Multimodal Emotion Detection for Autistic Children with GAN-Based Occluded Face Completion and 3D CNN

Under the guidance of

Dr . J. Dhalia Sweetlin

Associate Professor, MIT

RAHUL PRASANTH D - 2020506070

BALA NATESH R M- 2020506018

SANJAY G - 2020506080

Objective

- Develop a system that can comprehensively identify and understand emotions exhibited by autistic children, with the help of the facial expressions, hand gestures and other body movements.

- Utilize GAN for handling of the occluded face, implement landmark detection, and predict facial action units along with the body movements to enhance emotion prediction accuracy

- Helps to understand how autistic children experience emotions by using information about where and when emotions happen on their faces, along with specific facial expressions, to develop ways to support and help them more effectively.

Problem Statement

- Address the issue of occluded facial expressions using Generative Adversarial Networks (GANs) to complete missing facial features, ensuring that the model can effectively analyse facial expressions even when parts of the face are hidden from view.

- To develop the 3D CNN model capable of finding and analysing the facial expression, hand gestures, body movements, foot movement, etc.

- To fuse the information from multiple modalities to enhance the accuracy and performance of emotion prediction for the autistic children.

- To evaluate the overall performance of the model by comparing with the classical emotional detection techniques for the autistic children..

Need for the System

Emotion detection for autistic children is essential for:

1. **Improved Communication:**Helping those with limited verbal skills express emotions.

2. **Early Diagnosis:** Identifying autism indicators for early intervention.

3. **Personalized Support:** Tailoring interventions to unique emotional needs.

4. **Timely Assistance:** Recognizing emotional distress for immediate support.

5. **Advancing Understanding:** Contributing to autism research and therapies.

6. **Empowerment:**Equipping autistic individuals to manage their emotions effectively.

Literature Survey

S.no	Title	Methodology	Performance Metrics	Limitations
1	**Facial Action Unit Detection via Adaptive Attention and Relation** IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 32, 2023	● Proposes an adaptive attention regression network to generate global attention maps for each AU. ● This is done while considering predefined attention regions and guidance from AU detection. ● The goal is to capture dependencies in both strongly correlated and weakly correlated facial regions.	**Average F1-frame Score:** The AAR method achieved an average F1-frame score of approximately 63.8 on the BP4D benchmark.	The Adaptive Attention Regression (AAR) network was tested on input images with misalignment errors and occlusions.

Literature Survey (Contd.)

S.no	Title	Methodology	Performance Metrics	Limitations
2	**Action unit classification for facial expression recognition using active learning and SVM** Yao, L., Wan, Y., Ni, H. *et al.* Action unit classification for facial expression recognition using active learning and SVM. *Multimed Tools Appl* 80, 24287–24301 (2021).	• They utilized active learning and support vector machine (SVM) algorithms to classify facial action units (AU) for human facial expression recognition. • SVM was utilized to classify different AUs and ultimately map them to their corresponding facial expressions.	Different facial expressions, regardless of being female or male, had different recognition rates ranging from 90% to 95% for females and from 83% to 95% for males.	• Of the seven facial expressions, five expressions (joy, sadness, anger, hate, and neutral) were recognized • But some expressions (anger, hate, and neutral) seemed to be difficult to recognize

Literature Survey (Contd.)

S.no	Title	Methodology	Performance Metrics	Limitations
3	**Are 3D Face Shapes Expressive Enough for Recognising Continuous Emotions and Action Unit Intensities?** IEEE TRANSACTIONS ON AFFECTIVE COMPUTING DOI 10.1109/TAFFC.2023.3280530	● The 3D face models are employed to estimate valence and arousal, which are dimensions of emotional expression ● The method also investigates how 3D face shapes perform in estimating the intensity of individual AUs. ● It evaluates the performance on datasets such as BP4D.	**MSE:** 2D Shape : 0.36 3D Shape : 0.53	● 3D face models require good quality images for alignment ,which may limit their use cases. ● Datasets are imbalanced in gender, ethnicity, age, and other appearance relevant traits

Literature Survey (Contd.)

S.no	Title	Methodology	Performance Metrics	Limitations
4	**Multitask, Multilabel, and Multidomain Learning With Convolutional Networks for Emotion Recognition** IEEE TRANSACTIONS ON CYBERNETICS, VOL. 52, NO. 6, JUNE 2022	• A multitask learning loss function to share a common feature representation with other related tasks. • Showed that emotion recognition benefits from jointly learning a model with a detector of facial action units (collective muscle movements).	**Accuracy:** • Single RestNet-50 = 54.3% • SJMT RestNet-50 =84.2%	This proposal addresses one of the challenges with discrete emotion recognition in the wild, which is the lack of large public labeled datasets

Literature Survey (Contd.)

