Facial Micro-Expression Analysis on ADOS Videos

Na Zhang

- Facial Expression Analysis has attracted great interest over the past years

Digital Health

Human Machine Interaction

Behavior Analysis

Video Communication

- Socio-Emotion Interaction difficulties in ASD
  - Communication Disorder
  - Emotional Dysregulation
  - with 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

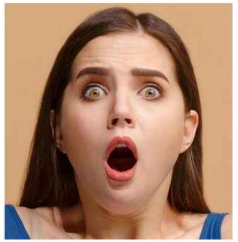
Rashidan, Mohammad Ariff, et al. "Technology-assisted emotion recognition for autism spectrum disorder (ASD) children: a systematic literature review." *IEEE Access* 9 (2021): 33638-33653.

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**



- Much research that have been embarked around recognizing human emotions, particularly for autistic children and individuals
- This study focuses on **analyzing emotions felt by the persons**

Micro- vs. Macro- Expression

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

Spontaneous

Posed

Macro-expression

Happy Disgust

Micro-expression

Happy Disgust

FACS, AU: Action Unit

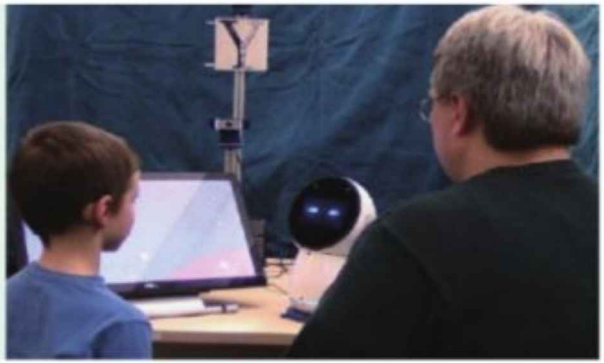
Allaert, B., Bilasco, I.M. and Djeraba, C., 2019. Micro and macro facial expression recognition using advanced Local Motion Patterns. *IEEE Transactions on Affective Computing*, *13*(1), pp.147-158.
Ben, X., Ren, Y., Zhang, J., Wang, S.J., Kpalma, K., Meng, W. and Liu, Y.J., 2021. Video-based facial micro-expression analysis: A survey of datasets, features and algorithms. *IEEE transactions on pattern analysis and machine intelligence*.

Motivation

- Micro-expression often reflects the true emotions that a person try to hide, suppress, mask, or conceal
- Especially important in high-risk situations

Lie Detection

Criminal Investigation



Clinical Diagnosis

Our work:

Analyze facial micro-expressions of participants in ADOS interview videos for autism 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.

ADOS Video

- Same scenes: 5-7 and 11-14
- **Two Categories**
  - **ASD: 42 videos**
  - **Control: 9+27 videos**
    - Raw: 9
    - Horizontal flipping: 9
    - Change brightness: 9
    - Histogram Equalization: 9

