Generating Synthetic Images with Asymmetric Facial Expressions for Inclusive Facial Expression Recognition

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

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

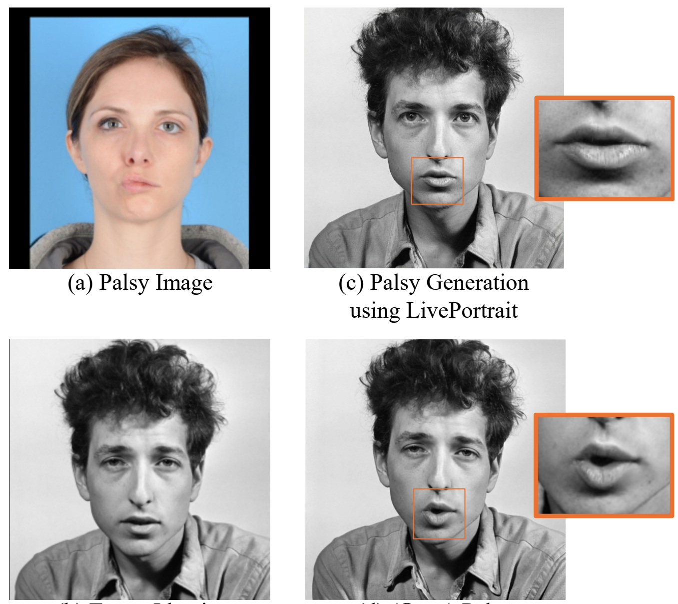
***Abstract*—Facial Expression Recognition (FER) systems often fail to generalize across diverse populations due to significant biases in existing datasets; particularly the under-representation of atypical facial morphologies such as facial palsy. This lack of representation leads to both technical shortcomings and ethical concerns, limiting the inclusiveness and fairness of FER in real-world applications. In this work, we present a novel synthetic data generation method —AFET (Asymmetric Facial Expression Transfer)— designed to simulate asymmetric facial expressions such as those observed in facial palsy. Our method blends halves of neutral and expressive faces from the same identity using a combination of facial landmark detection, locally weighted regression, and mesh-based warping for intermediate image generation. The resulting images maintain anatomical plausibility, making them suitable for model training and evaluation. Qualitative and quantitative evaluations demonstrate that our approach produces realistic facial asymmetry. By enabling the synthetic generation of underrepresented facial conditions, our method contributes to more equitable FER systems and offers a step forward in the ethical deployment of affective computing technologies.**

***Index Terms*—Affective computing, facial expressions, bias, synthetic data, facial palsy**

I. Introduction

Facial Expression Recognition (FER) plays a critical role in numerous applications, ranging from human-computer interaction and social robotics to driver safety and medical diagnosis [1]. By interpreting a person's facial expressions, such methods can enhance user experiences in virtual environments, improve accessibility tools for individuals with disabilities, and support healthcare professionals in monitoring patient conditions.

However, despite its widespread use, FER systems suffer from significant limitations in terms of performance and inclusivity, largely due to bias in existing datasets [2], [3]. Many of these datasets underrepresent demographic factors such as race [4], gender [5], [6], and age [7], [8], as well as physiological variations including facial deformities or conditions like facial palsy [9]. Models trained on such biased data frequently fail to generalize well to underrepresented groups [3], particularly individuals with facial deformities or conditions [9], [10]. Consequently, the lack of representation leads to poor feature learning, which leads to reduced model performance in realistic, nuanced scenarios, limiting their applicability in real-world contexts. This not only impairs model accuracy but also underscores deeper ethical concerns related to inclusivity and bias in AI systems.



(a) Palsy Image (c) Palsy Generation using LivePortrait

(b) Target Identity (d) **(Ours)** Palsy Generation

Fig. 1. Illustration of AFET: Asymmetric Facial Expression Transfer.

Additionally, with the rising emphasis on privacy protection and legal regulations such as the General Data Protection Regulation (GDPR) [11] and the AI Act [12], non-targeted scraping of facial images online or from CCTV cameras is prohibited. This intensifies the need for new data collection strategies that respect both individual privacy and regulatory frameworks, justifying a move toward synthetic datasets.

