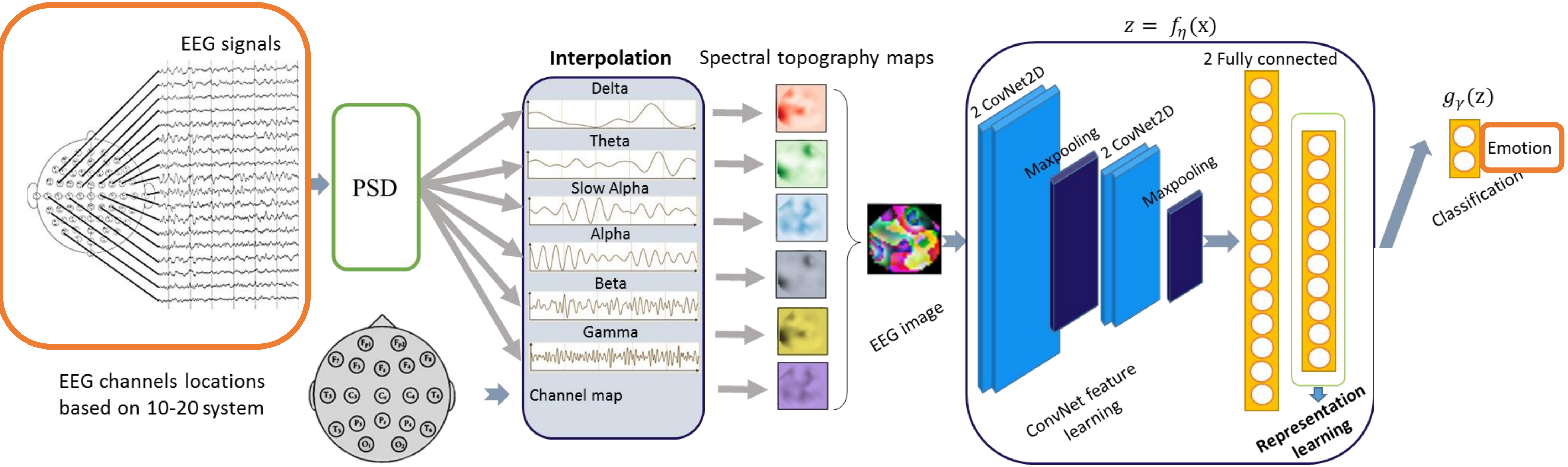
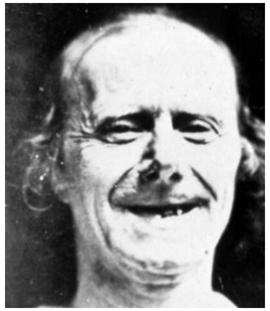
Expression-Guided EEG Representation Learning for Emotion Recognition

Soheil Rayatdoost[*], David Rudrauf[*] and Mohammad Soleymani[†]
[*]University of Geneva
[†] University of Southern California

Overview


EEG signals
EEG channels locations
based on 10-20 system
PSD
Interpolation
Delta
Theta
Slow Alpha
Alpha
Beta
Gamma
Channel map
Spectral topography maps
EEG image
$z = f_\eta(\mathrm{x})$
2 CovNet2D
Maxpooling
2 CovNet2D
Maxpooling
2 Fully connected
ConvNet feature learning
Representation learning
$g_\gamma(\mathrm{z})$
Emotion
Classification