5-7. Conversation on School, Work, Social Difficulties & Emotions

11-14. Conversation on Daily Living, Relationships, Plans

ASD

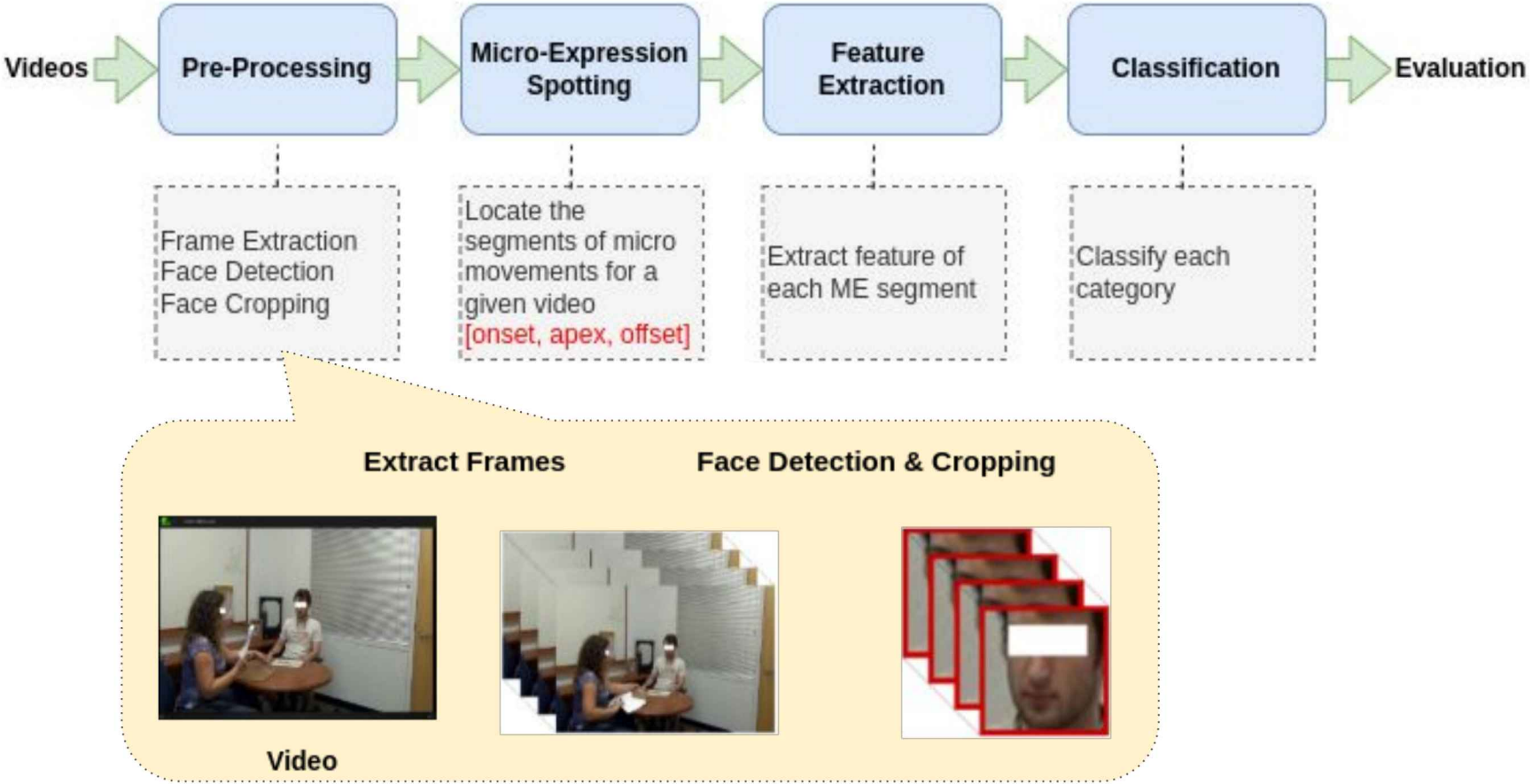
Control

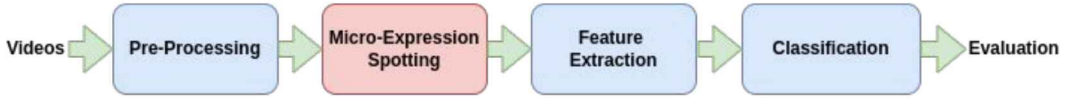
Horizontal flipping

Change brightness

Histogram Equalization

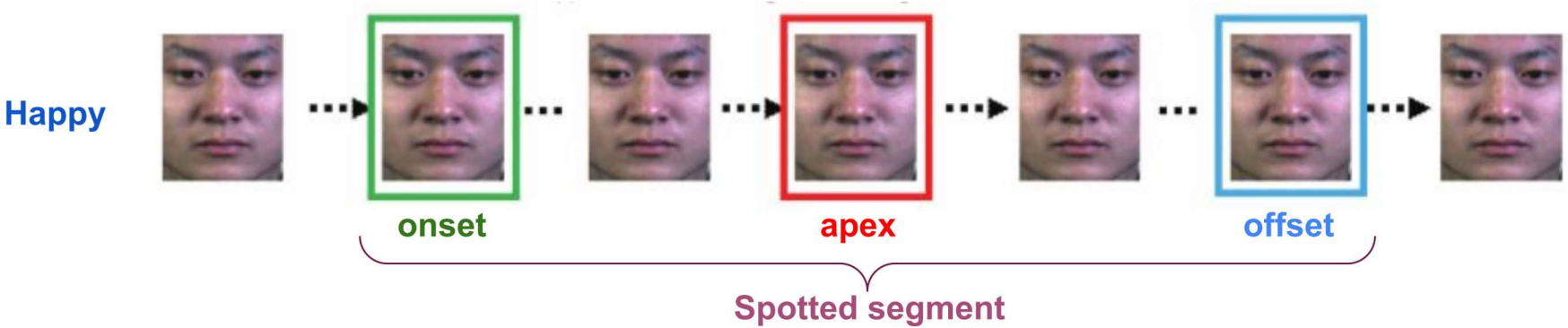
Pipeline


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



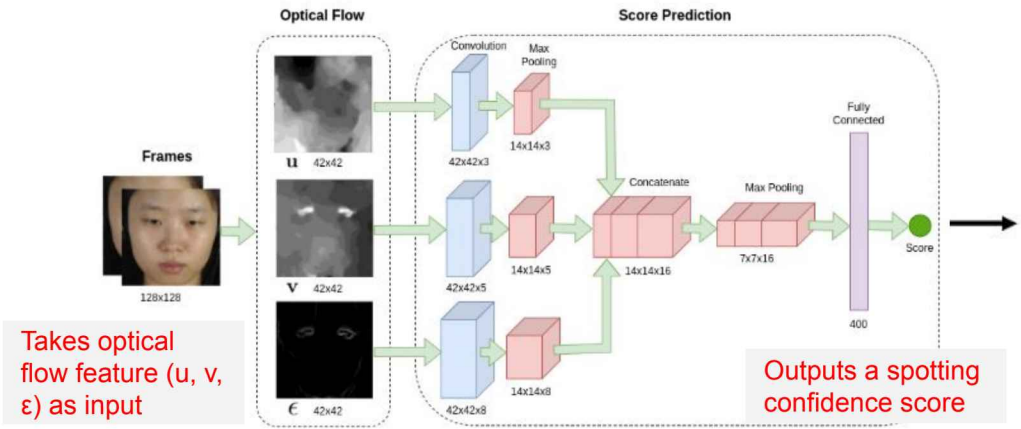
Micro-Expression Spotting

- 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



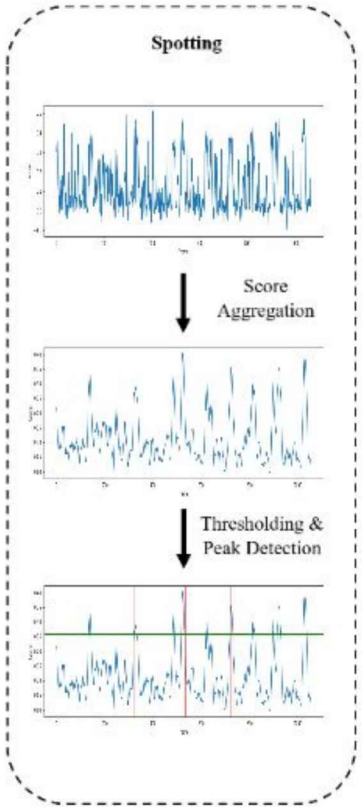
Pan, H., Xie, L., Wang, Z., Liu, B., Yang, M. and Tao, J., 2021. Review of micro-expression spotting and recognition in video sequences. *Virtual Reality & Intelligent Hardware*, *3*(1), pp.1-17.

