A survey of automatic Facial Micro-Expression Analysis: Databases, Methods, and Challenges

ME spotting

finding the occurance of MEs in a video sequence

一、pre-processing

1. facial landmark detection and tracking

- detection

  locate the facial points on the facial images

  - ASM
  - **DRMF**
  - SCMS
  - **Face++**
- Tracking

  very few works applied

  - APF
  - KTL

2. face registration

to remove large head translations and rotation that might *affect the spotting task.*

- Area-based

  2D-DFT
- Feature-based

  当局部结构比图像强度所携带的信息更重要时，通常采用基于特征的方法

  - LWM

3. face marking

to remove noise caused by undesired facial motions that might *affect the performance of the spotting task.*

- Remove the middle of the face(eyes/nose/mouth)

masking is necessary(some research showed that eye blinking motion is more intense than that of micro-expression motion)

NOTED:a majority of works in the literature still do not include a masking pre-processing step.

4. face region retrieval

to split of the face into important regions for automatic facial ME spotting.

- Split the face into m*n blocks
- **split the face into a few ROIs**

二、detection

classifier-based

each frame is classified to neutral, onset, apex and offset

- GentleSVM
- k mean cluster

rule-based

Threshold technique

binary search(for peck detection)

三、performance matrics

- TPR
- FPR
- MAE
- ASR

ME recognition

pre-processing

除了spotting包括的pre-processing外，还包括：

- magnification

  exaggerate or magnify these ME movements----enlarge the difference among different categories

  - EMM\A-EMM\P-EMM
  - GLMM
- interpolation 插值

  the uneven length of ME video samples

  - TIM
  - DMDSP

features

Appearance-based features from raw input

- LBP-TOP
- OF maps
- BI-WOOF
- HOOF

classifier

- KNN
- SVM
- RF

performance matrics

- F1-score
- CCC

The OMG-Emotion Behavior Dataset

- audio vision language
- contextual info.
- In-the-wild videos

evaluation protocol

- categorical emotion recognition

  F1-score

- Continuous estimation(based on arousal and valence)

  Consider it as a regression problem

  MSE(Mean squared root)

  CCC(concordance correlation coefficient)

  based on the model's response and the gold standard(根据每个utterence有5个annotrator然后根据求平均值等来确定的)

Multimodal Utterance-level Affect Analysis using Visual, Audio and Text Features

dataset

OMG-Emotin

Model

图-early fusion

result

	arousal	Valence
CCC	0.400	0.353

Multimodal Emotion Recognition for One-Minute-Gradual Emotion Challeng

dataset

OMG-Emotion

model

图 p3

data augmentation

!!attention

result

	arousal	Valence
CCC	0.397	0.520

A Multi-component CNN-RNN Approach for Dimensional Emotion Recognition in-the-wild

focus on the visual part of the provided videos(没有考虑contextual info.)

dataset

OMG-Emotion

approach

P2-p3

based on a CNN-RNN deep neural architecture which has produced best performance on the AffWild database.

result

	arousal	Valence
CCC	0.310	0.490

1. Post-processing 可以提高准确率

From Hard to Soft:Towards more Human-like Emotion Recognition by Modelling the Perception Uncertainty

not only the emotional state but also its corresponding perception uncertainty are jointly provided for each observed sample.

每个输入对应两个输出：emotional state and perception uncertainty

Dataset

SEWA:

This database was collected by undertaking spontaneous video-chats with 64 subjects (32 pairs), leading to a total duration of approximately 178 minutes. Specifically, each pair of subjects had a remote discussion after watching four given advertisement, and each discussion session lasted about three minutes. The discussions were recorded at a video sampling rate between 20 and 30 fps and at an audio sampling rate of either 44.1 or 48.0 KHz, depending on the recording devices used by the subjects. The dataset is available to researchers for non-commercial use at https://db.sewaproject.eu. Along with the audiovisual episodes and the annotations, acoustic and visual features are provided as well.

model

BLSTM

Multi-tasking => emotional state + perception uncertainty

emotional state:

- 单独训练vision feature
- 单独训练audio feature
- 然后late fusion

evaluation protocol

- ccc
- RMSE(root mean squared error)

experiment

1. 单独的audio/video和uncertainty v.s. mutli-tasking

   multi-tasking的效果好，说明可以学习他们的关系
2. 单独的audio和video v.s. late fusion

   late fusion可以提高performance

Training Deep Neural Networks with Different Datasets In-the-wild:The Emotion Recognition Paradigm

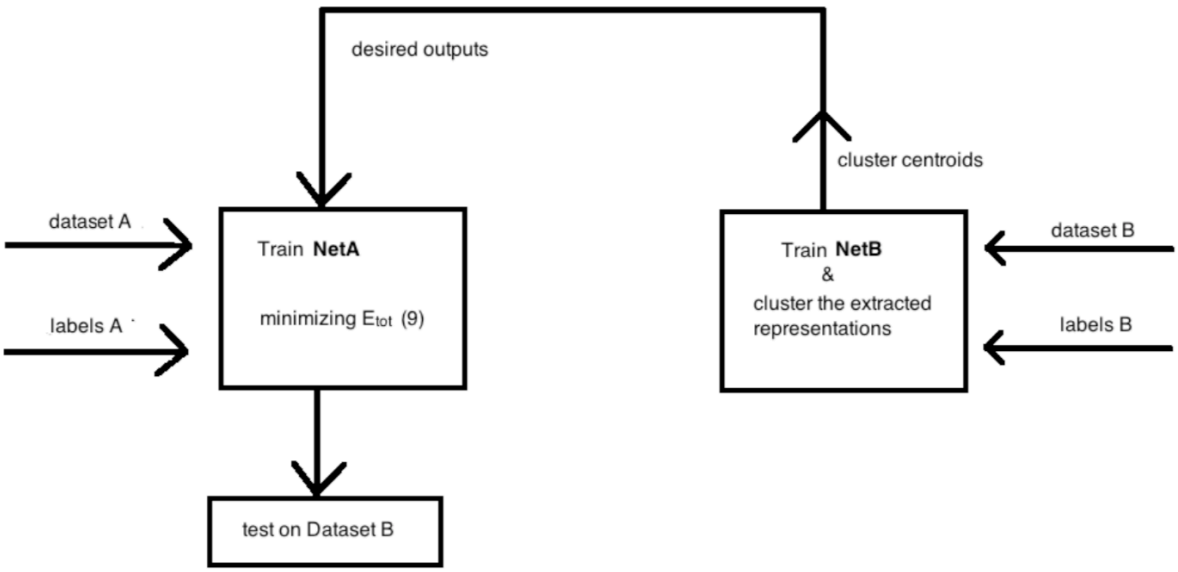
对数据没有进行任何处理，且只使用video frames

dataset

AFEW

The series of EmotiW challenges make use of data from the Acted Facial Expression in-the-wild (AFEW) dataset. This dataset is a dynamic temporal facial expressions data corpus consisting of close to real world scenes extracted from movies and reality television shows. In total it contains 1809 videos. The whole dataset is split into three sets: training set (773 video clips), validation set (383 video clips) and test set (653 video clips). It should be emphasized that both training and validation sets are mainly composed of real movie records, however 114 out of 653 video clips in the test set are real TV clips, increasing, therefore, the difficulty of the challenge. The number of subjects is more than 330, aged 1-77 years. The annotation is according to 7 facial expressions (Anger, Disgust,Fear, Happiness, Neutral, Sadness and Surprise)

structure



使用validation data来训练NetB，将NetB的hidden layer的最后一层利用k-mean来聚类成7个cluster，然后将聚类中心作为desired output来训练NetA（使用的数据是train data）

target

provide an as good as possible performance on the validation set while achieving the best possible performance on the training set as well

Facial Expression Recognition Using Weighted Mixture Deep Neural Network Based on Double-Channel Facial Images

dataset

CK+

JAFFE

Oulu-CASIA