Unilateral Facial Action Unit Detection: Revealing Nuanced Facial Expressions

Deniz Iren
Department of Information Science
Open Universiteit
Heerlen, The Netherlands
deniz.iren@ou.nl

Daniel Stanley Tan
Department of Computer Science
Open Universiteit
Heerlen, The Netherlands
daniel.tan@ou.nl

***Abstract*—Facial expressions are essential for non-verbal human communication as they convey behavioral intentions and emotional states. While facial action units (AUs) can occur bilaterally or unilaterally, the existing research in affective computing predominantly concentrates on bilateral expressions, largely due to the lack of datasets with unilateral AU labels. In this study, we present a method for generating unilateral AU labels and assess its efficacy against expert-labeled facial images. Furthermore, we introduce a dedicated model trained on the generated data and evaluate its performance across multiple datasets. Our findings offer insights into feature extraction for unilateral facial expression recognition. This research contributes to advancing the understanding and recognition of nuanced facial expressions, with potential applications in various domains such as healthcare and human-computer interaction.**

***Index Terms*—Affective computing, facial expression recognition, action units, unilateral facial expressions**

I. Introduction

Facial expressions play a crucial role in non-verbal human communication by serving as a vehicle for conveying behavioral intentions [1] and expressing emotional states [2]. Facial expressions comprise actions of one or more muscle structures on the face, such as raising eyebrows and pulling the corners of the lips. Various methods were proposed to study facial expressions systematically. Facial Action Coding System (FACS) [3], [4] stands as the predominant framework for analyzing facial expressions. It provides a comprehensive catalog of facial movements categorized into Action Units (AU).

AUs are discrete components describing observable movements of facial muscles that relate to both overt and subtle expressions. Furthermore, AUs may exhibit bilateral symmetry or unilateral manifestation [5] which adds complexity to facial expression dynamics. The diverse configurations of AUs on the face offer a broad spectrum of expressions, essential for effective communication in social contexts. Unilateral AUs provide signals distinct from their bilateral counterparts. For instance, in Western societies, raising a single eyebrow signifies curiosity, whereas the simultaneous elevation of both eyebrows typically indicates surprise or fear. Similarly, a smile characterized by raised cheeks commonly denotes happiness, while the elevation of a single cheek may suggest a smirk. Furthermore, individuals affected by conditions like facial palsy often experience unilateral impairment which leaves only one side of their face functional [6]. Consequently, the non-verbal cues they can exhibit are restricted to unilateral expressions.

Facial Expression Recognition (FER) [7] falls within the domain of affective computing, with the goal of identifying and categorizing expressions exhibited on human faces. Despite significant advancements in algorithms and the availability of pre-trained models for automated AU detection, unilateral AU detection remains largely unexplored within existing literature. This gap is particularly striking considering the potential significance of unilateral expressions in non-verbal human communication.

This paper seeks to bridge this gap by introducing a novel model for unilateral AU detection. Several challenges need to be addressed to achieve this goal. First, there is a notable scarcity of labeled data specifically focusing on unilateral AUs [8], which presents an obstacle in training and evaluating such models. Secondly, the algorithms and feature extraction methods commonly employed for facial analysis are not inherently designed to handle unilateral facial expressions which necessitates new approaches to representation and analysis.

Therefore, our research questions are as follows.

- How can facial image data annotated with unilateral AUs be effectively generated?
- What novel methods can be employed to represent unilateral facial features for the identification of unilateral AUs?

This study aims to contribute to the advancement of facial expression recognition by offering insights into the detection and interpretation of unilateral facial movements, thus, facilitating a deeper understanding of non-verbal human communication. Our specific proposed contributions include a method for generating unilateral AU labels on facial imagery, a model for detecting unilateral facial expressions, and insights into the prevalence and characteristics of unilateral AUs. These contributions hold potential applications in various domains, such as healthcare for diagnosing and treating facial palsy [8], training actors in facial expressiveness [9], and developing virtual agents with more authentic and nuanced facial expressions [10], [11].