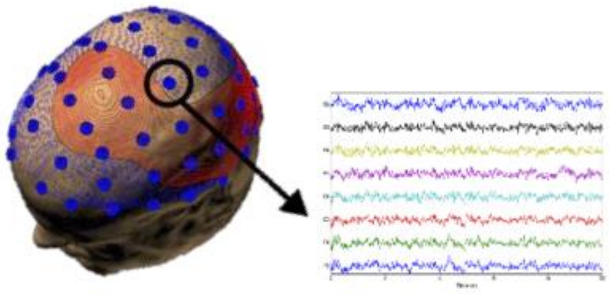


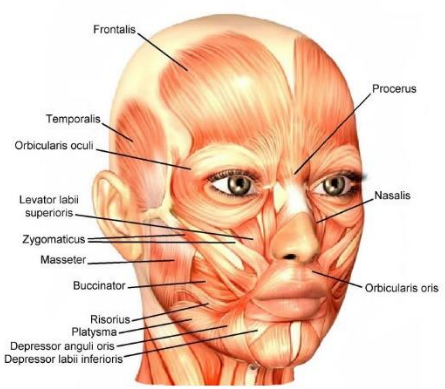
Multimodal emotion recognition

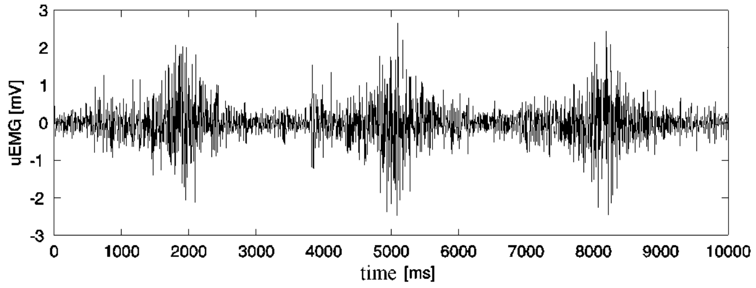
- Translating raw signals to lower dimensional representations
- Different modalities capturing different proxies of emotions
  - Facial and vocal expressions
  - Physiological signals (GSR, respiration effort,…)
  - Electroencephalogram (EEG) signals
- Multimodal emotion recognition is more robust and efficient
  - Feature-level fusion
  - Decision-level fusion
  - Joint representation
  - Coordinated representation

EEG, EMG & emotion

- EEG
  - Weak electrical potentials originated from cerebral activities
  - Frontal asymmetry
- EMG:
  - Behavioral activities







Problem definition & our approach

- Cross-modal association for emotion recognition
- Development of a cross-modal deep encoder
  - A representation guided by facial and ocular activities (expression)
  - Database for interactions between modalities (DAI-EF)
  - Transferred across corpora without EMG/EOG signal
    - MAHNOB
  - Adversarial Discriminative Domain Adaptation (ADDA)

Related work

- EEG-based emotion recognition
  - Zheng *et al.*: Identified stable patterns over time
  - Zheng *et al.*: Investigated critical frequency bands
  - Rayatdoost and Soleymani: Identified learnable EEG activity for emotion recognition (Sensory, emotional activities and muscular)
  - Soleymani *et al.:* Reported that emotion-related EEG activities are highly correlated with facial muscular activity
- EEG in multimodal emotion recognition
  - Soleymani *et al. and* Koelstra and Patras: Fused modalities for emotion recognition
    - Decision-level fusion to be superior to the early fusion ones