This paper addresses a specific and often overlooked challenge in FER: the underrepresentation of atypical facial morphologies, particularly those involving facial asymmetry, such

979-8-3315-8015-5/25/$31.00 ©2025 European Union

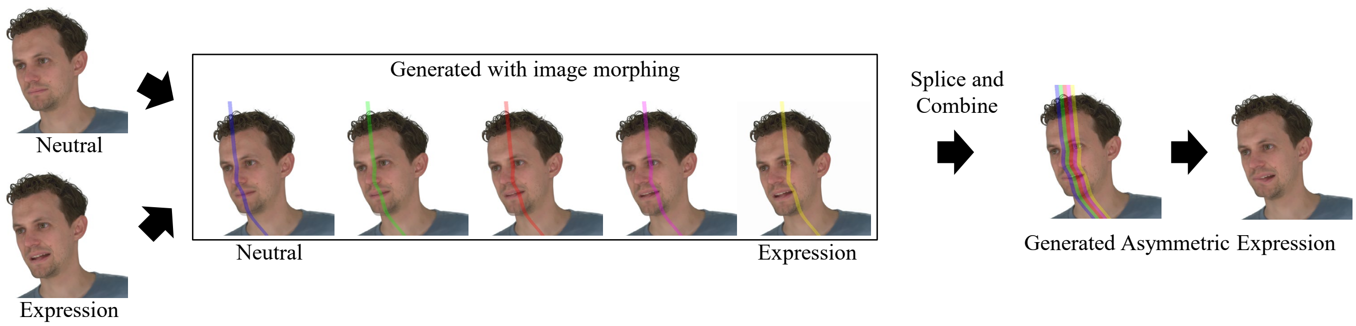


Fig. 2. The target facial image is synthesized with asymmetric expression by first generating morphed images between an input neutral face and expression face followed by column-wise splicing and combining of the corresponding morphed images.

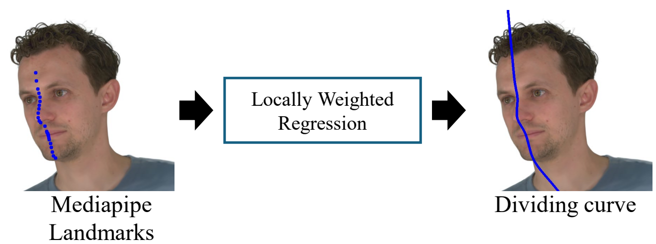


Fig. 3. An illustration of estimating the dividing curve that splits the face into two halves from Mediapipe landmarks using locally weighted regression.

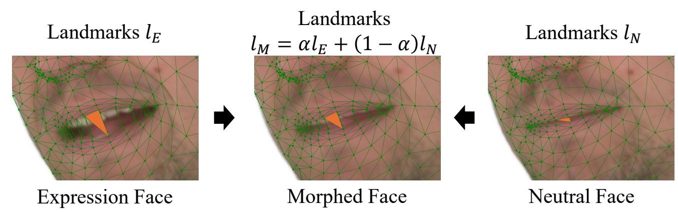


Fig. 4. An intermediate morphed image is generated by warping towards an interpolated landmark between the expression face and the neutral face.

as facial palsy [10]. Thus, this study seeks to answer the following question:

How can synthetic data generation be used to realistically and ethically model facial deformities for FER systems?

To address this, we propose a novel data generation method —AFET (Asymmetric Facial Expression Transfer)— that simulates asymmetric facial expressions such as those seen in facial palsy using morphing and blending techniques [13]. Our approach operates on facial images of healthy subjects and produces realistic facial palsy variants by stitching halves of differently expressed versions of the same face.

We demonstrate that our method produces realistic visualizations of facial palsy, outperforming direct face-reenactment [14] baselines in terms of continuity and asymmetric expression generation. The generated images are derived from manipulating facial landmarks of healthy faces, therefore, we can also derive the facial landmark annotations of the generated facial palsy images. This enables the training of models that could perform better in cases of facial deformity. This not only enhances accuracy for the cases of atypical facial morphologies but also contributes to a more ethical and inclusive deployment of AI technologies.

The subsequent sections of this paper are structured as follows. Section II covers background and related work. Section III provides the details of our method. Section IV discloses the results of the evaluation of our approach. Section V discusses how we address the research question and finally concludes the paper.