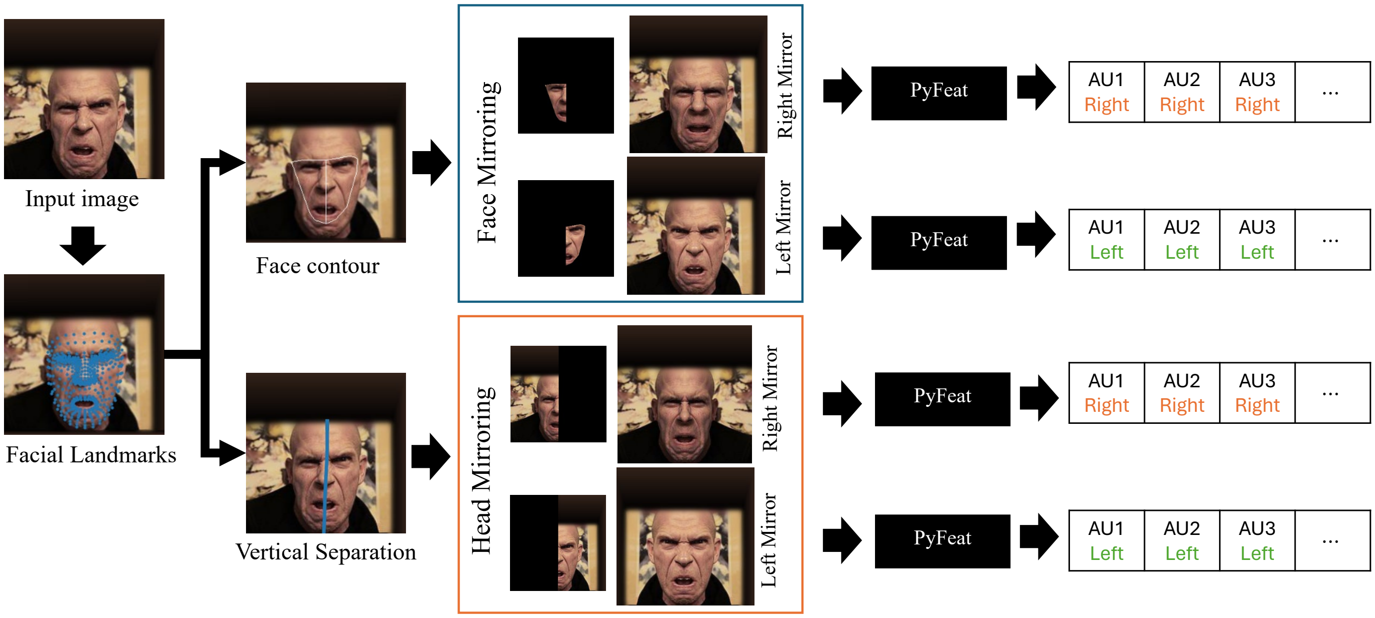

Fig. 1. Given a face image, we derive unilateral AUs by splitting the face in half and mirroring them. This produces two full face images coming from the left mirrored and right mirrored face. We can then use existing full face models to obtain the facial AUs of each side.

The subsequent sections of this paper are structured as follows. Section II comprises the background and related work on unilateral facial expressions and computational approaches to their detection. Section III covers the details of our unilateral AU label generation method. Section IV presents our approach to unilateral AU detection modeling. Section V discloses the results. Finally, Section VI discusses the theoretical and practical implications of our contributions, and concludes the paper.

II. Background and Related Work

A. Unilateral Facial Expressions

Facial actions are considered bilateral when muscle activity associated with a specific AU is observable on both sides of the face. Conversely, a unilateral action occurs when evidence of an AU is present on one side of the face but absent on the other side [5]. While most AUs can manifest bilaterally, by definition, some AUs can only appear unilaterally. For example, *AU46* denotes a *wink*, a movement inherently performed with only one eye. Furthermore, the unilateral manifestation of an expression often conveys a distinct meaning compared to its bilateral counterpart. Additionally, the display of subtle facial movements unilaterally can convey important information about an individual's disposition [12]. For instance, the so-called dominant smile is characterized by the asymmetrical manifestation of AU12 - Lip Corner Puller [13].

Studies examining unilateral facial expressions generally fall into two categories. The first group focuses on neurological or facial muscle control disorders, such as facial palsy, where one side of the face experiences a loss of voluntary muscle control. Diagnosis and grading of facial palsy typically involve assessing facial symmetry and comparing the range of motion of both sides of the face [6].

The systematic review of physical rehabilitation of facial palsy [6] covers examples of techniques based on the functional symmetry of facial muscles. For instance, recent studies in automated palsy grading include the comparison of facial function using FER [14], [15].

The second group of studies in the literature examines variations in emotional expressiveness between the two sides of the face [16], [17], as well as the cultural significance of specific unilateral facial expressions [18]. These studies highlight the prevalence of unilateral AUs as a distinct area of research alongside bilateral AUs, indicating a need for further research to develop methods that enable the examination of unilateral facial expressions.

B. Related Work on Unilateral AU Detection

Despite the significance of unilateral AUs in conveying nuanced non-verbal cues, they are rarely labeled in commonly available datasets. For instance, the Affectiva-MIT Facial Expression Dataset [19] includes unilateral labels for a limited subset of AUs, such as *AU12 (Lip Corner Puller)* and *AU14 (Dimpler)*. However, training supervised unilateral AU models requires more extensive labeled datasets, which are challenging and expensive to manually curate. To alleviate this issue, researchers have explored unconventional methods for AU labeling. One such approach is demonstrated in FaraPy [8], an augmented reality system designed for mirror therapy in facial palsy treatment. This method involves comparing the intensity of AUs on both sides of the face. To estimate the intensity of unilateral AUs, FaraPy utilizes intermediate visualizations from another pre-trained model [20]. The authors