S.no	Title	Methodology	Performance Metrics	Limitations
5	**Exploring Complexity of Facial Dynamics in Autism Spectrum Disorder** IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, VOL. 14, NO. 2, APRIL-JUNE 2023	• The front-facing camera of the tablet(in which movie scenes were played) was used to capture the face. • Facial landmarks' dynamics were then automatically computed using computer vision algorithms • They comparing the facial dynamics of the Autistic and Non-Autistic Children	**Accuracy:** (Cross-Validation using Decision Tree model) • Video1 = 77.5% • Video2 =74.3% • Video3 =73.8% • Video4 =67.2%	• Other measures of complexity might be more robust than the MSE • The study sample, has a limited number of ASD participants and did not have sufficient power to determine the impact of demographic characteristics on the results

Literature Survey (Contd.)

S.no	Title	Methodology	Performance Metrics	Limitations
6	**Real-time facial emotion recognition system among children with autism based on deep learning and IoT** Talaat, F.M. Real-time facial emotion recognition system among children with autism based on deep learning and IoT. *Neural Comput & Applic* 35, 12717–12728 (2023).	• This system proposes an enhanced deep learning(EDL) technique to classify the emotions using convolutional neural network. • The proposed emotion detection framework takes the benefit from using fog and IoT to reduce the latency for real-time detection with fast response and to be a location awareness.	**Training Accuracy:**0.963 **Validation Accuracy:**0.88	The limitation of the proposed technique is that it uses small dataset (limited scale) as the large number of real dataset is not available.

Literature Survey (Contd.)

S.no	Title	Methodology	Performance Metrics	Limitations
7	**Facial Action Unit Detection With Transformers** 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)	This paper proposes a method for detecting Facial Action Units (FAU), which define particular face muscle activity, from an input image	It shows the best average F1-score of 61.5	The major challenge with this dataset is the severity of the class imbalance and the variation in the head pose and expression
8	**Discriminative Few Shot Learning of Facial Dynamics in Interview Videos for Autism Trait Classification** IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, VOL. 14, NO. 2, APRIL-JUNE 2023	This paper attempts to fill this gap by developing a novel discriminative few shot learning method to analyze hour-long video data and exploring the fusion of facial dynamics for the trait classification of ASD	It achieve the best performance with an accuracy of 91.72% by fusing the seven selected scenes that are comparable to the standardized diagnostic scales	This Experiment adopts a scene-level feature fusion strategy, which requires manually splitting entire hour-long videos into 15 separate scenes by time markers and extracting facial-dynamics features of each scene.

Literature Survey (Contd.)

S.no	Title	Methodology	Performance Metrics	Limitations
9	**Deep Learning Approach for Emotion Recognition from Human Body Movements with Feedforward Deep Convolution Neural Networks** Rajaram, Santhoshkumar & Geetha, M..Procedia Computer Science. 152. 158-165. 10.1016/j.procs.2019.05.038.	● This system uses the FDCNN(FeedForward Deep CNN) to recognize the emotions. ● The action can be performed mostly by the head, hands and arm. These cues together and convey information of emotional states and the content in the interactions. ● With the support from psychological studies, identifying emotions from human body movement has plenty of applications.	**True Positive Rate:** Anger - 0.816 Joy - 0.887 Fear - 0.911 Sad - 0.962 Pride - 0.909	This system only aims to recognize the emotions of the adults, not for the children with autism spectrum disorder (ASD)

Literature Survey (Contd.)

S.no	Title	Methodology	Performance Metrics	Limitations
10	**An Immersive Computer-Mediated Caregiver-Child Interaction System for Young Children With Autism Spectrum Disorder** IEEE TRANSACTIONS ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING, VOL. 29, 2021.	• An immersive Computer-mediated Caregiver Child Interaction (C3I) system to help children with ASD practice IJA skills. • An immersive Computer-mediated Caregiver Child Interaction (C3I) system to help children with ASD practice initiation of joint attention (IJA) skills	• The time required for familiarization became less as the dyad participated more in the trials. • The increased response time of trial 10 as compared to trial 6 might have come from the combination of minor perturbation and tiredness.	The feasibility study had only one session per dyad with repetitive trials. In addition, the sample size was small and there was no control group. As such, the results of this feasibility study need to be considered with caution until a larger study verifies its generalizability

