ICPR 2021 Paper

Detecting Manipulated Facial Videos: A Time Series Solution

Zhewei Zhang, Can Ma, Meilin Gao, Bowen Ding

The Third Research Laboratory,
Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China

zhangzhewei@iie.ac.cn

Manipulated Video Detection Problems

Challenges:

- Huge datasets with various manipulate generation methods
- Poor robustness on unseen cases, slow inference speed

Related Works: Facial Forgery Detection Method

- [Pan et al., 2007] Eye-Blinking Model
- [Afchar et al., 2018] CNN-Based Model
- [Ferrara et al., 2012] Color-based Model
- [Li et al., 2018] LSTM-CNN Model

Our approaches: FA/DFA-LSTM
Simple yet Effective

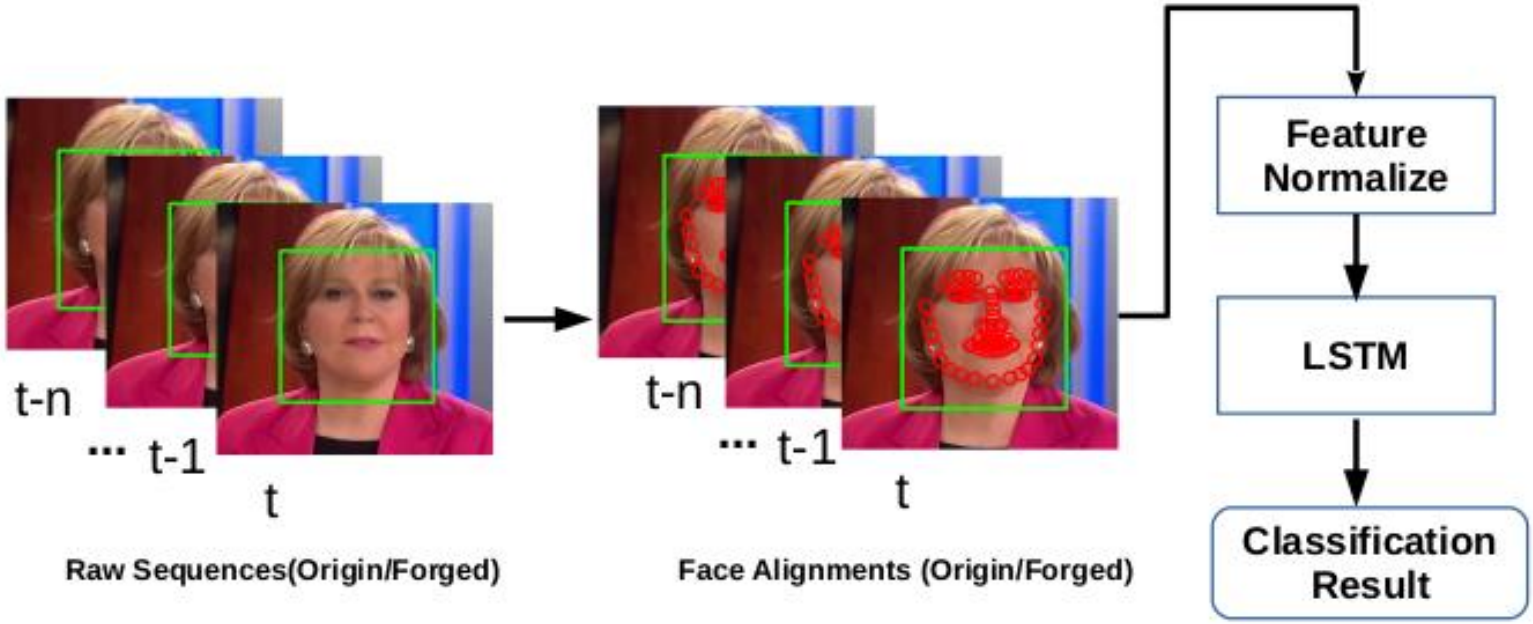


Figure 1: Framework of the proposed FA-LSTM model.

$I \in R^{n \times 136}$, in which n denotes the window length of LSTM. Note that n also equals to one slice of the i

Feature Project: Dense Face Alignment

3D Morphable Model: $\boldsymbol{S} = \bar{\boldsymbol{S}} + \boldsymbol{A}_{shp}\boldsymbol{\alpha}_{shp} + \boldsymbol{A}_{exp}\boldsymbol{\alpha}_{exp}$

Texture Model: $\boldsymbol{T} = \bar{\boldsymbol{T}} + \boldsymbol{B}_{shp}\boldsymbol{\beta}_{shp}$

Projected 3D facial vertices: $V(\boldsymbol{p}) = s \cdot \mathbf{P}_{\boldsymbol{r}} \cdot \boldsymbol{R}(\bar{\boldsymbol{S}} + \boldsymbol{A}_{shp}\boldsymbol{\alpha}_{shp} + \boldsymbol{A}_{exp}\boldsymbol{\alpha}_{exp}) + \boldsymbol{t}_x$