II. Background and Related Work

A. Bias and fairness in FER

FER systems have consistently demonstrated performance disparities across demographic groups, raising significant concerns about fairness and inclusivity. Multiple studies have shown that state-of-the-art FER models perform disproportionately poorly on faces from underrepresented groups, particularly those varying in race [4], gender [5], [6], and age [7], [8]. Training facial expression recognition requires large datasets. The datasets that are commonly used for this task are not necessarily composed of an even distribution regarding attributes such as race, gender, and age [15]. These disparities severely impact the performance of the models, yielding lower performances for the minority groups [3]. A plethora of methods have been used by researchers to mitigate bias [16]–[18].

The consequence of biases extends beyond technical performance; they may propagate existing societal inequities when FER is deployed in real-world applications such as education or healthcare [19]. For instance, misinterpretation of emotional states in clinical settings could lead to diagnostic errors or unequal treatment. Addressing bias in FER is therefore not only a matter of accuracy but also of ethical responsibility and social impact.

B. Underrepresentation of atypical facial morphologies

While demographic bias is increasingly recognized, another critical dimension remains underexplored: the representation of individuals with atypical facial morphologies. These include congenital conditions (e.g., cleft lip), acquired deformities

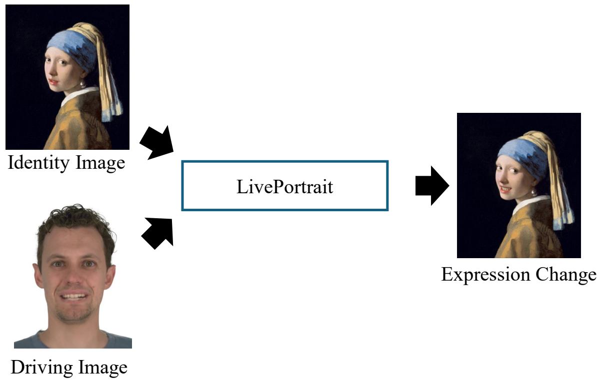


Fig. 5. LivePortrait changes the expression of an input identity image based on an input driving image.

(e.g., due to trauma or surgery), and neurological or muscular impairments (e.g., facial palsy) that affect facial expression symmetry or movement [9].

Existing FER datasets rarely include such faces, leading to a double marginalization: both underrepresentation in data and poor recognition accuracy by trained models. In many cases, these individuals are systematically excluded from training and evaluation, resulting in systems that fail to detect or correctly classify expressions for an underrepresented segment of the population.

There are a few datasets specifically curated for facial palsy analysis, however, they vary significantly in coverage and annotation richness. The YouTube Facial Palsy dataset [20] offers in-the-wild videos from 21 patients but lacks standardized landmarks or AU labels. The MEEI dataset [21] provides high-resolution, lab-controlled recordings with severity scores but no dense facial annotations. The Toronto NeuroFace dataset [22] includes clinical videos from 36 patients, annotated with 68 landmarks. Finally, the AFLFP dataset [9] contains images from 88 subjects with 68 manually annotated facial landmarks. Overall, no existing dataset combines diverse facial palsy expressions with both dense landmarks and AU annotations required to train FER models that could generalize to atypical facial morphologies.

C. *Synthetic facial image generation*

In response to data scarcity and privacy challenges, synthetic data generation has emerged as a powerful tool for augmenting and diversifying training datasets in FER. Techniques such as generative adversarial networks (GANs), 3D morphable models, and parametric facial simulators have been widely used to create artificial faces with varying expressions and attributes [23]–[26]. These methods enable controlled manipulation of facial identity, pose, lighting, and emotion.

More recently, diffusion models [27] have shown state-of-the-art performance in generating high-fidelity and diverse facial images, outperforming GANs in terms of image realism and mode coverage. However, since these models are trained on normal healthy faces, their ability to create medical conditions such as facial deformities are limited.

Fig. 6. Examples of real-life facial palsy patients from the MEEI dataset [21].

Synthetic data generation of facial deformities not only helps mitigate bias by improving representation but also offers a privacy-preserving alternative to scraping real images from the web or surveillance footage.