Novelty of Proposed Work

- **Multimodal Approach:** Utilizes a combination of facial expressions and body gestures for emotion detection, offering a more comprehensive view of emotions in autistic children.
- **GAN-Based Face Completion**: Introduces GAN (Generative Adversarial Network) for handling occluded faces, addressing a common challenge in emotion recognition.
- **3D CNN Analysis:** Employs 3D Convolutional Neural Networks to provide a deeper and more refined understanding of emotions, enhancing accuracy.
- **Tailored for Autism:** Customizes the approach to cater specifically to the unique challenges of emotion recognition in autistic children, contributing to a deeper understanding of their emotional experiences.
- **Early Diagnosis Potential:** The project's potential for early diagnosis of autism based on emotion recognition is a novel and valuable application of machine learning in the healthcare field.

Architecture

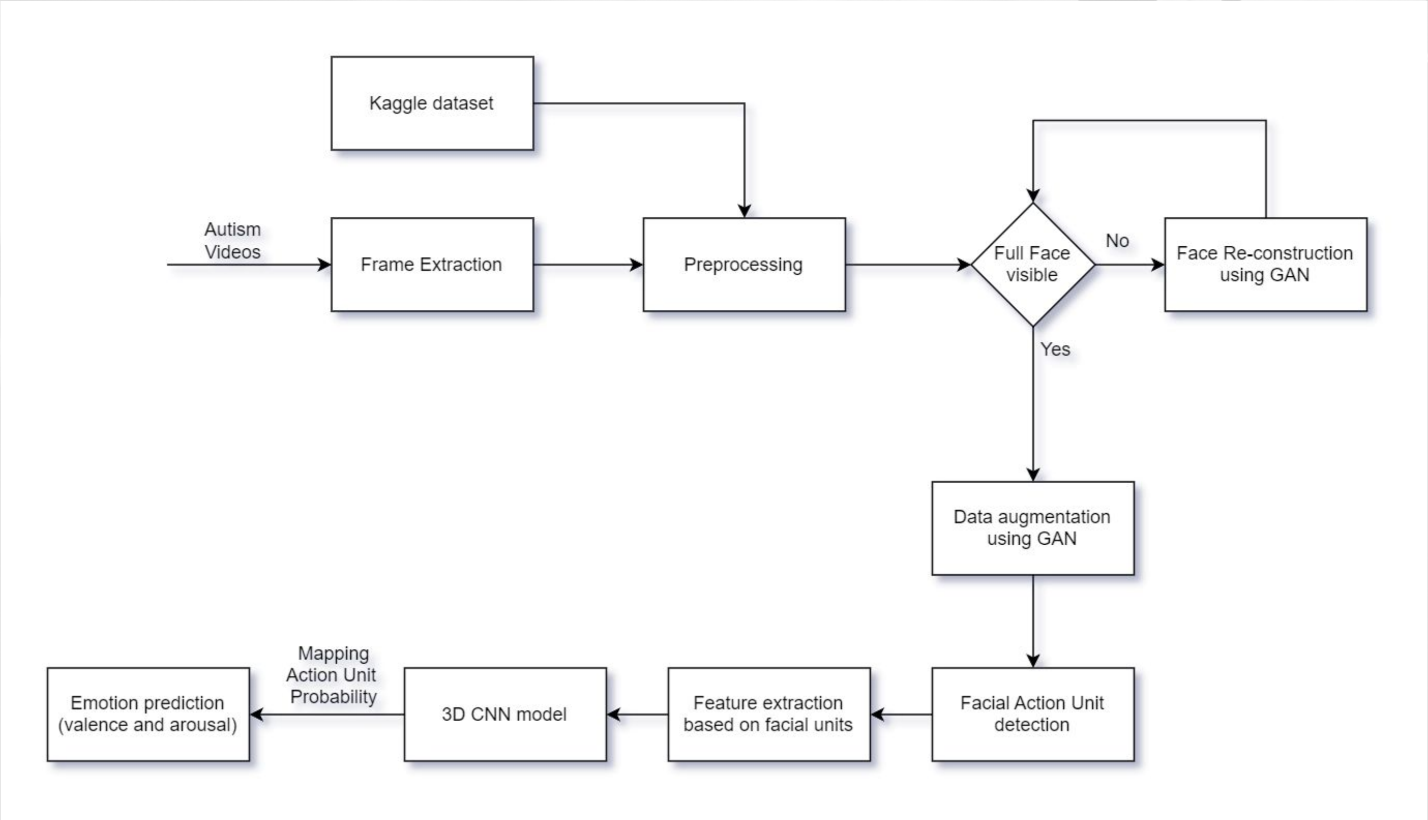


Fig 1: Facial Emotion Detection

Architecture (contd)