- SOFTNet: a shallow optical flow three-stream CNN model
  - Regression problem
  - Predict a **score** indicating how likely a **frame** belongs to a micro-expression



Takes optical flow feature (u, v, ε) as input

Outputs a spotting confidence score



Get smoothed scores of all frames

Thresholding and peak detection

  - The micro-expression is **determined** if the frame's score is **above the threshold** set for peak detection

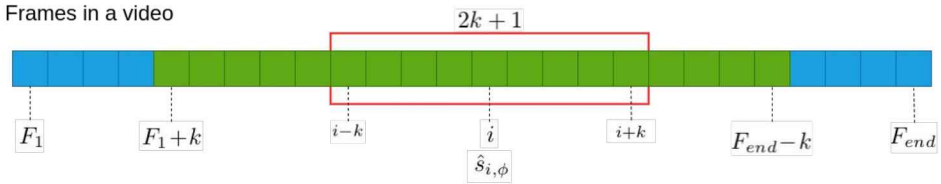
Liong, G.B., See, J. and Wong, L.K.. **Shallow optical flow three-stream CNN for macro-and micro-expression spotting from long videos**. *ICIP* 2021.
M. Shreve, J. Brizzi, S. Fefilatyev, T. Luguev, D. Goldgof, and S. Sarkar, "**Automatic expression spotting in videos**," Image and Vision Computing, 2014