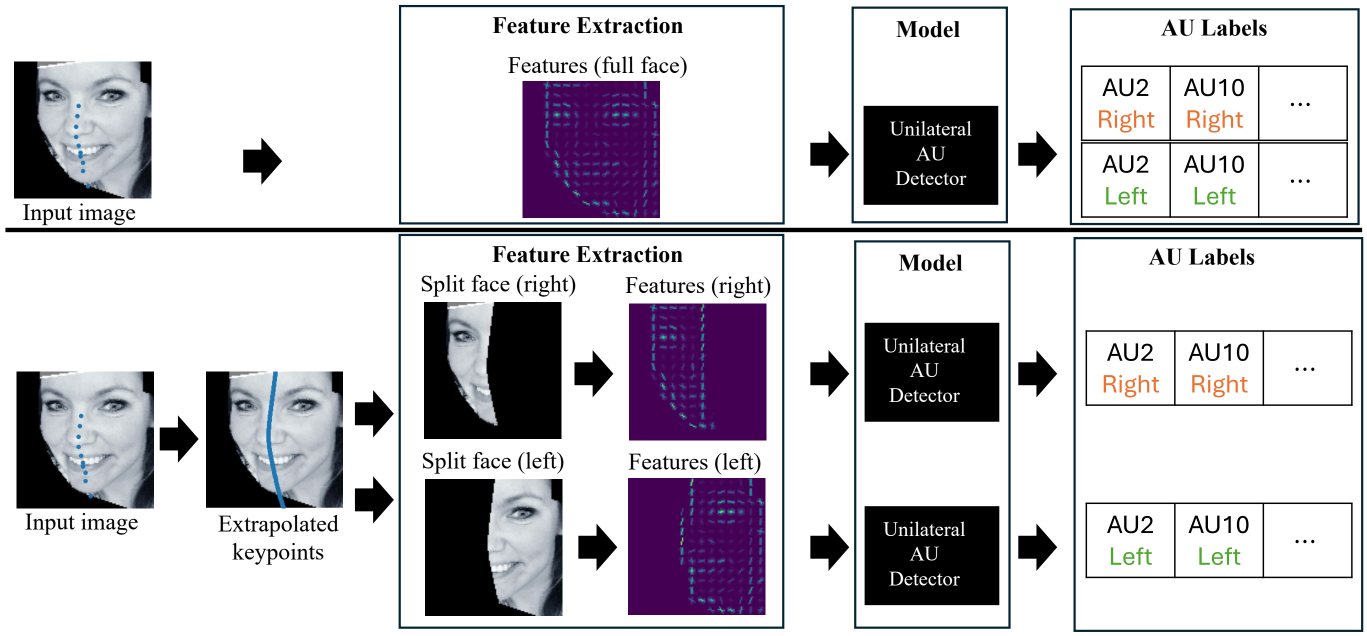

Fig. 2. Our unilateral AU detector first extracts facial features from either the full face (top) or the half face (bottom) and trains a classifier to predict left and right facial AUs.

assess performance by benchmarking against the outcomes of the same pre-trained model. In another approach, researchers aiming to develop avatars with dynamic facial expressions have utilized vertically split facial landmarks as features for unilateral AU classification [11]. These related works highlight the need for a robust and dedicated unilateral AU detection model.

III. Deriving unilateral AU labels

Current datasets commonly used in FER predominantly include bilateral AUs, and they rarely indicate laterality of the expression. Therefore, to effectively train and evaluate our model, we derived unilateral AU labels by generating mirrored versions of each side of the face and automatically labeling AUs on the mirrored images. This method ensures that the presence or absence of AUs is solely attributed to one side of the face. An overview of our approach is depicted in Figure 1, comprising the following steps.

Initially, we vertically split the face into two halves, resulting in separate images representing the right and left sides of the face (see Section III-A). Subsequently, we explore two approaches to reconstructing the complete facial images: either by mirroring the entire head (see Section III-B1) or mirroring only the facial regions (see Section III-B2). Finally, we employ a facial AU detector on each mirrored side (see Section III-C).

A. Vertical split

To determine the splitting point of the face, we initially detect facial landmarks using MediaPipe [21]. This results in a face mesh comprising 478 points, representing facial landmarks in detail such as those along the eyes, nose, mouth, eyebrows, and jawline (see Figure 1, left). Utilizing these facial landmarks, we normalize the face by performing rotations to straighten and center it within the image. To establish the vertical separation line, we identify landmarks keypoints corresponding to the forehead, center of the nose, middle of the lips, and middle of the chin. Subsequently, we employ a linear regression model to fit a line through these points.

B. Mirroring

The objective of mirroring is to synthesize new complete facial images from the split halves so that the synthesized facial image can be processed by existing AU detection models. Two alternative design choices are considered for mirroring: mirroring the entire head or mirroring only the face.

1) Head mirroring: Mirroring the entire head results in a new facial image with symmetrical details. While this ensures that all facial details and AUs originate from one side of the face, it may alter aspects such as hairstyle and face shape, particularly if they were initially asymmetrical.

To perform head mirroring, the face is divided into left and right halves along the separating line. These halves are then horizontally flipped and merged with the original halves along the separating line (see Figure 1, bottom). As a result, two complete facial images are generated comprising mirrored halves as shown in Figure 1 (bottom).

2) Face mirroring: Mirroring only the face aims to preserve details of the original image such as hairstyle and face shape.

From the facial landmarks, the region defined by the contour around landmark keypoints traversing the eyebrows, cheeks, lips, chin, and nose is carved out. The left and right sides of the face region are then split and horizontally flipped. To mitigate sharp seams or discontinuities around the edge of the

TABLE I
PERFORMANCE OF UNILATERAL AUS DERIVED FROM MIRRORING THE ENTIRE HEAD, EVALUATED ON THE EXPERT ANNOTATED DATA. THE NOTATION (_L) DENOTES LEFT, AND (_R) DENOTES RIGHT.