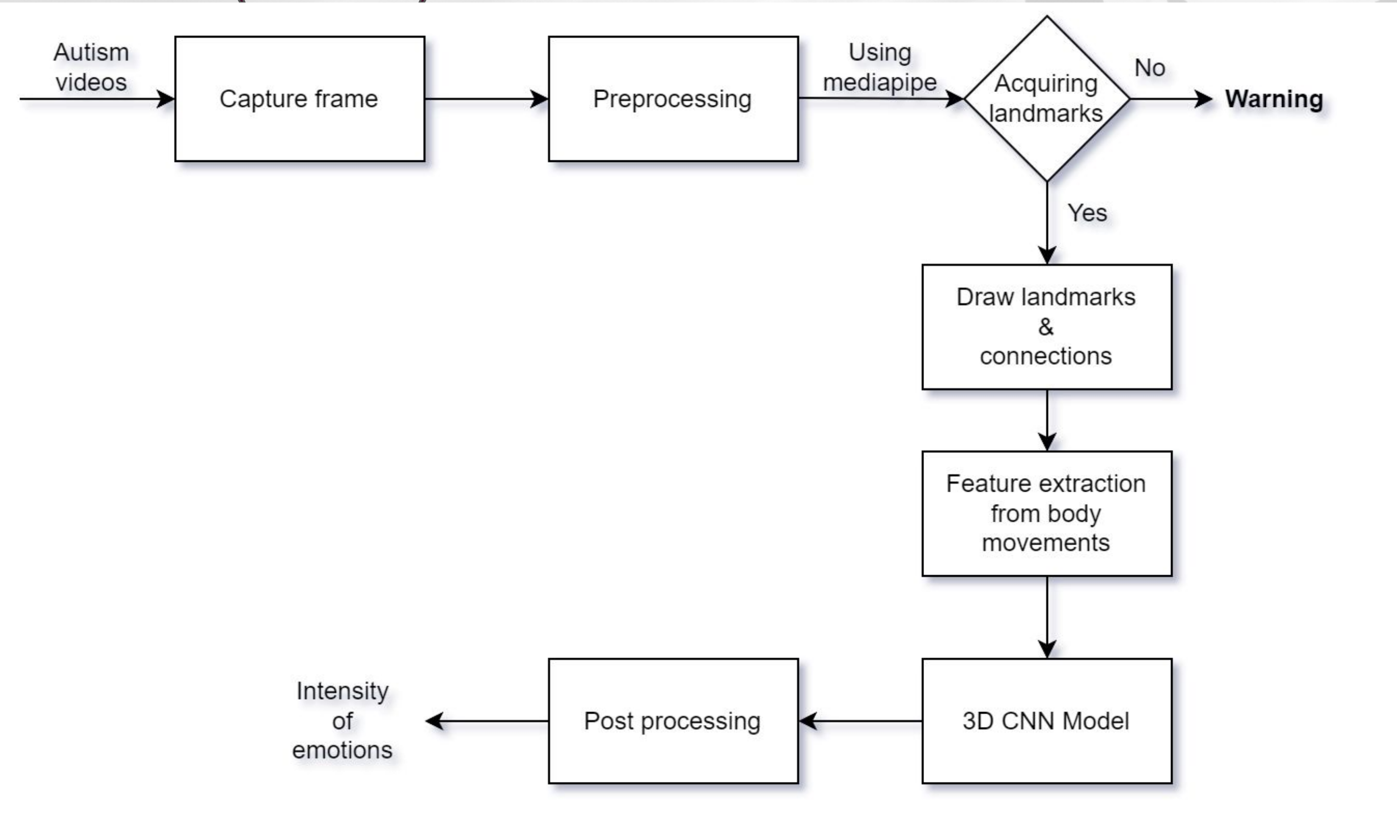


Figure:2 (Emotion detection from body languages, hand gestures, etc.)

