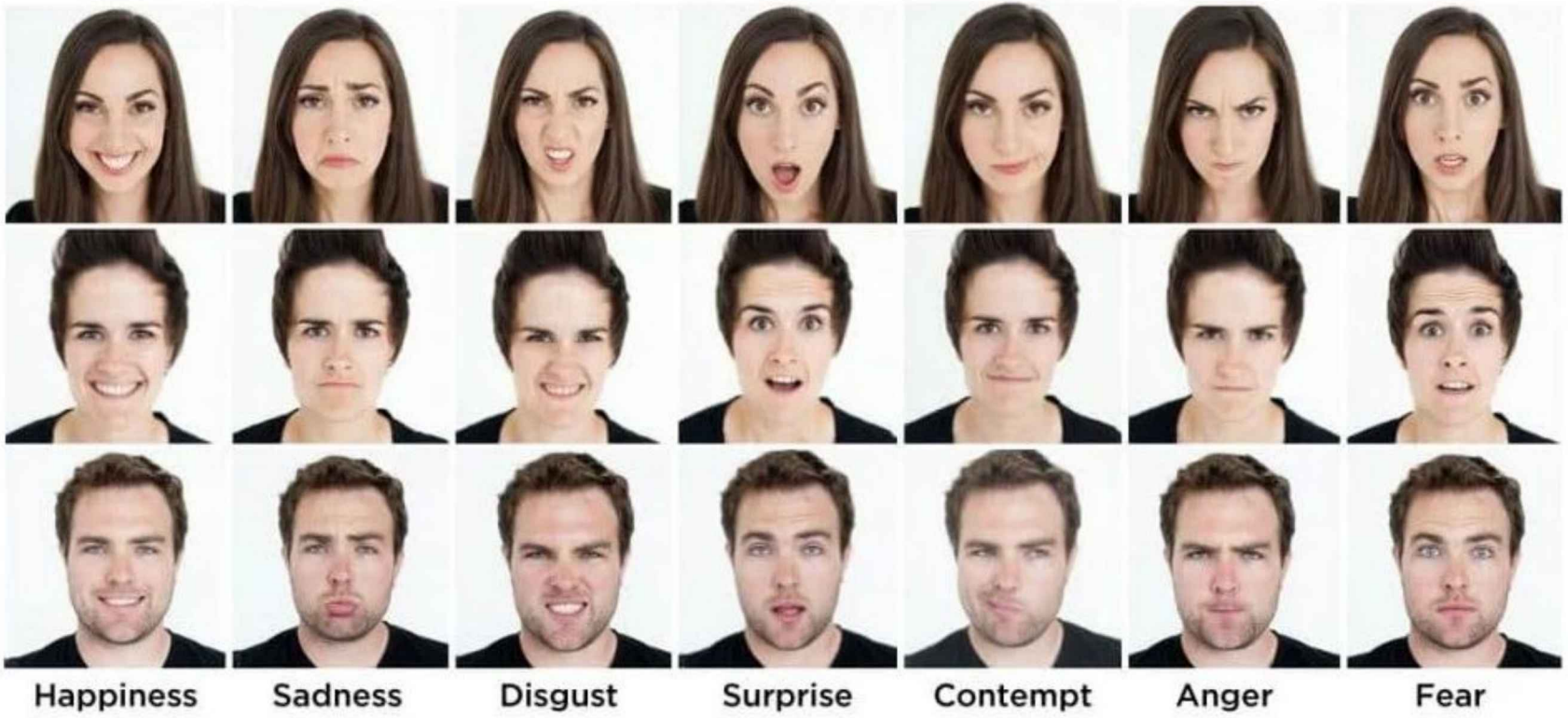
Facial Micro-Expression Analysis in ADOS Videos

Na Zhang

Face and Facial Expression



Facial Expression Analysis

- Has attracted great interest over the past years
  - Electronic / digital health
  - Human machine interaction
  - Human behavior analysis
  - Video communication

Digital Health [1]

Human Machine Interaction

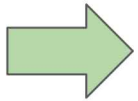
Video Communication [3]

Behavior Analysis [2]

[1] Figure from https://www.hospitaltimes.co.uk/digital-access-a-cause-and-solution-to-health-inequality/
[2] Figure from https://hertasecurity.com/news/a-new-unprecedented-solution-for-human-behavior-analysis/
[3] Figure from https://www.youtube.com/watch?v=1XyJNPa8Fwc

Socio-emotion Interaction in ASD

- **Socio-emotion interaction** difficulties
  - **communication disorder**
  - **emotional dysregulation**
  - **rigid and repetitive behaviors**



- **Problems** related to performance of
  - **Expressive language**
  - **Social**
  - **Emotional adaptive skills**

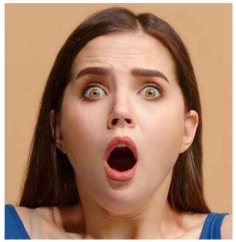
All individuals diagnosed with ASD, experience either **one or more aforementioned difficulties,** regardless of the severity levels of diagnosis

Rashidan, Mohammad Ariff, et al. "Technology-assisted emotion recognition for autism spectrum disorder (ASD) children: a systematic literature review." *IEEE Access* 9 (2021): 33638-33653.