Metric	Mean	†AU02_L	AU04_L	AU10_L	AU15_L	AU25_L	AU26_L	AU28_L	AU43_L
precision	0.76	0.00	1.00	0.96	1.00	0.89	0.33	1.00	1.00
recall	0.92	0.00	1.00	0.85	0.71	0.96	1.00	1.00	1.00
f-score	0.75	0.00	1.00	0.90	0.83	0.92	0.50	1.00	1.00
accuracy	0.97	0.97	1.00	0.92	0.97	0.94	0.94	1.00	1.00
		AU02_R	AU04_R	AU10_R	AU15_R	AU25_R	AU26_R	AU28_R	AU43_R
		0.83	0.71	1.00	0.75	0.84	0.60	0.20	1.00
		1.00	1.00	0.96	1.00	1.00	1.00	1.00	1.00
		0.91	0.83	0.98	0.86	0.91	0.75	0.33	1.00
		0.98	0.97	0.98	0.98	0.92	0.94	0.94	1.00

†Note that for the small annotated samples, there were no instances of AU02_L, hence precision, recall, and f-score are 0.

TABLE II
PERFORMANCE OF UNILATERAL ACTION AUS DERIVED FROM MIRRORING THE ONLY THE FACE, EVALUATED ON THE EXPERT ANNOTATED DATA. THE NOTATION (_L) DENOTES LEFT, AND (_R) DENOTES RIGHT.

Metric	Mean	†AU02_L	AU04_L	AU10_L	AU15_L	AU25_L	AU26_L	AU28_L	AU43_L
precision	0.38	0.00	0.75	0.78	0.50	0.96	0.22	0.00	1.00
recall	0.57	0.00	1.00	0.81	0.29	0.92	1.00	0.00	1.00
f-score	0.38	0.00	0.86	0.79	0.36	0.94	0.36	0.00	1.00
accuracy	0.93	0.94	0.98	0.83	0.89	0.95	0.89	0.97	1.00
		AU02_R	AU04_R	AU10_R	AU15_R	AU25_R	AU26_R	AU28_R	AU43_R
		0.50	0.67	0.88	0.33	0.83	0.30	0.00	0.67
		0.40	0.40	0.85	0.33	0.96	0.50	0.00	1.00
		0.44	0.50	0.87	0.33	0.89	0.38	0.00	0.80
		0.92	0.94	0.89	0.94	0.91	0.84	0.95	0.98

†Note that for the small annotated samples, there were no instances of AU02_L, hence precision, recall, and f-score are 0.

copied facial region, Gaussian blurring is applied (see Figure 1, top).

C. AU Prediction

As a result of the mirroring step, two facial images are obtained: one mirrored from the left half of the face and another mirrored from the right half. At this point, a standard facial AU detection model can be used to label the left and right facial AU. For our experimental setup, we employ the open-source library PyFeat [22].

IV. UNILATERAL AU DETECTOR

While the mirroring approach introduced in the previous section is effective for generating unilateral AU labels, it operates under the assumption of symmetric and front-facing images. However, this assumption may restrict its utility, particularly for individuals with asymmetrical features or conditions like facial palsy. To address this limitation, we have developed a unilateral AU detector that analyzes the facial structure of the original image without resorting to mirroring. This dedicated model is trained to predict AUs from both the left and right sides of the face, which enhances its applicability. Moreover, by eliminating the need for mirroring, our dedicated model reduces the unintentional error cascades stemming from mirrored faces. An overview of our unilateral AU detector is presented in Figure 2, and the steps are described in this section.

A. Facial Feature Extraction

To reduce the influence of background noise in the image, we initially extract the head region by computing the convex hull of the facial landmark keypoints. Following the methodology proposed by Cheong et al. [22], we then extract Histogram of Oriented Gradients (HOG) [23] features.

This process comprises dividing the image into 8×8 blocks and computing the edge gradients in both the x-axis and y-axis directions within these blocks. Subsequently, the orientations of the edge gradients are computed, and HOG is generated with 8 bins (representing 8 different orientations). From the edge gradients, we can compute their orientations. Finally, the histograms of each block are concatenated to form a feature vector of the input facial image.

We explore two alternative feature map representations:

- Full face features: In this representation, we compute the HOG across the entire face (see Figure 2, (top)).
- Half face features: This alternative splits the face in half and computes the HOG for each half of the face (see Figure 2, (bottom)).

Both of these feature representations yield a large feature vector. To compress this representation, we employ a standard Principal Component Analysis (PCA) and retain the number of components that capture 95% of the variance.

