

# An Experimental Evaluation of Recent Face Recognition Losses for Deepfake Detection

Yu-Cheng Liu\*, Chia-Ming Chang<sup>†</sup>, I-Hsuan Chen<sup>‡</sup>, Yu-Ru Ku<sup>§</sup> and Jun-Cheng Chen\*

\*Research Center for Information Technology Innovation, Academia Sinica, Taiwan

<sup>†</sup>Department of Electrical Engineering, National Taiwan University, Taiwan

<sup>‡</sup>Department of Engineering Science and Ocean Engineering, National Taiwan University, Taiwan

<sup>§</sup>Department of Statistics, National Taipei University, Taiwan

# Introduction

- With the rise of deep generative models in recent years, convolutional neural networks (CNN) can be applied to generate faces which do not exist.
- The forged face images or videos can deliver wrong messages or damage human reputation through social multimedia network.

# Motivation and Contribution

- Due to the threats of fake images, it is important to develop protection technology to recognize them from the pristine faces.
- Contribution
  - Quantitative evaluation with different loss functions
  - Qualitative analysis using Grad-Cam and t-SNE

# Pipeline

## Training

Input video frames



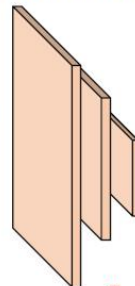
Face Detection/  
IoU Tracking



Face  
Alignment



Deep Classification Model



Loss functions

A-softmax loss  
L-softmax loss  
Softmax loss  
Arcface loss  
Focal loss  
 $L_2$  softmax loss  
Center loss  
COCO loss

## Inference

Input video frames



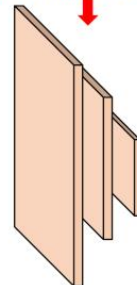
Face Detection/  
IoU Tracking



Face  
Alignment



Deep Classification Model



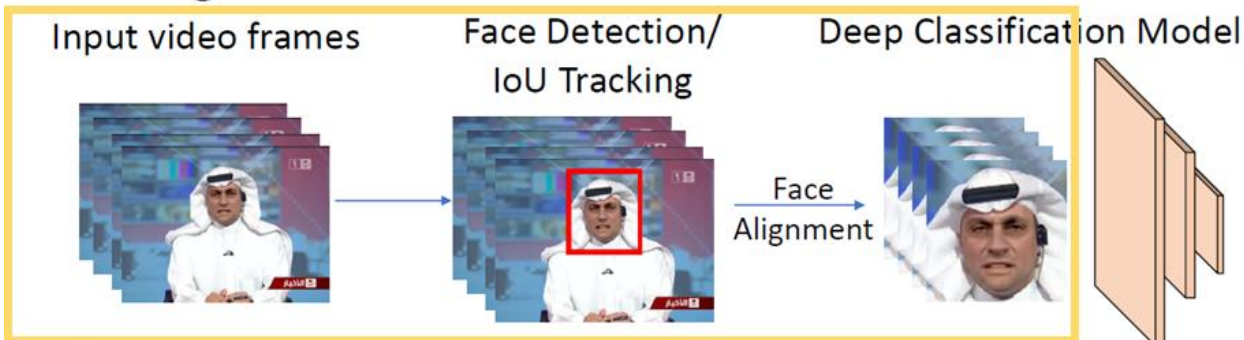
Temporal  
Aggregation

Real  
Fake

# Pipeline (Cont'd)

## Training

### Pre-processing

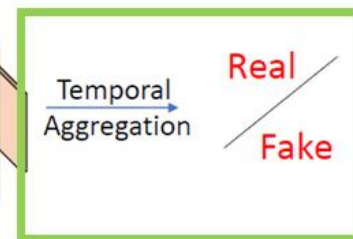


### Different losses

#### Loss functions

- A-softmax loss
- L-softmax loss
- Softmax loss
- Arcface loss
- Focal loss
- $L_2$  softmax loss
- Center loss
- COCO loss

