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

Xingxun Jiang*, Yuan Zong*, Wenming Zheng#, Chuangao Tang,
Wanchuang Xia, Cheng Lu, Jiateng Liu

Southeast University

Outline

- Introduction
- DFEW Database
- EC-STFL
- Experiments
- Conclusion

Outline

- Introduction
- DFEW Database
- EC-STFL
- Experiments
- Conclusion

Introduction - Background

Expression of Emotion = 7% Text + 38% Voice + **55%** Facial Expression & Action

Facial Expression Recognition is an important topic!

Polygraph

Social Interaction

Education and health

Introduction - In-the-lab V.S. In-the-wild

(a) in-the-lab

Research Interest

Challenges
- occlusion
- illumination
- pose

(b) in-the-wild

Introduction - In-the-wild Database

Database	Modality	#Sample	Expression Distribution	Annotation Time	From
EmotioNet	image	1,000,000	23 emotions	Automatically based on AU	Web
AffectNet	image	450,000 (labeled)	8 basic expressions & Valence-Arousal	1	Web
RAF-DB	image	29,672	7 basic expressions	About 40 Times	Web
CAER-S	image	70,000	7 basic expressions	3	79 TV shows
Aff-Wild	clip	298	Valence-Arousal	8	Web
AFEW 7.0	clip	1,809	7 basic expressions	2	54 Movies
AFEW-VA	clip	600	Valence-Arousal	2	AFEW database
CAER	clip	13,201	7 basic expressions	3	79 TV shows

Lack Large-scale well-annotated Dynamic Facial Expression Database!

Introduction - EmotiW Competition

Competition	Rank	Accuracy	Database	# Sample	Train/Val/Test
EmotiW 2019	Champion	62.78%	AFEW 7.0	1,809	773/383/653
EmotiW 2018	Champion	61.87%	AFEW 7.0	1,809	773/383/653
EmotiW 2017	Champion	60.34%	AFEW 7.0	1,809	773/383/653
EmotiW 2016	Champion	59.02%	AFEW 6.0	1,749	773/383/593

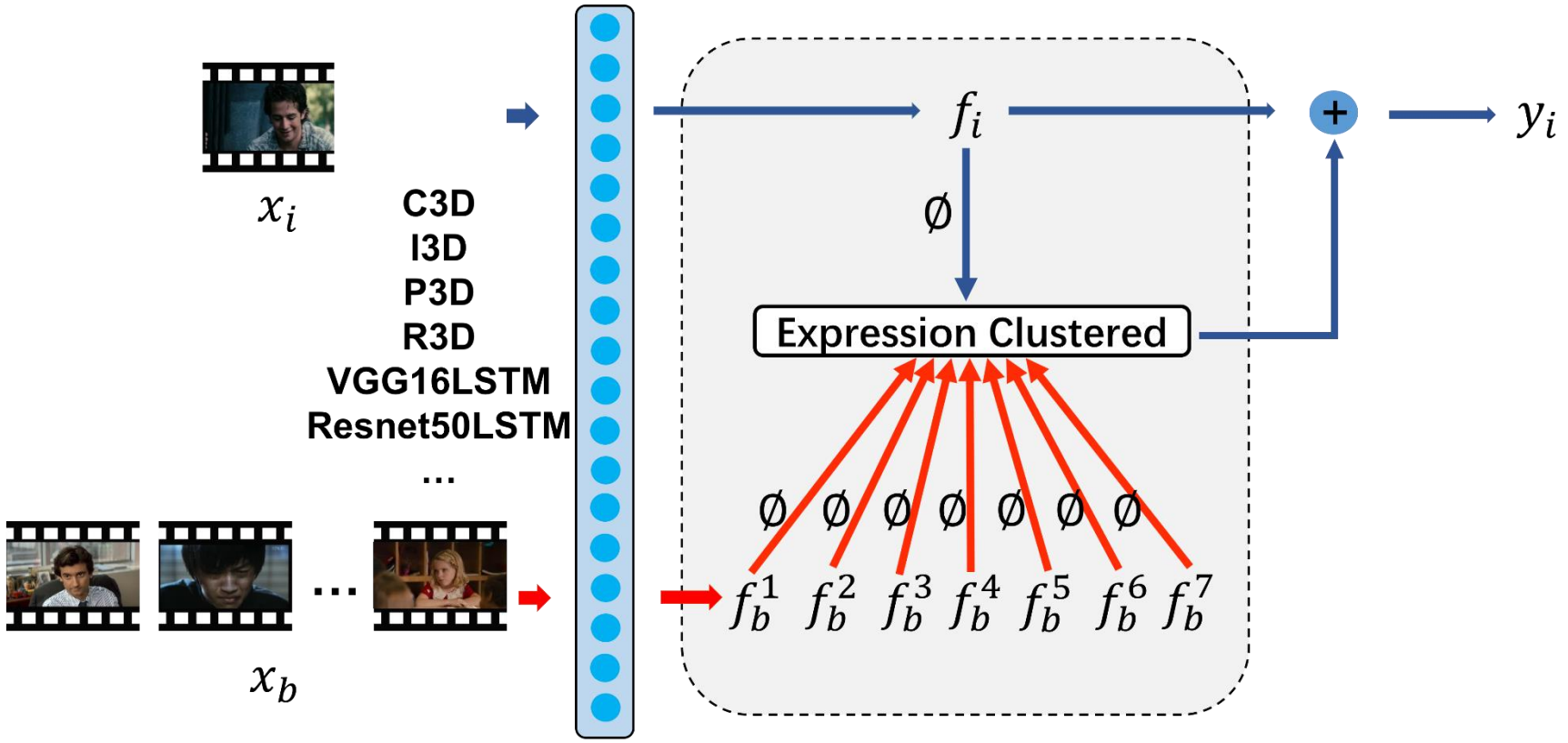
Champion of the 7th Audio-Video based FERW, EmotiW2019 : accuracy only 62.78%!

