[1] USTC [2]Shopee
[3]Alibaba [4]UB

Adaptive Face Forgery Detection in Cross Domain

[1]Luchuan Song [2]Zheng Fang [3]Xiaodan Li [1]Xiaoyi Dong
[1]Zhenchao Jin [3]Yuefeng Chen [4]Siwei Lyu

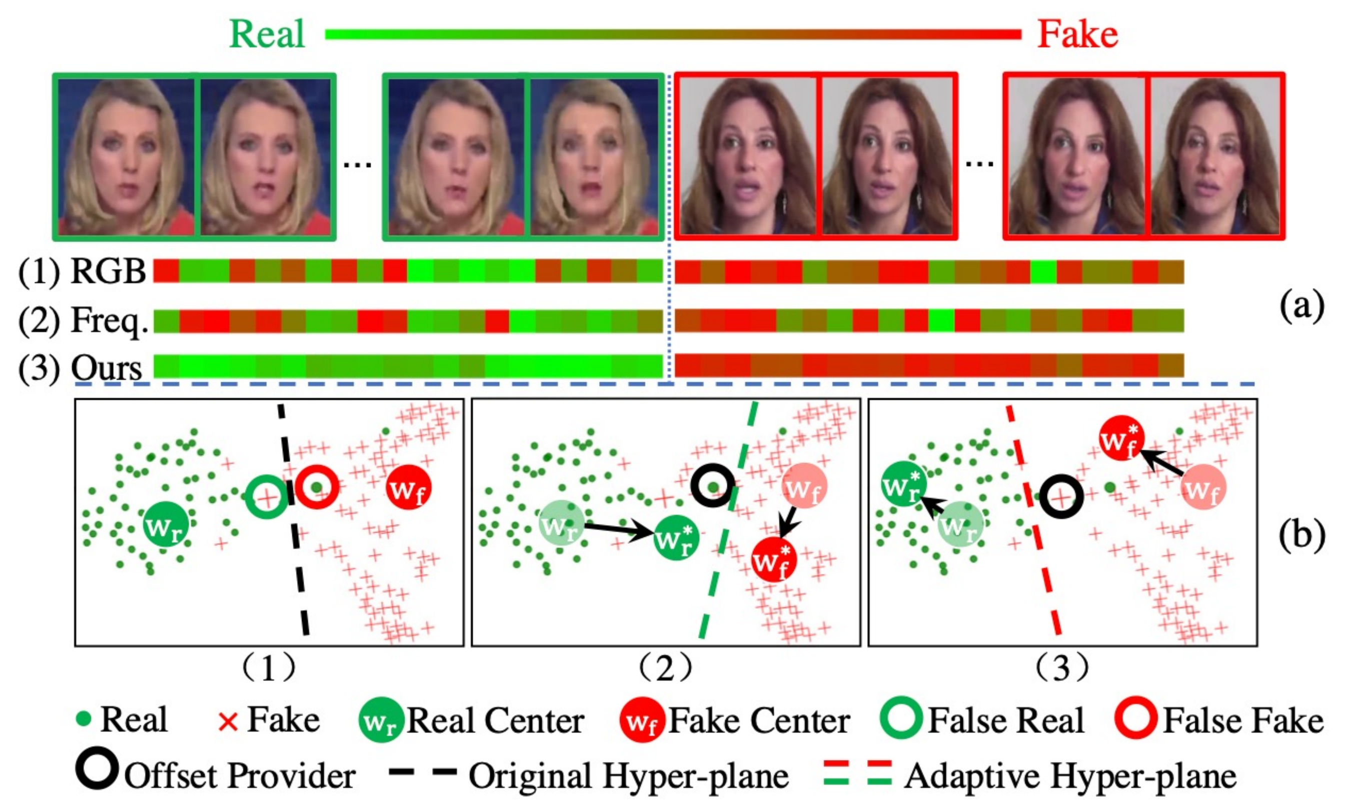
ECCV TEL AVIV 2022

1. Introduction

Motivation:

➢ The existing deepfake detection work originates from the large intra-class distance caused by various artifacts on fake faces.

➢ We want to improve the cross-frame detection consistency for spatial and frequency domain.