Data

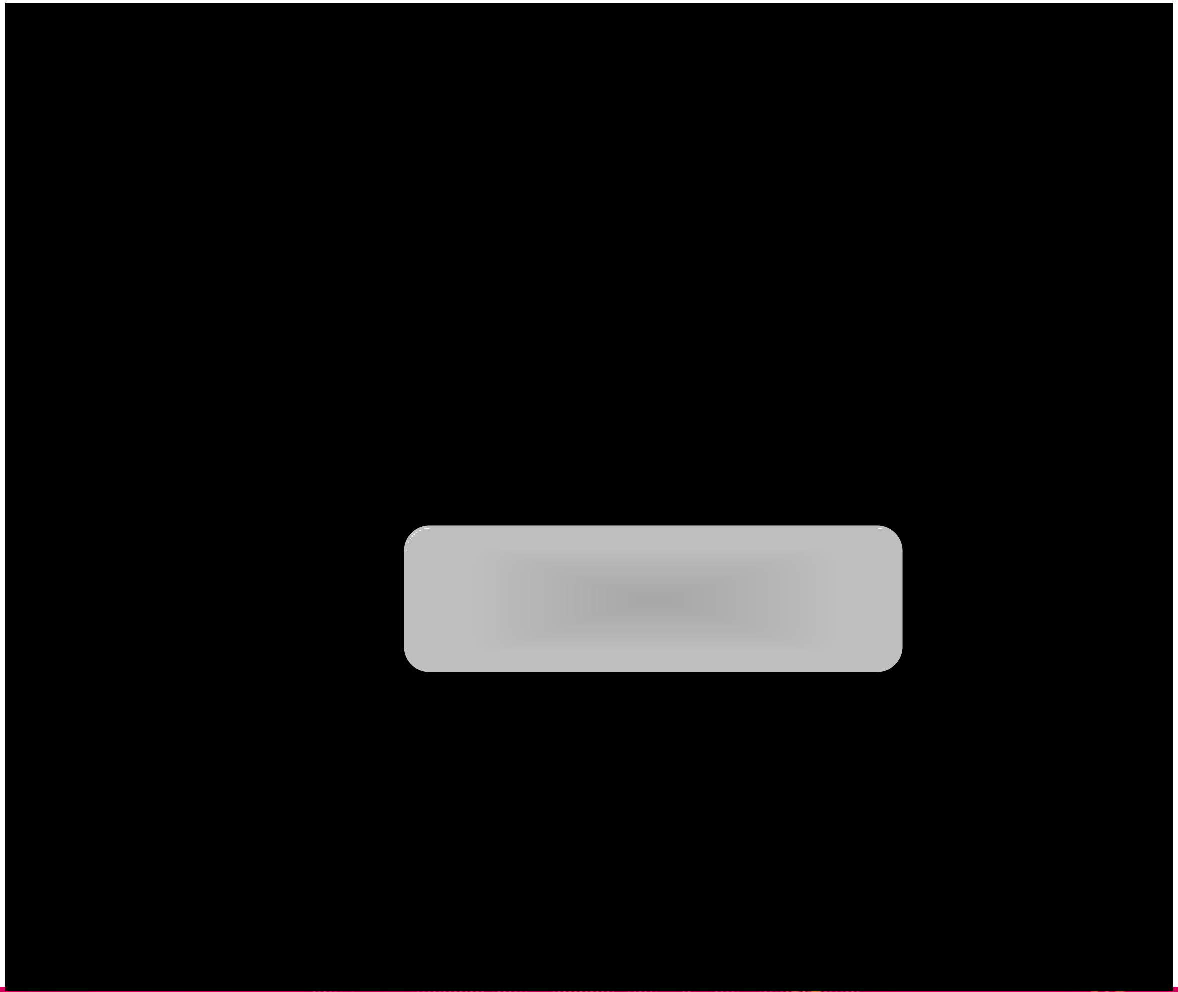
- Multimodal database for affect recognition and implicit tagging (MAHNOB)
  - EEG signal from 32 channels
  - 30 subjects
  - 20 video clips
- Database for Analysis of affective Interaction between EEG and Facial expression (DAI-EF)
  - EEG signal from 64 channels
  - Face video
  - 60 subjects (31 male, 17-67 years old, mean: 26.9 and SD: 8.4)
  - 40 video clips from commercial Movies