D. *Regulatory and Ethical Considerations*

The use of facial data is subject to increasing ethical concerns and legal regulations, particularly within the context of the European Union. The GDPR classifies biometric data as sensitive personal information, placing strict limits on its collection and use without explicit consent [11]. Additionally, the AI Act prohibits untargeted scraping of facial imagery from online sources or CCTV footage [12]. As a result, there is a growing need to develop FER training datasets that are both ethically sourced and legally compliant. Synthetic facial image generation offers a promising path forward.

III. Method

Our goal is to synthesize realistic facial palsy expressions by modifying standard facial images of healthy individuals. This enables the creation of diverse synthetic datasets from existing facial image collections. Our approach, AFET, simulates facial palsy by blending one half of a face displaying a neutral expression with the corresponding half of the same face showing an expressive state. This process involves first estimating the vertical dividing curve that separates the two facial halves (Section III-A), followed by splicing and stitching the corresponding halves of the two face images (Section III-B). Initially, our method requires access to two distinct facial expressions of the same individual which is feasible to extract from videos but not for static images. We overcome this challenge by generating different expressions from a single input image (Section III-C).

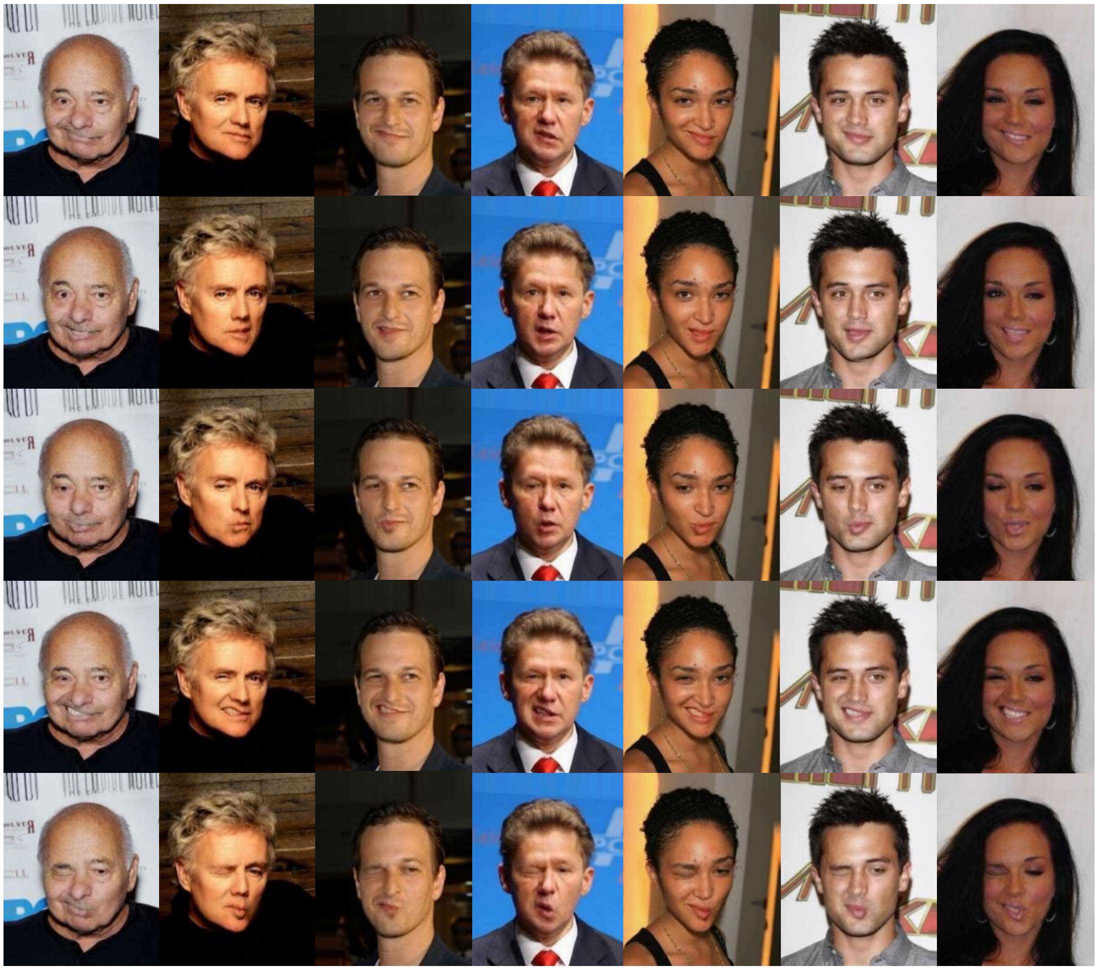


Fig. 7. Examples of our palsy face generations on different face identities.

