Face Forgery Detection by 3D Decomposition

Xiangyu Zhu[1,2*] Hao Wang[1*] Hongyan Fei[3] Zhen Lei[1,2†] Stan Z. Li[4]

[1]CBSR & NLPR, Institute of Automation, Chinese Academy of Sciences,

[2]School of Artificial Intelligence, University of Chinese Academy of Sciences,

[3]School of Automation and Electrical Engineering, University of Science and Technology Beijing,

[4]School of Engineering, Westlake University

CVPR VIRTUAL JUNE 19-25

Highlights

To solve the increasingly demanding digital face manipulation detection, we propose a novel framework, FD²Net, combining multi-modality learning and 3D decomposition, to extract more critical forgery clues from the constituent elements of an RGB face image, which, in turn, improves the efficacy, robustness and explainability of the detector.

- Introducing 3D decomposition into forgery detection
- Constructing facial detail to amplify subtle artifacts
- Proposing a two-stream structure FD²Net containing a supervised attention module
- Remarkable elevation on both detection performance and generalization ability compared with the SOTA

Ablation study

3D components

input	shape	amb	dir	ctex	itex	AUC
In-a	✓	✓	✓	✓	✓	99.13
In-b	✓	✓		✓		50.00
In-c	✓		✓		✓	99.29
In-d	✓				✓	99.14
In-e		✓	✓	✓	✓	98.93
In-f		✓		✓		50.00
In-g			✓		✓	**99.56**
In-h					✓	99.27

Two-stream Network

Structure	FFpp			DFD			DFDC		
	AP	AUC	EER	AP	AUC	EER	AP	AUC	EER
Img	99.44	99.31	5.39	88.07	65.57	38.38	85.60	62.17	39.99
Detail	99.40	99.12	5.51	87.24	64.29	40.87	85.02	61.80	40.37
Img (x2)	99.67	99.38	5.37	89.45	74.14	34.07	86.70	63.22	38.77
Img + Detail (SF)	97.84	92.91	11.07	83.71	72.82	36.44	81.91	62.10	47.32
Img + Detail (FF)	**99.72**	**99.45**	**5.31**	89.56	78.55	26.80	87.16	65.36	36.17
Img + Detail (HF)	99.42	98.73	5.63	**89.61**	**78.65**	**26.03**	**87.31**	**66.09**	**35.46**

Detail-guided Attention

Attention on Stream		FFpp			DFD			DFDC		
Img	Detail	AP	AUC	EER	AP	AUC	EER	AP	AUC	EER
		99.42	98.73	5.63	89.61	78.65	26.03	87.31	66.09	35.46
✓		99.45	98.73	5.62	89.65	78.71	25.94	87.58	66.51	35.33
	✓	99.47	98.74	5.51	89.37	78.69	26.01	87.56	66.48	35.33
✓	✓	**99.48**	**98.76**	**5.59**	**89.84**	**79.08**	**25.18**	**87.93**	**67.70**	**34.91**
✓+	✓+	99.44	98.68	5.88	88.55	78.37	27.46	87.02	65.46	37.23

Facial Detail in FD²Net

S1	S2		FFpp			DFD			DFDC		
Image	Tex Norm	Shape Norm	AP	AUC	EER	AP	AUC	EER	AP	AUC	EER
✓			99.46	99.47	4.48	88.14	72.51	36.77	85.64	62.28	39.56
✓		✓	99.57	99.59	4.30	84.06	76.09	29.11	86.13	64.43	38.40
✓	✓		**99.61**	**99.68**	**4.28**	85.10	76.84	27.08	87.73	66.01	38.32
✓	✓	✓	99.48	98.76	5.59	**89.84**	**79.08**	**25.18**	**87.93**	**67.70**	**34.91**