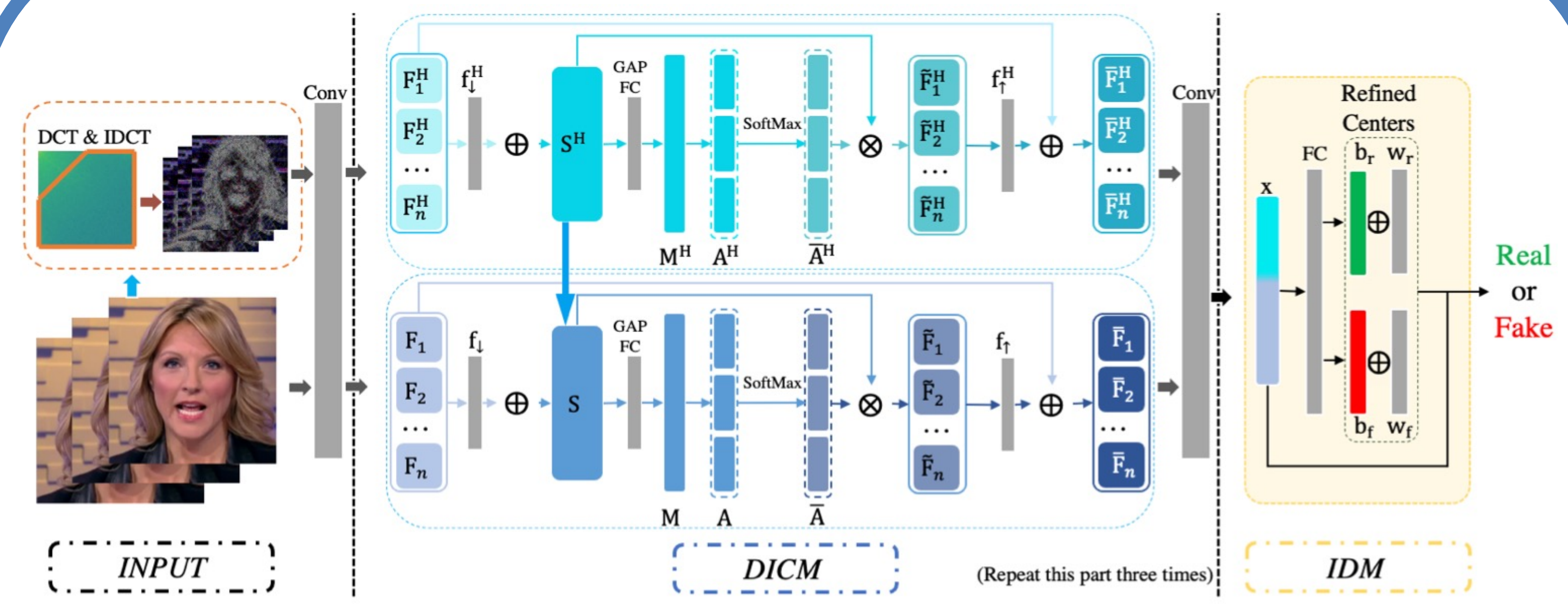


Contribution:

➢ We introduce a Dual-domain Intra-Consistency Module (DICM) to improve consistency and stability of instance representation, which is extracted based on multiple frames in various domains, *i.e.* RGB and frequency patterns.

➢ We introduce an Instance-Discrimination Module to adjust the discriminative centers. It can dynamically adjust the position of the hyperplane according to the input instance, which can help to improve the detection performance further.

