Face-Focused Cross-Stream Network for Deception Detection in Videos – Supplementary Material –

Mingyu Ding[1,*] An Zhao[1,*] Zhiwu Lu[1,†] Tao Xiang[2,3] Ji-Rong Wen[1]
[1]Beijing Key Laboratory of Big Data Management and Analysis Methods
School of Information, Renmin University of China, Beijing 100872, China
[2]Department of Electrical and Electronic Engineering, University of Surrey, UK
[3]Samsung AI Centre, Cambridge, UK
zhiwu.lu@gmail.com t.xiang@surrey.ac.uk

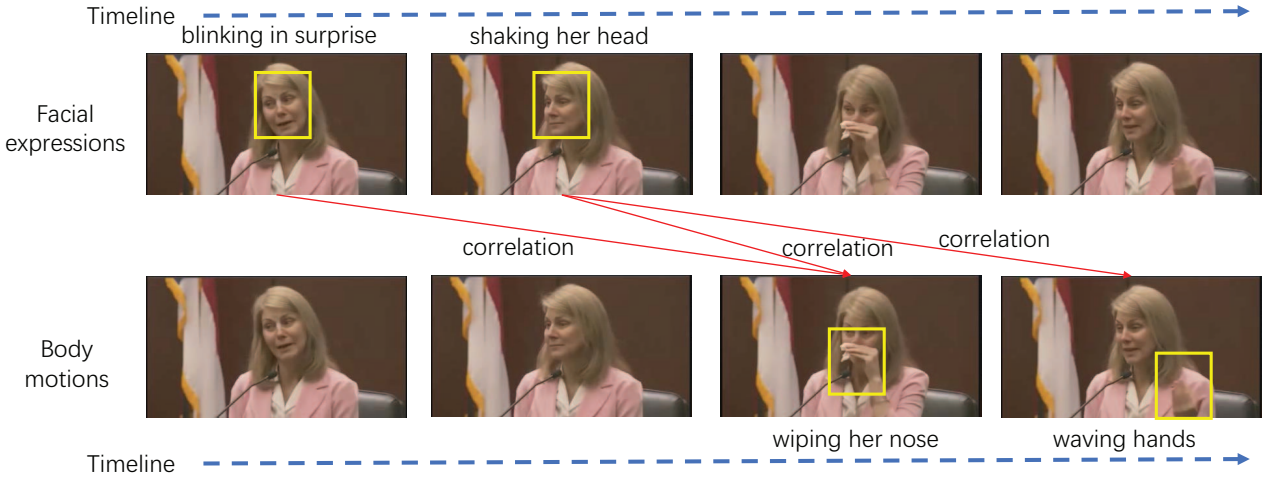

Figure 1. One example to show the asynchronization between facial expressions and body motions in deception detection.

1. More Examples for Asynchronization in Deception Detection

In Figure 1 of the main paper, we have presented an example to show the asynchronization between facial expressions and body motions in deception detection. In this suppl. material, two more examples are presented in Figures 1 and 2. We can see that there indeed exists asynchronization between facial expressions and body motions. In this work, a novel face-focused cross-stream network (FFCSN) model is proposed to overcome this challenge.

2. Meta Learning Results for Deception Detection

In the experiment part of the main paper, we have focused on evaluating the performance of deception detection in itself. Given that meta learning is utilized to address the data scarcity problem for training our FFCSN model, we directly evaluate the performance of meta learning as follows. First, we generate 408 quintuple pairs from the test set, where each quintuple pair includes one deceptive-deceptive binary pair and four deceptive-truthful binary pairs. Second, given the ground-truth

*Equal contribution.
†Corresponding author.

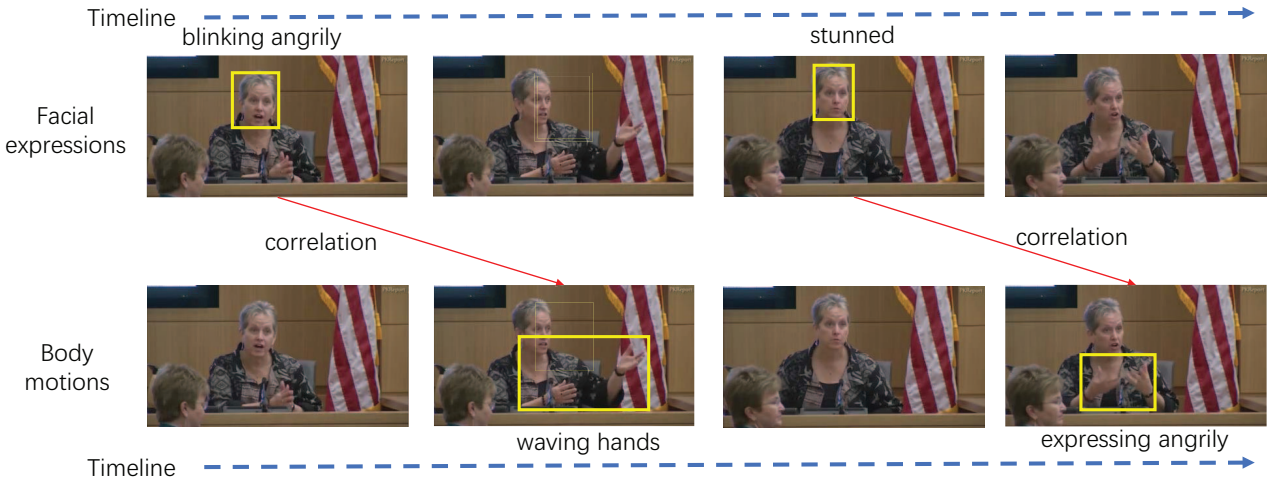

Figure 2. Another example to show the asynchronization between facial expressions and body motions in deception detection.

scores $[gs_1, ..., gs_5] \in \{0, 1\}^5$ and the predicted scores $[ps_1, ..., ps_5]$ for each quintuple pair, the accuracy for meta learning over all quintuple pairs is defined as follows:

$$\text{ACC} = \#(\arg\max_j gs_j = \arg\max_j ps_j)/N_t. \tag{1}$$

In this work, we have $N_t = 408$, and then compute the accuracy for meta learning as: ACC = **91.42%**. This result shows that our FFCSN model can also achieve a high accuracy for meta learning in deception detection.

3. Per-Class Accuracies for Emotion Recognition

We provide the per-class accuracies for emotion recognition in Table 1. It can be seen that our model not only achieves the highest overall accuracy, but also outperforms the state-of-the-art models on 7 out of 8 emotion classes.

Model	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust	Overall
[1]	53.0	7.6	44.6	47.3	48.3	20.0	76.9	28.5	46.1
[2]	48.5	0.0	**53.8**	52.7	54.2	32.4	78.7	**43.8**	51.1
Ours	**63.7**	**32.6**	52.1	**58.6**	**57.4**	**68.8**	**85.6**	**43.8**	**57.8**

Table 1. Per-class accuracies (%) for emotion recognition from videos on the YouTube-8 dataset.

References

[1] Yu-Gang Jiang, Baohan Xu, and Xiangyang Xue. Predicting emotions in user-generated videos. In *AAAI*, volume 14, pages 73–79, 2014. 2

[2] Lei Pang, Shiai Zhu, and Chong-Wah Ngo. Deep multimodal learning for affective analysis and retrieval. *IEEE Trans. Multimedia*, 17(11):2008–2020, 2015. 2