Implementation

Data Augmentation

```
[15]: import albumentations as alb
```

```
[17]: augmentor = alb.Compose([ alb.HorizontalFlip(p=0.5),
                              alb.RandomBrightnessContrast(p=0.2),
                              alb.RandomGamma(p=0.2),
                              alb.RGBShift(p=0.2),
                              alb.VerticalFlip(p=0.5)],
                             bbox_params=alb.BboxParams(format='albumentations',
                                                        label_fields=['class_labels']))
```

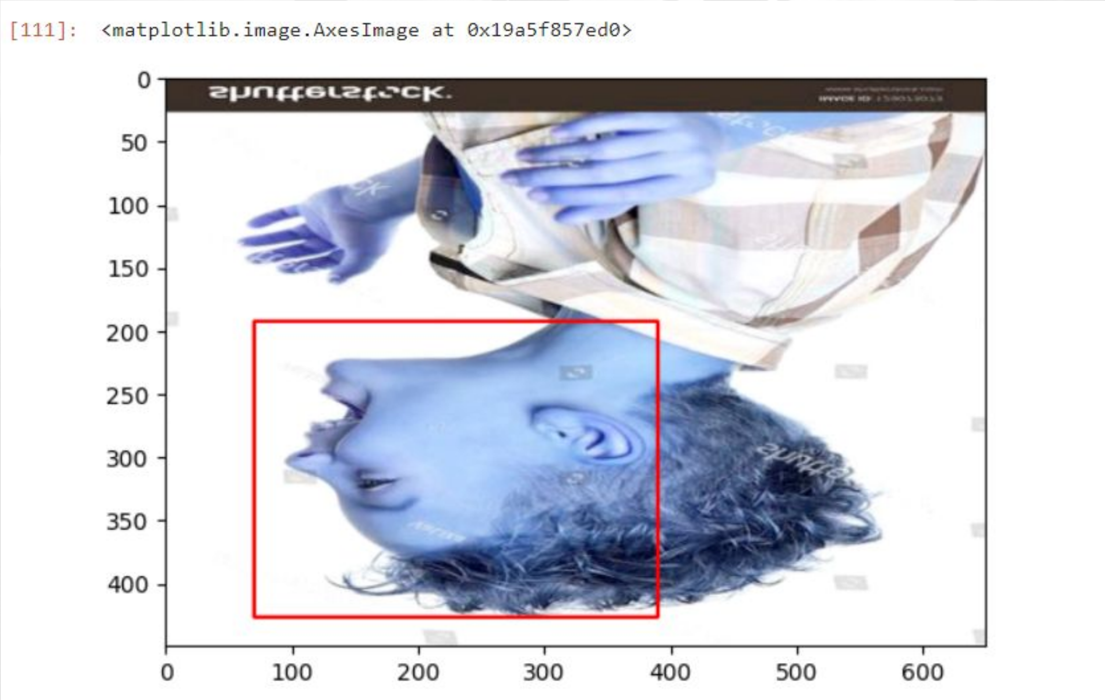
```
[108]: augmented = augmentor(image=img, bboxes=[coords], class_labels=['face'])
```

```
[110]: augmented['bboxes']

[110]: [(0.15727257853403137,
         0.430900959860384,
         0.8681691753926699,
         0.950261780104712)]
```

```
[111]: cv2.rectangle(augmented['image'],
                     tuple(np.multiply(augmented['bboxes'][0][:2], [450,450]).astype(int)),
                     tuple(np.multiply(augmented['bboxes'][0][2:], [450,450]).astype(int)),
                           (255,0,0), 2)

       plt.imshow(augmented['image'])
```

Implementation

Data Augmentation Output:



Implementation

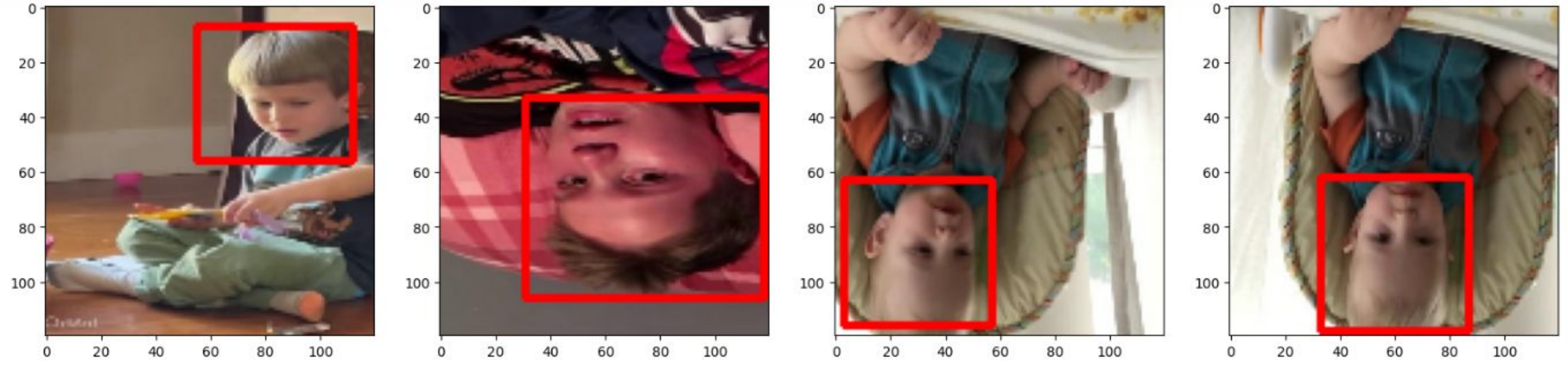
Face Recognition:

```
[136]: test_data = test.as_numpy_iterator()

[137]: test_sample = test_data.next()

[138]: yhat = facetracker.predict(test_sample[0])

1/1 [==============================] - 3s 3s/step

[139]: fig, ax = plt.subplots(ncols=4, figsize=(20,20))
for idx in range(4):
    sample_image = test_sample[0][idx]
    sample_coords = yhat[1][idx]
    #
    if yhat[0][idx] > 0.9:
        cv2.rectangle(sample_image,
                      tuple(np.multiply(sample_coords[:2], [120,120]).astype(int)),
                      tuple(np.multiply(sample_coords[2:], [120,120]).astype(int)),
                            (255,0,0), 2)

    ax[idx].imshow(sample_image)
```