Emotions in ASD

- Usually do not show the emotions in a way that normal people would be able to recognize and understand
  - ➢ either they do **not respond** emotionally

  - ➢ or their emotional responses might sometimes seem **over-reaction**



- With the advancement of CV and ML technology, there are much research that have been embarked around recognizing human emotions, particularly for autistic children and individuals
- This study focuses on showing **how to analyze emotions felt by the autistic persons**, not the capacity of the person to recognize emotions in others

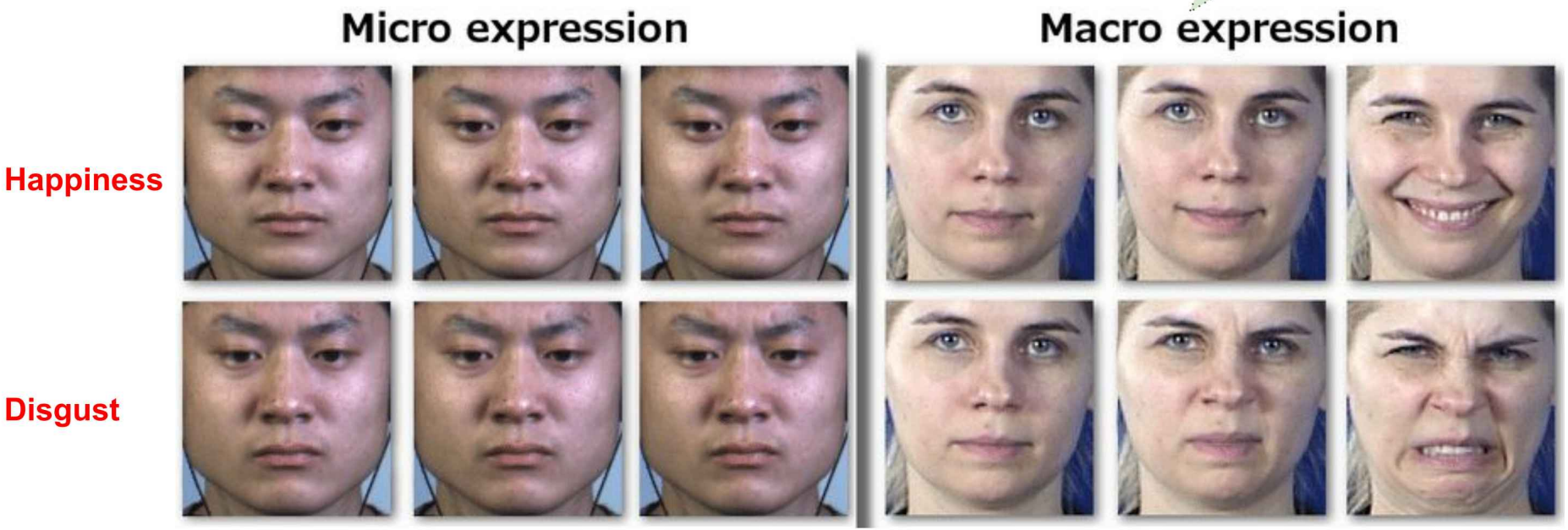
Existing Study on Emotion Recognition for ASD

- Types of stimuli to invoke emotions
  - **Picture**
    - a series of images that invoke the desired emotions
  - **Video**
    - bi-sensory stimulation that combines audio and visual stimulus
  - **Task**
    - directive real-life situation that elicit the desired emotions
- The evoked emotions includes some basic emotions, like
  - fear, anger, happiness, sadness, disgust, surprise, etc.
- Types of extracted feature about emotions
  - Optical flow, spatio-temporal, thermal intensity value(TIV) from thermal images, EEG, texture, etc.
- Types of classifier
  - K-NN, K-means, SVM, Decision Tree, GMM, CNN, etc.

Rashidan, Mohammad Ariff, et al. "Technology-assisted emotion recognition for autism spectrum disorder (ASD) children: a systematic literature review." *IEEE Access* 9 (2021): 33638-33653.

Micro- vs. Macro- Expression

Traditional Emotion Recognition

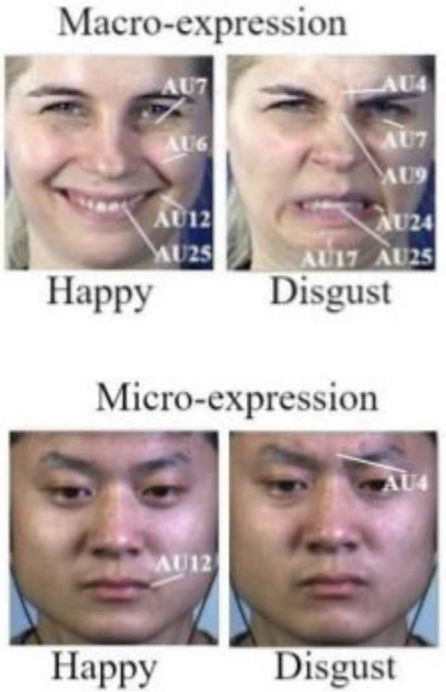


Difference of motion intensity between micro and macro expression

Allaert, B., Bilasco, I.M. and Djeraba, C., 2019. Micro and macro facial expression recognition using advanced Local Motion Patterns. *IEEE Transactions on Affective Computing*, *13*(1), pp.147-158.

