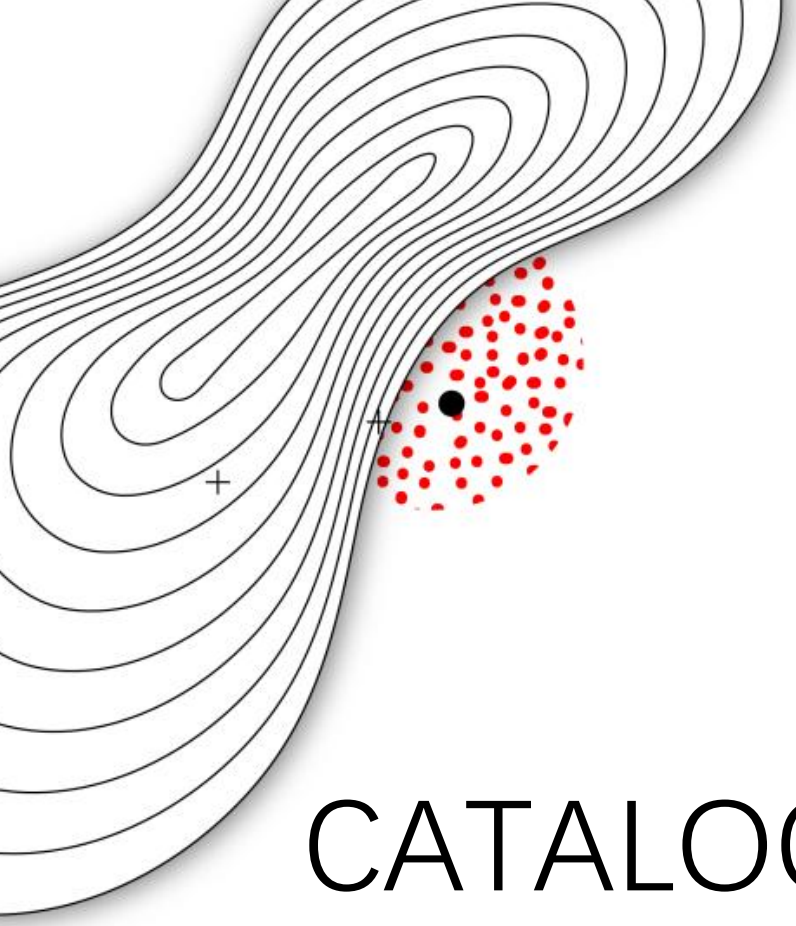
Dynamic Graph Learning with Content-guided Spatial-Frequency Relation Reasoning for Deepfake Detection

Yuan Wang[1,4], Kun Yu[2], Chen Chen[1*], Xiyuan Hu[3], Silong Peng[1,4,5]

[1]Institute of Automation, Chinese Academy of Sciences
[2]Alibaba Group
[3]School of Computer Science and Engineering,
Nanjing University of Science and Technology
[4]University of Chinese Academy of Sciences, [5]Beijing Visystem Co.Ltd