ICASSP2020 Barcelona

USC

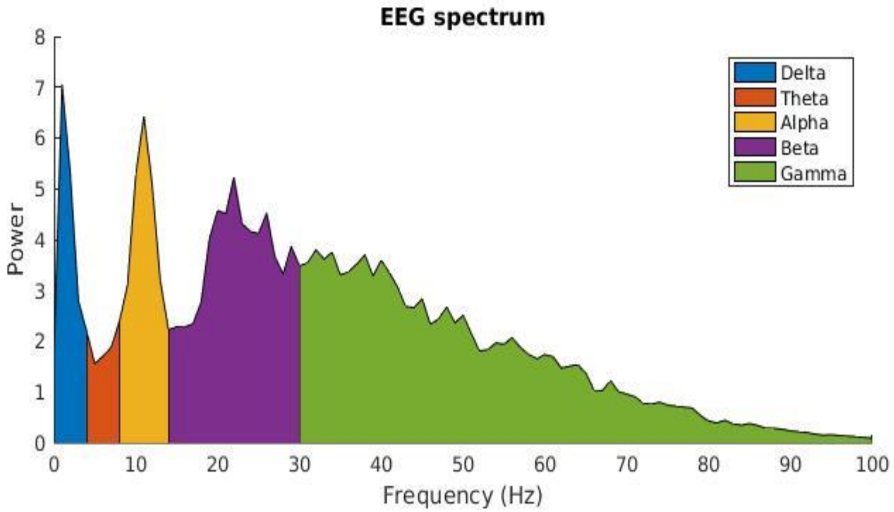
UNIVERSITÉ DE GENÈVE

Data



Features

- EEG
  - The power spectral density (PSD) in different frequency bands
    - Delta (0Hz < f < 4Hz), Theta (4Hz < f < 8Hz), Slow alpha (8Hz < f < 10Hz),
    - Alpha (8Hz < f < 13Hz),Beta (13Hz < f < 30Hz), Gamma (30Hz < f)
- EMG/EOG
  - Statistical moments
  - Energy and power features



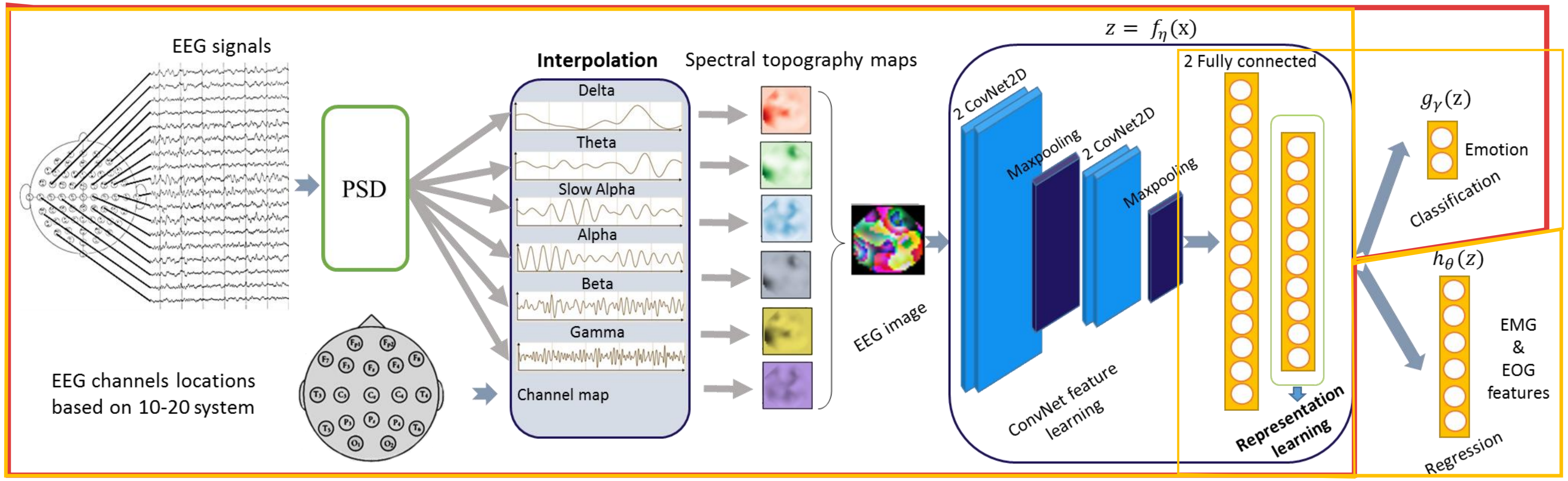
ICASSP 2020 Barcelona
USC
UNIVERSITÉ DE GENÈVE