Main differences between macro- and micro- expressions

Difference	Micro-expression	Macro-expression
Noticeability	Easy to ignore	Easily noticed
Time interval	Short duration (0.065-0.5 seconds)	Long duration (0.5-4 seconds)
Motion intensity	Slight variation	Large variation
Subjectivity	Involuntary (uncontrollable)	Voluntary (under control)
Action areas	Fewer	Almost all areas



FACS, AU: Action Unit

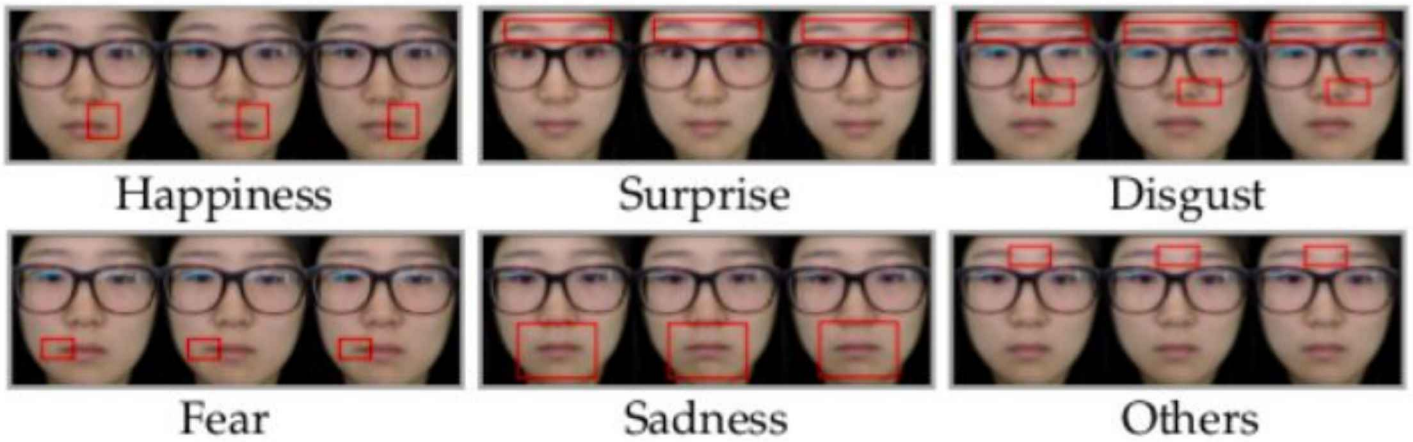
➔ posed facial expressions (i.e., voluntary facial actions)

➔ spontaneous facial expressions (i.e., involuntary emotion)

Ben, X., Ren, Y., Zhang, J., Wang, S.J., Kpalma, K., Meng, W. and Liu, Y.J., 2021. Video-based facial micro-expression analysis: A survey of datasets, features and algorithms. *IEEE transactions on pattern analysis and machine intelligence*.

Micro-Expressions

- Are characterized by short durations (less than 0.5 seconds), involuntary generation and low intensity
- Regarded as unique cues showing one's hidden real emotions
- Only a few facial action units are involved



Ben, X., Ren, Y., Zhang, J., Wang, S.J., Kpalma, K., Meng, W. and Liu, Y.J., 2021. Video-based facial micro-expression analysis: A survey of datasets, features and algorithms. *IEEE transactions on pattern analysis and machine intelligence*.

Emotion Classes

- There are many different basic emotion definitions

Characteristics	Datasets								
	MEVIEW	SMIC			CASME	CASME II	CAS(ME)^2	SAMM	MMEW
		HS	VIS	NIR					
Num of samples*	40	164	71	71	195	247	57	159	300
Participants*	16	16	8	8	35	35	22	32	36
Frame rate	25	100	25	25	60	200	30	200	90
Resolution	1280×720	640×480			640×480 & 1280×720	640×480	640×480	2040×1088	1920×1080
Facial resolution	N/A	190×230			150×190	280×340	N/A	400×400	400×400
Emotion classes	7 categories: Happiness (6) Anger (2) Disgust (1) Surprise (9) Contempt (6) Fear (3) Unclear (13)	3 categories: Positive (107) Negative (116) Surprise (83)			8 categories: Amusement (5) Disgust (88) Sadness (6) Contempt (3) Fear (2) Tense (28) Surprise (20) Repression (40)	5 categories: Happiness (33) Repression (27) Surprise (25) Disgust (60) Others (102)	4 categories: Positive (8) Negative (21) Surprise (9) Others (19)	7 categories: Happiness (24) Surprise (13) Anger (20) Disgust (8) Sadness (3) Fear (7) Others (84)	7 categories: Happiness (36) Anger (8) Surprise (89) Disgust (72) Fear (16) Sadness (13) Others (102)
Available labels	Emotion/FACS	Emotion			Emotion/FACS	Emotion/FACS	Emotion/FACS/ Video type	Emotion/FACS	Emotion/FACS

- Most popular specific emotions: 7 types
  - anger, disgust, fear, happiness, sadness, surprise, and contempt

P. Ekman, “Facial expressions of emotion: an old controversy and new findings,” Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, vol. 335, no. 1273, pp. 63-69, 1992.

FACS: Facial Action Units of Different Micro-Expressions

- Happiness
  - Eye area
    - Cheeks are raised
    - Lower eyelid may show wrinkles or be tense
  - Mouth area
    - Corners of the lips are pulled
    - A wrinkle runs from outer nose to outer lip

- Sadness
  - Eye area
    - Inner corners of the eyebrows are drawn in and then up
    - Skin below the eyebrows is triangulated, with inner corner up
    - Drooping upper eyelids
    - Losing focus in eyes
  - Mouth area
    - Corner of the lips are drawn down
    - Jaw comes up

- Surprise
  - Eye area
    - The eyebrows are raised and curved
    - Skin below the brow is stretched
    - Horizontal wrinkles show across the forehead
    - Eyes are widened; Eyelids are opened, white of the eye showing above and below.
  - Mouth area
    - Jaw drops open and teeth are parted but there is no tension or stretching of the mouth.
    - Lip is Puckered