## Inference



### Classifier

# Pre-Processing

- Face detection, alignment and tracking



# Loss functions

- Types
  - Softmax Loss
  - L2-constrained Softmax Loss (L2s)
  - Large-margin Softmax Loss (L-softmax)
  - Angular Softmax Loss (A-softmax)
  - Arcface Loss
  - Center Loss
  - COCO Loss
  - Focal Loss

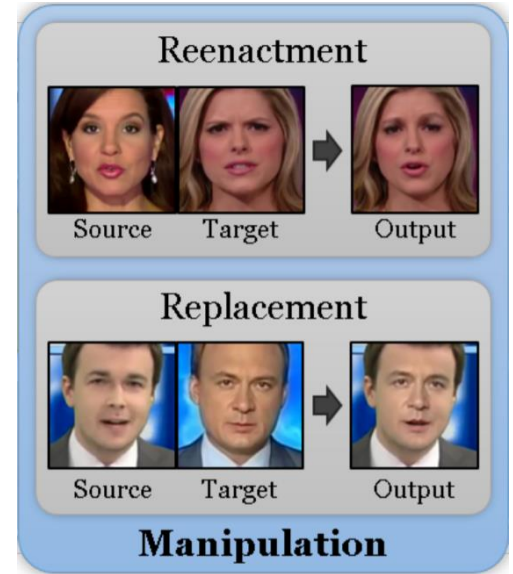
# Temporal Aggregation

- Max-voting through faces (MF)
- Max-voting between tracks (MT)
- Average-voting between tracks (AT)



# Datasets

- FaceForensics++
  - Contains 5000 videos with four manipulation methods
  - Three different video qualities including raw, c23 and c40



# Metrics

- Accuracy
- AUC
- Logloss

$$\mathcal{L}_{\log} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

# Results

- Intra-dataset (FaceForensics++)

Raw									
Class	All			logloss			AUC		
Loss	MF	MT	AT	MF	MT	AT	MF	MT	AT
Center	95.86	95.00	96.29	0.1824	0.2092	0.1931	0.9939	0.9926	0.9924
COCO	97.57	95.71	98.14	1.4150	0.1929	0.1365	0.9810	0.9928	0.9956
Areface	98.57	<b>98.28</b>	<b>98.57</b>	<b>0.0969</b>	<b>0.1217</b>	<b>0.1183</b>	0.9967	0.9963	0.9959
A-softmax	98.57	96.86	<b>98.57</b>	0.1473	0.3936	0.2212	<b>0.9986</b>	<b>0.9948</b>	<b>0.9971</b>
L-softmax	98.57	95.71	98.43	0.1590	0.3859	0.2272	0.9980	0.9946	0.9966
Softmax	<b>98.71</b>	97.29	<b>98.57</b>	0.1429	0.2648	0.1748	0.9972	0.9930	0.9951
Focal	97.43	95.71	98.43	0.1116	0.1858	0.1243	0.9973	0.9922	0.9960
L2s	<b>98.71</b>	97.57	98.43	0.1679	0.2833	0.2360	0.9964	0.9930	0.9934

c23									
Class	All			logloss			AUC		
Loss	MF	MT	AT	MF	MT	AT	MF	MT	AT
Center	92.71	91.14	92.86	1.5530	0.4304	0.3426	0.9271	0.9635	0.9786
COCO	95.71	93.57	95.86	1.3318	<b>0.3876</b>	0.3029	0.9804	0.9830	0.9874
Areface	95.71	94.57	<b>96.57</b>	<b>0.2991</b>	0.4149	0.2878	0.9901	0.9765	0.9883
A-softmax	95.71	94.14	95.57	0.6486	0.8040	0.6468	0.9914	0.9878	0.9902
L-softmax	<b>96.29</b>	<b>95.14</b>	96.14	0.5299	0.6980	0.5945	<b>0.9928</b>	<b>0.9891</b>	<b>0.9916</b>
Softmax	95.86	94.71	95.71	0.5878	0.6573	0.5058	0.9883	0.9848	0.9892
Focal	95.14	93.00	95.29	0.3002	0.6433	<b>0.2463</b>	0.9872	0.9669	0.9889
L2s	95.71	94.85	96.14	0.8083	0.9913	0.8405	0.9919	0.9885	0.9902