Classification method

- Classes (two classes): positive/negative valence and high/low arousal
- Cross-modal encoder
  - Within-database
  - Between-database
  - Transfer learning
    - Supervised: Fine-tuning
    - Unsupervised: ADDA

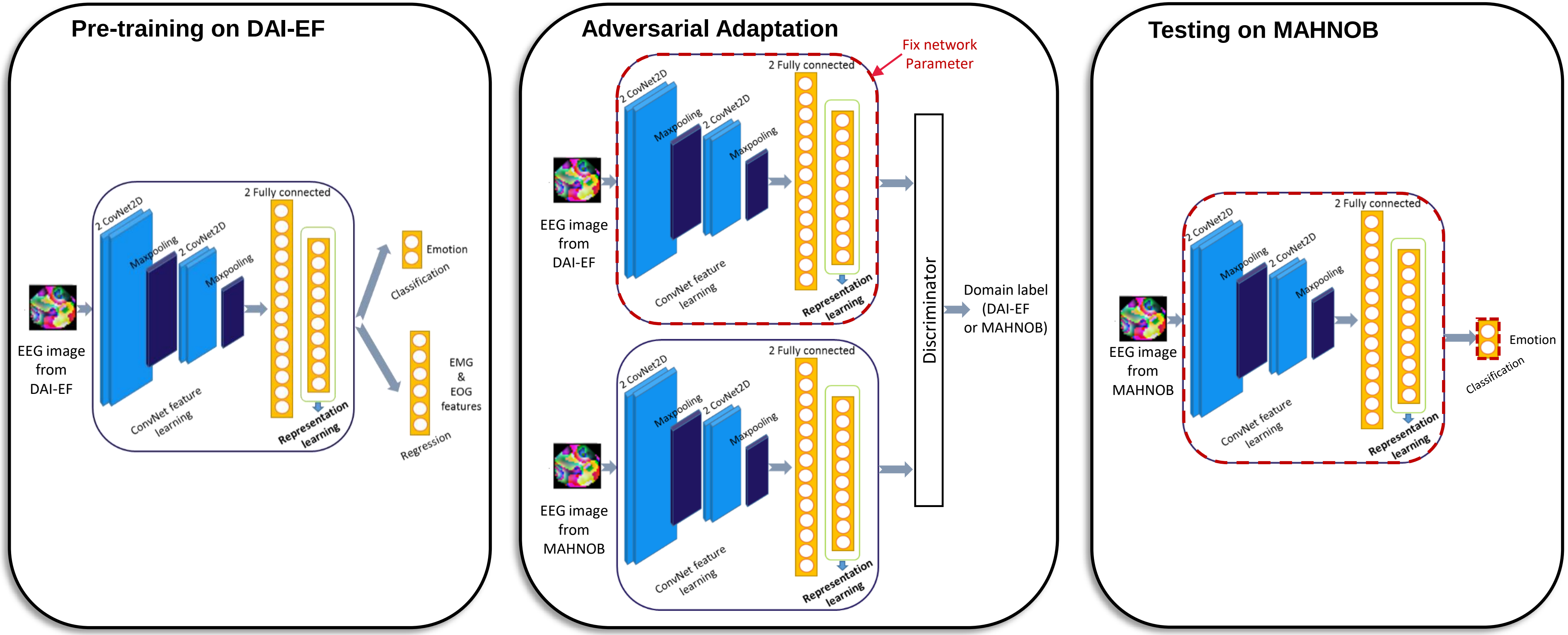
Cross-modal encoder

- EEG deep CNN
  - Solely trained on EEG
- Expression encoder
- Emotion-expression encoder (Emo-expression)



Domain adaptation

- Adversarial discriminative domain adaptation (ADDA)



