



DFEW: A Large-Scale Database for Recognizing Dynamic Facial Expression Recognition in the Wild

Xingxun Jiang[#], Yuan Zong[#], Wenming Zheng^{*}, Chuangao Tang, Wanchuang Xia, Cheng Lu, Jiateng Liu
Southeast University, Nanjing, China

{jiangxingxun, xhzongyuan, wenming_zheng, tcg2016, xiawanchuang, cheng.lu, Jiateng_Liu}@seu.edu.cn

1. Introduction

- **Facial Expression Recognition (FER)** is important, but development limited in the real world, or **in-the-wild** condition.
- **Databases** are investigated, we lack a **large-scale well-annotated** dynamic facial expression database.
- To solve this problem, we present our database, **Dynamic Facial Expression in-the-Wild**, called **DFEW** for short.

3. EC-STFL

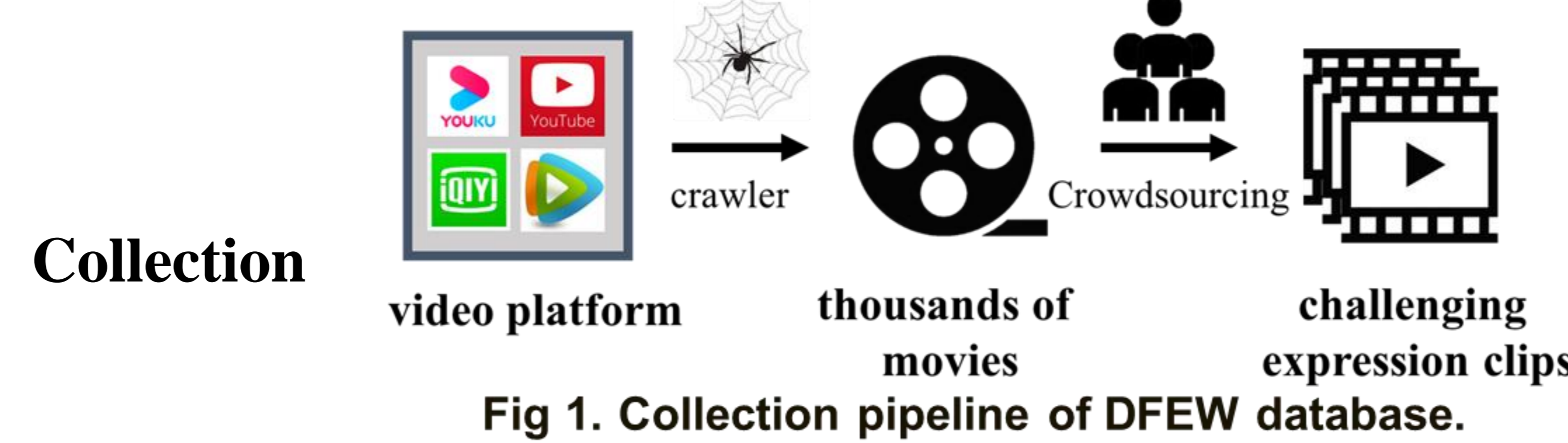
- To clarify the blurred margins of unconstrained faces, and alleviate the unbalanced problem, we proposed **EC-STFL**.

$$L = L_s + \lambda L_{EC-STFL}$$

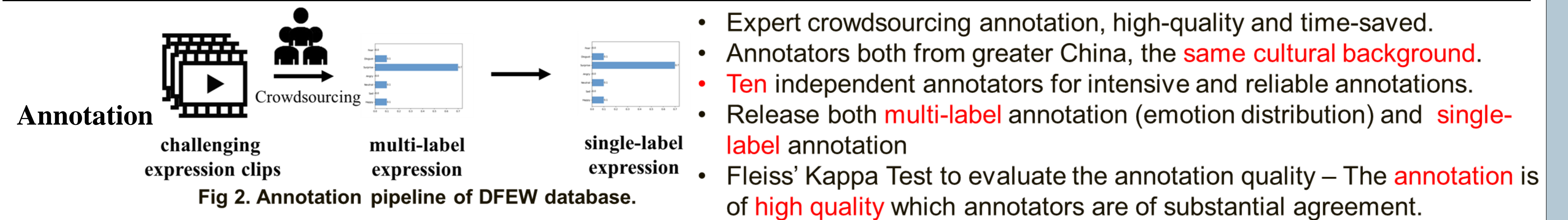
$$L_{EC-STFL} = \frac{\sum_{1 \leq i, j \leq n, x_i \in N\{x_i\}} \frac{\|x_i - x_j\|}{N_{x_i}}}{\sum_{1 \leq i, j \leq n, x_j \notin N\{x_i\}} \frac{\|x_i - x_j\|}{N_{x_j}}}$$

- L_s : Softmax loss
- $N\{x_i\}$: the same labeled set of sample x_i in mini-batch.
- N_{x_i} : the set size of $N\{x_i\}$.
- n : mini-batch size.

