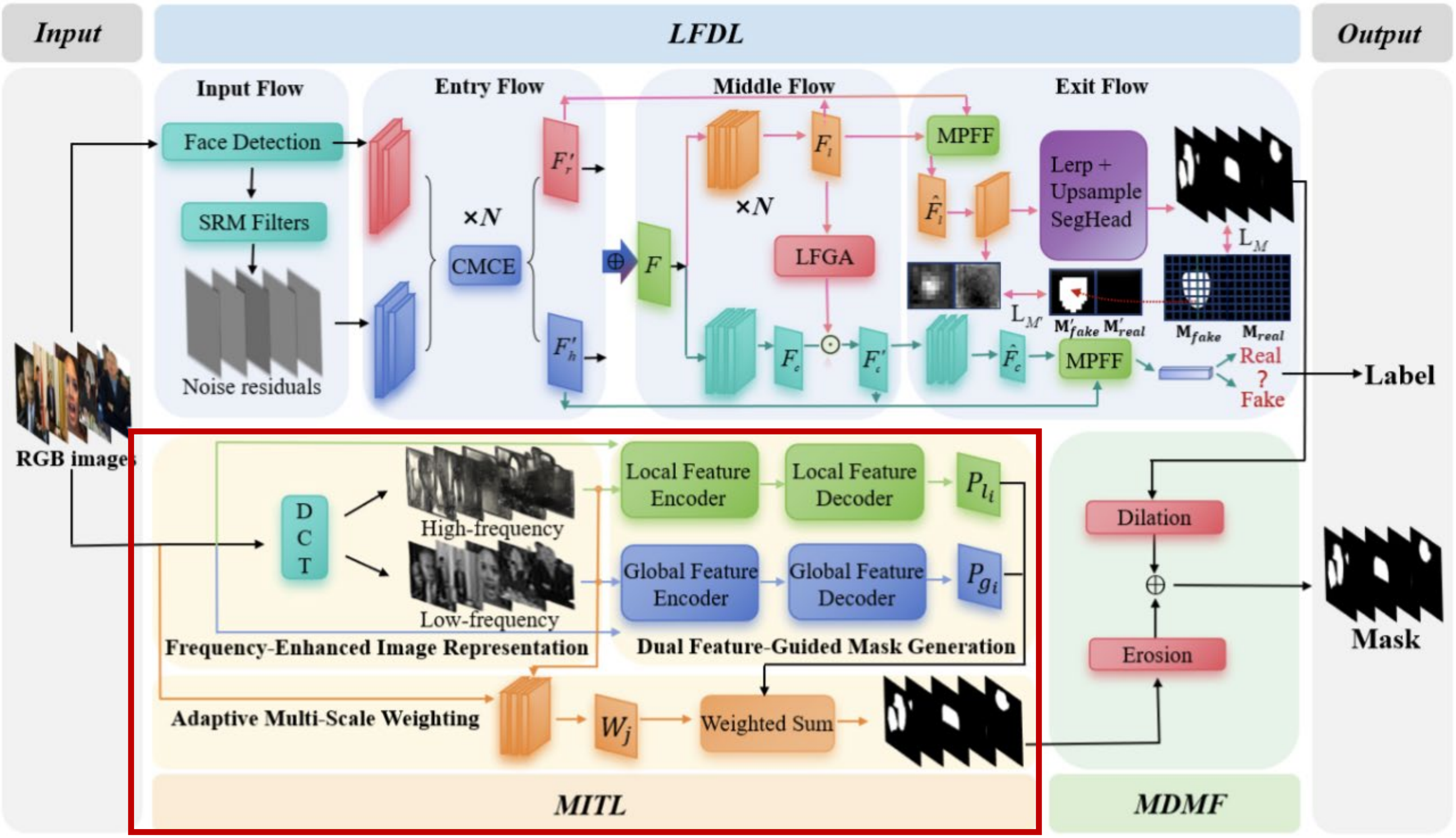
- Align the **resolution** of the input faces
- Fuse the **RGB-view** and **SRM-view** features
- Localization branch **enhances** the classification branch