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

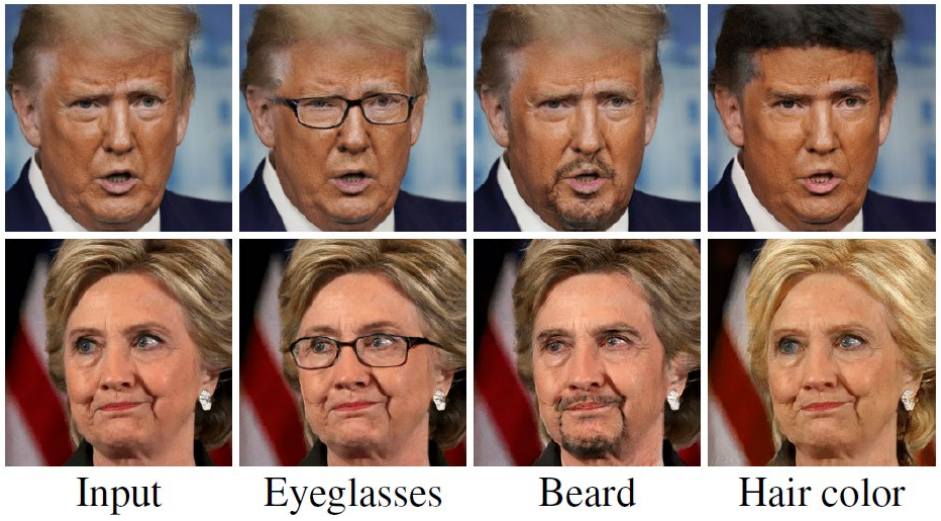


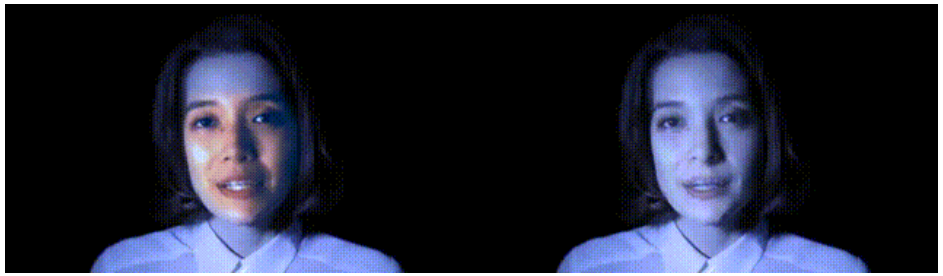
CATALOG

Introduction: Deepfake Detection

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Face forgery technology has stimulated many applications whereas can be abused by malicious intentions to make pornographic movies, fake news and political rumors.





Technical Bottlenecks

- ✓ **Poor Generalization:** Rely heavily on the specific dataset and generation algorithm.
- ✓ **Bad Robustness:** Restricted to the visual compression and noise disturbance.
- ✓ **Multi-domain Interaction:** Incapable of capturing the high-order relationships.

Introduction: Contribution

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

- We propose a **Spatial-Frequency Dynamic Graph network** which is qualified to exploit relation-aware spatial-frequency features to promote generalized forgery detection.

- A CAFÉ module for **content-aware frequency feature extraction** and an MDAML scheme to excavate multiscale spatial-frequency attention maps for understanding the contextual forgeries.

- A DG-SF^{3}Net to discover the multi-domain relationships **with a graph-based relation-reasoning approach**.

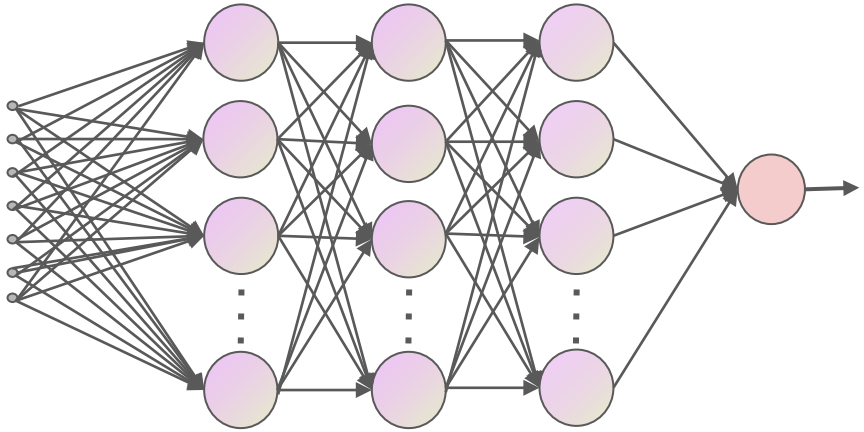
- Our method **achieves state-of-the-art performance** on six benchmarks.

Introduction: Related Works

- Obvious Physiological Forgeries
  - Inconsistent head pose, unnatural eye blinding, ……
  - Off-the-shelf backbones: **ResNet, Xception, EfficientNet**, ……

catastrophic overfitting
Content bias

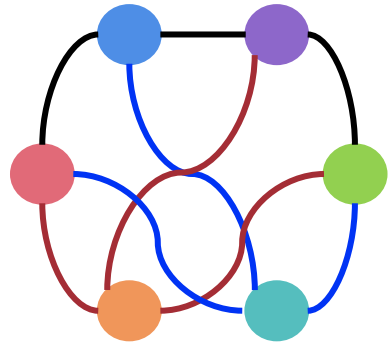
- Generalize Deepfake Detection
  - Multi-task learning strategy
  - Boundary generation
  - Knowledge distillation
  - ……

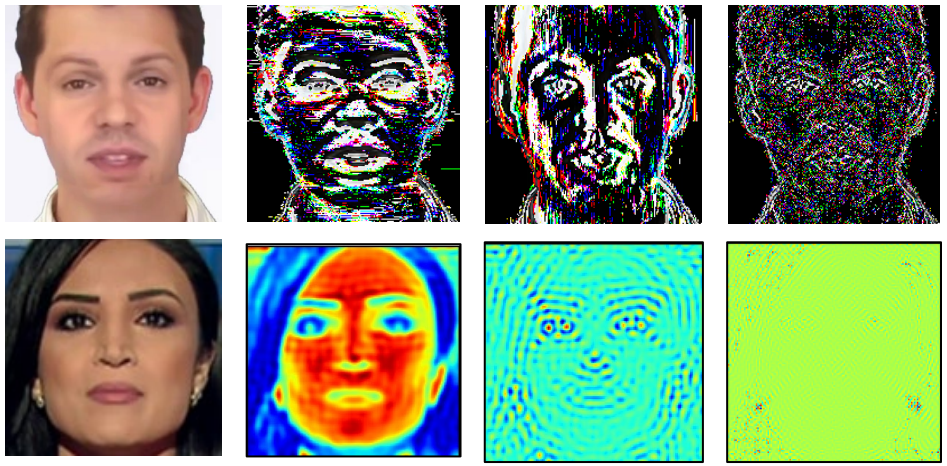


Introduction: Related Works

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

- Specific Multi-domain forgery patterns
  - Frequency clues (Qian et al. **F^{3}Net**, Gu et al., **PEL**)
  - Reconstruction difference (Cao et al. **RECCE**)
  - Local texture (Zhao et al. **MADD**)
  - High-frequency noise (Li et al. **GAFF**)