c40									
Class	All			logloss			AUC		
Loss	MF	MT	AT	MF	MT	AT	MF	MT	AT
Center	79.00	80.14	80.29	<b>0.8116</b>	<b>0.8191</b>	<b>0.7965</b>	0.9179	0.9152	0.9194
COCO	80.14	82.57	82.43	0.9582	1.0054	0.9692	0.9364	0.9305	0.9343
Areface	82.57	83.00	83.43	0.9103	0.9758	0.8921	0.9223	0.9138	0.9233
A-softmax	83.00	82.85	83.29	1.3762	1.4758	1.4109	0.9446	0.9390	0.9424
L-softmax	80.43	<b>83.43</b>	<b>83.71</b>	1.2185	1.6849	1.5672	0.9319	0.9334	0.9382
Softmax	<b>83.14</b>	82.85	83.29	1.2853	1.4066	1.3021	<b>0.9495</b>	<b>0.9435</b>	<b>0.9482</b>
Focal	81.00	81.28	81.71	0.8187	0.8479	0.8221	0.9174	0.9124	0.9169
L2s	80.57	80.85	81.29	1.4834	1.6671	1.5224	0.9475	0.9396	0.9452

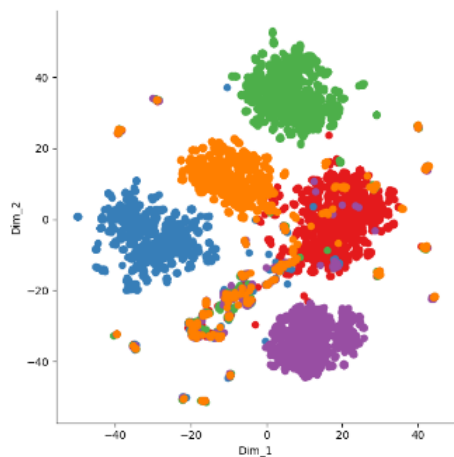
# Results (Cont'd)

- Fusion

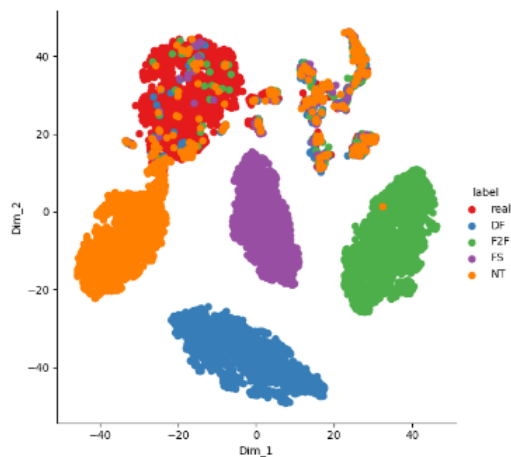
<b>raw</b>								
Loss	NT	DF	F2F	FS	P	All	logloss	AUC
A-softmax+Arcface	95.00	100.00	97.85	100.00	100.00	98.57	0.1061	0.9965
Softmax+L2s	95.00	100.00	98.57	100.00	99.29	98.57	0.1685	0.9947
<b>c23</b>								
Arcface+L-softmax	95.00	98.57	97.85	98.57	92.14	96.42	0.2535	0.9897
Softmax+L2s	94.29	97.85	97.14	98.57	92.14	96.00	0.4264	0.9904
<b>c40</b>								
A-softmax+Arcface	76.43	95.71	85.00	92.14	76.42	85.14	0.6703	0.9472
Softmax+L-softmax	76.43	92.86	83.57	91.42	76.42	84.14	1.0977	0.9448

