Tutorial

ICCV 2020

Recent Advances and Challenges in Facial Micro-Expression Analysis

Challenges & Future Directions

John See Multimedia University, Malaysia

Moi-Hoon Yap Manchester Metropolitan University, UK

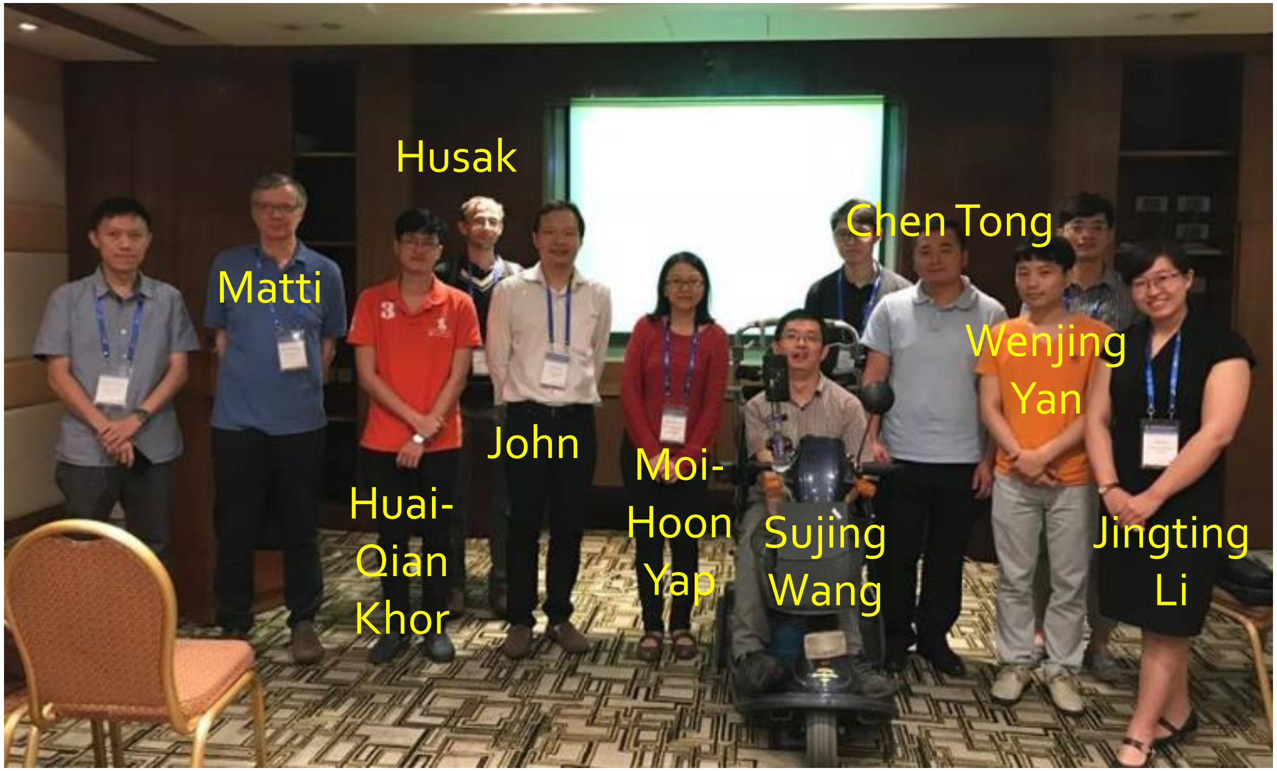
Su-Jing Wang Chinese Academy of Sciences, China

Jingting Li Chinese Academy of Sciences, China

Sze-Teng Liong Feng Chia University, Taiwan

1st Micro-Expression Grand Challenge 2018 @ FG 2018

So, we decided that we should meet up and have a "real-world" look at each other's expressions…