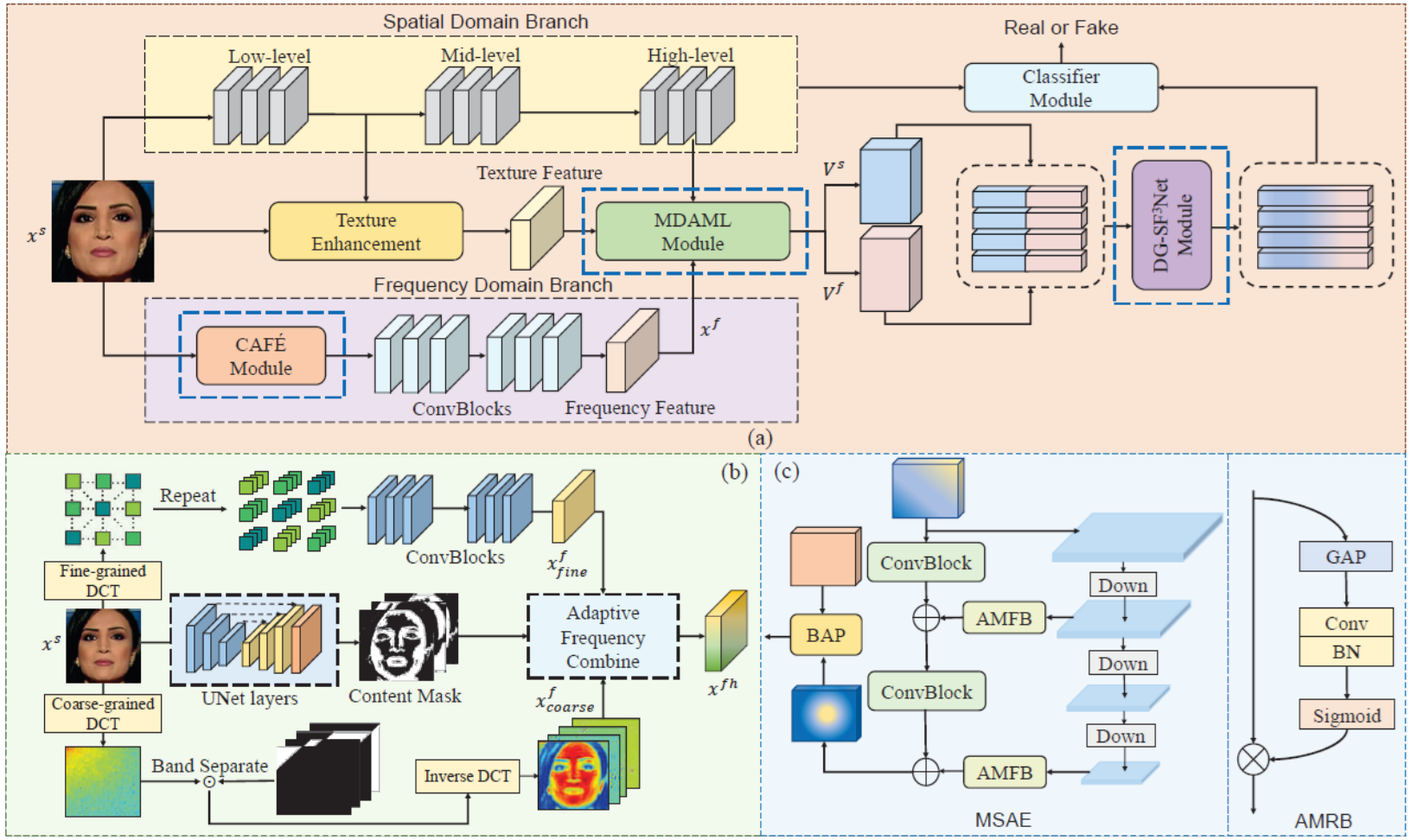


The specific patterns are almost **content-irrelevant**.

The vanilla fusion approach fails to exploit the **high-order** semantic relationships.