A. Estimating the dividing curve of the face

To accurately separate the two halves of a face, we first estimate a dividing curve that follows the natural contour of the face formed by the different facial parts. We use MediaPipe [28], to extract facial points located along the midline, including the nose bridge, center of the forehead, middle of the mouth, and chin. These landmarks serve as reference points for determining a vertical partition.

We further extrapolate a curve from these small set of landmark points in order to get the dividing curve of the entire face. A simple piecewise linear fit between these landmarks would result in sharp discontinuities at the forehead and chin, leading to an unnatural segmentation. To ensure a smooth, contour-following boundary, we instead employ locally weighted regression (LOWESS) [29]. This method assigns higher weights to nearby points, allowing the estimated curve to smoothly adapt to the natural curvature of the face, particularly at the forehead and chin. By doing so, we achieve a more anatomically accurate partitioning of the face, which is crucial for realistic synthesis in later stages. Figure 3 shows an example of how the curve fitting looks like. We can observe that it fits the landmark points well and extrapolates nicely outside of the set of landmark points, following the contour of the face.

B. Stitching the two half faces

A direct stitching of the two face images, i.e., one with a neutral expression and the other with a different expression, often results in visible discontinuities. These discontinuities are particularly prominent in regions where the expressions do not align well, such as when one image has an open mouth while the other has a closed mouth.

To bridge the gap between the two expressions and create a more seamless transition between the two sides, we employ a progressive warping approach and generate intermediate images by applying a mesh-based warping technique [13], [30]. Figure 2 shows an overview of this process. This warping gradually morphs the face with an arbitrary expression into a neutral expression face while preserving structural coherence.

The warping process is guided by facial landmarks, ensuring that intermediate images gradually aligns the landmarks between the two face images, maintaining coherency between the expressions and minimizing visible discontinuities. Specif-

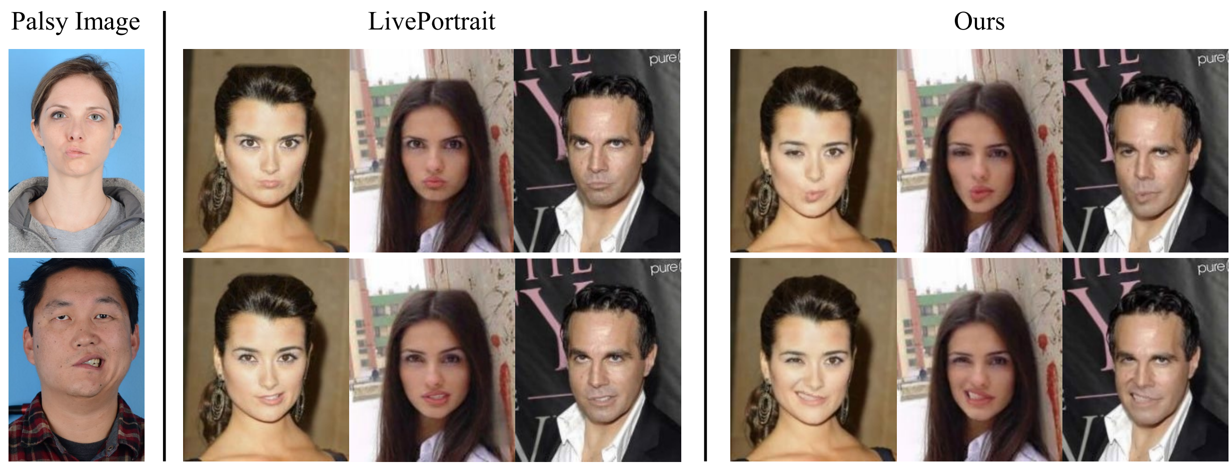


Fig. 8. Comparison between LivePortrait [14] and AFET in generating facial palsy expressions. We use the palsy image as the driving image for LivePortrait.