Implementation

Face Recognition Output:

```
Clipping input data to the valid range for imshow with RGB data ([0..1] for floats or [0..255] for integers).
Clipping input data to the valid range for imshow with RGB data ([0..1] for floats or [0..255] for integers).
Clipping input data to the valid range for imshow with RGB data ([0..1] for floats or [0..255] for integers).
Clipping input data to the valid range for imshow with RGB data ([0..1] for floats or [0..255] for integers).
```



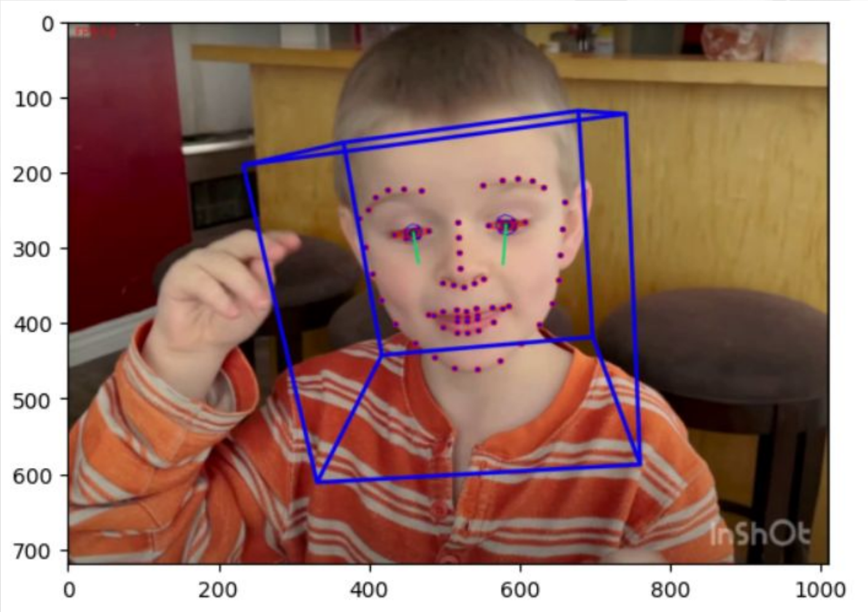
Implementation

Facial Action unit:

```
import cv2
import matplotlib.pyplot as plt
!FeatureExtraction.exe -f "J:\AUTISM DETECTION PROJECT\Action Unit Mapping\sample_images\Autism1.mp4"
video_file = 'processed/Autism1.avi'
cap = cv2.VideoCapture(video_file)
fig = plt.figure()
frame_number = 0
while True:
    ret, frame = cap.read()
    if not ret:
        break
    frame_number += 1
    if frame_number % 30 == 0:
        plt.imshow(cv2.cvtColor(frame, cv2.COLOR_BGR2RGB))
        plt.pause(0.01)
        plt.clf()
    if cv2.waitKey(25) & 0xFF == ord('q'):
        break
cap.release()
plt.close()
```

Implementation

Facial Action unit output:




```
1:
 AU02 0.0
 AU12 0.1
 AU15 0.1
 AU10 0.0
 AU06 0.0
 AU14 0.14
 AU25 1.33
 AU17 0.18
 AU23 0.0
 AU07 0.0
Value not found for  AU28
 AU09 0.0
 AU05 0.16
 AU20 0.0
 AU04 0.03
 AU01 0.23
 AU26 0.0
 AU45 0.0
```