Introduction: Overall Framework

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA



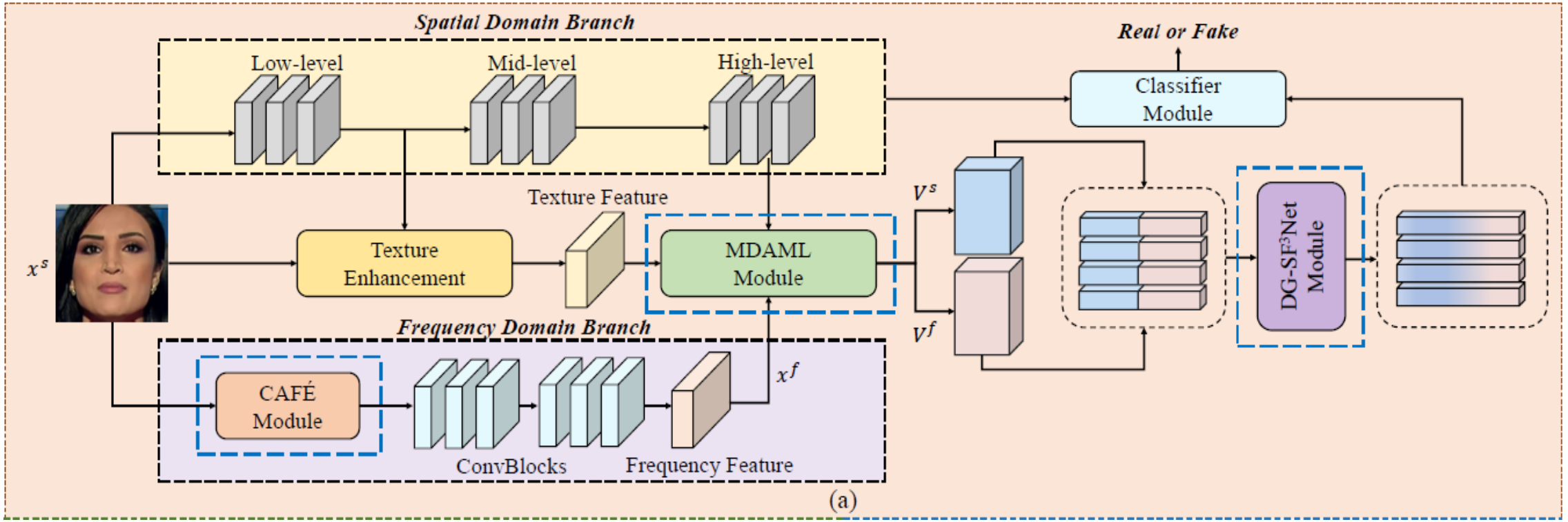
Three Primary Module

CAFE: Content-aware Adaptive Frequency Extraction

MDAML: Multi-Domain Attention Maps Learning

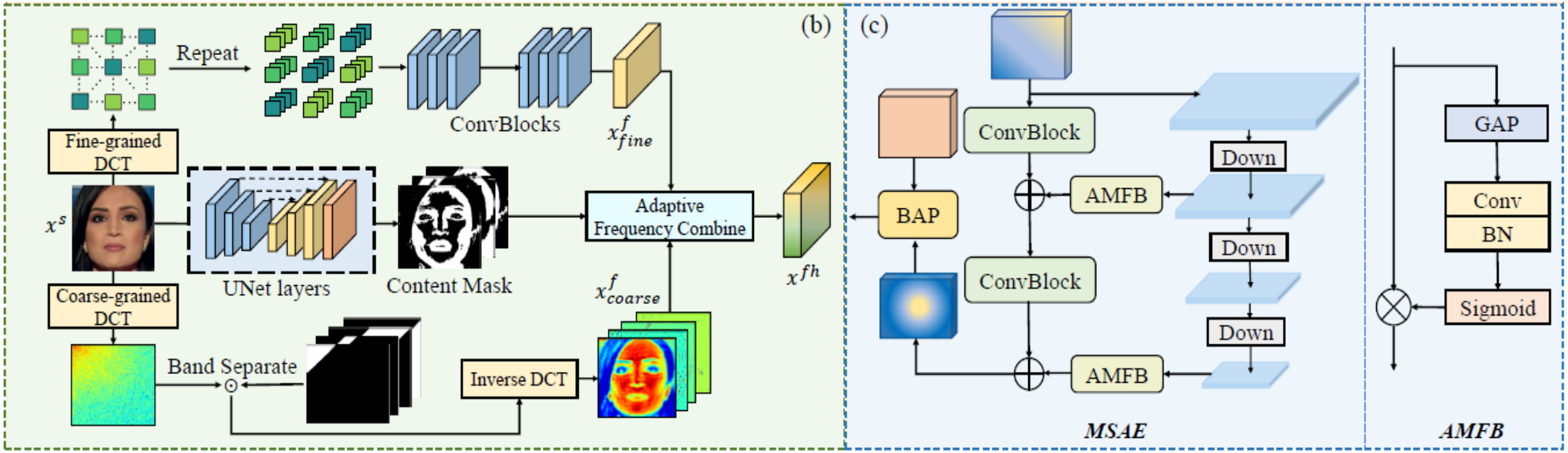
DG-SF³Net: Dynamic Graph-based Spatial Frequency Feature Fusion Network