Scores Aggregation: sliding window approach

- Predicted score for frame *i*:

$$\hat{s}_{i,\phi} = \frac{1}{2k+1} \sum_{j=i-k}^{i+k} \hat{s}_{j,\phi} \text{ for } i = F_1+k,\ldots,F_{end}-k$$

$k = (N+1)/2$ is half the average length of expression

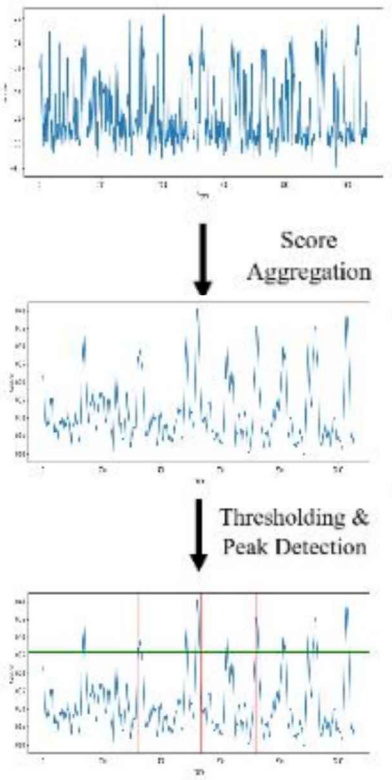


Thresholding & Peak Detection

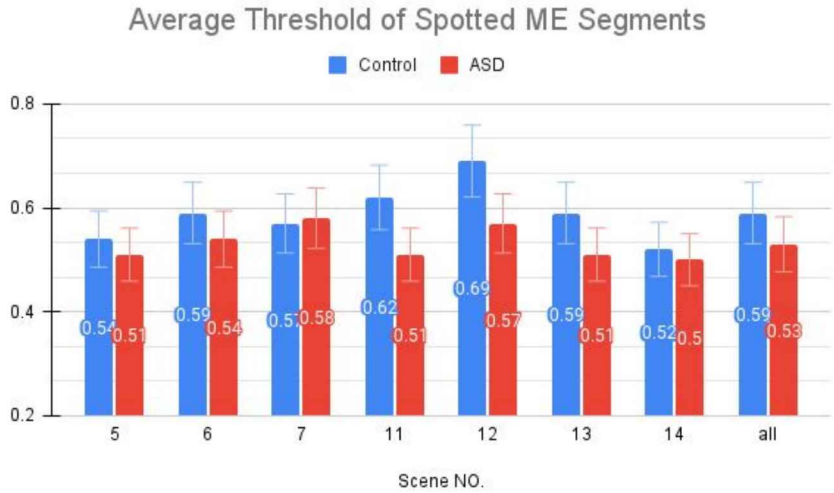
- Different input subvideos have different thresholds
- Threshold T:

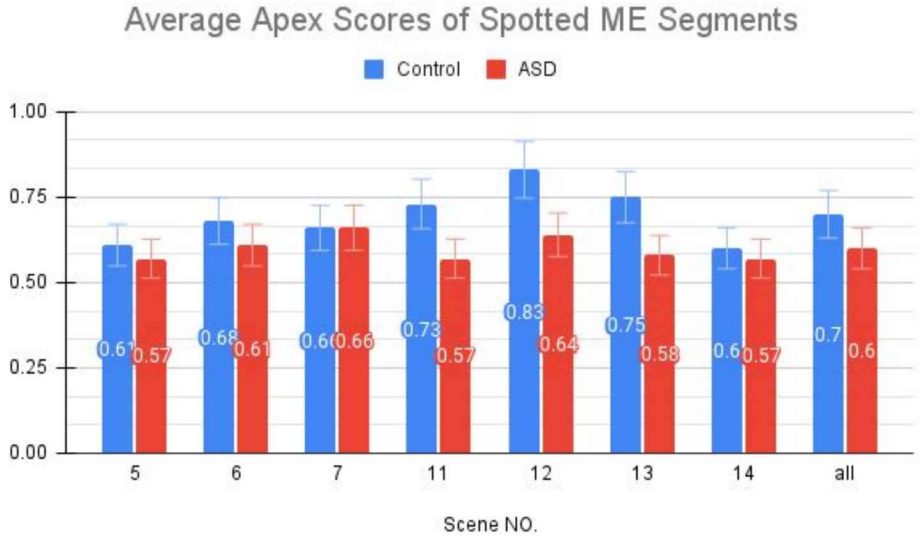
$$T = \hat{S}_{mean} + p \times (\hat{S}_{max} - \hat{S}_{mean})$$