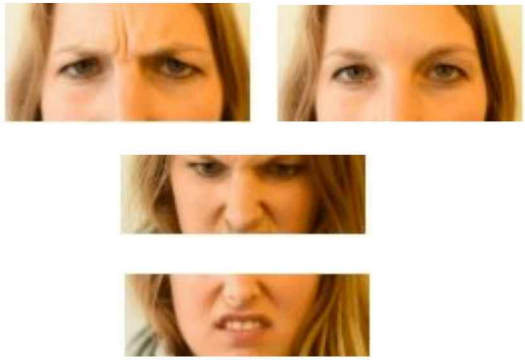
- Fear
  - Eye area
    - Eyebrows are drawn together, usually in a flat line
    - Wrinkles in the forehead are in the center between the eyebrows, not across
    - Upper eyelid is raised, but the lower lid is tense and drawn up
    - Eyes have the upper white showing, but not the lower white
  - Mouth area
    - Mouth is open or lips are slightly tensed or stretched and drawn back
    - Nasolabial area is deepened




- Anger
  - Eye area
    - Eyelid is tightened
    - The eyebrows are lowered and drawn together
    - Vertical lines appear between the eyebrows
    - Eyes are in hard stare or bulging
  - Nose area
    - Nose is wrinkled
    - Nostrils may be dilated
  - Mouth area
    - Lower lip is tensed
    - Lips can be pressed firmly together, with corners down, or in a square shape as if shouting
    - The lower jaw juts out




- Disgust
  - Eye area
    - Eyes are narrowed
    - Eyebrows are lowered
    - Eyelids are tightened
  - Nose area
    - Nose is wrinkled
  - Mouth area
    - Upper lip is raised
    - Upper teeth may be exposed
  - Cheeks
    - Cheeks are raised



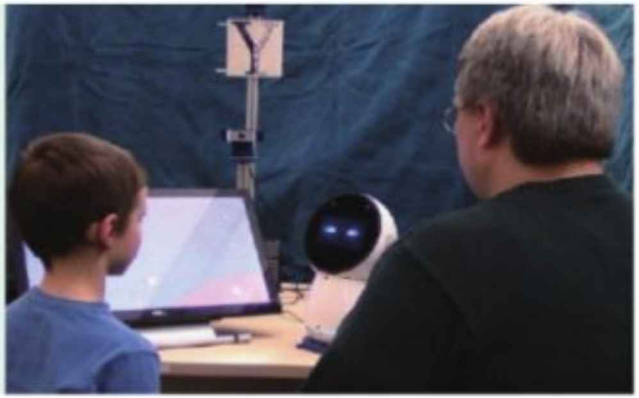
- Contempt
  - Mouth area
    - Lip corner is tightened and raised on only one side of the face

Applications of Micro-expression (ME) Analysis

- Micro-expressions are involuntary and transient facial expressions
- Can reflect the true emotions that a person try to suppress, hide, mask, or conceal
- Provide more important information than macro-expression

Lie Detection

Criminal Investigation



Clinical Diagnosis

Pan, H., Xie, L., Wang, Z., Liu, B., Yang, M. and Tao, J., 2021. Review of micro-expression spotting and recognition in video sequences. *Virtual Reality & Intelligent Hardware*, *3*(1), pp.1-17.

Public Micro-Expression Databases

- Due to the intrinsic characteristics, such as
  - involuntariness
  - short duration
  - slight variation
- Eliciting micro-expressions in a controlled environment is challenging
- So far, a few micro-expression datasets have been developed
- However, most of these still have some deficiencies as regards their
  - **elicitation paradigms**
  - **labelling** methods
  - or **small data size**

Seven Publicly Released Micro-Expression Datasets

Characteristics	Datasets								
	MEVIEW	SMIC			CASME	CASME II	CAS(ME)^2	SAMM	MMEW
		HS	VIS	NIR					
Num of samples*	40	164	71	71	195	247	57	159	300
Participants*	16	16	8	8	35	35	22	32	36
Frame rate	25	100	25	25	60	200	30	200	90
Mean age	N/A	N/A			22.03	22.03	22.59	33.24	22.35
Ethnicities	N/A	3			1	1	1	13	1
Resolution	1280×720	640×480			640×480 & 1280×720	640×480	640×480	2040×1088	1920×1080
Facial resolution	N/A	190×230			150×190	280×340	N/A	400×400	400×400
Emotion classes	7 categories: Happiness (6) Anger (2) Disgust (1) Surprise (9) Contempt (6) Fear (3) Unclear (13)	3 categories: Positive (107) Negative (116) Surprise (83)			8 categories: Amusement (5) Disgust (88) Sadness (6) Contempt (3) Fear (2) Tense (28) Surprise (20) Repression (40)	5 categories: Happiness (33) Repression (27) Surprise (25) Disgust (60) Others (102)	4 categories: Positive (8) Negative (21) Surprise (9) Others (19)	7 categories: Happiness (24) Surprise (13) Anger (20) Disgust (8) Sadness (3) Fear (7) Others (84)	7 categories: Happiness (36) Anger (8) Surprise (89) Disgust (72) Fear (16) Sadness (13) Others (66)
Available labels	Emotion/FACS	Emotion			Emotion/FACS	Emotion/FACS	Emotion/FACS/ Video type	Emotion/FACS	Emotion/FACS
Download URL	http://cmp.felk.cvut.cz/cechj/ME/	http://www.cse.oulu.fi/SMICDatabase			http://fu.psych.ac.cn/CASME/casme-en.php	http://fu.psych.ac.cn/CASME/casme2-en.php	http://fu.psych.ac.cn/CASME/cas(me)2-en.php	http://www2.docm.mmu.ac.uk/STAFF/M.Yap/dataset.php	http://www.dpailab.com/database.html

Here, samples* *refers to the samples of micro-expressions. Note that CAS(ME)^2 and MMEW also contain samples of macro-expressions. Not all participants produce micro-expressions; here,* Participants* *refers to the number of participants who generate micro-expressions. The parentheses enclose the number of emotional video clips in that category.*

HS: high-speed camera; VIS: normal visual camera; NIR: near-infrared camera

Ben, Xianye, et al. "Video-based facial micro-expression analysis: A survey of datasets, features and algorithms." *IEEE transactions on pattern analysis and machine intelligence* (2021).