Lack Large-scale well-annotated Dynamic Facial Expression Database!

Introduction - DFEW Database

Database	Modality	#Sample	Expression Distribution	Annotation Time	From
EmotioNet	image	1,000,000	23 emotions	Automatically based on AU	Web
AffectNet	image	450,000 (labeled)	8 basic expressions & Valence-Arousal	1	Web
RAF-DB	image	29,672	7 basic expressions	About 40 Times	Web
CAER-S	image	70,000	7 basic expressions	3	79 TV shows
Aff-Wild	clip	298	Valence-Arousal	8	Web
AFEW 7.0	clip	1,809	7 basic expressions	2	54 Movies
AFEW-VA	clip	600	Valence-Arousal	2	AFEW database
CAER	clip	13,201	7 basic expressions	3	79 TV shows
DFEW	clip	16,372	7 basic expressions	10	1500 movies
		Largest!		Largest!	Largest!

Introduction - EC-STFL Loss

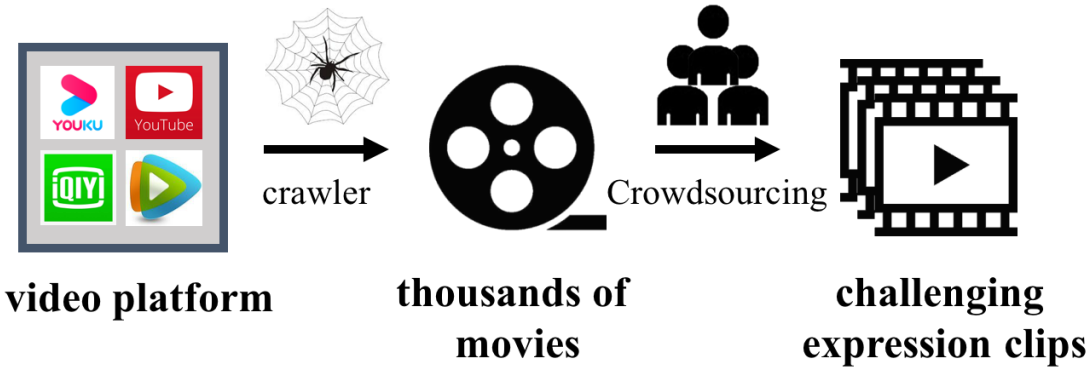


EC-STFL : Expression Clustered - SpatioTemporal Feature Learning

Outline

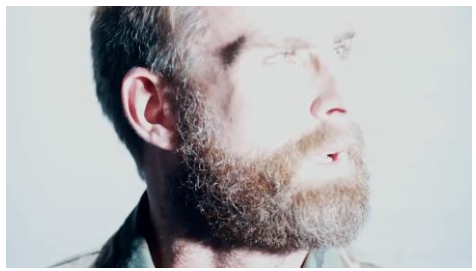
- Introduction
- **DFEW Database**
- EC-STFL
- Experiments
- Conclusion

DFEW Database - Collection

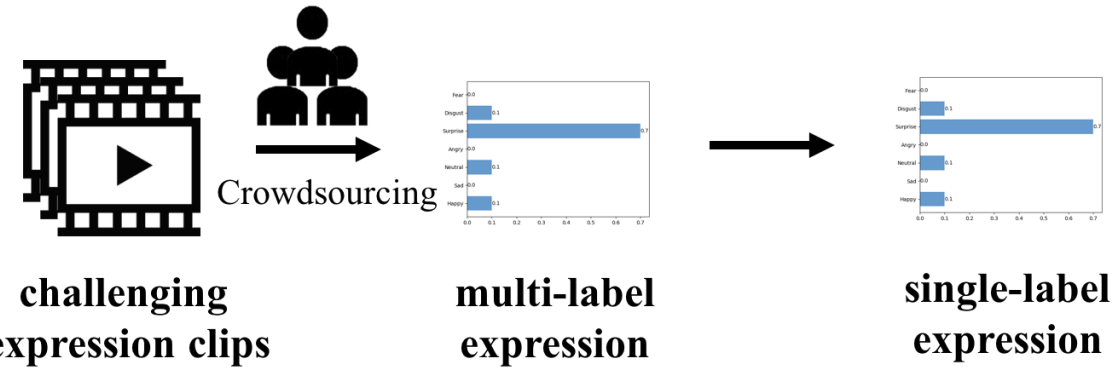


- Clips from movies to mimic our real life