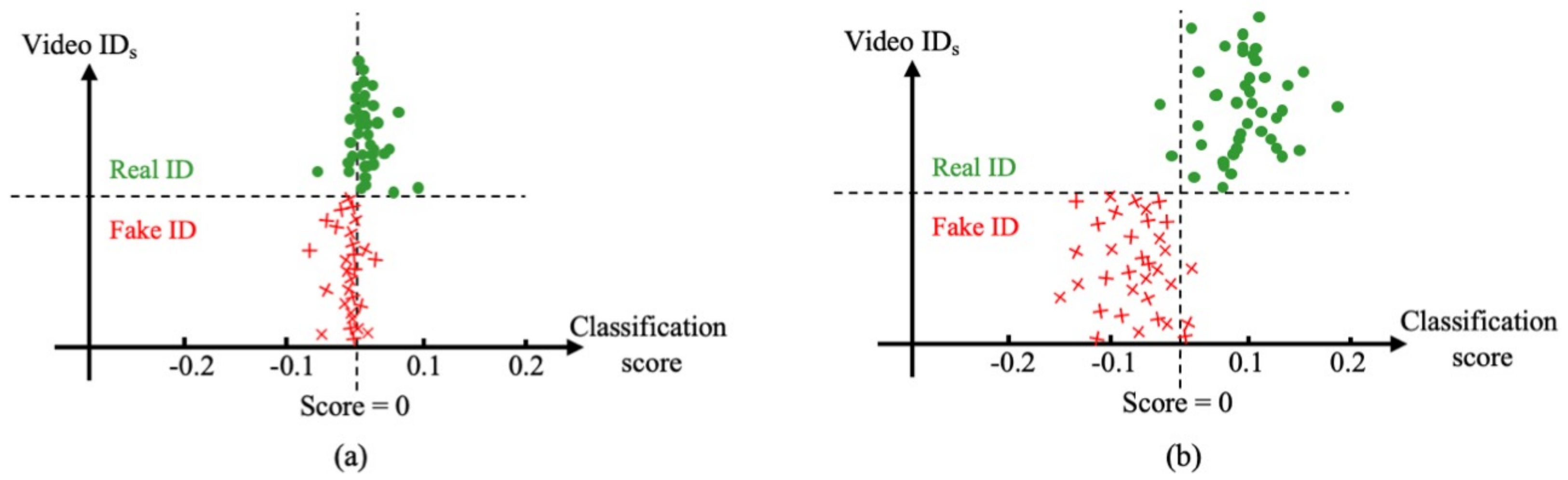
➢ We verify that our approach can achieve state-of-the-art performance on several datasets under both in-domain and out-domain settings.



2. Dual-domain Intra-Consistency Module

➢ We propose a Dual-domain Intra-Consistency Module to extract consistent representations in both the RGB and frequency domain from the input multiple n frames to interact with each other.

➢ The structure of Dual-domain Intra-Consistency Module is in the DICM in above.



3. Instance-Discrimination Module

➢ We propose a novel Instance-Discrimination Module to adaptively adjust the discriminative center based on the instance itself to make robust and efficient predictions.

➢ We propose the IDM to adaptively adjust the discriminative centers based on the instance itself.

$$P(Y=y|\mathbf{x}) = \frac{\exp(\tau \frac{\mathbf{w}_y^\top + \mathbf{b}_y^\top(\mathbf{x})}{\left\|\mathbf{w}_y^\top + \mathbf{b}_y^\top(\mathbf{x})\right\|_2} \frac{\mathbf{x}}{\|\mathbf{x}\|_2})}{\sum_j^N \exp(\tau \frac{\mathbf{w}_j^\top + \mathbf{b}_j^\top(\mathbf{x})}{\left\|\mathbf{w}_j^\top + \mathbf{b}_j^\top(\mathbf{x})\right\|_2} \frac{\mathbf{x}}{\|\mathbf{x}\|_2})},$$

➢ The IDM adjust discriminative centers based on each individual instance. To give insight into it, we compare the difference of Cosine Similarity between $\mathbf{T}_{\mathbf{Norm}}$ and $\mathbf{T}_{\mathbf{IDM}}$, specifically,

$$T_{\text{Norm}} = \frac{\mathbf{w}^\top}{\|\mathbf{w}^\top\|_2} \frac{\mathbf{x}}{\|\mathbf{x}\|_2}, T_{\text{Bias}} = \frac{\mathbf{b}^\top(\mathbf{x})}{\|\mathbf{b}^\top(\mathbf{x})\|_2};$$