Xi'an, China
May 2018

1st Micro-Expression Grand Challenge 2018 @ FG 2018, Xi'an, China

- 5 "Objective classes" (grouped by Facial AU) instead of emotion classes
- Cross-database protocols
  - **Holdout Database Evaluation (HDE)**
    - Train on one dataset, Test on the other. Swap, repeat. (WAR, UAR)
  - **Composite Database Evaluation (CDE)**
    - Combine both datasets, evaluate by LOSO (F1-score)

Class	CASME II	SAMM	Composite
I	25	24	49
II	15	13	28
III	99	20	119
IV	26	8	34
V	20	3	23
Total	185	68	253

Facial Micro-Expressions Grand Challenge 2018 Summary

Moi Hoon Yap[1], John See[2], Xiaopeng Hong[3], Su-Jing Wang[4]
[1] School of Computing, Mathematics and Digital Technology, Manchester Metropolitan University, UK
[2] Faculty of Computing and Informatics, Multimedia University, Malaysia
[3] Center for Machine Vision and Signal Analysis, University of Oulu, Finland
[4] CAS Key Laboratory of Behavioral Science, Institute of Psychology, Beijing, 100101, China

***Abstract*—This paper summarises the Facial Micro-Expression Grand Challenge (MEGC 2018) held in conjunction with the 13th IEEE Conference on Automatic Face and Gesture Recognition (FG) 2018. In this workshop, we aim to stimulate new ideas and techniques for facial micro-expression analysis by proposing a new cross-database challenge. Two state-of-the-art datasets, CASME II and SAMM, are used to validate the performance of existing and new algorithms. Also, the challenge advocates the recognition of micro-expressions based on AU-centric objective classes rather than emotional classes. We present a summary and analysis of the baseline results using LBP-TOP, HOOF and 3DHOG, together with results from the challenge submissions.**

inconsistencies adds further justification for the introduction of new classes based on AUs only [3].

This challenge aims to stimulate the micro-expressions researchers in developing new techniques for the AU-centric objective classes. A summary of the objective classes are as illustrated in Table I. A single composite database for this experiment has a total of 253 micro-expressions.

TABLE I
THE TOTAL NUMBER OF MOVEMENTS ASSIGNED TO THE NEW OBJECTIVE CLASSES FOR CASME II AND SAMM.

1st Micro-Expression Grand Challenge 2018 @ FG 2018, Xi'an, China

THE RESULTS OF HOLDOUT-DATABASE EVALUATION (TASK A).

Method	WAR			UAR		
	@SAMM	@CASME II	Average	@SAMM	@CASME II	Average
LBP-TOP	0.338	0.232	0.285	0.327	0.316	0.322
3DHOG	0.353	0.373	0.363	0.269	0.187	0.228
HOOF	0.441	0.265	0.353	0.349	**0.346**	0.348
Peng et al.	**0.544**	**0.578**	**0.561**	**0.440**	0.337	**0.389**
Khor et al.	0.485	0.384	0.435	0.382	0.322	0.352

THE RESULTS OF COMPOSITE DATABASE EVALUATION (TASK B) BASED ON LOSO CROSS VALIDATION.

Method	F1-Score	Weighted F1-score
LBP-TOP	0.400	0.524
3DHOG	0.271	0.436
HOOF	0.404	0.527
Peng at al.	**0.639**	**0.733**
Merghani et al.	0.454	0.579
Khor et al.	0.393	0.523

- UAR results were very close
- 6 papers accepted (50%) – 3 challenge, 3 non-challenge

Transfer learning of macro-trained deep models

- Train deep models on macro-expression apex samples ➔ Transfer learning on micro-expression apex samples
  - **ResNet10** pre-trained on 4 macro-exp. datasets using apex frames

    CK+ (852 images)

    Oulu CASIA NIR & VIS (1200 images)

    Jaffe (151 images)

    MUGFE (8228 images)

    TOTAL: 10,431 images → oversample to 5,000 images/expression
  - Fine-tuning on micro-exp datasets using apex frames → oversample to 200 images/expression
  - Assumption: That apex information is available!