- 1500+ high-definition movies
- Extract clips manually for accurate samples
- Extract at most 20 clips each movie
- Pre-annotation: Check clips whether containing one of the seven typical emotions.
- Additional reward for rare expression samples, i.e., disgust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



DFEW Database - Annotation



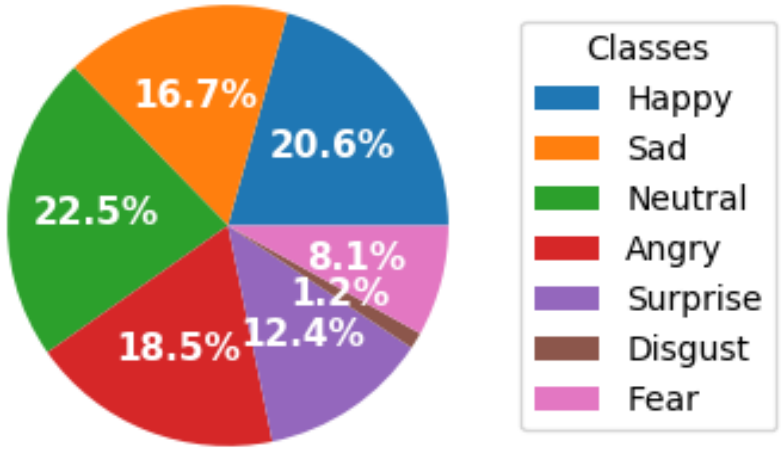
- Expert crowdsourcing annotation, high-quality and time-saved.
- Annotators both from greater China, the same cultural background.
- Ten independent annotators for intensive and reliable annotations.
- Release both multi-label annotation (emotion distribution) and single-label annotation

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 !**

DFEW Database - Annotation

Emotions	Clips				Percent
	0-2s	2-5s	5s+	Total	
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



Single-labeled:

- **Selected** from the multi-labeled, i.e., all 16,372 clips with 7-dim emotion ground truth.
- At least **6** annotators (**10** totally) believe this clip belong to one specific emotion.