*Equal contribution. †Corresponding author. +unsupervised

3D Decomposition

- Z-Buffer

$$\mathbf{I}_{syn} = \textit{Z-Buffer}(\mathbf{S}, \mathbf{C})$$

- Lambertian assumption

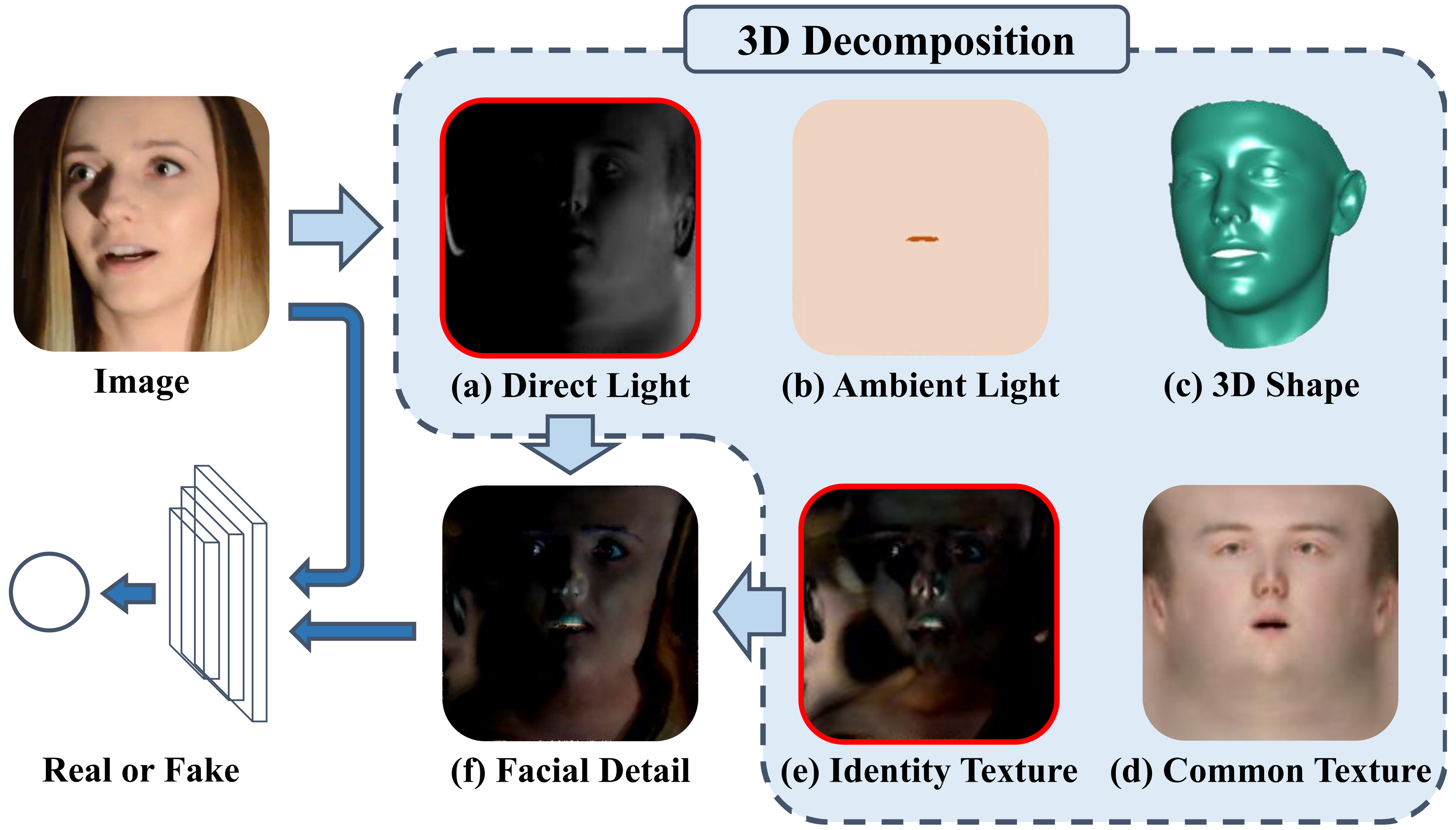
$$\mathbf{T} = \overline{\mathbf{T}} + \mathbf{B}\boldsymbol{\beta} + \mathbf{T}_{id}$$

- Model the common texture by the BFM PCA texture model

$$\mathbf{C}_i = \mathbf{Amb} * \mathbf{T}_i + \langle \mathbf{n}_i, \mathbf{l} \rangle \cdot \mathbf{Dir} * \mathbf{T}_i$$

- 3D face decomposition

$$\arg \min_{\mathbf{S}, \mathbf{Amb}, \mathbf{Dir}, \boldsymbol{\beta}, \mathbf{T}_{id}} \| \mathbf{I} - \mathbf{I}_{syn}(\mathbf{S}, \mathbf{Amb}, \mathbf{Dir}, \boldsymbol{\beta}, \mathbf{T}_{id}) \|$$