Evaluation

- Emotion classes: positive/negative valence and high/low arousal
  - Datasets: MAHNOB, DAI-EF
    - Number of subjects: MAHNOB(30), DAI-EF(47)
  - Normalization: per subject normalization (z-score)
  - Evaluation metrics: Classification Rate and F1-score
  - Cross validation: corpus independent, 10-fold

Experiment

Within-database

Classifier Type	Test	Valence		Arousal	
		CR	F1	CR	F1
Deep CNN	DAI-EF	69.5	67.2	61.4	61.3
Expression	DAI-EF	71.0	70.1	61.2	61.1
Emo-expression	DAI-EF	**72.8**	**72.1**	**63.2**	**62.5**
Deep CNN	MAHNOB	62.5	60.1	55.6	53.8

Experiment

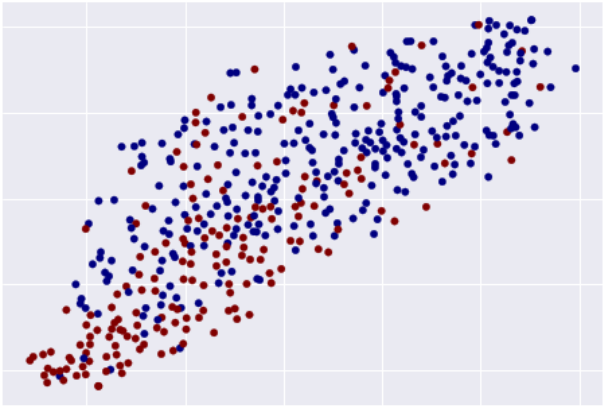
Between-database

Classifier Type	Test	Valence		Arousal	
		CR	F1	CR	F1
Deep CNN	MAHNOB	64.8	62.9	50.3	50.0
Expression	MAHNOB	64.6	63.4	53.0	52.7
Emo-expression	MAHNOB	**65.4**	**64.5**	**55.8**	**55.3**

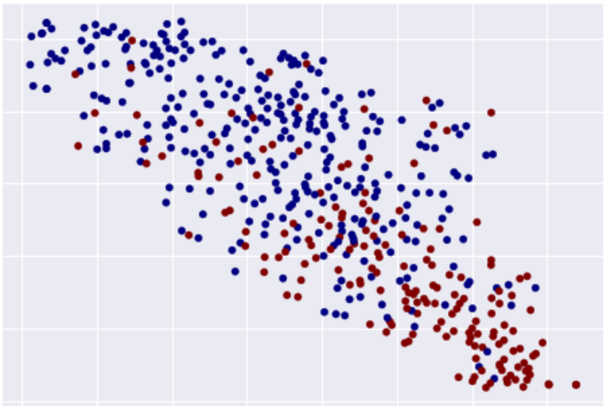
Transfer learning and domain adaptation

ADDA	MAHNOB	67.2	65.3	54.0	53.9
Deep CNN	MAHNOB	71.8	70.7	57.9	51.8
Expression	MAHNOB	71.8	70.8	55.1	54.4
Emo-expression	MAHNOB	**73.5**	**72.5**	**56.1**	**55.7**

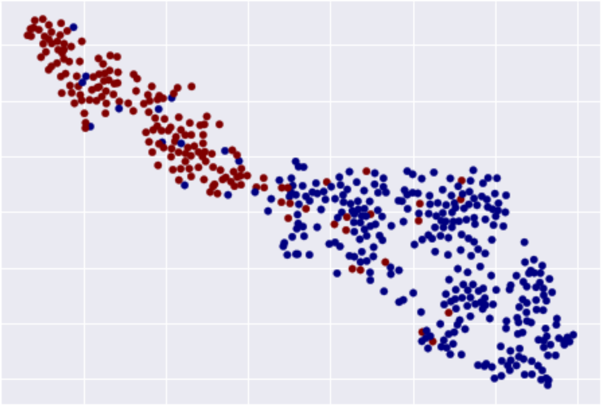
Representation's t-SNE embedding



The Deep CNN was trained on DAI-EF
and was tested on MAHNOB



The Deep CNN was trained on DAI-EF and was fine-tuned on MAHNOB (transfer learning