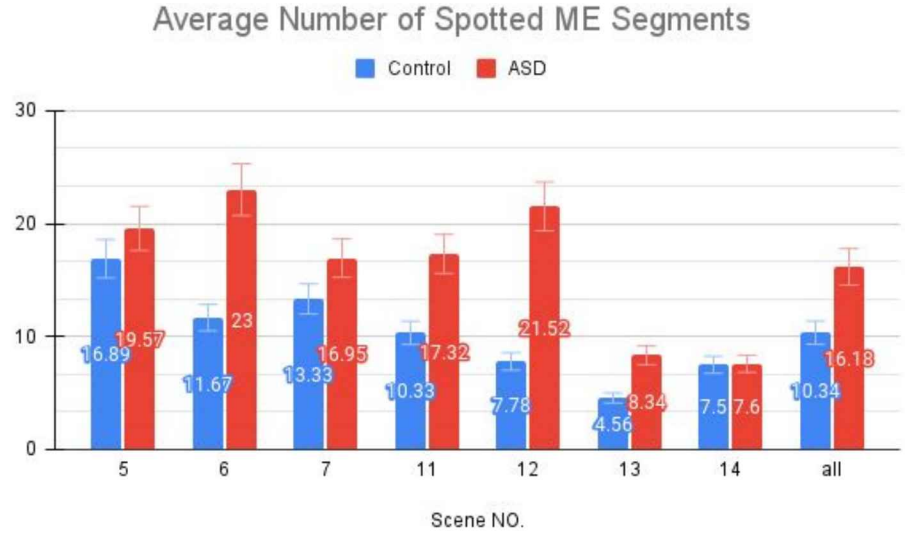
  - Ŝmean: the average predicted score over the given video
  - Ŝmax: the maximum predicted score over the given video
  - p: a tuning parameter in the range of [0, 1]



- Spotted Intervals:
  - A peak frame s is spotted by finding a local maxima
  - And extends by k frames to obtain the final spotted interval
    - **Ê = [s−k, s+k]**
    - 2k+1 frames in total







- Control group shows higher apex scores than ASD
- Control group shows higher threshold scores than ASD
- More segments in ASD group
- Seems ASD participants show more micro-expression

Samples of Spotted ME

Control

Onset Apex Offset

Corners of the lips are pulled; A wrinkle runs from outer nose to outer lip;

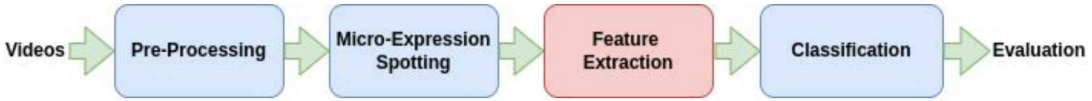
The eyebrows are raised and curved; Lip is Puckered; Eyes are widened