Facial Detail Generation

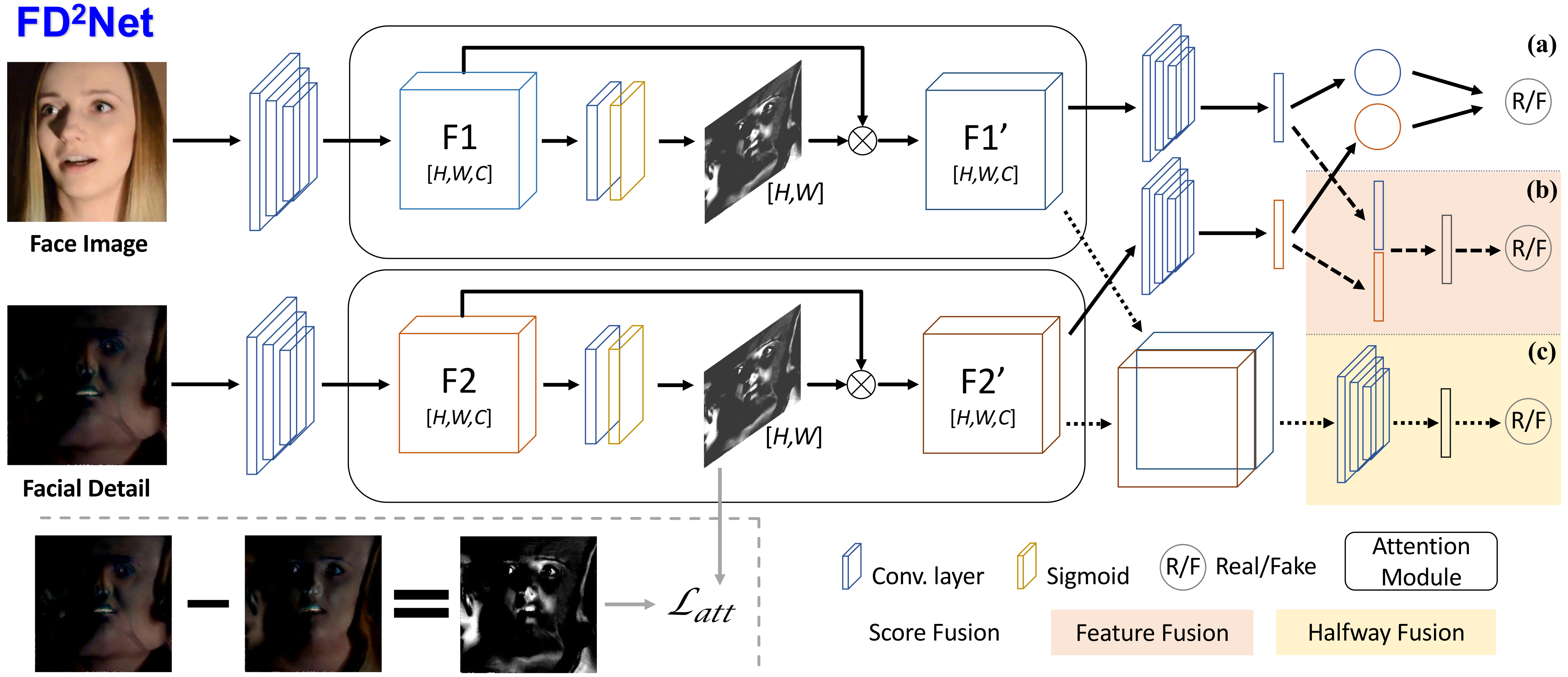
- Common texture based on spherical harmonics

$$\mathbf{I}(\mathbf{S}) = (\mathbf{H}\boldsymbol{\gamma}).*(\overline{\mathbf{T}} + \mathbf{B}\boldsymbol{\beta})$$

- Facial detail

$$\mathbf{FD} = UV(\mathbf{I} - (\mathbf{h}_1\gamma_1).*(\overline{\mathbf{T}} + \mathbf{B}\boldsymbol{\beta}), \mathbf{S})$$

FD²Net



Comparison with other methods

Cross-data Evaluation

Model	Training dataset	DFD			DFDC		
		AP	AUC	EER	AP	AUC	EER
Xception [50]	FFpp	88.07	65.57	38.38	85.60	62.17	39.99
EfficientNetB4 Ensemble [9]	FFpp	89.35	72.82	34.86	85.71	63.03	38.86
FD²Net	FFpp	**89.84**	**79.08**	**25.18**	**87.93**	**67.70**	**34.91**

Different Manipulation Methods

Model	Training data	Acc	
		F2F	FS
MesoInception4 [1]	F2F	84.56	56.71
VA-LogReg [39]		83.62	59.45
LAE [19]		90.34	62.51
Multi-task [40]		91.27	55.04
Face X-ray [36]		97.73	85.69
Xception + HP Filter		97.98	57.46
FD²Net		**98.22**	**86.54**