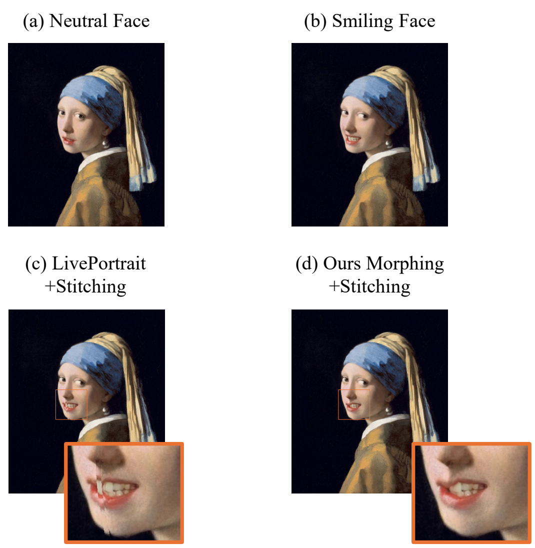


Fig. 9. Our approach takes a (a) neutral face and a (b) expressive face as inputs and stitches them together to create a facial palsy image. If we use (c) LivePortrait to generate the intermediate images, we can observe visible discontinuities. In contrast, (d) our morphing approach produces smoother stitching of the two images with different facial expressions.

ically, we use the landmarks provided by Mediapipe [28] and generate a Delaunay triangulation out of the landmark points.

Next, we create intermediate images by warping to an interpolated landmark l_M between the expression face and the neutral face, as defined below

$$l_M = \alpha l_E + (1 - \alpha) l_N, \tag{1}$$

where l_M is the interpolated landmark, l_E denotes the landmark of the expression face, l_N denotes the landmark of the neutral face, and α is a hyperparameter that controls the interpolation weights. Figure 4 shows an example of an intermediate image generated from the morphing process. The value of α depends on the number of intermediate images we want to generate, which we set to be equivalent to the number of pixels between the center of the face and the edge of the face.

Once the intermediate images are generated, we stitching the two face halves using a column-wise replacement strategy. Starting from the dividing curve, we iteratively replace vertical columns of pixels from the neutral face with corresponding columns from the intermediate images, as defined below using a python-like indexing notation to denote the assignment of the columns:

$$x^N[:, d + i] = x_i^M[:, d + i], \tag{2}$$

where x^N denotes the neutral face image, x_i^M denotes the i-th morphed image, d denotes the position of the dividing curve, $d+i$ refers to an offset of i from the dividing curve d. Figure 2 visualizes this column-wise replacement strategy with color strips denoting the dividing curve with offsets.

This progressive transition ensures that the blending appears natural and smooth, preventing abrupt changes in texture or shape. The final synthesized image retains the structural integrity of the original face while realistically simulating the asymmetry characteristic of facial palsy. Moreover, since the synthesized image came directly from controlling facial landmarks, we also get ground truth facial landmark annotations for all the synthesized facial palsy images.

C. *Generating facial expression image pairs*

At this point, we have the method for stitching half of a neutral face and half of an expression face. The last piece of the puzzle is to generate an arbitrary expression face and neutral face from an input face image in order to use as our image pair for the face stitching. To do this, we utilize an existing face re-enactment model namely LivePortrait [14], which takes in an input identity image and a driving image. LivePortrait will then generate an image that follows the identity of the input image but with the expression of the driving image, as shown in Figure 5. This allows us to generate facial palsy image of any arbitrary face image.

IV. RESULTS

We evaluate our method using a combination of qualitative visual inspection, comparison with an existing model, and quantitative analyses. For identity images, we use the CelebA dataset [31], which includes celebrity faces annotated with 40 binary attributes. The dataset offers substantial diversity in gender, skin tone, hair color, and head pose, making it well-suited for generating a variety of synthetic expressions. For reference expressions, we use facial expression images from the MEEI facial palsy dataset [21], which provides controlled examples of typical facial palsy cases. We implement our models with PyTorch framework and use an RTX 3090 GPU for our experiments. The execution time for a 512×512 image is around 3.4 seconds for blending + 6.2 seconds for facial expression warping. We believe this is a reasonable run time considering we only have to generate the dataset once.