Paradigm of Elicitation

An effective method

- Watching an emotional video while maintaining a neutral expression (i.e. suppressing emotions)

A typical data acquisition setup

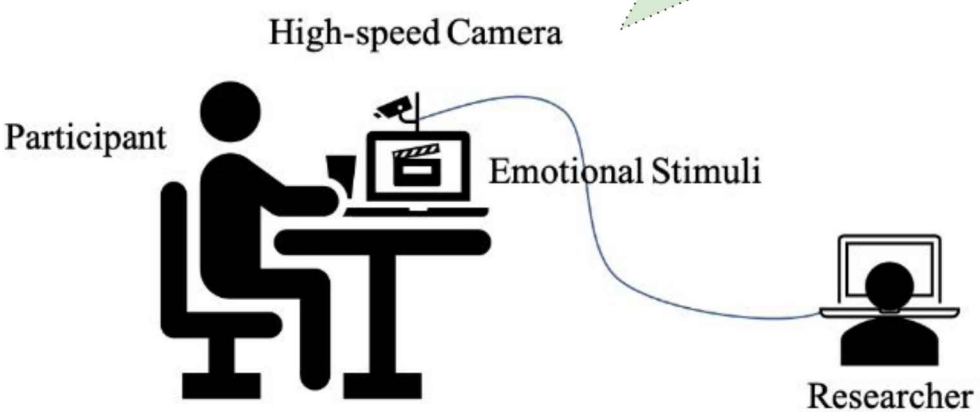


Figure 13. For the purposes of data collection, participants watch an emotional video while their faces are imaged by high-speed camera.

Five datasets — SMIC, CASME, CASME II, SAMM and CAS(ME)2 — used this elicitation paradigm.

Zhang, Liangfei, and Ognjen Arandjelović. "Review of automatic microexpression recognition in the past decade." *Machine Learning and Knowledge Extraction* 3.2 (2021): 414-434.

Micro-Expression Spotting

- to find the time interval([onset, apex, offset]) at which micro-expression are detected
  - onset: the first frame at which a ME starts
    - i.e., changing from the baseline, which is usually the neutral facial expression
  - apex: the frame at which the highest intensity of the facial expression is reached
  - offset: the last frame at which a ME ends
    - i.e., returning back to the neutral facial expression

Pan, H., Xie, L., Wang, Z., Liu, B., Yang, M. and Tao, J., 2021. Review of micro-expression spotting and recognition in video sequences. *Virtual Reality & Intelligent Hardware*, *3*(1), pp.1-17.

- A prerequisite for advanced facial ME analysis
- Proper spotting can decrease the redundant information for further analysis

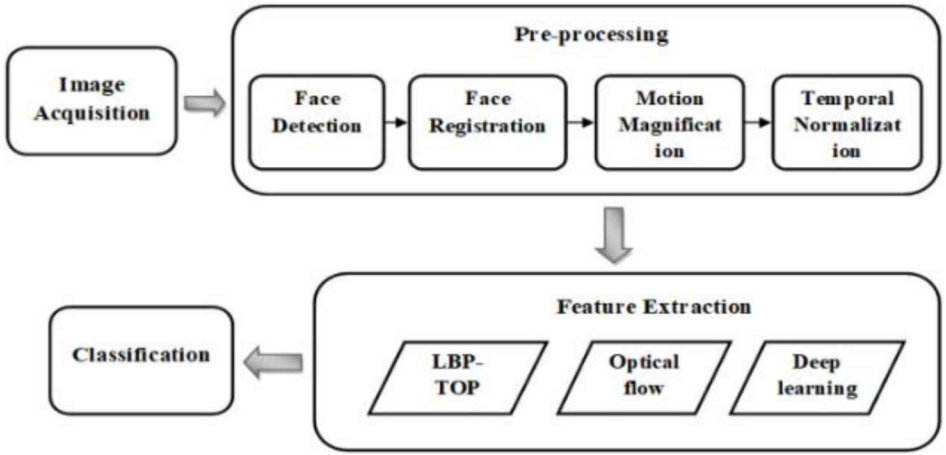
- Three major challenges
  - (1) relies on setting the optimal thresholds to detect ME for any given feature
  - (2) different people may perform different extra facial actions
    - some people blink habitually, while other people sniff more frequently
  - (3) when recording videos, many comprehensive factors may significantly influence the micro-expression spotting
    - head movement
    - physical activity
    - recording environment
    - Lighting condition

- A discriminative feature
  - should be able to capture the **difference in both spatial and temporal domains**
  - should be able to capture the **micro-level differences**
- Existing spotting algorithms can be broadly divided into two classes:
  - **Hand-crafted Feature**
    - Optical flow based
    - Feature descriptor based
      - Gradient based Feature: histogram of gradients (HOG), HOG-TOP
      - Local Binary Pattern (LBP) based Feature
      - LBP-TOP variants
  - **Learning based Feature**
    - Single stream learning: 2D CNN, 3D CNN (motion)
    - Shallow learning: limited and imbalanced data
    - Two-step learning: spatio+temporal feature

Xie, Hong-Xia, et al. "An overview of facial micro-expression analysis: Data, methodology and challenge." *arXiv preprint arXiv:2012.11307* (2020).

Micro-Expression Recognition

- Attempts to categorize a video of micro-expressions into one of the expression classes
  - happiness, sadness, disgust, fear, contempt, anger, surprise, etc.