Evaluation

Loss in Face recognition

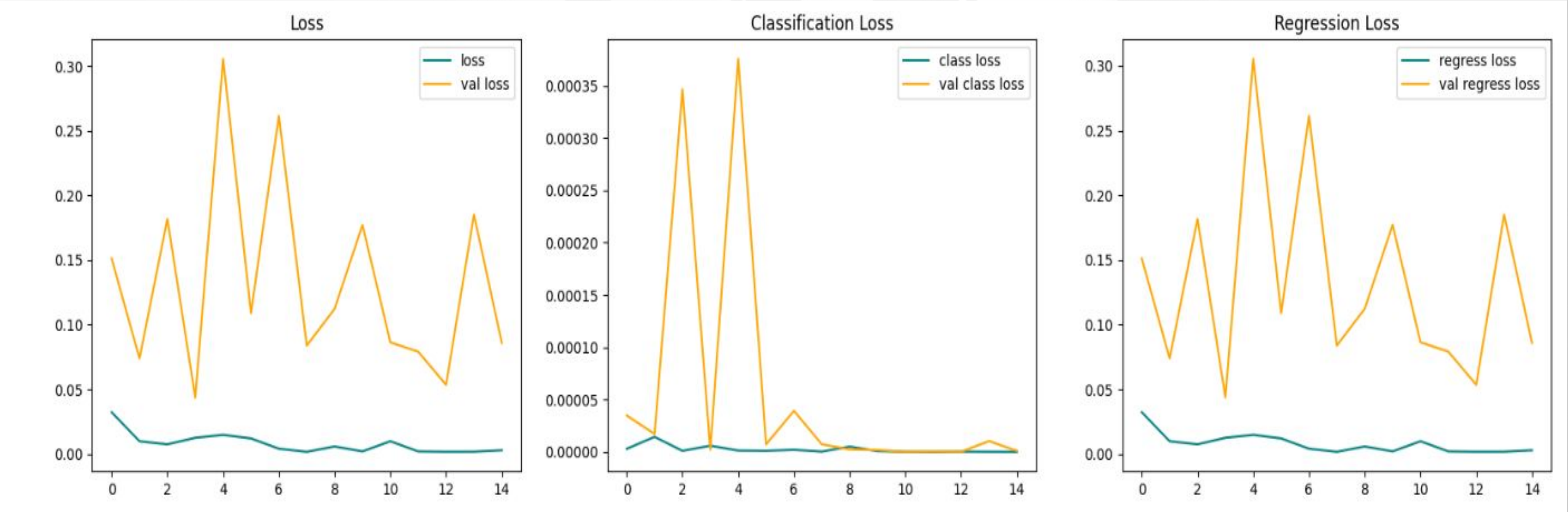
```
{'total_loss': [0.03217591717839241,
  0.009736192412674427,
  0.007351981475949287,
  0.012408425100147724,
  0.014659960754215717,
  0.011912941932678223,
  0.003950148820877075,
  0.0015429991763085127,
  0.005682547576725483,
  0.001880866358987987,
  0.009747179225087166,
  0.0018575640860944986,
  0.001604691380634904,
  0.001641246723011136,
  0.002755643799010086],
```

```
 'class_loss': [2.868484898499446e-06,
  1.4298380847321823e-05,
  9.387757700096699e-07,
  5.848765340488171e-06,
  1.2218966958243982e-06,
  9.685780923973653e-07,
  1.974414999494911e-06,
  2.0861644145497849e-07,
  4.999395059712697e-06,
  6.482011940533994e-07,
  8.940700269022273e-08,
  0.0,
  2.0116583243634523e-07,
  6.705523958316917e-08,
  0.0],
```

```
 'regress_loss': [0.0321744829416275,
  0.009729043580591679,
  0.007351512089371681,
  0.012405500747263432,
  0.0146593498060677,
  0.011912457644939423,
  0.003949161618947983,
  0.0015428948681801558,
  0.005680047906935215,
  0.0018805422587320209,
  0.009747134521603584,
  0.0018575640860944986,
  0.0016045907977968454,
  0.00164121319539845,
  0.002755643799010086],
```

Evaluation

Graphical representation of loss in Face recognition



Evaluation

Loss function in Face recognition

$$\lambda_{\mathbf{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right]$$

$$+ \lambda_{\mathbf{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right) \right.$$

Challenges

Developing a system to predict emotions in autistic children using facial images presents challenges and solutions as follows:

- **Limited Data:** Gather diverse data manually and use data augmentation
- **Complex Expression:** Employ transfer learning to accommodate individual expression styles and account for varied emotional patterns.
- **GAN Realism:** Fine-tune GAN architectures and apply perceptual loss to enhance generated face realism.
- **Landmark and Pose Detection:** Utilize advanced models for landmark and pose detection and preprocess images to improve accuracy.
- **Model Training:** Leverage cloud resources to manage 3D CNN complexity.