Problem:

- Strongly degraded faces are not **recognized** and **extracted !!!**

Method

✓ MITL: Mesoscopic Image Tampering Localization



Key points:

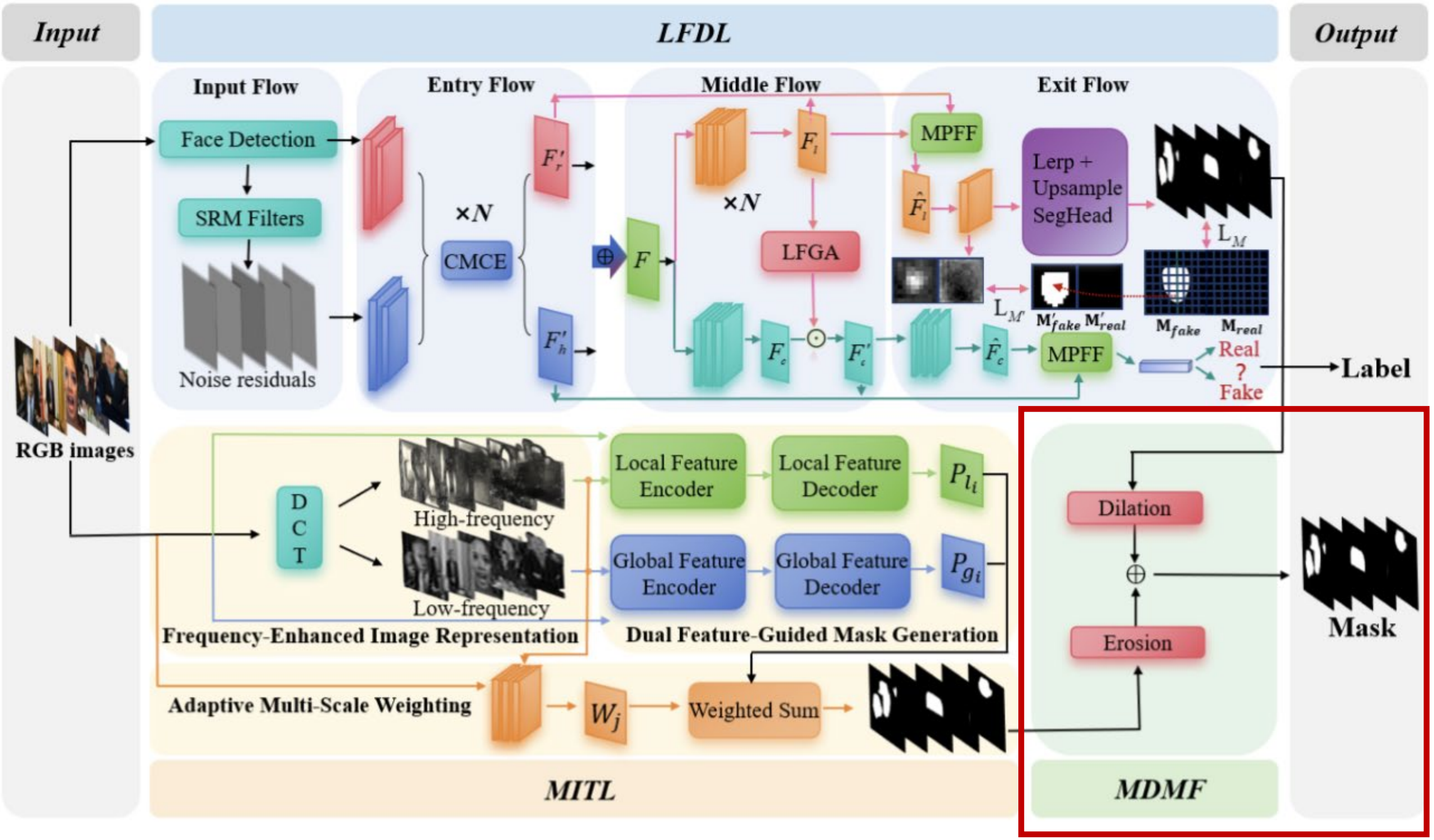
- **End-to-end training** is more suitable for multi-scale multi-face image
- Adaptive **multi-scale** and **dual feature-guided** mask generation

Problem:

- Error Detection and Missed Detection !!!