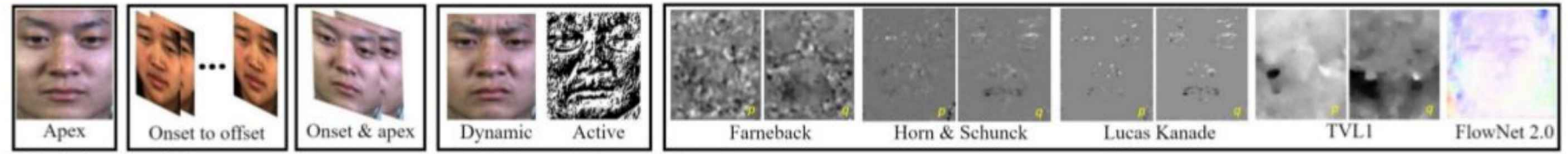


Pan, Hang, et al. "Review of micro-expression spotting and recognition in video sequences." *Virtual Reality & Intelligent Hardware* 3.1 (2021): 1-17.

- Various inputs

Input	Strength	Shortcoming
Apex	Efficient; Take advantage of massive facial images	Require magnification and apex detection
Sequence	Process directly	Not efficient; Redundancy
Frame aggregation	Efficiently leverage key temporal information	Require apex detection
Image with dynamic information	Efficiently embedding spatio-temporal information	Required dynamic information computation
Optical flow	Remove identity in some degree	Optical flow computation is necessary
Combination	Fully exploring spatial and temporal information	High computation

multiple inputs

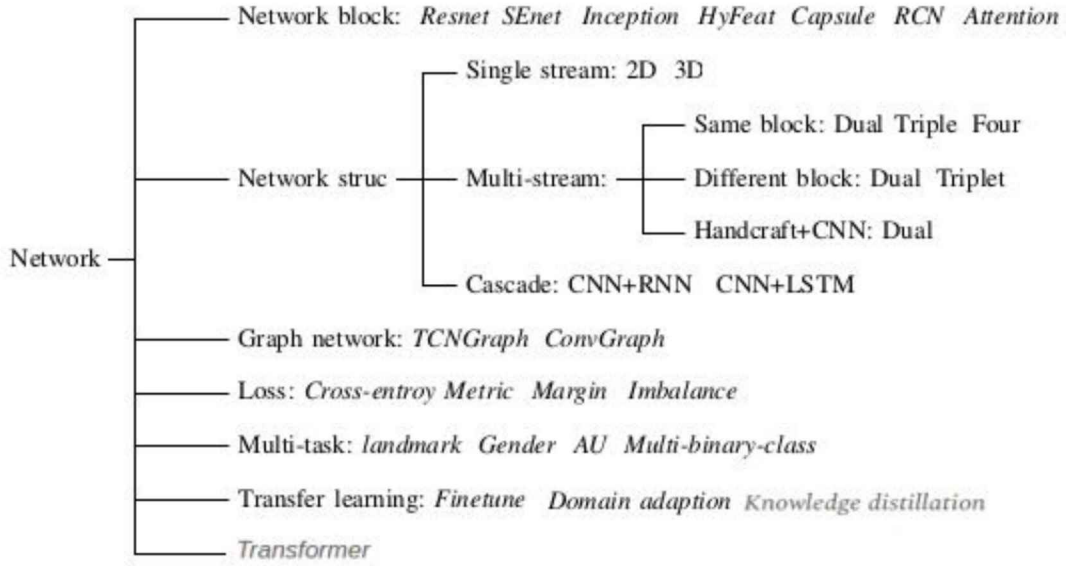


(a) Static image (b) Sequence (c) Frame aggereation (d) Image with dynamic infomation (e) Optical flow

* Image with dynamic information: a standard image that holds the dynamics of an entire video sequence in a single instance.

Li, Yante, et al. "Deep learning for micro-expression recognition: A survey." *IEEE Transactions on Affective Computing* (2022).

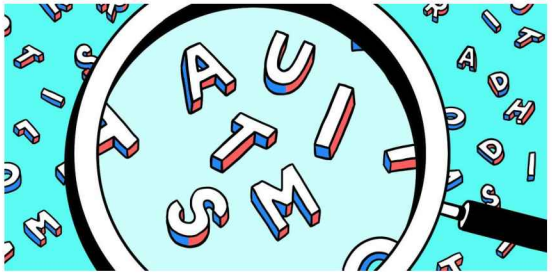
- Existing methods
  - Traditional classifiers
    - SVM, KNN, extreme learning machine (ELM), etc.
  - Deep learning based



Li, Yante, et al. "Deep learning for micro-expression recognition: A survey." *IEEE Transactions on Affective Computing* (2022).

Our Work

- Previous study on facial expression analysis for ASD mainly focus on macro-expression analysis
- However, micro-expression can reflect more true feelings of individuals



In our work:

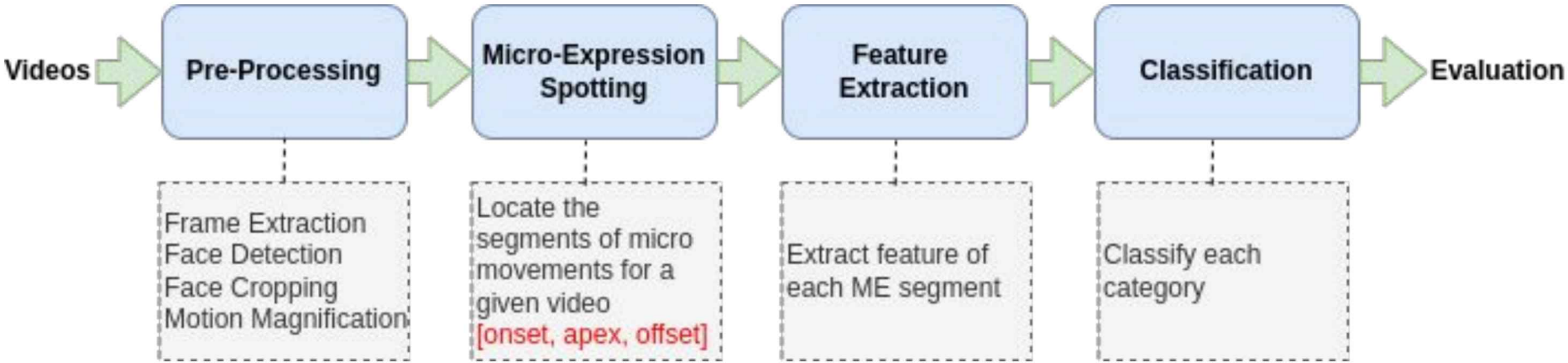
- Use computer vision and machine learning technology
- to analyze facial micro-expressions of participants
- in hour-long ADOS video sequences
- for the diagnosis of ASD

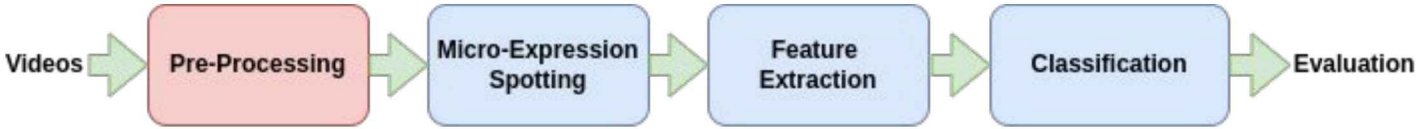
ADOS Videos

- Structured but natural discussion
- Participants with ASD underwent the ADOS interview
- ADOS: Autism Diagnostic Observation Schedule
  - 15 observation activities
  - Response to each activity is inspected

1. Construction Task
2. Telling A Story
3-4. Describing A Picture & Talking
5-7. Conversation on School, Work, Social Difficulties & Emotions
8. Demonstration Task
9. Cartoons
10. Break
11-14. Conversation on Daily Living, Relationships, Plans
15. Creating A Story

Pipeline

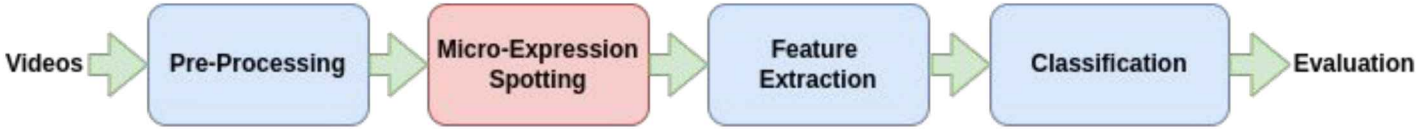

Videos
Pre-Processing
Micro-Expression Spotting
Feature Extraction
Classification
Evaluation
Frame Extraction
Face Detection
Face Cropping
Motion Magnification
Locate the segments of micro movements for a given video
[onset, apex, offset]
Extract feature of each ME segment
Classify each category



Pre-process

- Split the whole video into 15 separate subvideos based on the 15 tasks
- For the chosen tasks (5-7, 11-14)
  - Extract frames
  - Detect the face of the participant
  - Crop square face regions

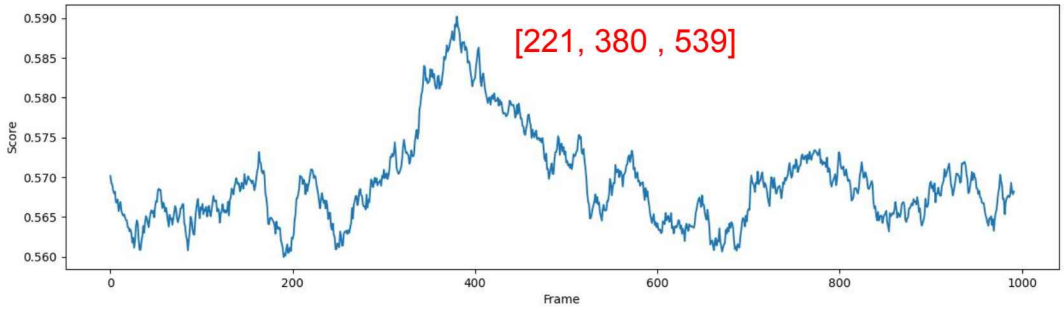
Dlib toolbox



- In our experiment
  - Spotting model is trained on CAS(ME)^2
    - 4 categories (more general terms):
      - Positive: Happiness
      - Negative: Anger, Disgust, Sadness, Fear
      - Surprise: Surprise
      - Others: Contempt, sympathy, confusion, helplessness, etc.
    - The emotion labeling criteria is based on the FACS coding results
  - For each spotted segment in ADOS videos
    - 35 frames for 30 fps (frame rate) videos
      - Calculated by averaging segment length of CAS(ME)^2