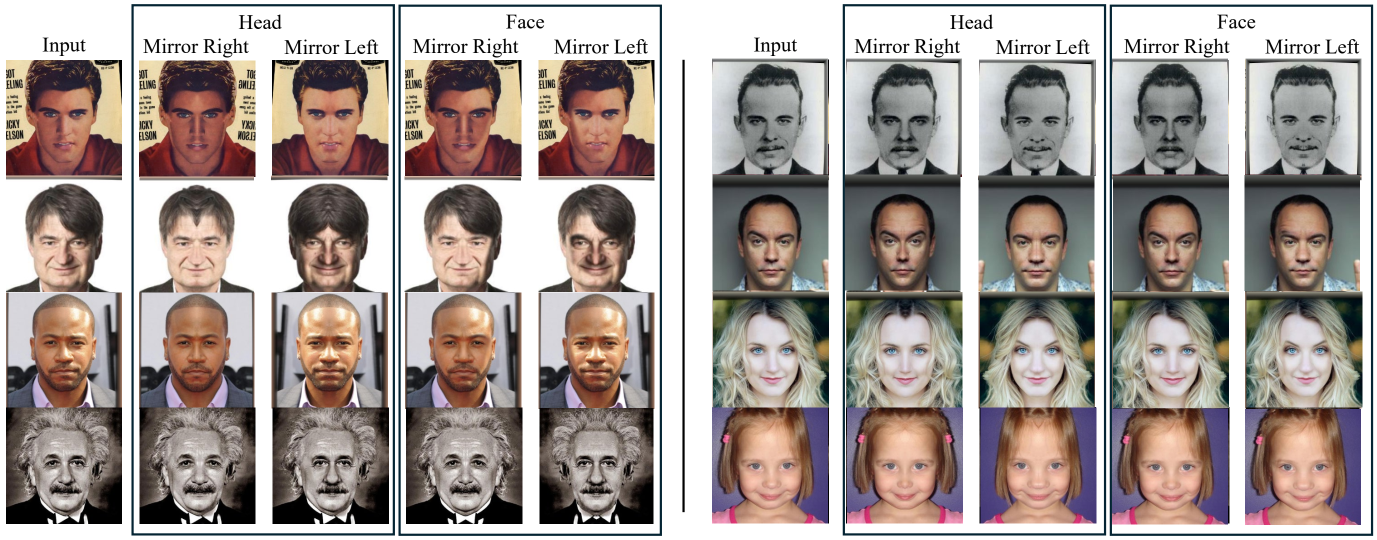

Fig. 3. Example results of both face mirroring and head mirroring. Face mirroring preserves the background, face shape, and hair style, while head mirroring simply replicates everything from one side.

TABLE III
PERFORMANCE OF OUR UNILATERAL AU DETECTOR MODEL, EVALUATED ON THE EXPERT ANNOTATED DATA. PERFORMANCE IS MEASURED IN TERMS OF F-SCORE. THE NOTATION (_L) DENOTES LEFT, AND (_R) DENOTES RIGHT.

Model	Features	Mean	†AU02_L	AU04_L	AU10_L	AU15_L	AU25_L	AU26_L	AU28_L	AU43_L
SVM	Full face	0.45	0.00	0.55	0.66	0.36	0.86	0.25	0.22	0.55
XGB	Full face	0.41	0.00	0.40	0.62	0.35	0.84	0.13	0.15	0.53
SVM	Half faces	0.49	0.00	0.50	0.74	0.52	0.88	0.25	0.21	0.67
XGB	Half faces	0.42	0.00	0.33	0.68	0.30	0.82	0.22	0.08	0.62
			AU02_R	AU04_R	AU10_R	AU15_R	AU25_R	AU26_R	AU28_R	AU43_R
			0.42	0.44	0.86	0.29	0.87	0.42	0.08	0.33
			0.47	0.40	0.77	0.26	0.89	0.40	0.07	0.31
			0.50	0.40	0.89	0.32	0.91	0.45	0.07	0.50
			0.50	0.33	0.83	0.29	0.94	0.37	0.07	0.36

†Note that for the small annotated samples, there were no instances of AU02_L, hence precision, recall, and f-score are 0.

B. AU Classifier Training

Lastly, we train a classifier on the facial features to predict left and right AUs. We investigate two widely used algorithms in the literature: Support Vector Machines (SVM) [24] and XGBoost [25].

C. Dataset and Ground Truth

EmotioNet Dataset: For training, we mainly utilized the EmotioNet [26] dataset, which includes facial images collected from the internet through querying emotional keywords. We performed filtering based on the head's pose, retaining images with an absolute value of the roll, pitch, and yaw to be less than five. In doing so, we ensured that the facial images were front-facing, i.e., directly facing the camera with both sides of the face visible. Subsequently, mirroring was applied as described in the previous section, and both frontal and mirrored images were processed using PyFeat, resulting in images labeled with 20 AUs. Images that failed processing due to obstructions of the face, incomplete frames, or anomalies were removed, leaving 5,221 original frontal images. Along with their mirrored variants, the number of images reached 26,106. Out of 20 AUs provided by PyFeat, we selected a subset of eight AUs (*i.e., AU02, AU04, AU10, AU15, AU25, AU26, AU28, and AU43*) that appeared in at least ten percent of the cases unilaterally.

For evaluation, we curated a sample of 66 images. Half of these images were randomly selected from a subset displaying unilateral facial expressions, while the other half was randomly chosen from the remaining images. Each selected sample underwent manual annotation by two researchers independently, who indicated unilateral AU labels for the left and right sides, and bilateral AU labels for the complete face separately. Both annotators studied the FACS guide prior to the annotation task and agreed on the annotation procedure. This procedure involved annotators identifying the presence or absence of specific AUs on either the left or right side of the face. The ground truth was established by taking the intersection of the sets of AUs for the left and right sides as determined by both annotators. To assess the inter-rater agreement, we utilized Cohen's Kappa [27]. The agreement on the unilateral labels

TABLE IV
PERFORMANCE OF OUR UNILATERAL AU DETECTOR MODEL, EVALUATED ON THE FACIAL PALSY DATASET. PERFORMANCE IS MEASURED IN TERMS OF F-SCORE. THE NOTATION (_L) DENOTES LEFT, AND (_R) DENOTES RIGHT.