2. DFEW Database



- Clips from movies to mimic our real life
- **1500+** high-definition movies
- Extract clips **manually** for accurate samples
- Extract at most **20** clips each movie
- Pre-annotation: Check clips whether containing one of the seven typical emotions.
- **Additional reward** for rare expression samples, i.e., disgust and fear.



- Expert crowdsourcing annotation, high-quality and time-saved.
- Annotators both from greater China, the **same cultural background**.
- **Ten** independent annotators for intensive and reliable annotations.
- Release both **multi-label** annotation (emotion distribution) and **single-label** annotation
- Fleiss' Kappa Test to evaluate the annotation quality – The **annotation** is of **high quality** which annotators are of substantial agreement.

Table 1: Summary of existing databases of dynamic facial expression in the wild.

Database	#Sample	Source	Expression Distribution	#Annotation Times	Available?
Aff-Wild	298	Web	Valence-arousal	8	Yes
AFEW 7.0	1,809	54 Movies	7 basic expressions	2	Yes
AFEW-VA	600	AFEW database	Valence-arousal	2	Yes
CAER	13,201	79 TVshows	7 basic expressions	3	Yes
DFEW	16,372	1500 movies	7 basic expressions	10	Yes

Largest ! Largest ! Largest !

Table 2: The basic information of single-labeled DFEW.

Emotions	Clips			Total	Percent
	0-2s	2-5s	5s+		
Happy	852	1252	384	2488	20.63
Sad	440	915	653	2008	16.65
Neutral	832	1335	542	2709	22.46
Angry	762	1091	376	2229	18.48
Surprise	691	648	159	1498	12.42
Disgust	71	58	17	146	1.22
Fear	408	435	138	981	8.14
Total	4056	5734	2269	12059	100.00

4. Experiments and Results

- **Protocol: 5-fold cross-validation**
- **Evaluation metrics: weighted average recall (WAR) and unweighted average recall (UAR).**
- **Benchmark Experiments: C3D, P3D, R3D18, 3D Resnet18, I3D-RGB, VGG11+LSTM, VGG11+LSTM.**
- **EC-STFL Experiments: 1. compare with center loss. 2. hyper-parameter discussion.**
- **Transfer Task: From action database/ DFEW database to AFEW database.**

Table 3: Benchmark results of single-labeled DFEW.

Model	Emotions							Metric	
	Happy	Sad	Neutral	Angry	Surprise	Disgust	Fear	UAR	WAR
C3D	75.17	39.49	55.11	62.49	45.00	1.38	20.51	42.74	53.54
P3D	74.85	43.40	54.18	60.42	50.99	0.69	23.28	43.97	54.47
R3D18	79.67	39.07	57.66	50.39	48.26	3.45	21.06	42.79	53.22
3D Resnet18	73.13	48.26	50.51	64.75	50.10	0.00	26.39	44.73	54.98
I3D-RGB	78.61	44.19	56.69	55.87	45.88	2.07	20.51	43.40	54.27
VGG11+LSTM	76.89	37.65	58.04	60.70	43.70	0.00	19.73	42.39	53.70
Resnet18+LSTM	78.00	40.65	53.77	56.83	45.00	4.14	21.62	42.86	53.08

Table 4: The results of EC-STFL Experiments.

Model	Emotions							Metric	
	Happy	Sad	Neutral	Surprise	Disgust	Fear	UAR	WAR	
C3D	75.17	39.49	55.11	62.49	45.00	1.38	20.51	42.74	53.54
C3D, center loss	75.62	44.67	54.18	63.14	42.21	2.07	22.17	43.44	54.17
C3D, EC-STFL	75.87	49.26	54.81	61.53	45.95	3.45	24.83	45.10	55.50
3D Resnet18	73.13	48.26	50.51	64.75	50.10	0.00	26.39	44.73	54.98
3D Resnet18, center loss	78.49	44.30	54.89	58.40	52.35	0.69	25.28	44.91	55.48
3D Resnet18, EC-STFL	79.18	49.05	57.85	60.98	46.15	2.76	21.51	45.35	56.51

Table 5: The results of Transfer Task.

Pretrained	Finetuned models			
	C3D	C3D, EC-STFL	3D Resnet18	3D Resnet18, EC-STFL
Sports 1M	41.78	44.91	-	-
UCF101	41.25	42.34	-	-
Kinect700	-	-	49.35	49.61
Kinect700+Moments In Time	-	-	49.35	49.35
DFEW, fd2	44.91	45.56	53.00	53.26
DFEW, fd5	49.87	49.87	49.61	49.66

Better!

Conclusion and Discussion

- 1) We present a large-scale **Dynamic Facial Expression in-the-Wild** database, called **DFEW** for short.
- 2) We give the **benchmark** of DFEW, and proposed **EC-STFL** to improve them. Experiments show the stability.
- 3) Transfer tasks verify the **necessity** of DFEW database.