Parameters: $\boldsymbol{p} = [s, \boldsymbol{R}, \boldsymbol{t}_x, \boldsymbol{\alpha}_{shp}, \boldsymbol{\alpha}_{exp}]$

PAF: 3D vertex => 64 × 64 × 2 feature anchors

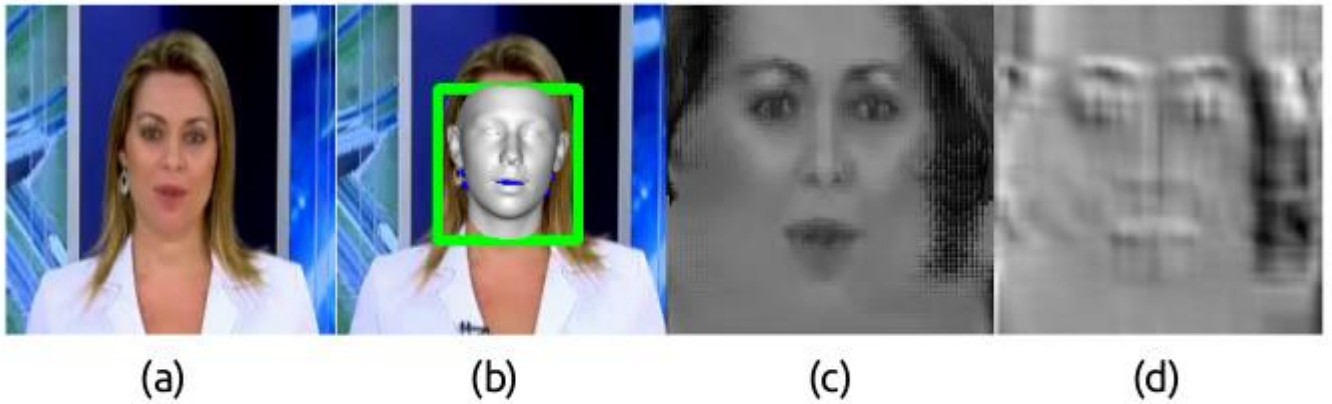
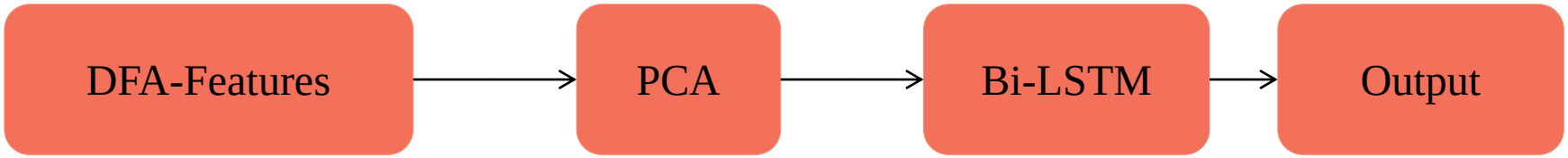


Figure 4: Illustrations of different feature types. (a) raw sequence image. (b) 3D vertices mask of a facial surface. (c) PAF of a facial surface. (d) PAC of a facial surface.

DFA-LSTM



$$J(\boldsymbol{W}, \boldsymbol{Z}) = \| \boldsymbol{X} - \boldsymbol{W}\boldsymbol{Z} \|_F^2$$

$$\mathbf{Z} \in \mathbb{R}^{L \times N} \; \boldsymbol{W} \in \mathbb{R}^{D \times L}$$