ASD

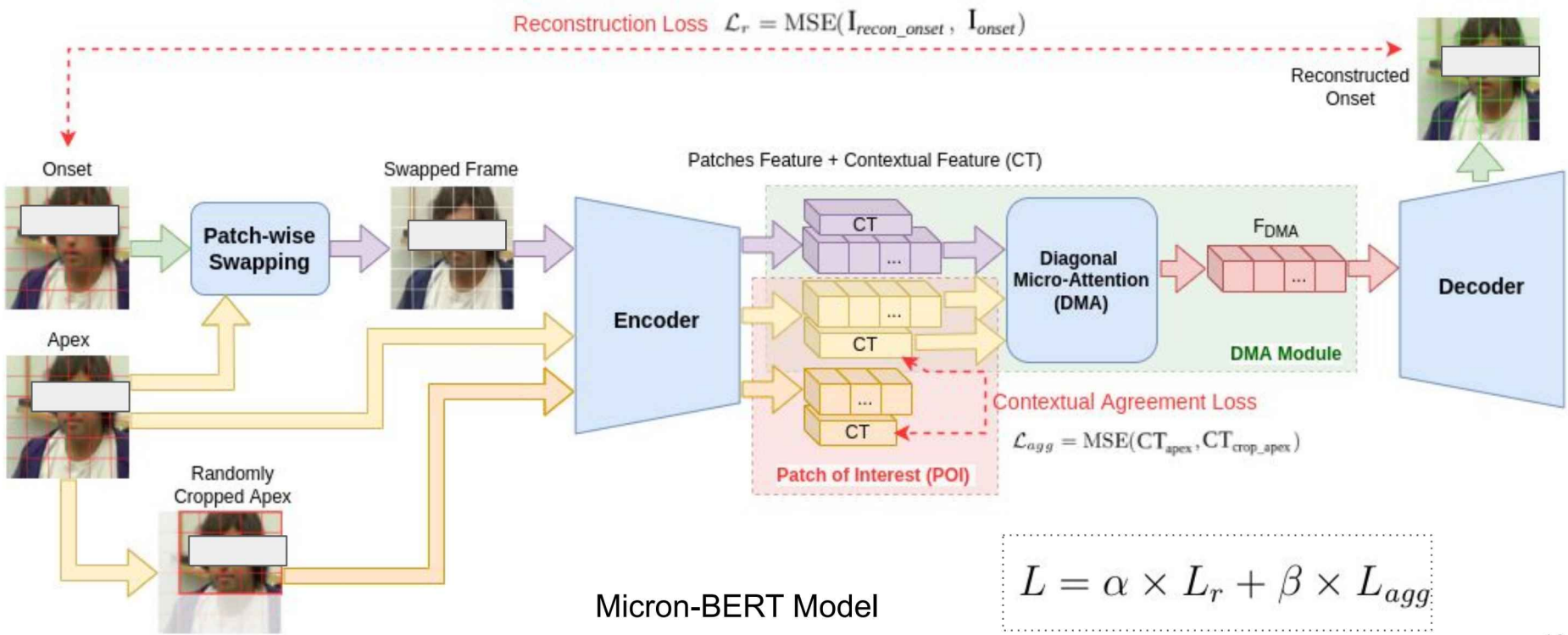
Onset Apex Offset

Lip corner tightened and raised on only one side of the face; Drooping upper eyelids