# Results (Cont'd)

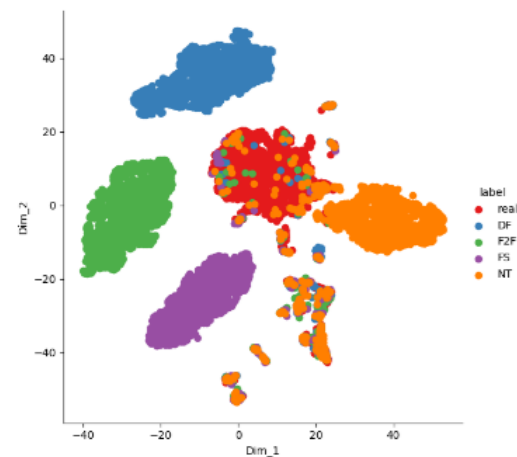
- t-SNE



(a) Feature distribution of raw L2s on t-SNE



(b) Feature distribution of raw Softmax on t-SNE



(c) Feature distribution of fusion of L2s and softmax on t-SNE

# Results (Cont'd)

- Intra-dataset cross-class (FaceForensics++)

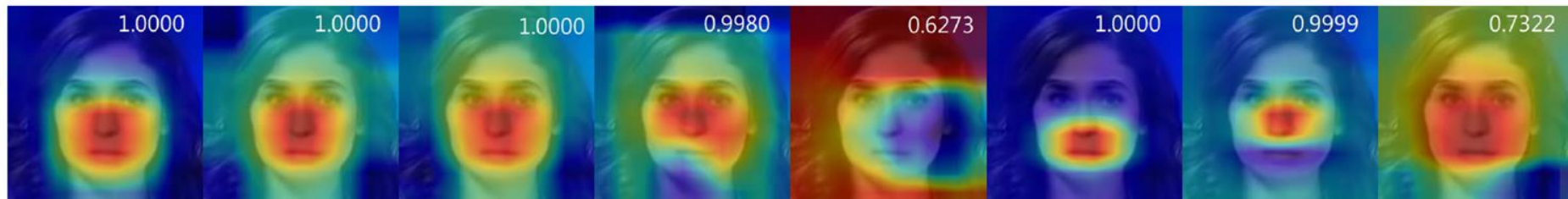
Accuracy									
Train	Test	Softmax	L2s(s=64)	L-softmax	A-softmax	Arcface(s=64)	Center	COCO	Focal
NT	NT	94.29	89.64	95.00	93.57	95.00	88.93	<b>95.36</b>	94.29
NT	DF	66.43	70.00	71.07	62.14	<b>74.64</b>	60.71	71.07	71.79
NT	FS	48.57	47.14	47.86	48.21	48.21	48.21	<b>50.00</b>	48.21
NT	F2F	57.86	57.86	56.43	55.36	<b>61.79</b>	58.57	55.71	56.43
DF	DF	99.29	98.21	<b>99.64</b>	<b>99.64</b>	98.21	98.57	<b>99.64</b>	98.57
DF	NT	51.07	51.43	50.71	50.36	50.36	51.43	<b>53.21</b>	51.43
DF	FS	<b>50.00</b>	49.64	<b>50.00</b>	<b>50.00</b>	49.64	49.29	<b>50.00</b>	49.64
DF	F2F	51.07	51.43	51.07	50.36	50.00	50.71	<b>52.14</b>	50.00
FS	FS	99.29	99.29	99.29	<b>99.64</b>	99.29	99.29	99.29	99.29
FS	NT	<b>50.00</b>	<b>50.00</b>	<b>50.00</b>	<b>50.00</b>	<b>50.00</b>	<b>50.00</b>	<b>50.00</b>	<b>50.00</b>
FS	DF	50.36	51.07	50.36	51.43	<b>51.79</b>	50.71	50.36	50.71
FS	F2F	50.00	<b>51.43</b>	50.00	50.71	50.36	50.00	50.36	50.71
F2F	F2F	<b>99.64</b>	98.57	99.64	98.93	98.93	99.29	98.93	99.29
F2F	NT	50.00	<b>51.79</b>	50.71	50.71	50.00	50.36	50.00	50.00
F2F	DF	53.57	56.43	54.29	52.86	51.07	<b>59.29</b>	53.21	53.21
F2F	FS	50.00	<b>50.71</b>	50.00	50.00	50.00	50.00	50.36	<b>50.71</b>