A. Visual Inspection

Figure 7 presents a selection of generated facial palsy expressions across a range of identities, including individuals of different gender presentations and skin tones. We can observe that our method effectively captures the characteristic asymmetry of facial palsy, with realistic expression blending and minimal visible discontinuities, illustrating the robustness and generalization ability of AFET.

B. Comparison with an existing model

We compared our approach against LivePortrait [14], a state-of-the-art facial reenactment model capable of generating arbitrary expressions from a driving image. In this setup, we used real facial palsy images as driving inputs to guide the expression synthesis. Figure 8 shows the comparison results. Because LivePortrait was trained exclusively on healthy, able-bodied individuals, it consistently failed to replicate the asymmetrical characteristics of facial palsy. Instead, it generated symmetric, healthy-looking expressions, effectively overriding the facial palsy features present in the driving image.

In contrast, AFET explicitly models asymmetry by blending expressive and neutral facial halves. This strategy allows us to explicitly introduce the asymmetric features typical of facial palsy, while still utilizing the power of existing models to generate healthy-looking facial expressions. As a result, our generated expressions more faithfully reproduce the nuanced presentation of facial palsy, including uneven muscle activation and localized expression suppression, while still preserving photorealism and identity.

We also experimented on using LivePortrait for generating the intermediate images. As shown in Figure 9, it produces visible discontinuities. In contrast, our image morphing approach generates smoother transitions with minimal discontinuities.

C. Quantitative Evaluation

To perform a quantitative evaluation of our method, we used the MEEI dataset [21]. It comprises images of facial palsy patients exhibiting neutral and enacted facial expressions. We used the neutral face for one half of the face, and the enacted facial expression as the driving image for LivePortrait in generating the other half of the face. This gives us comparable pairs of actual and generated expression faces with facial palsy.

TABLE I
RESULTS OF THE USER EVALUATION WITH 3 DIFFERENT RATERS.

Actual		Labeled				
Generated	Real	Generated	Real	R1	R2	R3
	✓	✓		7	21	16
	✓		✓	132	118	123
✓		✓		6	34	22
✓			✓	133	105	117

TABLE II
PER ACTION UNIT COMPARISONS AND OCCURRENCE BETWEEN GENERATED AND ACTUAL FACIAL PALSY IMAGES.

AU	Model	Prec	Recall	F-score	Real occ.	Gen Occ.
brow_lowerer	AFET	0.443	0.711	0.545	38	61
	LivePort.	0.337	0.725	0.460		86
inner_brow_raiser	AFET	0.782	0.812	0.797	128	133
	LivePort.	0.717	0.662	0.688		120
cheek_raiser	AFET	0.721	0.388	0.504	80	43
	LivePort.	0.475	0.354	0.406		59
lip_corner_puller	AFET	0.930	0.606	0.733	109	71
	LivePort.	0.711	0.622	0.663		97
upper_lip_raiser	AFET	0.904	0.671	0.770	140	104
	LivePort.	0.771	0.755	0.763		140
lip_tightener	AFET	0.571	0.412	0.479	68	49
	LivePort.	0.542	0.471	0.504		59
chin_raiser	AFET	0.758	0.690	0.723	168	153
	LivePort.	0.763	0.695	0.727		152
lid_tightener	AFET	0.970	0.785	0.868	205	166
	LivePort.	0.909	0.820	0.862		186
lip_corner_depressor	AFET	0.651	0.368	0.471	76	43
	LivePort.	0.527	0.397	0.453		55
outer_brow_raiser	AFET	0.731	0.731	0.731	67	67
	LivePort.	0.606	0.642	0.623		71
dimpler	AFET	0.438	0.206	0.280	34	16
	LivePort.	0.500	0.243	0.327		18
lip_pressor	AFET	0.300	0.375	0.333	24	30
	LivePort.	0.265	0.375	0.310		34

1) User Evaluation: We conducted a user evaluation study with the task of detecting non-convincing facial palsy images. Three participants were shown a mix of 139 real and 139 synthetic facial palsy images (278 in total) in random order and asked to classify each as either real or generated. Table I summarizes the results of the user evaluation study. A large portion of the generated images were labeled as real, indicating that participants had difficulty distinguishing between real and synthetic palsy expressions. This supports the visual plausibility of our method.