	Hold-database Evaluation (HDE)	
	WAR	*UAR*
LBP-TOP	0.285	0.332
HOOF	0.353	0.348
HOG3D	0.363	0.228
Our method	**0.561**	**0.389**

TABLE VII. RECOGNITION ACCURACY AND F1 SCORE OF OUR METHODSON THE COMPOSITE DATASETS IN CDE

	Leave-One-subject-Out (LOSO)	
	Accuracy (%)	*F1 score*
Our method	**74.70**	**0.64**

Peng, M., Wu, Z., Zhang, Z., & Chen, T. (2018). From macro to micro expression recognition: Deep learning on small datasets using transfer learning. In 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018) (pp. 657-661).

1st Micro-Expression Grand Challenge 2018 @ FG 2018, Xi'an, China

Insights:

- Cross-database task is challenging
  - Leveraging macro-expression samples seem to work reasonably well
  - Lack of data -> LSTMs not suitable
- We needed more people to work on this area, and
  - Annotation services/tools
  - Datasets

2nd Micro-Expression Grand Challenge 2019 @ FG 2019

Lille, France
May 2019

2nd Micro-Expression Grand Challenge 2019 @ FG 2019, Lille, France

- 2 Challenges were held:
  - **Spotting Challenge**
    - CAS(ME)^2 & SAMM Long videos
    - Metrics: TP, FP, FN, F1-score
    - Only 1 team participated
      - LTP-ML method: spots maximal movement based on local temporal patterns;
      - Beats baseline performance (LBP-χ^2) by a margin

TABLE II: F1-Score of LTP-ML and LBP-χ^2 for ME spotting from long videos.

Database	SAMM^c	SAMM^f	CAS(ME)^2
LTP-ML	**0.0316**	0.0229	**0.0179**
LBP-χ^2	0.0055	N/A†	0.0035

†This method requires cropped faces, so SAMM^f is not applicable.

MEGC 2019 – The Second Facial Micro-Expressions Grand Challenge

John See[1], Moi Hoon Yap[2], Jingting Li[3], Xiaopeng Hong[4], Su-Jing Wang[5]
[1] Faculty of Computing and Informatics, Multimedia University, Malaysia
[2] Faculty of Science and Engineering, Manchester Metropolitan University, UK
[3] CentraleSuplec, CNRS, IETR, UMR 6164, F-35000 Rennes, France
[4] Center for Machine Vision and Signal Analysis, University of Oulu, Finland
[5] Key Laboratory of Behavior Sciences, Institute of Psychology, Chinese Academy of Sciences, Beijing, China

Abstract— **Automatic facial micro-expression (ME) analysis is a growing field of research that has gained much attention in the last five years. With many recent works testing on limited data, there is a need to spur better approaches that are both robust and effective. This paper summarises the 2nd Facial Micro-Expression Grand Challenge (MEGC 2019) held in conjunction with the 14th IEEE Conference on Automatic Face and Gesture Recognition (FG) 2019. In this workshop, we proposed challenges for two micro-expression (ME) tasks—spotting and recognition, with the aim of encouraging rigorous evaluation and development of new robust techniques that can accommodate data captured across a variety of settings. In this paper, we outline the evaluation protocols for the two challenge tasks, the datasets involved, and an analysis of the best performing works from the participating teams, together with a summary of results. Finally, we highlight some possible future directions.**

or features are typically constructed individually for each dataset. As such, the rigour and realism in evaluation is most thoroughly lacking in current mode of practices. This edition of the recognition challenge pushes the boundaries of current evaluation schemes towards a comprehensive composite database that comprises of samples collected from different environments, from a diverse range of subjects.