Corners of the lips are pulled; A wrinkle runs from outer nose to outer lip



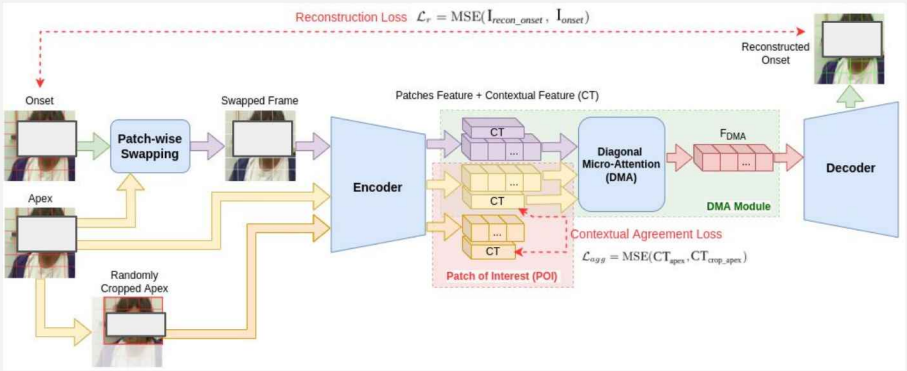
Micro-Expression Feature Extraction

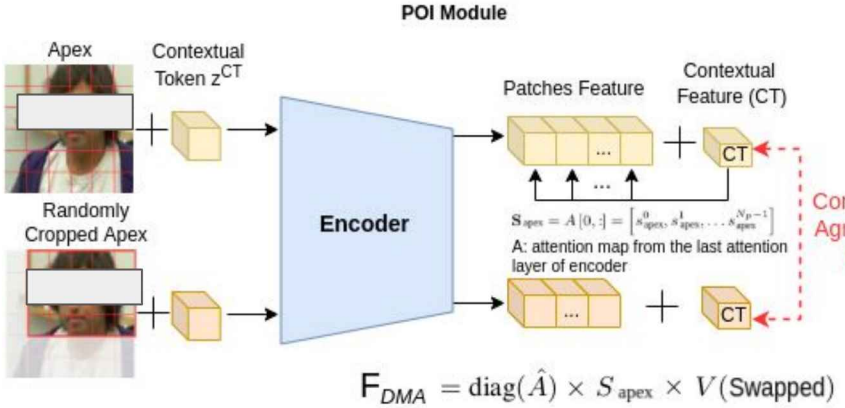


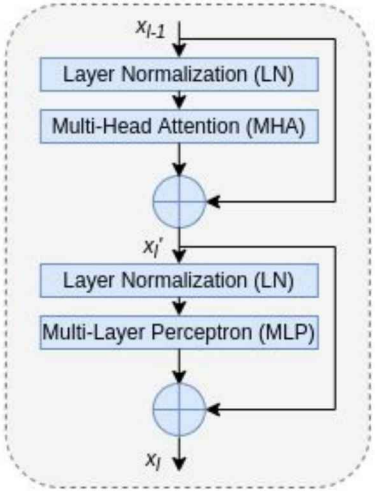
Micron-BERT Model

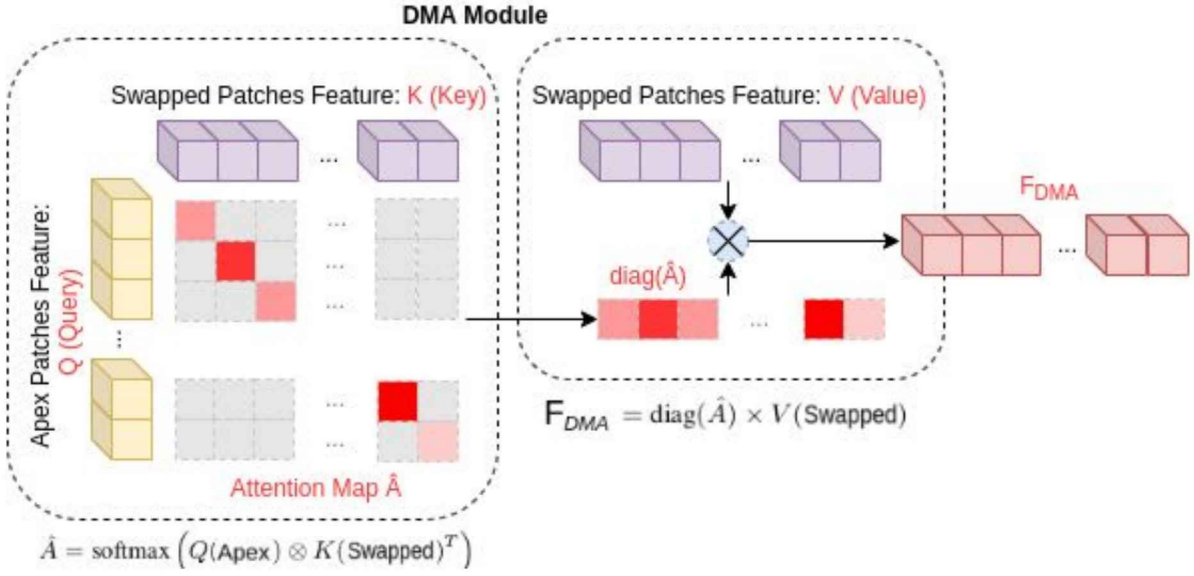
$$L = \alpha \times L_r + \beta \times L_{agg}$$

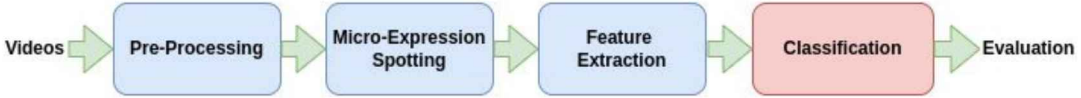
Nguyen, X. B., Duong, C. N., Li, X., Gauch, S., Seo, H. S., & Luu, K. (2023). Micron-BERT: BERT-based Facial Micro-Expression Recognition. *arXiv preprint arXiv:2304.03195*.