AUC									
Train	Test	Softmax	L2s(s=64)	L-softmax	A-softmax	Arcface(s=64)	Center	COCO	Focal
NT	NT	0.9616	0.9526	0.9674	0.9591	0.9671	0.8613	0.9584	<b>0.9680</b>
NT	DF	0.7719	0.7894	0.8159	0.7809	0.8129	0.6933	<b>0.8576</b>	0.8196
NT	FS	0.4688	0.4457	0.4566	0.4586	0.4711	0.4402	<b>0.4953</b>	0.4667
NT	F2F	0.6700	0.6898	0.7001	0.6714	0.7001	0.6457	<b>0.7246</b>	0.6860
DF	DF	0.9990	0.9982	0.9991	<b>0.9994</b>	0.9991	0.9612	0.9971	0.9987
DF	NT	0.7391	<b>0.7613</b>	0.6875	0.6764	0.7484	0.7186	0.7307	0.7496
DF	FS	0.4349	0.3083	0.4132	<b>0.4388</b>	0.3638	0.3527	0.3425	0.3784
DF	F2F	0.7247	0.7216	0.6827	0.6495	<b>0.7649</b>	0.6652	0.6923	0.7521
FS	FS	<b>0.9989</b>	0.9967	0.9977	<b>0.9989</b>	0.9967	0.9935	0.9972	0.9981
FS	NT	0.4616	0.3942	0.4715	0.4793	<b>0.5221</b>	0.4668	0.4651	0.4657
FS	DF	0.6009	0.5153	0.5720	0.5695	<b>0.6403</b>	0.5551	0.5218	0.5263
FS	F2F	0.6725	0.6438	0.6455	0.6548	0.6567	0.6300	0.6264	<b>0.6940</b>
F2F	F2F	0.9916	0.9897	0.9929	0.9936	0.9925	0.9885	0.9911	<b>0.9938</b>
F2F	NT	0.6187	0.6484	0.6257	0.6116	<b>0.6691</b>	0.5386	0.5869	0.5575
F2F	DF	0.7763	0.7831	0.7587	0.7522	<b>0.8216</b>	0.7853	0.5938	0.7678
F2F	FS	0.5847	0.5803	0.5577	0.5669	<b>0.5932</b>	0.4144	0.4975	0.5594

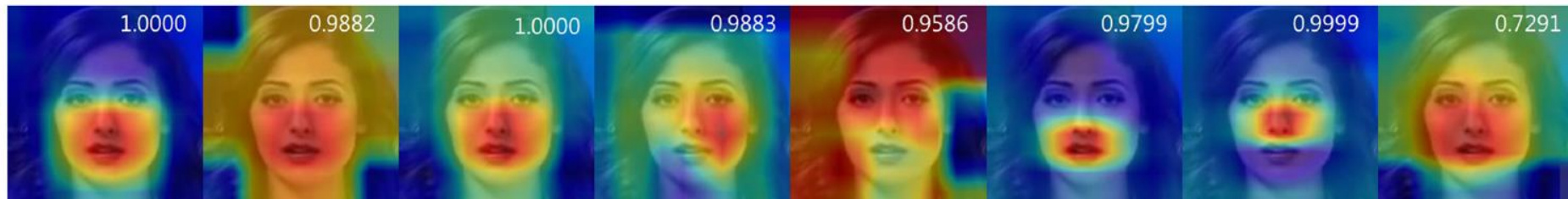
# Results (Cont'd)

- Grad-Cam

DF Image 1



NT Image 1



(a) Softmax

(b) L-softmax

(c) A-softmax

(d) COCO

(e) Center

(f) Arcface

(g) Focal

(h) L2s

# Conclusion

- Models trained on specific losses reach best performances in different qualities of data in FaceForensics++.
- Model trained on specific class of data (NeuralTextures) has transferability to another class (DeepFakes).
- Performances can be enhanced by fusing different models