DFEW Database – Agreement Test

Fleiss's Kappa test:

to discuss the annotation's quality

Order	Happy	j-th Emotion	…	Fear
1	10	0	…	0
2	8	1	…	0
i-th	0	0	…	3
…	…	…	…	…

$$\begin{cases} p_j = \frac{1}{N \times n} \sum_{i=1}^{N} n_{ij} \\ \sum_{j=1}^{K} p_j = 1 \end{cases}$$

$$P_i = \frac{1}{n \times (n-1)} \left[\left(\sum_{j=1}^{K} n_{ij}^2 \right) - n \right]$$

$$\bar{P} = \frac{1}{N} \sum_{i=1}^{N} P_i \qquad \bar{P_e} = \sum_{j=1}^{k} p_j^2$$

$$\kappa = \frac{\bar{P} - \bar{P_e}}{1 - \bar{P_e}}$$

n_{ij} : the number of annotators who assigned the i-th clip and the j-th emotion.

p_j : the proportion of all assignments which were to the j-th emotion.

DFEW Database – Agreement Test

Fleiss's Kappa test:

to discuss the annotation's quality

Order	Happy	j-th Emotion	…	Fear
1	10	0	…	0
2	8	1	…	0
i-th	0	0	…	3
…	…	…	…	…

κ	Interpretation
<0	Poor agreement
0.01-0.20	Slight agreement
0.21-0.40	Fair agreement
0.41-0.60	Moderate agreement
0.61-0.80	Substantial agreement
0.81-1.00	Almost perfect agreement

n_{ij} : the number of annotators who assigned the i-th clip and the j-th emotion.

p_j : the proportion of all assignments which were to the j-th emotion.

Types	κ
Single-labeled DFEW	0.63
Multi-labeled DFEW	0.70

DFEW Database - Demo Samples

Happy

Sad

Neutral

Angry

Surprise

Disgust

Fear

Outline

- Introduction
- DFEW Database
- **EC-STFL**
- Experiments
- Conclusion

Experiments - EC-STFL Experiments

EC-STFL Loss: **E**xpression **C**lustered **S**patio**T**emporal **F**eature **L**earning

Target:

$$\min_{W} \sum_{i,j} \frac{P_{ij}\phi(x_i, x_j)}{Q_{ij}\phi(x_i, x_j)}$$

$$\phi(x_i, x_j) = \|x_i - x_j\|$$

$$P_{ij} = \begin{cases} 0, & \text{if } x_i \text{ and } x_j \text{ has the same label} \\ 1, & \text{otherwise} \end{cases}$$

$$Q_{ij} = \begin{cases} 0, & \text{if } x_i \text{ and } x_j \text{ has the different label} \\ 1, & \text{otherwise} \end{cases}$$

$x \in \mathbb{R}^d$: extracted from the final hidden fully connected layers

Loss:

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

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

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.

Outline

- Introduction
- DFEW Database
- EC-STFL
- **Experiments**
- Conclusion

Experiments - Experimental Setup

Data Protocol:

- For single-labeled DFEW database with 12,059 video clips.
- Using 5-fold cross-validation protocol for single-labeled benchmarks.

Evaluation:

- UAR: Unweighted Average Recall, i.e., the accuracy per class divided by the number of classes without considerations of instances per class.
- WAR: Weighted Average Recall, i.e., accuracy

Preprocessing:

- Acquire Face region and landmarks: face++ API
- Remove the non-face/undetected frames: manually
- Remove the clips which useful frames less than 50%: remove 362 clips
- Face affine transformation: Seetaface toolbox and face++ landmarks
- Fixed temporal length: Time Interpolation method

Experiments - Baseline Experiments

Model	Emotions							Metric	
	Happy	Sad	Neutral	Angey	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