The emo-expression encoder was trained on DAI-EF and was fine-tuned on MAHNOB

ICASSP 2020 Barcelona USC UNIVERSITÉ DE GENÈVE

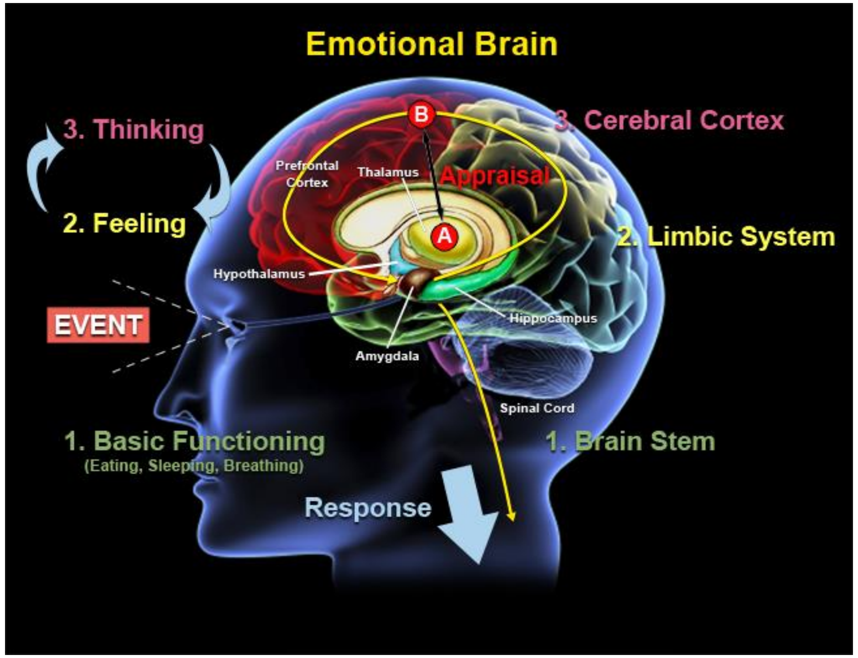
Conclusions

- Developed a novel cross-modal representation for emotion recognition
- Cross-modal representation learning enhances the performance of emotion recognition
- The representation generalizes when transferred across corpora
  - Supervised fine-tuning
  - Unsupervised domain adaptation
- Muscular activities component of EEG signals have contributed to the learnt emotional patterns

Thanks

Emotional responses

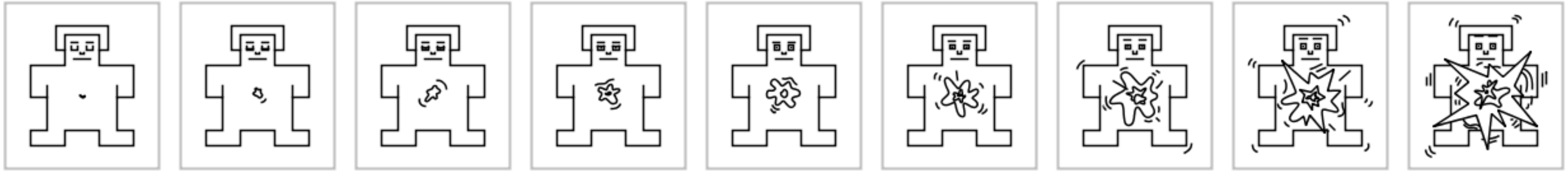
- Emotion:
  - Facial and vocal expressions
  - Physiological signals
  - Electroencephalogram (EEG)



Self-reports

Self Assessment Manikin (SAM):

Arousal (activation)



Valence (positive/negative)