This workshop aims to narrow the gaps found in these tasks and to continuously promote interactions between researchers and scholars from within this area of study, and also those from broader areas of psychology and physiology. Besides the two challenges, we also solicited original works that address a variety of challenges in the computational aspect of ME research, including that of other related fields

See, J., Yap, M. H., Li, J., Hong, X., & Wang, S. J. (2019). Megc 2019–the second facial micro-expressions grand challenge. In 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019) (pp. 1-5)

2nd Micro-Expression Grand Challenge 2019 @ FG 2019, Lille, France

- **Recognition Challenge**
  - Cross-database mode: SMIC, CASME II, SAMM are combined on the basis of 3 general classes
    - **Negative** (i.e. 'Repression', 'Anger', 'Contempt', 'Disgust', 'Fear' and 'Sadness')
    - **Positive** ('Happiness'), and
    - **Surprise** ('Surprise')

TABLE III: 3-class sample distribution of all datasets for CDE

Emotion Class	SMIC	CASME II	SAMM	3DB-combined
Negative	70	88†	92‡	250
Positive	51	32	26	109
Surprise	43	25	15	83
TOTAL	164	145	133	442

† Negative class of CASME II: Disgust and Repression.
‡ Negative class of SAMM: Anger, Contempt, Disgust, Fear and Sadness.

  - CDE mode, LOSO (68 subjects)
  - Metrics: Unweighted-F1, Unweighted Average Recall (UAR)

$$\mathrm{F1}_c = \frac{2 \cdot TP_c}{2 \cdot TP_c + FP_c + FN_c}$$

$$\mathrm{UF1} = \frac{F1_c}{C}$$

$$\mathrm{UAR} = \frac{1}{C} \sum_c \frac{TP_c}{n_c}$$

See, J., Yap, M. H., Li, J., Hong, X., & Wang, S. J. (2019). Megc 2019–the second facial micro-expressions grand challenge. In 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019) (pp. 1-5)

2nd Micro-Expression Grand Challenge 2019 @ FG 2019, Lille, France

- **Recognition Challenge**
  - 7 submissions were received (papers from the top 4 results were accepted for publication)
  - Output logs from submissions were verified
  - Teams are required to share code implementations in GitHub

Method	Full		SMIC		CASME II		SAMM	
	UF1	UAR	UF1	UAR	UF1	UAR	UF1	UAR
LBP-TOP [25]	0.5882	0.5785	0.2000	0.5280	0.7026	0.7429	0.3954	0.4102
Bi-WOOF [20]	0.6296	0.6227	0.5727	0.5829	0.7805	0.8026	0.5211	0.5139
OFF-ApexNet [26]	0.7196	0.7096	0.6817	0.6695	**0.8764**	0.8681	0.5409	0.5392
Quang et al. [19]	0.6520	0.6506	0.5820	0.5877	0.7068	0.7018	0.6209	0.5989
Zhou et al. [18]	0.7322	0.7278	0.6645	0.6726	0.8621	0.8560	0.5868	0.5663
Liong et al. [17]	0.7353	0.7605	0.6801	0.7013	0.8382	**0.8686**	0.6588	0.6810
Liu et al. [16]	**0.7885**	**0.7824**	**0.7461**	**0.7530**	0.8293	0.8209	**0.7754**	**0.7152**

See, J., Yap, M. H., Li, J., Hong, X., & Wang, S. J. (2019). Megc 2019–the second facial micro-expressions grand challenge. In 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019) (pp. 1-5)