Model	Features	Mean	AU02_L	AU04_L	AU10_L	AU15_L	AU25_L	AU26_L	AU28_L	AU43_L
SVM	Full face	0.68	0.70	0.59	0.80	0.61	0.84	0.62	0.43	0.75
XGB	Full face	0.64	0.70	0.54	0.77	0.57	0.82	0.58	0.35	0.72
SVM	Half face	0.69	0.73	0.63	0.79	0.64	0.84	0.68	0.47	0.78
XGB	Half face	0.65	0.71	0.56	0.73	0.59	0.82	0.61	0.37	0.74
			AU02_R	AU04_R	AU10_R	AU15_R	AU25_R	AU26_R	AU28_R	AU43_R
			0.68	0.70	0.81	0.65	0.82	0.70	0.38	0.80
			0.63	0.62	0.77	0.60	0.79	0.66	0.39	0.79
			0.68	0.68	0.82	0.65	0.82	0.70	0.39	0.76
			0.66	0.61	0.82	0.61	0.80	0.64	0.36	0.80

TABLE V
5-FOLD CROSS-VALIDATION PERFORMANCE OF OUR UNILATERAL AU DETECTOR MODEL, EVALUATED ON THE EMOTIONET DATASET. PERFORMANCE IS MEASURED IN TERMS OF F-SCORE. THE NOTATION (_L) DENOTES LEFT, AND (_R) DENOTES RIGHT.

Model	Feature	Mean	AU02_L	AU04_L	AU10_L	AU15_L	AU25_L	AU26_L	AU28_L	AU43_L
SVM	Full face	0.81	0.82	0.80	0.85	0.75	0.88	0.84	0.72	0.83
XGB	Full face	0.79	0.80	0.77	0.84	0.69	0.89	0.82	0.68	0.84
SVM	Half face	0.81	0.82	0.80	0.85	0.75	0.89	0.84	0.72	0.85
XGB	Half face	0.80	0.81	0.78	0.85	0.70	0.89	0.81	0.70	0.85
			AU02_R	AU04_R	AU10_R	AU15_R	AU25_R	AU26_R	AU28_R	AU43_R
			0.82	0.80	0.86	0.75	0.87	0.81	0.73	0.85
			0.81	0.76	0.85	0.71	0.86	0.79	0.69	0.87
			0.82	0.81	0.86	0.75	0.87	0.81	0.72	0.85
			0.81	0.78	0.85	0.72	0.87	0.79	0.70	0.87

was K=0.8 for the right side, K=0.74 for the left side, and K=0.64 for the entire face, indicating substantial agreement strength [28]. Subsequently, we utilized the intersection of the AU labels provided by the annotators as the ground truth for further analysis.

Facial Palsy Datasets: In addition to EmotioNet dataset, we also aimed to evaluate the performance of our model on images depicting edge cases of unilateral facial expressions. To achieve this, we curated a dataset as the combination of three different data sources featuring images of facial palsy patients: (1) MEEI Standard dataset [29], (2) Neuroface Toronto [30], and (3) YouTube Facial Palsy dataset [31]. Both MEEI Standard and Neuroface Toronto datasets were originally collected under controlled experimental setting, while the YouTube Facial Palsy dataset was obtained from YouTube videos.

Following the filtering step for the frontal images, a total of 408 images remained for evaluation purposes. To obtain unilateral AU labels, we used the same mirroring procedure as previously described. We opted only for head mirroring due to its superior performance in our previous observations.

V. RESULTS

A. Qualitative Evaluation of the Mirroring

Visual examples of the mirroring results for both face mirroring and head mirroring are presented in Figure 3. Overall, both head and face mirroring techniques produce coherent and plausible facial images, which supports the validity of this approach. While there are some noticeable artifacts, such as inconsistent lighting in face mirroring and changes in facial shape in head mirroring, our observations indicate that these artifacts do not significantly degrade the recognition of AUs.

B. Comparison between Face and Head Mirroring

We assess the efficacy of our two strategies for deriving unilateral AUs by comparing their outputs with expert-annotated samples (as detailed in Section IV-C). Table I presents the performance of mirroring the head, while Table II displays the performance of mirroring the face. Evidently, mirroring the entire head yields significantly better results than mirroring only the face. This observation suggests that features near the contour of the head and face, such as the jawline, cheeks, and forehead, play crucial roles in inferring AUs.

C. Comparison between full face and half face features

Next, we assess two alternative strategies for extracting facial features when training a unilateral AU detector. Table III presents the f-score on the expert annotated data that shows the performance of both full-face and half-face features. We observe that half-face features outperform full-face features. This finding is consistent across Table IV and Table V as well.

This is surprising since the full-face features are expected to contain the information encoded in the half-face features. A potential explanation is that the improvements may be due to a clearer delineation of the separation between the left and right sides in the half faces compared to the full faces. In the latter,

TABLE VI
COMPARISON BETWEEN AUS IDENTIFIED ON THE ORIGINAL IMAGES AND THE UNIFIED SET OF UNILATERAL AUS DETECTED ON BOTH SIDES USING HEAD MIRRORING.

Metric	Mean	AU02	AU04	AU10	AU15	AU25	AU26	AU28	AU43
precision	0.93	0.98	0.92	0.87	0.93	0.90	0.92	0.94	0.97
recall	0.90	0.96	0.89	0.85	0.89	0.88	0.88	0.91	0.96
f-score	0.91	0.97	0.90	0.85	0.91	0.89	0.89	0.92	0.96
accuracy	0.90	0.96	0.89	0.85	0.89	0.88	0.88	0.91	0.96

the model needs to allocate additional capacity to determine the point of laterality change. Moreover, the model does not have to filter out details of the other side for the half faces, potentially reducing the side effects of spurious correlations.