Method: Spatial-Frequency Dynamic Graph



(a) The proposed **Spatial-Frequency Dynamic Graph** (SFDG) is a two-branch network, exploiting the relation-aware fine-grained spatial-frequency features in an **Adaptive Extraction-Contextual Enhance-Graph Fusion** protocol

Method: Spatial-Frequency Feature Learning



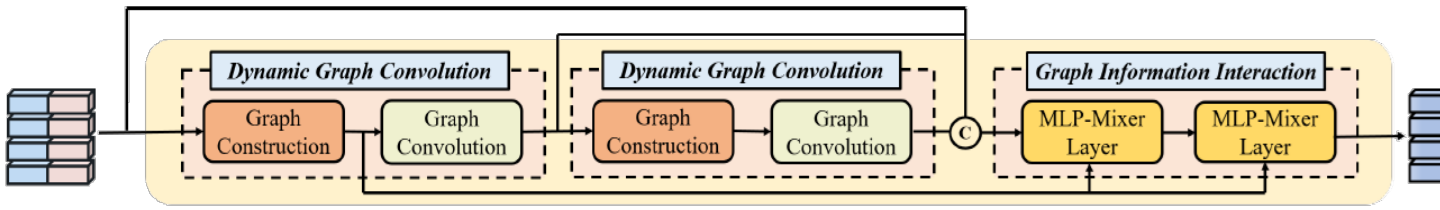
$$\boldsymbol{x}^{f,i}_{coarse} = \mathcal{H}^{-1}[\mathcal{H}(\boldsymbol{x}^s) \odot \boldsymbol{f}^i]$$

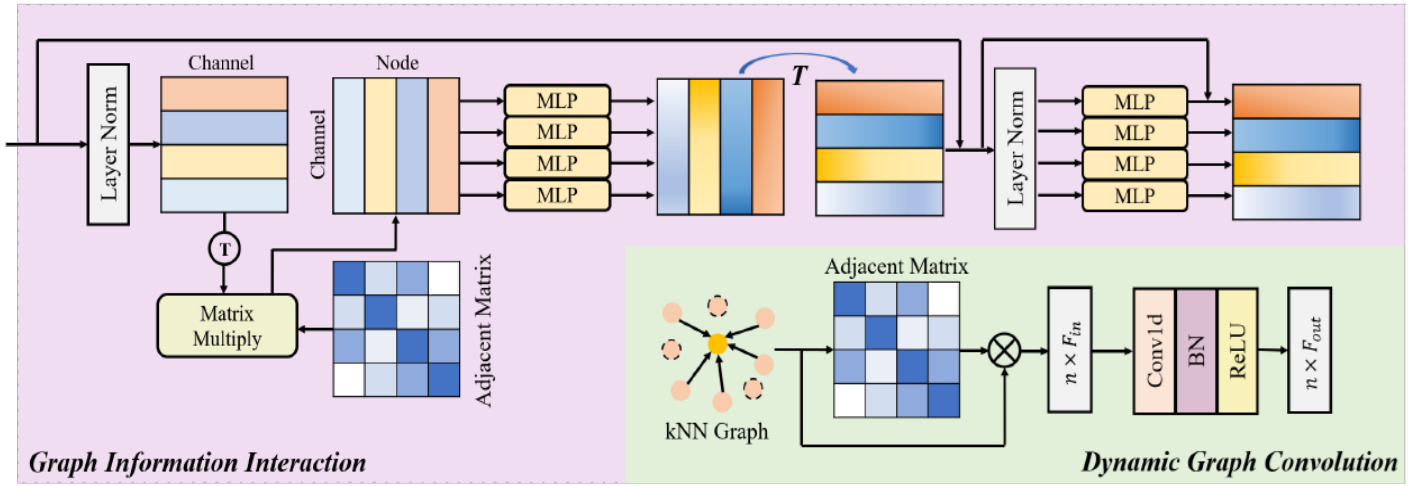
$$\boldsymbol{x}^{fh} = (1 - \mathbf{M}^s) \odot \boldsymbol{x}^f_{fine} + \mathbf{M}^s \odot \boldsymbol{x}^f_{coarse}$$