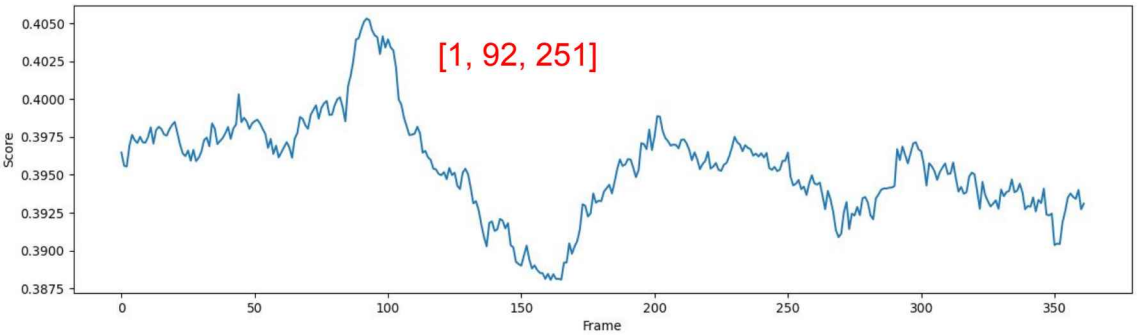
* 7 scenes

Qu, Fangbing, et al. "**CAS (ME) 2: A Database of Spontaneous Macro-expressions and Micro-expressions**." *18th International Conference, HCI International 2016, July 17-22, 2016.*



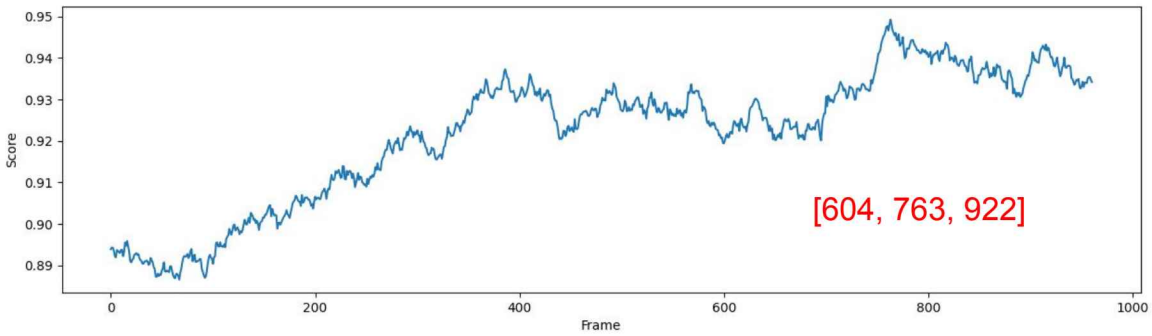
Onset Apex Offset

Eyebrows raised and pulled together



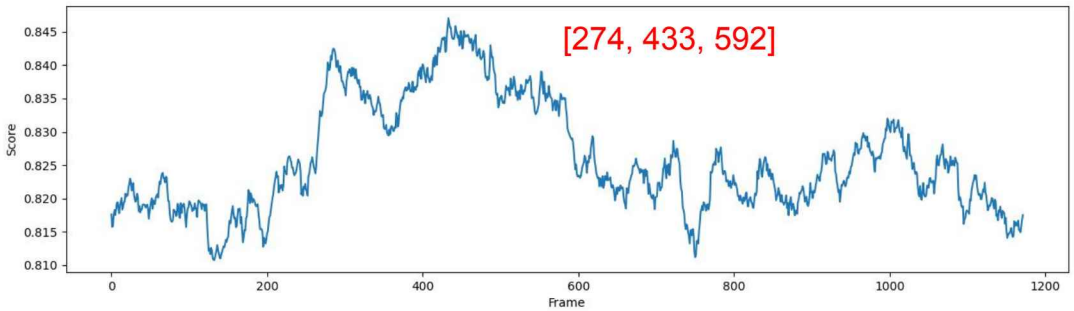
Onset Apex Offset

Narrowing of the lips; Drooping upper eyelids



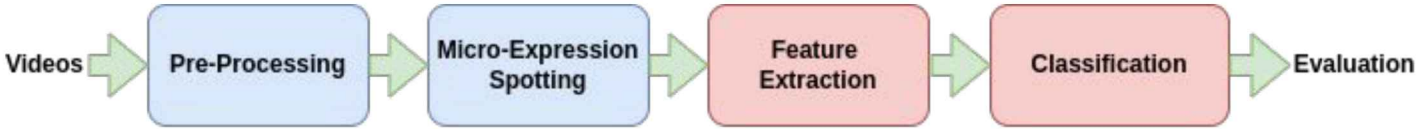
Onset Apex Offset

Narrowing of the lips; slight pulling down of lip corners; Drooping upper eyelids



Onset Apex Offset

Lose focus in eyes



Next work

- So far, we have obtained all the spotted intervals of ADOS videos
- Next step:
  - extract discriminative feature of each interval for autism diagnosis
- Basic Idea: use a transformer and optical flow
  - Transformer
    - inductive bias – modeling long-range relations between input patches
    - the pattern of ME is subtle and appears only in a part of each frame in the video
    - might seem more suitable for micro-expression analysis
  - optical flow
    - as an effective motion feature
    - can complement the lack of data

Transformer: Liu, Z.; Ning, J.; Cao, Y.; Wei, Y.; Zhang, Z.; Lin, S.; Hu, H. Video swin transformer. arXiv 2021, arXiv:2106.13230.
OF: Farnebäck, G. Two-frame motion estimation based on polynomial expansion. In Scandinavian Conference on Image Analysis; Springer: Berlin/Heidelberg, Germany, 2003; pp. 363–370.

Evaluation Protocol

- **10-fold cross-validation**
  - measures the performance of the model using 10% videos as a validation set and the rest as a training set.
  - We repeat the training and validation process 10 times
  - 10 **Accuracy/F1-score** values are averaged to produce the final result.

Summary

- Gave a brief overview of existing macro and micro expression analysis technology, and applications in ASD diagnosis field
- Reviewed the pipeline of the micro-expression analysis in ADOS videos
- Showed some spotting result
- Discussed a feasible plan for next step