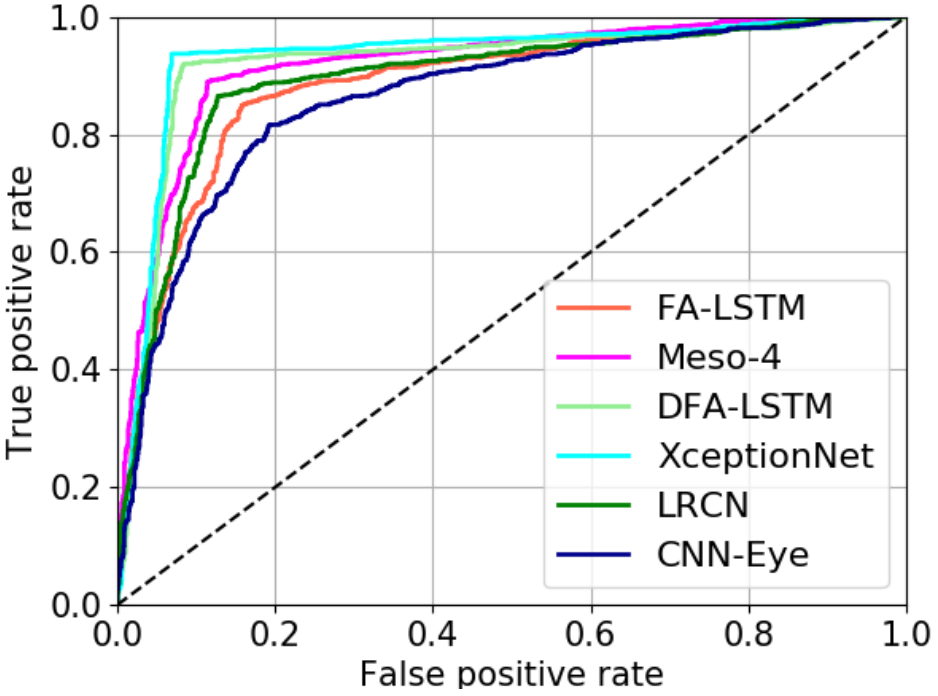
$$\begin{cases} \boldsymbol{\Sigma} = 1/N \sum_{i=1}^{N} \boldsymbol{x}_i \boldsymbol{x}_i^T \\ \boldsymbol{\Sigma} \boldsymbol{w}_i = \lambda_i \boldsymbol{w}_i. \end{cases}$$

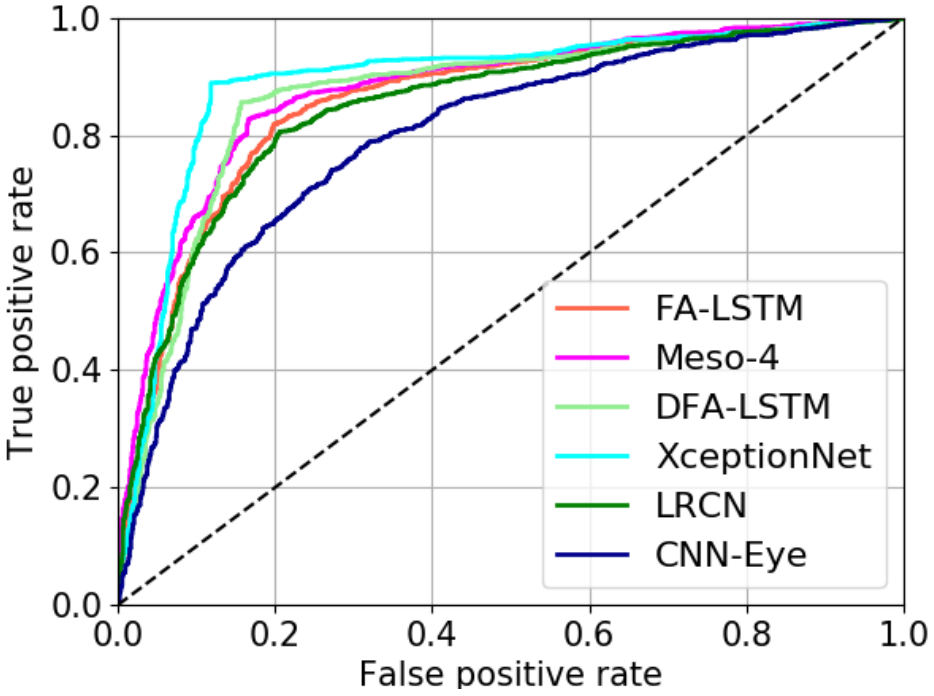
Wi , $0 \leq i \leq N$ is computed as an eigenvector of the covariance matrix Σ with the eigenvalue λi

Loss Function: $$\mathcal{L}(\widehat{y_i}, y_i) = -w_i [y_i \log \widehat{y_i} + (1 - y_i) \log(1 - \widehat{y_i})]$$

Experimental Results

ROC curve for forgeries detection method on FaceForensics++ dataSet.
Compression Level: 23 and 40