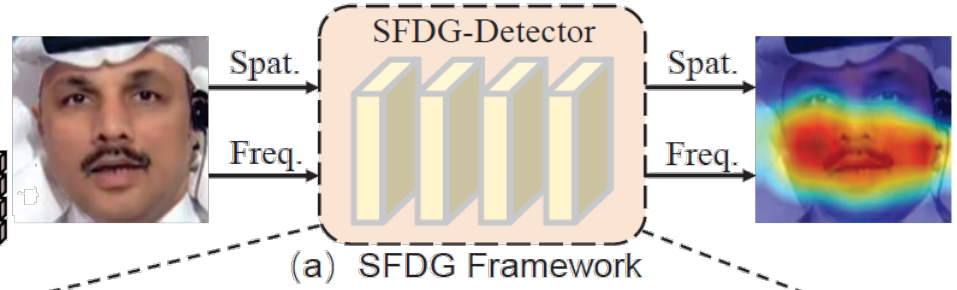
$$\boldsymbol{p}^s_k = \sum_{m=1}^{H_{tex}} \sum_{n=1}^{W_{tex}} \mathbf{F}^{tex}_{k,m,n} \Big/ \left\| \sum_{m=1}^{H_{tex}} \sum_{n=1}^{W_{tex}} \mathbf{F}^{tex}_{k,m,n} \right\|_2$$

(b) The CAFÉ module captures **content-aware adaptive frequency features** for deepfake detection. (c) The MDAML module **enhances the contextual semantics of spatial-frequency features** via multiscale attention maps.

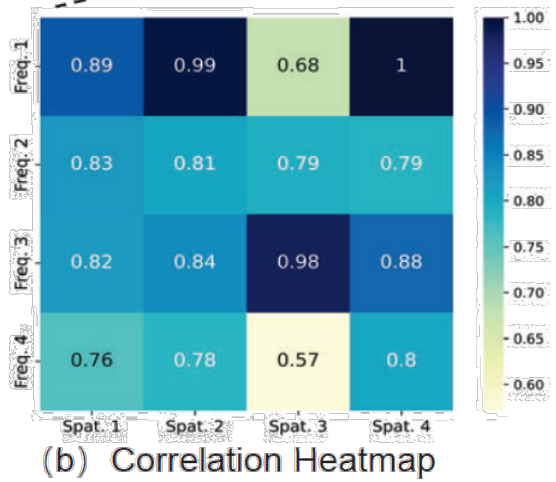
Method: Spatial-Frequency Feature Fusion



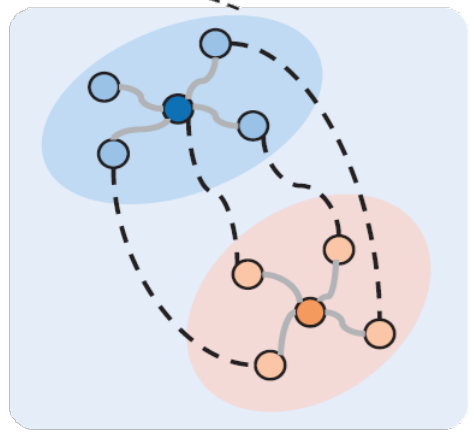




(a) SFDG Framework



(b) Correlation Heatmap



(c) Graph Reasoning

$$\mathcal{N}^{(t)}(i) = \left\{ \boldsymbol{v}_{j_{im}}^{(t)} | \boldsymbol{v}_{j_{im}}^{(t)} \in kNN(\boldsymbol{v}_i^{(t)}), m = 1, \ldots, k \right\}$$

$$\mathbf{V}^{(t+1)} = \mathrm{ReLU}\left(\mathbf{D}^{(t)-\frac{1}{2}} \mathbf{A}^{(t)} \mathbf{D}^{(t)-\frac{1}{2}} \mathbf{V}^{(t)} \mathbf{W}^{(t)} \right)$$

The DG-SF^{3}Net module formulates the interaction of the spatial and frequency domains **via a graph-based relation** discovery protocol.

Experiments: Intra-testing

Method	FF++ (LQ)		FF++ (HQ)		WildDeepfake		Celeb-DF	
	Acc	AUC	Acc	AUC	Acc	AUC	Acc	AUC
Xception	86.86	89.30	95.73	96.30	79.99	88.86	97.90	99.73
Ef-b4	86.67	88.20	96.63	99.18	82.23	90.12	98.19	99.83
Add-Net	87.50	91.01	96.78	97.74	76.25	86.17	96.93	99.55
SCL	89.00	92.40	96.69	99.30	—	—	—	—
MADD	88.69	90.40	97.60	99.29	82.62	90.71	97.92	99.94
F3Net	90.43	93.30	97.52	98.10	80.66	87.53	95.95	98.93
PEL	90.52	94.28	97.63	99.32	**84.14**	91.62	—	—
RECCE	91.03	95.02	97.06	99.32	83.25	92.02	98.59	99.94
Local Relation	91.47	**95.21**	97.59	99.46	—	—	—	—
M2TR	**92.35**	94.22	**98.23**	**99.48**	—	—	—	—
SFDG(Xcep.)	91.08	94.49	97.61	99.45	83.36	**92.15**	**98.95**	**99.94**
SFDG(Ours)	**92.28**	**95.98**	**98.19**	**99.53**	**84.41**	**92.57**	**99.22**	**99.96**

Our SFDG method consistently achieves admirably performance **on all quality settings** and trumps all reference methods by a considerable margin.