D. Comparison between SVM and XGBoost

An examination of Table III, Table IV, and Table V reveals that SVM consistently outperforms XGBoost. This discrepancy potentially arises from XGBoost's high sensitivity to imbalanced data, which is the case in our experiments. In contrast, SVM exhibits greater robustness to imbalanced datasets since it focuses solely on support vectors around the decision boundary.

E. Performance on facial palsy images

The results of the evaluation of our model on the facial palsy datasets are presented in Table IV. The best-performing model achieves an f-score of around 0.69. Considering that our proxy labels derived from the mirroring strategy align with human annotations at around 0.75 (as shown in Table V), and our model achieves an f-score of approximately 0.80 from the 5-fold cross-validations, the performance of the model on the facial palsy datasets is in line with expectations. This alignment is reasonable given the cascade of errors, which approximately yields $0.75 \times 0.8 = 0.60$.

F. Estimated degradation of head mirroring

Ideally, an AU present on the original facial image should also appear on one or both of the mirrored facial halves. Moreover, mirroring should not introduce any AUs that were absent on the original facial image. To assess the impact of head mirroring on AU detection, we compared the original set of AUs per image with the unified list of AUs identified on the left and right mirrored versions, and reported the results in Table VI. The average F1-Score for the eight AUs is 0.91, indicating the degree to which mirroring affects the accuracy of AU identification.

VI. DISCUSSION AND CONCLUSION

In this paper, we proposed a novel method for generating AU labels independently for each side of facial images, along with a dedicated model for detecting unilateral facial expressions. We explored two approaches for generating unilateral AU labels: mirroring both sides of the entire head and mirroring only the face. Our findings indicate that both approaches are feasible, with head mirroring yielding superior results. Using the generated AU labels, we trained a dedicated unilateral AU detection model, which we assessed by comparing against expert-annotated sample data and applying it to multiple datasets featuring images of patients with facial palsy.

This study has potential implications for research and practical applications. Researchers can employ our AU label generation method based on lateral mirroring to automatically produce unilateral AU labels, addressing an important gap in the literature. Moreover, our findings can inform the development of advanced systems for nuanced facial expression recognition. The dedicated unilateral AU detection model offers a valuable tool for gaining insights into the manifestation of nuanced facial expressions in humans.

In practice, our unilateral AU detection model can be utilized in the healthcare domain for diagnosing and treating disorders that lead to asymmetrical facial impairments. Additionally, it can aid in training actors to enhance their facial expressiveness and in designing virtual agents with realistic and nuanced facial expressions.

However, the primary limitation of this study lies in the size and diversity of the human-annotated data used for evaluation. To address this, our next research endeavor involves manually annotating a larger and more diverse dataset. Furthermore, the inability to compare our model against the state-of-the-art is due to the scarcity of available unilateral AU models in existing literature. This highlights an area for future exploration and benchmarking.

ETHICAL IMPACT STATEMENT

We evaluated the potential ethical implications of this study in accordance with guidelines established by the affective computing community [32] and through thematic analysis of previous literature [33].

In this research, we utilized publicly available datasets and conducted a thorough examination to identify potential biases and contextual limitations. However, it is important to acknowledge that the generalizability of our model may be compromised by the limited number and diversity of datasets used.

Facial expressions of emotions are inherently influenced by culture and context, with variations even within individuals. Therefore, Facial Expression Recognition (FER) studies must be diligent in interpreting expressions accurately. Our study makes an important contribution by facilitating nuanced facial expression analysis, thereby offering improved opportunities to account for contextual, cultural, and individual variability in expressions.

Our contributions extend beyond specific domains and hold potential across various applications. We recommend that future research and development endeavors that extend our study, particularly those involving sensitive applications such as emotion recognition in educational institutions, workplaces, and law enforcement, comply with the regulatory and ethical guidelines. In adherence to open science principles, the code used for data preparation and analysis is publicly available [34].

References

[1] A. J. Fridlund, *Human facial expression: An evolutionary view*. Academic press, 2014.

[2] K. R. Scherer, "Vocal affect expression: a review and a model for future research." *Psychological bulletin*, vol. 99, no. 2, p. 143, 1986.

[3] P. Ekman and W. V. Friesen, "Measuring facial movement," *Environmental psychology and nonverbal behavior*, vol. 1, pp. 56–75, 1976.

[4] ——, "Facial action coding system," *Environmental Psychology & Nonverbal Behavior*, 1978.

[5] P. Ekman, W. Friesen, and J. Hager, *Facial Action Coding System: The Manual on CD ROM*. Salt Lake City: A Human Face., 2002.

[6] A. Vaughan and D. Gardner, "A systematic review of physical rehabilitation of facial palsy," *Frontiers in Neurology*, vol. 11, p. 500858, 2020.

[7] S. Li and W. Deng, "Deep facial expression recognition: A survey," *IEEE transactions on affective computing*, vol. 13, no. 3, pp. 1195–1215, 2020.

[8] G. Barrios Dell'Olio and M. Sra, "Farapy: An augmented reality feedback system for facial paralysis using action unit intensity estimation," in *The 34th Annual ACM Symposium on User Interface Software and Technology*, 2021, pp. 1027–1038.