Method

✓ MDMF: Morphology-Driven Mask Fusion



Key points:

- Apply a **dilation operation** to the $\mathbf{M_{LFDL}}$ to smooth edges
- Apply an **erosion operation** to $\mathbf{M_{MITL}}$ to loss of complete details

$$M_{\mathrm{LFDL}} \oplus B = \{z \in \mathbb{Z}^2 \mid (B)_z \cap M_{\mathrm{LFDL}} \neq \emptyset\}$$

$$M_{\mathrm{MITL}} \ominus B = \{z \in \mathbb{Z}^2 \mid (B)_z \subseteq M_{\mathrm{LFDL}}\}$$

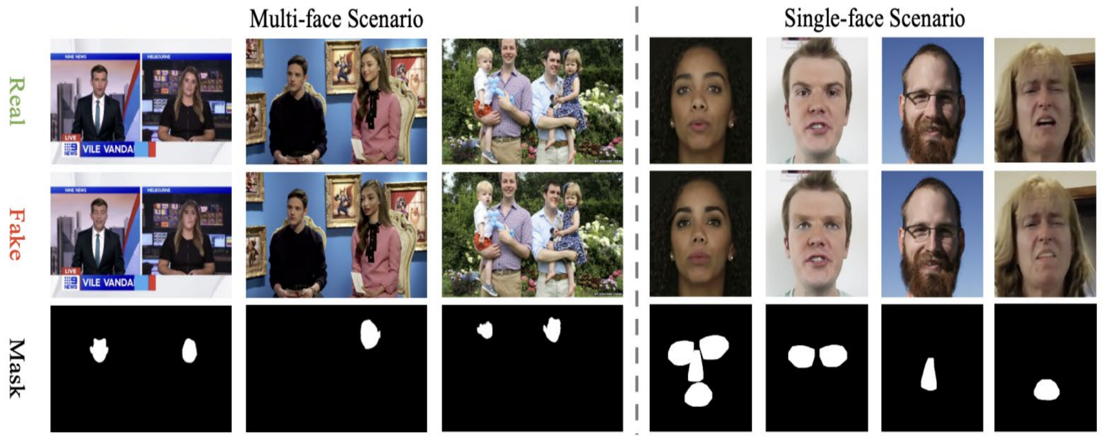
$$M_{\mathrm{final}} = (M_{\mathrm{LFDL}} \oplus B) \cup (M_{\mathrm{MITL}} \ominus B)$$

Experiment

- DDL-I dataset
  - 1.5 million samples with pixel-level an-notations
  - 61 latest deepfake methods
  - Four forgery types
  - Single-face and multi-face scenarios

Datasets	Publication	Tasks	Latest Deepfake	Methods	Image	Video	Audio	#Fake
FaceForensics++ [33]	ICCV' 19	Cla	NeuralTextures [37] (2019)	4	0	4	0	4K
Celeb-DF [23]	CVPR' 20	Cla	Unknown	1	0	1	0	5K+
DeeperForensics-1.0 [16]	CVPR' 20	Cla	DF-VAE [16] (2020)	1	0	1	0	10K
DFDC [8]	Arxiv' 20	Cla	StyleGAN [17] (2018)	8	1	6	1	0.1M+
FFIW [46]	CVPR' 21	Cla/SL	FSGAN [28] (2019)	3	0	3	0	10K
OpenForensics [22]	ICCV' 21	SL	InterFaceGAN (2020)	2	2	0	0	0.1M
FakeAVCeleb [20]	NeurIPS' 21	Cla	Wav2Lip [30] (2021)	4	0	1	3	19K+
ForgeryNet [13]	CVPR' 21	Cla/TL/SL	StarGANv2 [18] (2020)	15	7	8	0	1.4M+
LAV-DF [2]	DICTA' 22	Cla/TL	Wav2Lip [30] (2021)	2	0	1	1	0.1M+
DeepFakeFace [35]	ArXiv'23	Cla	Stable-Diffusion [32] (2021)	3	3	0	0	90K
DiffusionDeepfake [1]	ArXiv'24	Cla	Stable-Diffusion [32] (2021)	3	3	0	0	0.1M+
AV-Deepfake1M [3]	MM' 24	Cla/TL	TalkLip (2023)	3	0	1	2	0.8M+
DF40 [41]	NeurIPS' 24	Cla	PixArt-α [4] (2024)	40	17	23	0	1.1M+
DDL	**2025**	Cla/TL/SL	**Kling-2.1 (2025)**	**75**	**40**	**26**	**9**	**1.8M+**