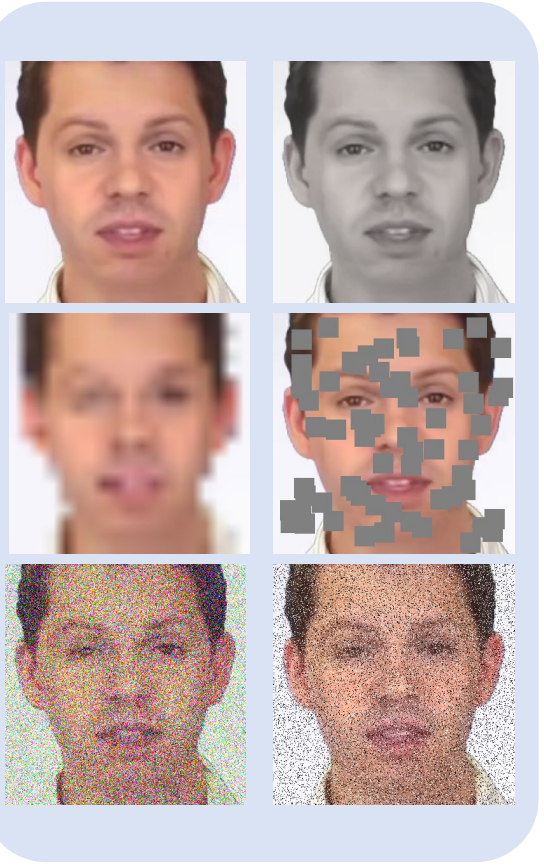
Experiments: Cross-testing

Training Dataset	Method	Testing Dataset									
		Celeb-DF		DFD		WildDeepfake		DFDC		DF-v1.0	
		AUC	EER	AUC	EER	AUC	EER	AUC	EER	AUC	EER
FF++	Xception	60.05	0.43	65.43	0.39	60.59	0.62	55.65	0.46	80.27	0.67
	Ef-b4	64.29	0.42	83.17	0.24	64.27	0.38	60.12	0.43	85.31	0.23
	Add-Net	57.83	0.44	57.16	0.45	54.21	0.46	51.60	0.55	—	—
	MADD	68.64	0.37	74.18	0.33	65.65	0.40	63.02	0.41	89.34	0.17
	F3Net	67.95	0.37	69.50	0.35	60.49	0.43	57.87	0.44	82.27	0.25
	PEL	69.18	0.36	75.86	0.31	67.39	0.39	63.31	0.40	—	—
	SFDG	**75.83**	**0.30**	**88.00**	**0.19**	**69.27**	**0.37**	**73.64**	**0.33**	**92.10**	**0.15**

SFDG explores the essential forgery with **content-aware** attention maps and reasons generalized forged cues via **graph-based** high-order relation discovery