2nd Micro-Expression Grand Challenge 2019 @ FG 2019, Lille, France

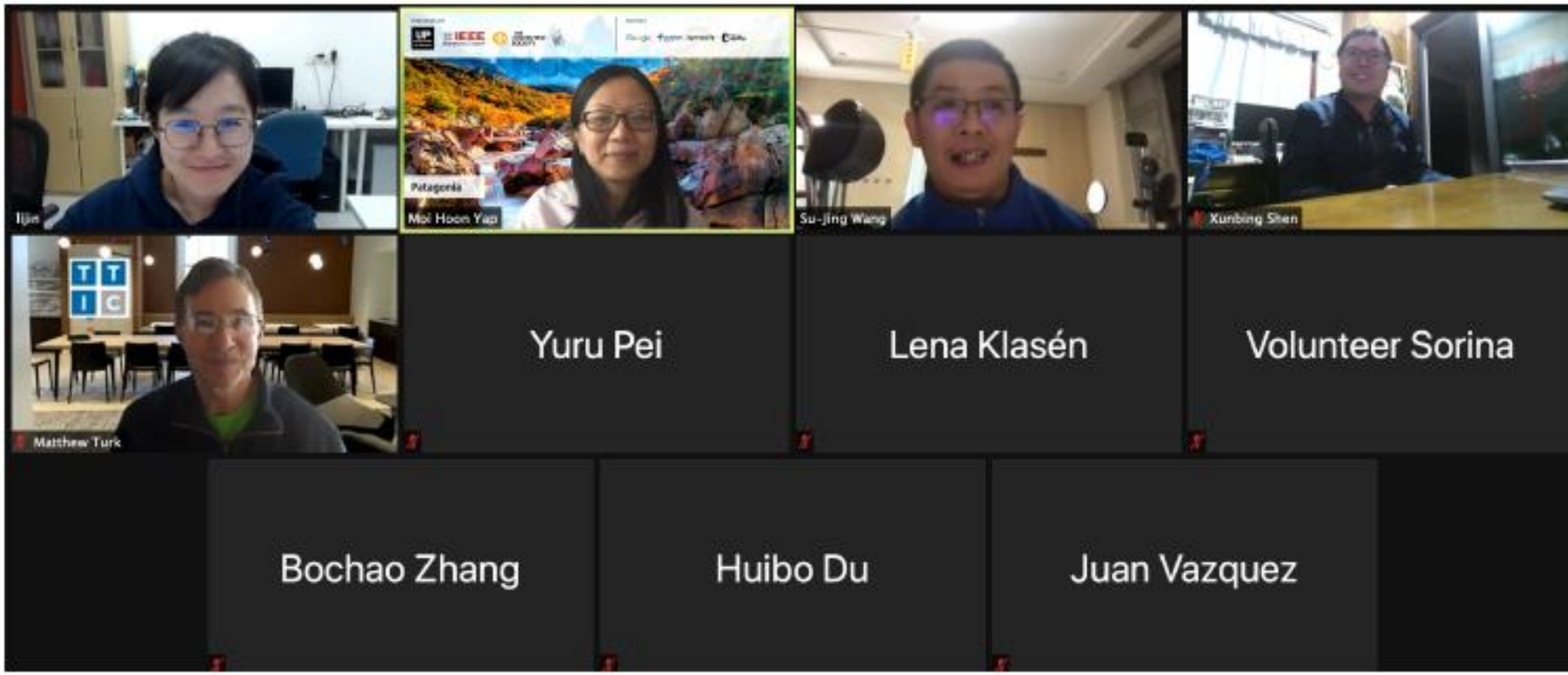
- **Insights (Recognition Challenge)**
  - Winning method (EMR with adversarial training) performed very well on SAMM but not on CASME II.
    - CASME II is predominantly Chinese subjects, while SAMM is the most diverse
  - Most submitted works opted to use the apex frame rather than the full sequences
  - Top 3 works all used optical flow as choice of input

See, J., Yap, M. H., Li, J., Hong, X., & Wang, S. J. (2019). Megc 2019–the second facial micro-expressions grand challenge. In 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019) (pp. 1-5)

3rd Micro-Expression Grand Challenge 2020 @ FG 2020, Buenos Aires, Argentina (Virtual)

Li, J., Wang, S., Yap, M. H., See, J., Hong, X., & Li, X. MEGC2020-The Third Facial Micro-Expression Grand Challenge. In 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)(FG) (pp. 234-237).