[9] K. Shingjergji, D. Iren, F. Böttger, C. Urlings, and R. Klemke, "Interpretable explainability in facial emotion recognition and gamification for data collection," in *2022 10th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 2022, pp. 1–8.

[10] J. Zhang, K. Chen, and J. Zheng, "Facial expression retargeting from human to avatar made easy," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 2, pp. 1274–1287, 2020.

[11] M. A. Witherow, C. Butler, W. J. Shields, F. Ilgin, N. Diawara, J. Keener, J. W. Harrington, and K. M. Iftekharuddin, "Customizable avatars with dynamic facial action coded expressions (cadyface) for improved user engagement," *arXiv preprint arXiv:2403.07314*, 2024.

[12] M. Moscovitch and J. Olds, "Asymmetries in spontaneous facial expressions and their possible relation to hemispheric specialization," *Neuropsychologia*, vol. 20, no. 1, pp. 71–81, 1982.

[13] M. Rychlowska, R. E. Jack, O. G. Garrod, P. G. Schyns, J. D. Martin, and P. M. Niedenthal, "Functional smiles: Tools for love, sympathy, and war," *Psychological science*, vol. 28, no. 9, pp. 1259–1270, 2017.

[14] X. Ge, J. M. Jose, P. Wang, A. Iyer, X. Liu, and H. Han, "Automatic facial paralysis estimation with facial action units," *arXiv preprint arXiv:2203.01800*, 2022.

[15] L. Knoedler, H. Baecher, M. Kauke-Navarro, L. Prantl, H.-G. Machens, P. Scheuermann, C. Palm, R. Baumann, A. Kehrer, A. C. Panayi *et al.*, "Towards a reliable and rapid automated grading system in facial palsy patients: facial palsy surgery meets computer science," *Journal of Clinical Medicine*, vol. 11, no. 17, p. 4998, 2022.

[16] W. R. Powell and J. A. Schirillo, "Asymmetrical facial expressions in portraits and hemispheric laterality: A literature review," *Laterality*, vol. 14, no. 6, pp. 545–572, 2009.

[17] M. K. Mandal and N. Ambady, "Laterality of facial expressions of emotion: Universal and culture-specific influences," *Behavioural Neurology*, vol. 15, no. 1-2, pp. 23–34, 2004.

[18] E. L. Rosenberg and P. Ekman, "Conceptual and methodological issues in the judgment of facial expressions of emotion," *Motivation and Emotion*, vol. 19, no. 2, pp. 111–138, 1995.

[19] D. McDuff, R. Kaliouby, T. Senechal, M. Amr, J. Cohn, and R. Picard, "Affectiva-mit facial expression dataset (am-fed): Naturalistic and spontaneous facial expressions collected," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2013, pp. 881–888.

[20] Y. Fan, J. Lam, and V. Li, "Facial action unit intensity estimation via semantic correspondence learning with dynamic graph convolution," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, 2020, pp. 12 701–12 708.

[21] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee *et al.*, "Mediapipe: A framework for building perception pipelines," *arXiv preprint arXiv:1906.08172*, 2019.

[22] J. H. Cheong, E. Jolly, T. Xie, S. Byrne, M. Kenney, and L. J. Chang, "Py-feat: Python facial expression analysis toolbox," *Affective Science*, vol. 4, no. 4, pp. 781–796, 2023.

[23] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, vol. 1. Ieee, 2005, pp. 886–893.

[24] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, pp. 273–297, 1995.

[25] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[26] C. Fabian Benitez-Quiroz, R. Srinivasan, and A. M. Martinez, "Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 5562–5570.

[27] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and psychological measurement*, vol. 20, no. 1, pp. 37–46, 1960.

[28] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *biometrics*, pp. 159–174, 1977.

[29] J. J. Greene, D. L. Guarin, J. Tavares, E. Fortier, M. Robinson, J. Dusseldorp, O. Quatela, N. Jowett, and T. Hadlock, "The spectrum of facial palsy: The meei facial palsy photo and video standard set," *The Laryngoscope*, vol. 130, no. 1, pp. 32–37, 2020.

[30] A. Bandini, S. Rezaei, D. L. Guarín, M. Kulkarni, D. Lim, M. I. Boulos, L. Zinman, Y. Yunusova, and B. Taati, "A new dataset for facial motion analysis in individuals with neurological disorders," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 4, pp. 1111–1119, 2020.

[31] G.-S. J. Hsu, J.-H. Kang, and W.-F. Huang, "Deep hierarchical network with line segment learning for quantitative analysis of facial palsy," *IEEE access*, vol. 7, pp. 4833–4842, 2018.

[32] D. Ong, J. Hernandez, and R. Picard, "MS Windows NT kernel description," https://acii-conf.net/2023/wp-content/uploads/2023/03/instructions-ethical-statement.pdf, accessed: 2024-03-22.

[33] D. Iren, E. Yildirim, and K. Shingjergji, "Ethical risks, concerns, and practices of affective computing: A thematic analysis," in *2023 11th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*. IEEE, 2023, pp. 1–4.

[34] Actionable-Data-Science, "Unilateral au detection code and data repository." [Online]. Available: https://github.com/Actionable-Data-Science/ACII24_Unilateral_AU_Detection