$$\begin{aligned}
T_{\text{IDM}} &= \frac{\mathbf{w}^\top + \mathbf{b}^\top(\mathbf{x})}{\|\mathbf{w}^\top + \mathbf{b}^\top(\mathbf{x})\|_2} \frac{\mathbf{x}}{\|\mathbf{x}\|_2} \\
&= \frac{\mathbf{w}^\top}{\|\mathbf{w}^\top + \mathbf{b}^\top(\mathbf{x})\|_2} \frac{\mathbf{x}}{\|\mathbf{x}\|_2} + \frac{\mathbf{b}^\top}{\|\mathbf{w}^\top + \mathbf{b}^\top(\mathbf{x})\|_2} \frac{\mathbf{x}}{\|\mathbf{x}\|_2} \\
&= \frac{\left\|\mathbf{w}^\top\right\|_2}{\|\mathbf{w}^\top + \mathbf{b}^\top(\mathbf{x})\|_2} \left(\frac{\mathbf{w}^\top}{\|\mathbf{w}^\top\|_2} \frac{\mathbf{x}}{\|\mathbf{x}\|_2}\right) + \frac{\left\|\mathbf{b}^\top(\mathbf{x})\right\|_2}{\|\mathbf{w}^\top + \mathbf{b}^\top(\mathbf{x})\|_2} \left(\frac{\mathbf{b}^\top(\mathbf{x})}{\|\mathbf{b}^\top(\mathbf{x})\|_2} \frac{\mathbf{x}}{\|\mathbf{x}\|_2}\right) \\
&= \frac{\left\|\mathbf{w}^\top\right\|_2}{\|\mathbf{w}^\top + \mathbf{b}^\top(\mathbf{x})\|_2} T_{\text{Norm}} + \frac{\left\|\mathbf{b}^\top(\mathbf{x})\right\|_2}{\|\mathbf{w}^\top + \mathbf{b}^\top(\mathbf{x})\|_2} T_{\text{Bias}}
\end{aligned}$$

4. Experiments

➢ In-domain results:

Methods	AUC (LQ)	Acc (LQ)	AUC (HQ)	Acc (HQ)	AUC (RAW)	Acc (RAW)
Steg.Features [21]	-	55.98%	-	70.97%	-	97.63%
LD-CNN [14]	-	58.69%	-	78.45%	-	98.57%
Constrained Conv [6]	-	66.84%	-	82.97%	-	98.74%
CustomPooling CNN [41]	-	61.18%	-	79.08%	-	97.03%
MesoNet [3]	-	70.47%	-	83.10%	-	95.23%
Face X-ray [27]	0.616	-	0.874	-	0.987	-
Two-branch RNN [35]	0.911	86.34%	0.991	96.43%	-	-
Xception [11]	0.925	84.11%	0.963	95.04%	0.992	98.77%
STIL† [22]	0.948	86.31%	0.986	98.57%	0.993	99.04%
PCL&I2G† [55]	0.939	87.02%	0.990	98.85%	0.997	99.78%
F^3-Net (Xception) [40]	0.933	86.89%	0.981	97.31%	0.998	99.84%
CD-Net (Xception)	**0.952**	**88.12%**	**0.999**	**98.75%**	**0.999**	**99.91%**
I3D [8]	-	87.43%	-	-	-	-
3D ResNet [23]	-	83.86%	-	-	-	-
3D ResNeXt [51]	-	85.14%	-	-	-	-
3D R50-FTCN [56]	0.966	92.35%	0.995	98.59%	0.997	99.84%
Slowfast [19]	0.936	88.25%	0.982	96.92%	0.994	99.34%
F^3-Net (Slowfast) [40]	0.958	92.37%	0.993	98.64%	0.999	99.91%
CD-Net (Slowfast)	**0.985**	**93.21%**	**0.999**	**98.93%**	**0.999**	**99.91%**

➢ Out-domain results:

Methods	DFDC	Celeb-DF v2	Methods	DFDC	Celeb-DF v2
Two-Branch [35]	-	0.767	PCL&I2G [55]	0.675	0.900
CNN-aug [50]	0.721	0.756	3DR50-FTCN [56]	0.740	0.869
CNN-GRU [43]	0.689	0.698	Multi-task [38]	0.681	0.757
FWA [29]	0.695	0.673	PatchForensics [9]	0.656	0.696
Face X-ray [27]	0.655	0.795	STIL† [22]	0.661	0.715
VA-LogReg [36]	0.680	0.651	DSP-FWA [29]	0.630	0.693
Xception-raw [11]	0.709	0.655	**CD-Net**[1]	**0.783**	0.877
Xception-c23 [11]	0.717	0.635	**CD-Net**[2]	0.770	0.885
Xception-c40 [11]	0.709	0.655	**CD-Net**[3]	0.753	**0.921**