3rd Micro-Expression Grand Challenge 2020 @ FG 2020



~~Buenos Aires, Argentina~~ **Virtual !**
November 2020

3rd Micro-Expression Grand Challenge 2020 @ FG 2020, Buenos Aires, Argentina (Virtual)

- **Challenge: Spotting Macro- and Micro-Expression from Long Sequences**
  - Similar metrics from 2nd MEGC were used, except that they are computed separately for micro- and macro- cases then combined
  - All videos are regarded as one long video.
  - 5 submissions were received
    - Method by Zhang et al. was best in CAS(ME)^2
    - Method by Yap et al. was best in SAMM Long
    - Only one method utilised deep learning; all other methods rely on feature difference computation based on descriptors!
    - On average, performance on macro better than micro

Database	CAS(ME)^2			SAMM Long Videos		
Method	Macro-expression	Micro-expression	Overall	Macro-expression	Micro-expression	Overall
Baseline [9]	0.1196	0.0082	0.0376	0.0629	0.0364	0.0445
Gan et al [8]	0.1436	0.0098	0.0448	-	-	-
Pan [15]	-	-	0.0595	-	-	0.0813
Zhang et al. [23]	0.0547	0.2131	**0.1403**	0.1331	0.0725	**0.0999**
Yap et al. [21]	-	-	-	0.4081	0.0508	**0.3299**

Li, J., Wang, S., Yap, M. H., See, J., Hong, X., & Li, X. MEGC2020-The Third Facial Micro-Expression Grand Challenge. In 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)(FG) (pp. 234-237).

Challenges

DATABASES

1. **Subjectivity in humans**
   - Certain emotions (e.g. happiness) are easier to elicit compared to others (e.g. fear, sadness, anger)
   - Some people are more "poker-faced" than others – they hide their emotions well!
2. **Sample distribution**
   - Bias learning → Imbalanced distribution of samples per emotion, samples per subject
3. **Creative strategies for inducement**
   - Complementary info from body region[1], or heart rate from skin variations[2]

1 Song, Y., Morency, L. P., & Davis, R. (2013). Learning a sparse codebook of facial and body microexpressions for emotion recognition. In Proceedings of the 15th ACM on International conference on multimodal interaction (pp. 237-244).
2 Gupta, P., Bhowmick, B., & Pal, A. (2018). Exploring the feasibility of face video based instantaneous heart-rate for micro-expression spotting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 1316-1323).

Challenges

DATABASES

4. **Subject diversity**

- Most datasets contain a majority of subjects from one particular country or ethnicity

5. **Environment and setting**

- Real-world scenarios are much needed: Job interviews, criminal interrogation, patient assessment etc. (but many cannot pass ethic committees!)
- How about “two truths and a lie” game?

Challenges

SPOTTING

1. Landmark detection

- Room for improvement in existing methods. ME requires very stable detection / robust against noise to capture minute changes in facial muscles.

2. Threshold or classify?

- Most existing works employ rule-based strategies
  ➔ Not robust and adaptable!
- Per-frame classification of ME occurrence
  ➔ Rigid and noisy!

3. Onset and offset detection

- Current works do not consider detecting the start and end frames, which could be useful to trim ME sequences before classification

Challenges

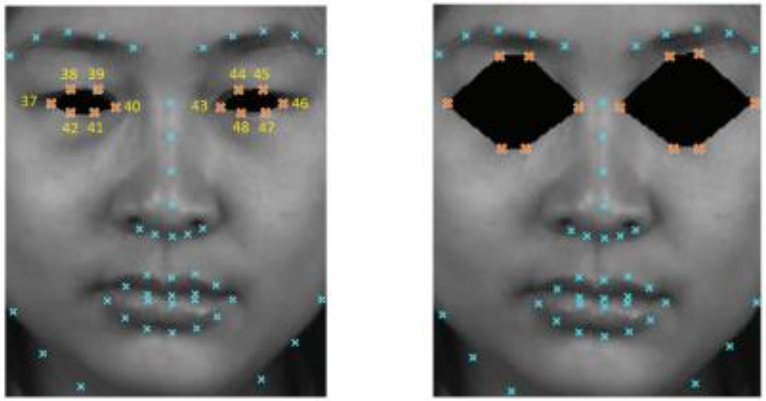
RECOGNITION

1. Block Selection

- Block-based methods of extracting features are quite popular
- Assignment of weights to blocks with key information
  ➔ Learning which blocks are discriminative[3]