Experiments: Robustness

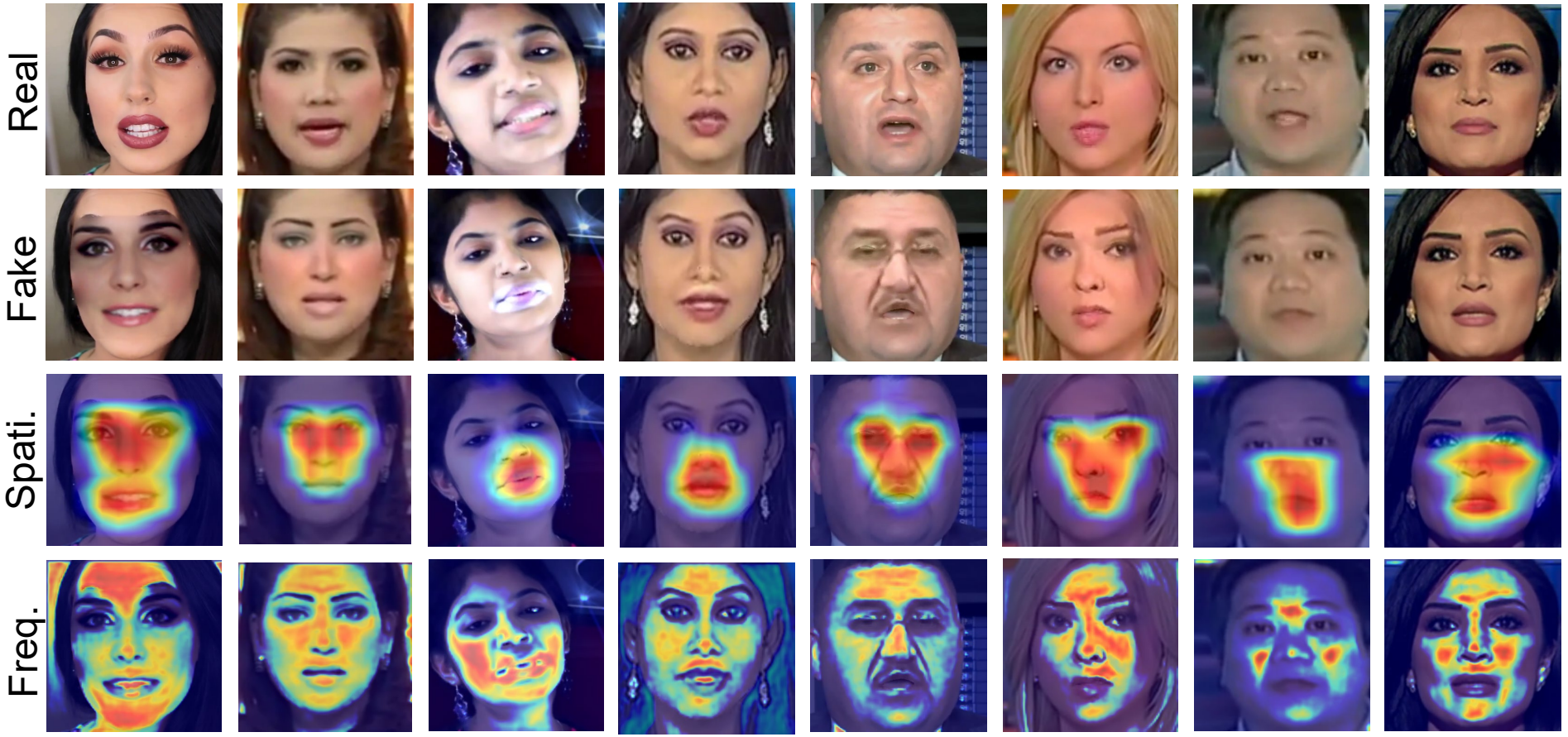
Our method **with the least performance decline** is more robust to perturbations



Methods	Compress	Contrast	Saturate	Pixelate	Average
Xception [2](CVPR'2017)	86.01	81.90	84.96	66.24	79.78
Ef-b4 [12](ICML'2019)	87.63	84.25	86.71	72.93	82.88
RFM [16](CVPR'2021)	83.74	79.77	82.59	71.25	79.35
Add-Net [18](MM'2020)	83.34	89.85	85.13	64.33	80.66
F3-Net [9](ECCV'2020)	86.71	86.53	87.67	73.23	83.54
MADD [17](CVPR'2021)	89.64	89.30	90.37	79.44	87.19
RECCE [1](CVPR'2022)	89.65	91.19	91.74	**83.88**	89.15
SFDG (Ours)	**91.43**	**92.02**	**92.11**	82.71	**89.56**

Method	+GaussianNoise		+SaltPepperNoise		+GaussianBlur	
	ΔAcc(FF)	ΔAcc(Wild)	ΔAcc(FF)	ΔAcc(Wild)	ΔAcc(FF)	ΔAcc(Wild)
Xception [4]	-2.65%	-0.98%	-32.44%	-27.80%	-6.22%	-12.71%
Add-Net [52]	-41.51%	-11.66%	-11.28%	-18.21%	-11.28%	-12.91%
F3Net [28]	-9.86%	-1.17%	-31.08%	-43.57%	-11.08%	-12.43%
MADD [50]	-1.79%	-0.99%	-49.30%	-29.47%	-12.23%	-14.86%
PEL [9]	-0.10%	-0.86%	**-9.39%**	-4.25%	-7.41%	-10.88%
SFDG (Ours)	**-0.10%**	**-0.75%**	-10.10%	**-3.74%**	**-3.76%**	**-5.12%**

Visualization: Grad-CAM



The spatial branch focuses on the **pronounced forgery traces**, while frequency streamline searches manipulated clues from wider areas, e.g., **hair or entire face**

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Visualization: Attention Maps

Feature maps with different scales **highlight distinctive activated intensities**

Denoiser & Aggregator

Enhanced attention maps via hierarchical pyramid can **resist noise disturbance**

Conclusion

- We propose a **Spatial-Frequency Dynamic Graph network** to exploit relations of spatial-frequency domains for spotting subtle forgery clues.

- The CAFÉ module to **mine the adaptive frequency clues** via content-aware frequency learning and the MDAML module captures rich contextual information of spatial-frequency feature.

- The DG-SF^{3}Net performs relation reasoning of spatial and frequency domains via **improved graph convolution**.

- Experiments and visualizations on widely-used benchmarks **confirm the effectiveness** of the SFDG method compared with other contenders.