Solving these challenges will result in a precise and impactful emotion recogniton system for autistic children.

HW / SW Requirements

- **Hardware**
    - CPU: Multi-core processor (Intel Core i5 or equivalent)
    - RAM: Minimum 8 GB (16 GB or more recommended)
    - GPU: NVIDIA GPU (GTX 1060 or higher) for training deep learning models efficiently

- **Software**
    - Windows or Linux operating system
    - Python 3.x: Core programming language for most deep learning frameworks.
    - Anaconda : Package managers to manage Python environments and dependencies.

Workplan

0 — Data Collection and preprocessing of data — 1 — Handling of the missing face and occluded face in the image — 2 — Model Training and Testing — 3

References

[1] Z. Shao, Y. Zhou, J. Cai, H. Zhu and R. Yao, "Facial Action Unit Detection via Adaptive Attention and Relation," in IEEE Transactions on Image Processing, vol. 32, pp. 3354-3366, 2023, doi: 10.1109/TIP.2023.3277794. Link

[2] G. Pons and D. Masip, "Multitask, Multilabel, and Multidomain Learning With Convolutional Networks for Emotion Recognition," in IEEE Transactions on Cybernetics, vol. 52, no. 6, pp. 4764-4771, June 2022, doi: 10.1109/TCYB.2020.3036935. Link

[3] Yao, L., Wan, Y., Ni, H. et al. Action unit classification for facial expression recognition using active learning and SVM. Multimed Tools Appl 80, 24287–24301 (2021). Link

[4] M. K. Tellamekala, Ö. Sümer, B. W. Schuller, E. André, T. Giesbrecht and M. Valstar, "Are 3D Face Shapes Expressive Enough for Recognising Continuous Emotions and Action Unit Intensities?," in IEEE Transactions on Affective Computing, doi: 10.1109/TAFFC.2023.3280530. Link

References (Contd.)

[5] P. R. K. Babu et al., "Exploring Complexity of Facial Dynamics in Autism Spectrum Disorder," in IEEE Transactions on Affective Computing, vol. 14, no. 2, pp. 919-930, 1 April-June 2023, doi: 10.1109/TAFFC.2021.3113876. Link

[6] M. Talaat, Fatma. (2023). Real-time facial emotion recognition system among children with autism based on deep learning and IoT. Neural Computing and Applications. 35. 10.1007/s00521-023-08372-9. Link

[7] Geethu Miriam Jacob, Bjorn Stenger; Facial Action Unit Detection With Transformers, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 7680-7689 Link

[8] N. Zhang, M. Ruan, S. Wang, L. Paul and X. Li, "Discriminative Few Shot Learning of Facial Dynamics in Interview Videos for Autism Trait Classification," in IEEE Transactions on Affective Computing, vol. 14, no. 2, pp. 1110-1124, 1 April-June 2023, doi: 10.1109/TAFFC.2022.3178946. Link

References (Contd.)

[9] Rajaram, Santhoshkumar & Geetha, M.. (2019). Deep Learning Approach for Emotion Recognition from Human Body Movements with Feedforward Deep Convolution Neural Networks. Procedia Computer Science. 152. 158-165. 10.1016/j.procs.2019.05.038. Link

[10] Nie G, Ullal A, Zheng Z, Swanson AR, Weitlauf AS, Warren ZE, Sarkar N. An Immersive Computer-Mediated Caregiver-Child Interaction System for Young Children With Autism Spectrum Disorder. IEEE Trans Neural Syst Rehabil Eng. 2021;29:884-893. doi: 10.1109/TNSRE.2021.3077480. Epub 2021 May 18. PMID: 33945481; PMCID: PMC8189254. Link

[11] Stef van der Struijk, Hung-Hsuan Huang, Maryam Sadat Mirzaei, Toyoaki Nishida
FACSvatar: An Open Source Modular Framework for Real-Time FACS based Facial Animation.
IVA '18: Proceedings of the 18th International Conference on Intelligent Virtual Agents. Link

[12] Losses https://wikidocs.net/167690

THANK YOU