2. Eyes: To Keep or Not To Keep?

- Some[4] works mask out eye regions to avoid eye blink motions, some[5] think otherwise



3 Zong, Y., Huang, X., Zheng, W., Cui, Z., & Zhao, G. (2018). Learning from hierarchical spatiotemporal descriptors for micro-expression recognition. *IEEE Transactions on Multimedia*, *20*(11), 3160-3172.
4 Liong, S. T., See, J., Wong, K., & Phan, R. C. W. (2016). Automatic micro-expression recognition from long video using a single spotted apex. In *Asian conference on computer vision* (pp. 345-360). Springer, Cham.
5 Duan, X., Dai, Q., Wang, X., Wang, Y., & Hua, Z. (2016). Recognizing spontaneous micro-expression from eye region. *Neurocomputing*, *217*, 27-36.

Challenges

RECOGNITION

3. Feature crafting / learning

- Most crafted features circa 2014-2016 are still holding reasonably strong results
- DL getting popular 2016 onwards – pushing the limits
  - There are obvious weaknesses and strengths in using DL
- Shallow DL a good choice?

4. Cross-DB recognition

- Mimics realistic setting (multi-environment enrolment)
- Generalizing and re-generating across different domains is fast gaining popularity[6]

6 Zong, Y., Zheng, W., Huang, X., Shi, J., Cui, Z., & Zhao, G. (2018). Domain regeneration for cross-database micro-expression recognition. IEEE Transactions on Image Processing, 27(5), 2484-2498.

Experiment-related issues

1. Evaluation Protocol

- Use LOSO cross-validation[7] instead of LOVO cross-validation (some works still do this! ☹)
  - LOVO exposes the training to samples belonging to the test sample subject

2. Performance Metrics

- Use F1-score instead of Accuracy
  - Accuracy tends to be bias in imbalanced datasets or heavily skewed data
  - Use unweighted metrics that give equal emphasis to rare classes

3. Class Labels

- A few works consider fewer number of classes than it should be ➔ problem benchmarking!
- Emotion classes vs. Objective classes[8]

7 Le Ngo, A. C., Phan, R. C. W., & See, J. (2014, November). Spontaneous subtle expression recognition: Imbalanced databases and solutions. In Asian conference on computer vision (pp. 33-48). Springer, Cham.

8 Davison, A. K., Merghani, W., & Yap, M. H. (2018). Objective classes for micro-facial expression recognition. Journal of Imaging, 4(10), 119.

Acknowledgements

Many thanks to the our institutions and sponsors throughout the years from 2014 till today!

Many thanks to our key collaborators, colleagues worldwide








The Not Face


thank you
شكراً جزيلاً
ngiyabonga
рахмат
danke
謝謝
teşekkür ederim
Баярлалаа
спасибо
faafetai lava
mersi
kia ora
barka
welalin
tack
spas
mahalo
tapadh leat
vinaka
спасибі
blagodaram
dank je
misaotra
matondo
paldies
grazzi
gracias
хвала
asante
manana
obrigada
tenki
murakoze
kiitos
dankie
nanni
nandri
dhanyavad
mauruuru
köszönöm
aciu
dankon
akun
bedankt
bayarlalaa
grácie
hvala
enkosi
dziękuję
djiere dieuf
tau
mochchakkeram
дякую
mamnun
chokrane
chnorakaloutioun
gratias ago
grácies
sulpáy
taiku
go raibh maith agat
sobodi
dēkuji
obrigado
sukriya
kop khun krap
sagolun
mèsi
didi madloba
ありがとう
najis tuke
kam sah hamnida
rahmat
terima kasih
tanemirt
rahmet
grazie
arigatô
takk
dakujem
trugarez
ευχαριστώ
diolch
dhanyavadagalu
shukriya
merce
мерси
তোমাকে ধন্যবাদ
감사합니다
xièxie
merci