Reconstruction Loss $\mathcal{L}_r = \text{MSE}(\text{I}_{recon\_onset}, \text{I}_{onset})$
Reconstructed Onset
Onset
Patch-wise Swapping
Swapped Frame
Patches Feature + Contextual Feature (CT)
Encoder
CT
Diagonal Micro-Attention (DMA)
F_{DMA}
DMA Module
Decoder
Apex
Randomly Cropped Apex
Contextual Agreement Loss
$\mathcal{L}_{agg} = \text{MSE}(\text{CT}_{apex}, \text{CT}_{crop\_apex})$
Patch of Interest (POI)


POI Module
Apex
Contextual Token z^{CT}
Randomly Cropped Apex
Encoder
Patches Feature
Contextual Feature (CT)
CT
$\text{S}_{apex} = A[0,:] = [s^0_{apex}, s^1_{apex}, \ldots s^{N_p-1}_{apex}]$
A: attention map from the last attention layer of encoder
Contextual Agreement Loss
$\text{F}_{DMA} = \text{diag}(\hat{A}) \times S_{apex} \times V(\text{Swapped})$


Building block of Encoder and Decoder
x_{l-1}
Layer Normalization (LN)
Multi-Head Attention (MHA)
x_l'
Layer Normalization (LN)
Multi-Layer Perceptron (MLP)
x_l


DMA Module
Swapped Patches Feature: K (Key)
Swapped Patches Feature: V (Value)
Apex Patches Feature: Q (Query)
Attention Map Â
diag(Â)
F_{DMA}
$\text{F}_{DMA} = \text{diag}(\hat{A}) \times V(\text{Swapped})$
$\hat{A} = \text{softmax}\left(Q(\text{Apex}) \otimes K(\text{Swapped})^T\right)$



Classification

- For each scene [5-7, 11-14]
  - Binary classification
  - 10-fold cross-validation
- Scene-level fusion
  - Top 3, 5, 7

Table 4.7: Performance of our method on different scenes. Accu. - Accuracy.

Swapping Ratio	0		0.3		0.5		0.7		1.0	
	Accu.	F1 score	Accu.	F1 score	Accu.	F1 score	Accu.	F1 score	Accu.	F1 score
Scene 5	0.9482	0.9421	0.9481	0.9421	0.9482	0.9482	0.9421	0.9421	**0.9482**	0.9421
Scene 6	0.8982	0.8857	0.8982	0.8857	0.8857	0.8730	0.8982	0.8857	0.8982	0.8857
Scene 7	0.9107	0.9027	0.8982	0.8887	0.8982	0.8887	0.9125	0.9030	0.8982	0.8887
Scene 11	0.9458	0.9435	0.9458	0.9435	0.9333	0.9265	0.9458	0.9435	0.9333	0.9265
Scene 12	0.9357	0.9337	0.9357	0.9337	**0.9607**	**0.9603**	0.9232	0.9167	0.9232	0.9197
Scene 13	0.9446	0.9433	0.9446	0.9433	0.9446	0.9432	0.9446	0.9433	0.9446	**0.9433**
Scene 14	0.8917	0.8960	0.8917	0.8960	0.9042	0.913	0.9042	0.9130	0.9042	0.9130
Top 3	**0.9607**	**0.9590**	**0.9607**	**0.9590**	0.9482	0.9463	**0.9732**	**0.9730**	**0.9482**	0.9421
Top 5	0.9482	0.9463	**0.9607**	**0.9590**	0.9482	0.9421	0.9607	0.9603	**0.9482**	0.9421
Top 7	0.9482	0.9463	0.9482	0.9463	0.9357	0.9294	0.9482	0.9463	0.9357	0.9294

ASD vs. Control

- ASD participants, in high severity level, have more trouble making natural spontaneous expressions
  - are less expressive;
  - tend to remain expressionless [low spotting threshold]
  - less smiling
  - produce looks that are odd or difficult to interpret
  - sometimes give ambiguous looks