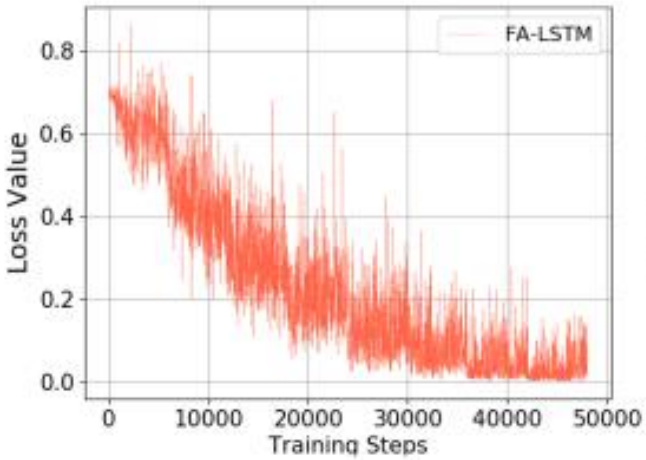
Experimental Results

- Inference Speed and Accuracy Test

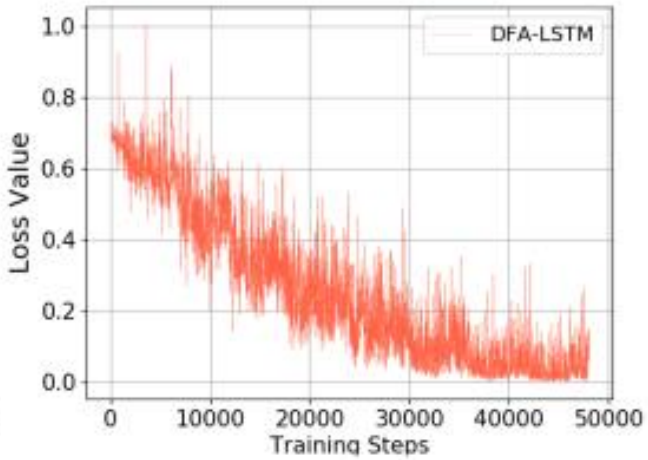
Methods	**CPU**	**GPU**
Meso-4[Afchar *et al.*, 2018]	43.251ms	16.344ms
XceptionNet[Chollet, 2017]	166.021ms	38.130ms
LRCN[Li *et al.*, 2018]	38.260ms	13.974ms
CNN-Eye[Kim *et al.*, 2017]	18.362ms	6.611ms
FA-LSTM	**16.125ms**	**5.314ms**
DFA-LSTM	**37.320ms**	**10.253ms**

Methods	**Accuracy**		**Runtime**	
	C_23	C_40	CPU	GPU
FA-LSTM	0.825	0.689	16.125ms	5.314ms
FA-LSTM + Attention	**0.879**	**0.721**	**20.025ms**	**7.221ms**
Meso-4[Afchar *et al.*, 2018]	0.830	0.702	43.251ms	16.344ms

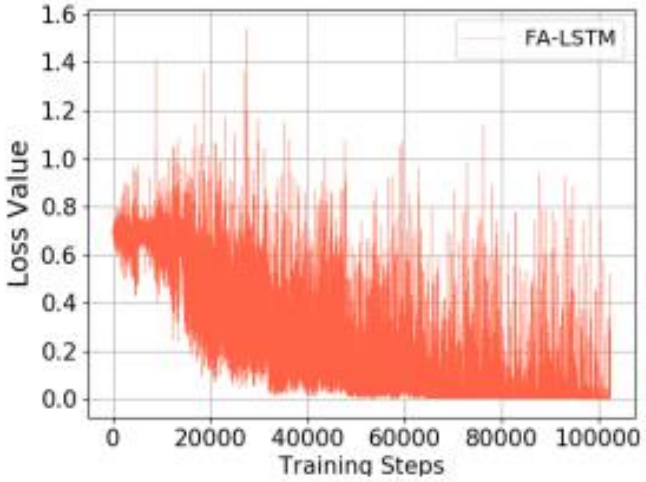
Experimental Results



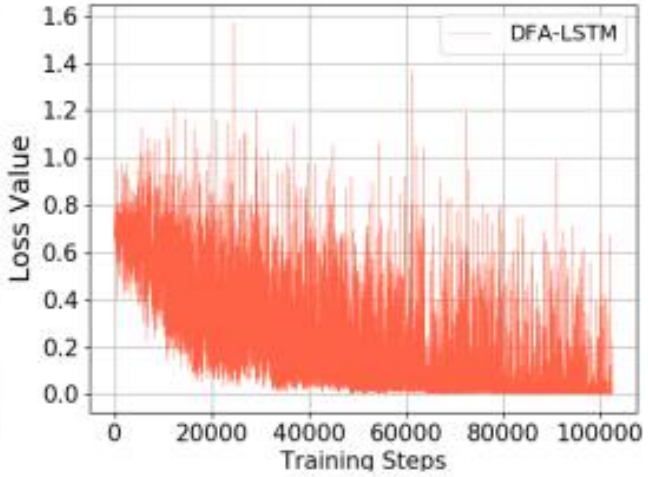
(a) FA-LSTM, BatchSize=8



(b) DFA-LSTM, BatchSize=8



(c) FA-LSTM, BatchSize=1



(d) DFA-LSTM, BatchSize=1

Training Convergence
Analysis for DFA/FA LSTM Metho

Experimental Results

- Visual Effect Demo for Facial Forgery Detection

Example sequences of failure dete
Misclassification samples of FA-I
#2, #4, #6: Improved results of #1
attention layer. #1, #2 are forged s
original samples.

THANKS