DDL-I dataset



Example of DDL-I dataset

Experiment

➢ **We compared the performance of different models and mask merge strategy.**

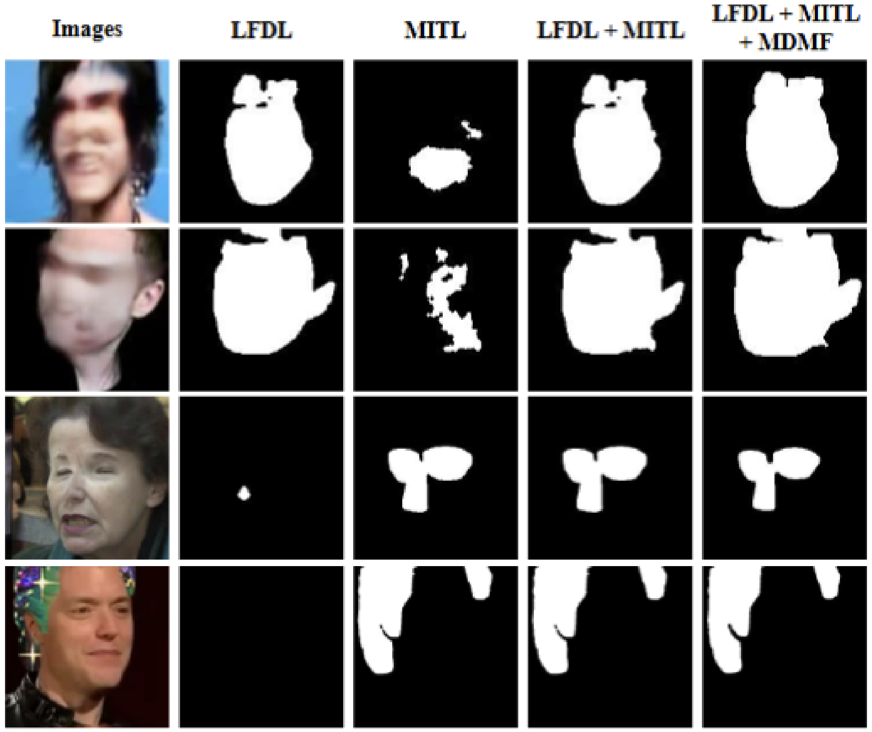
Method	Detection AUC	F1-score	IoU	Final Score
LFDL	0.9790	0.6840	0.5981	0.7497
MITL	-	-	-	0.2349
LFDL + MITL	-	-	-	0.3200
LFDL + MITL + Mask Naive Fusion	0.9790	0.7598	0.6657	0.8015
LFDL + MITL + MDMF	0.9790	0.7759	0.6902	0.8150

The LFDL module offers precise local forgery detection, while the MDMF module complements forgery masks by providing global contextual information.

Note that MITL does not converge until we submit our results.

Experiment

➢ **Visualization results of our methods. We apply dilation to LFDL masks and erosion to MITL masks, then combine them to achieve precise and coherent forgery localization.**



浙江大學
ZHEJIANG UNIVERSITY

Thank you for listening

Contact