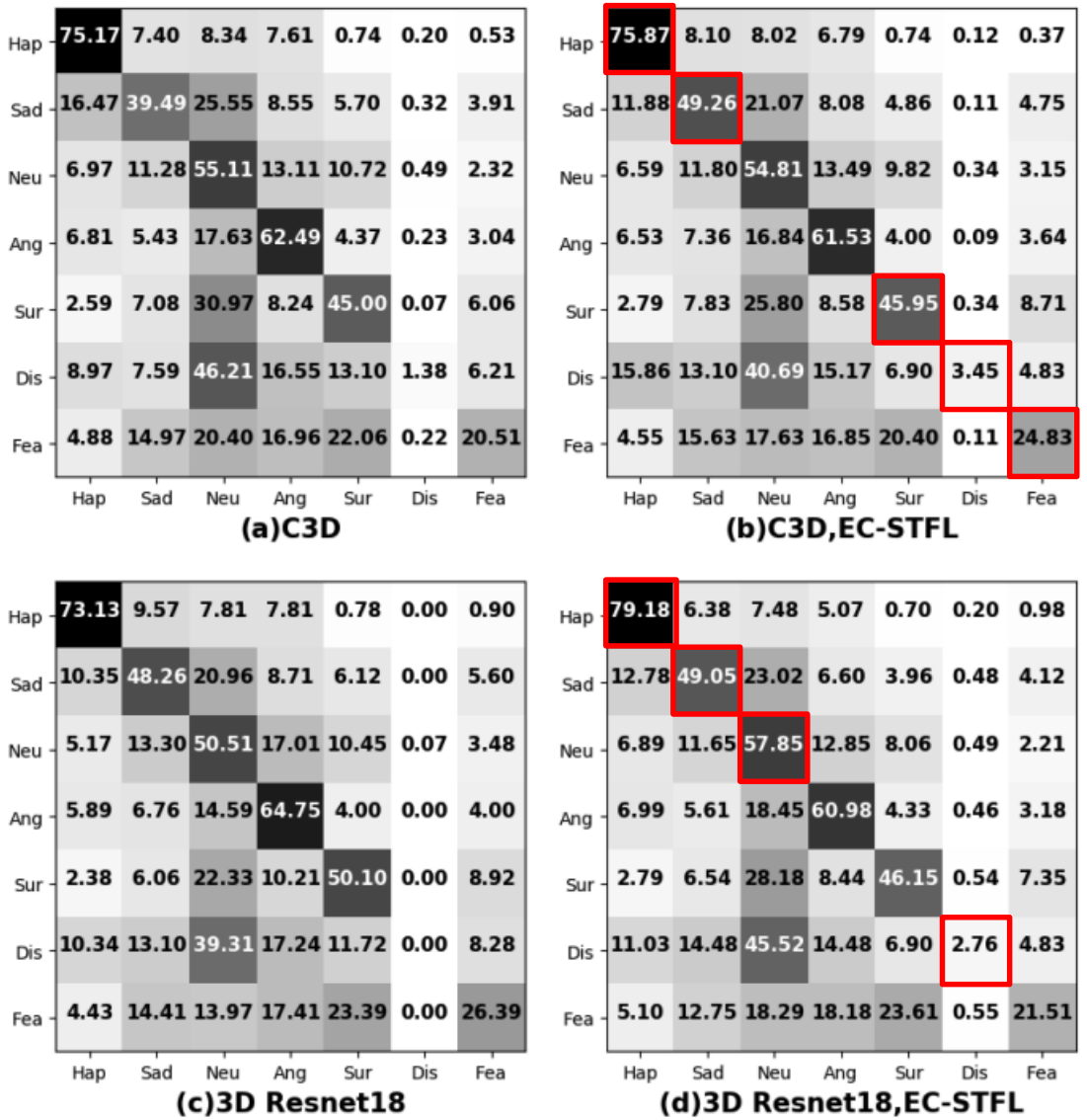
Get start easily:

some baseline for the evaluation of methods

Experiments - EC-STFL Experiments

Model	Metric	
	UAR	WAR
C3D	42.74	53.54
C3D,EC-STFL	**45.10**	**55.50**
P3D	43.97	54.47
P3D,EC-STFL	**45.22**	**56.48**
R3D18	42.79	53.22
R3D18,EC-STFL	**45.05**	**56.19**
3D Resnet18	44.73	54.98
3D Resnet18,EC-STFL	**45.35**	**56.51**
I3D-RGB	43.40	54.27
I3D-RGB,EC-STFL	**45.05**	**56.19**
VGG11+LSTM	42.39	53.70
VGG11+LSTM,EC-STFL	**44.78**	**56.25**
Resnet18+LSTM	42.86	53.08
Resnet18+LSTM,EC-STFL	**43.60**	**54.72**

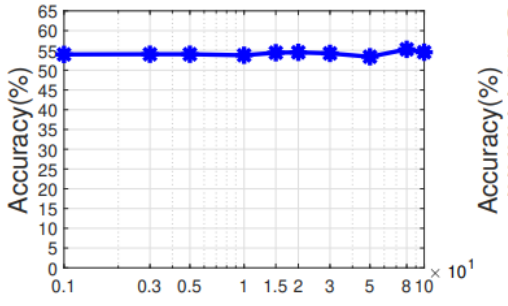
Better!



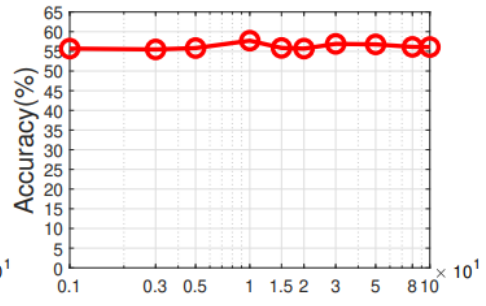
Experiments - EC-STFL Experiments

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
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 Renset18	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**

Better !

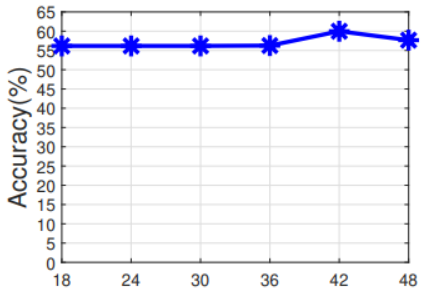


(a) C3D,EC-STFL

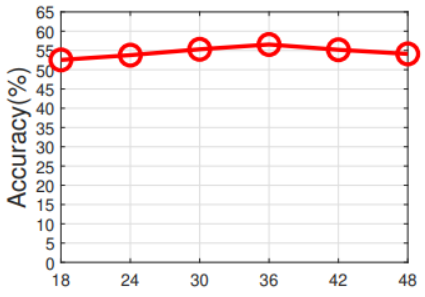


(b) 3DResnet18,EC-STFL

Perform largely stable at hyper-parameter λ !



(a) C3D,EC-STFL



(b) 3DResnet18,EC-STFL

Perform largely stable at batch size m !

Experiments - Transfer Learning 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!

DFEW, fd2: used the pre-trained models trained on the second data split.

Transfer Learning:

- From action database / DFEW database to AFEW database
- Initializing models with weights trained from source database, then go on training and test models on AFEW

Purpose:

- Verify the **necessity** of DFEW database for developing excellent emotion prediction models in real-life applications.

Outline

- Introduction
- DFEW Database
- EC-STFL
- Experiments
- Conclusion

Conclusion

- We present a new large-scale unconstrained Dynamic Facial Expression database in-the-Wild, DFEW.
  - 16372 video clips from over 1500 different movies.
  - Reliable distribution information of 7 basic expression annotated by 10 annotators, release both multi-labeled and single-labeled annotation.
- We propose a new EC-STFL loss to improve the performance of FERW.
- We conduct extensive experiments on DFEW
  - Extensive baseline experiments as well as EC-STFL to get DFEW database started easily.
  - Transfer tasks to verify the necessity of DFEW database.

Thanks for Listening !

DFEW project page

Full paper

Email: jiangxingxun@seu.edu.cn
Homepage: https://jiangxingxun.github.io/