2) Unilateral AU: We used unilateral AU detection [32] in combination with facetorch [33] to detect the AUs on both sides of the actual and generated facial images. For benchmarking purposes, we picked a subset of 12 AUs that are commonly reported in other studies, namely, AU1: inner brow raiser, AU2: outer brow raiser, AU4: brow lowerer, AU6: cheek raiser, AU7: lid tightened, AU10: upper lip raiser, AU12: lip corner puller, AU14: dimpler, AU15: lip corner depressor, AU17: chin raiser, AU23: lip tightener, and AU24: lip pressor. The mean number of AUs detected in the real facial palsy images was 4.5 (SD = 2.1), while the generated facial palsy images had a mean of 3.8 (SD = 1.8). Table II shows a more detailed breakdown of the comparisons per action unit between generated and real images. The results show that AFET can mimic the AUs present in actual facial palsy images and performs better than just using a facial reenactment baseline (LivePortrait).

Since our generation process always uses a neutral face for one side, we compute the Jaccard similarity separately: one between the generated neutral side and the reference neutral image, and one between the expressive side and the reference expression. Our method achieves a 0.524 Jaccard similarity score in this experimental setup. Using the expressive side only, we achieve a Jaccard similarity score of 0.551.

3) Degree of Asymmetry: We also estimated the degree of asymmetry by comparing the Jaccard similarity score of the AUs from the left and right side of the face. Specifically, we use compute it using the formula $1 - Jaccard(AU_{left}, AU_{right})$. Our generated facial palsy dataset has a 0.650 degree of asymmetry while the MEEI dataset has 0.583 degree of asymmetry. The higher asymmetry in our results is expected, given our design choice of using a fully neutral face on one side.

V. Discussion and Conclusion

We proposed AFET - Asymmetric Facial Expression Transfer; a novel approach for synthesizing facial imagery that exhibits asymmetric facial expressions. It simulates realistic facial asymmetry without requiring access to clinical datasets or facial palsy patients. This enables a scalable generation of diverse and privacy-preserving anatomically plausible asymmetrical facial expressions.

Our visual inspections and qualitative results show that AFET can produce realistic images that convincingly reflect the key visual attributes of facial palsy. In contrast, LivePortrait could only generate healthy subjects due to its training data. This indicates that explicit modeling of asymmetry is necessary when generating synthetic data for underrepresented facial morphologies.

The AU-based evaluation supports the visual quality of the generated images. This slight difference between the mean number of AUs detected in real facial images and the generated ones suggests that while the synthetic expressions capture a substantial portion of the muscular activity observed in real cases, they may exhibit slightly less AU diversity or intensity overall. This is expected, as the synthetic expressions are derived from blending a neutral and expressive half, which naturally limits the range of active AUs.

Our method achieves a Jaccard similarity score of 0.524 when comparing the generated expression halves against reference images, and 0.551 when focusing only on the expressive side. Looking at per AU comparison, we achieve promising precision on AUs that had large occurrences. These results suggest that our synthesized expressions maintain meaningful AU-level consistency. Moreover, the estimated degree of asymmetry in our dataset (0.650) is comparable to—and slightly higher than—that of the clinical MEEI dataset (0.583), further validating the asymmetry modeling capabilities of our approach. The higher asymmetry values are likely a result of our method's use of a fully neutral face on one side, which maximizes contrast and may be beneficial for training FER systems to recognize such conditions.

Our observation is that all synthesized images could be used by the AU detection model, and that no errors were generated simply because the generated image exhibits unexpected image anomalies.

This study has potential theoretical and practical implications. From a theoretical perspective, our results show that condition-specific constraints—such as asymmetric facial muscle control—require dedicated modeling strategies. Moreover, our method provides the means to systematically study underrepresented facial conditions.

AFET offers a scalable and privacy-preserving solution for augmenting FER training datasets with realistic examples of facial palsy. This is particularly useful in settings where acquiring labeled clinical data is difficult due to privacy concerns or data scarcity. Practitioners can use AFET to balance their training datasets, improving model robustness to atypical facial presentations in healthcare and assistive technologies. By facilitating the synthetic generation of underrepresented facial conditions, our method advances the development of more inclusive and equitable FER systems, while contributing to the responsible and ethical deployment of affective computing technologies.

This study is not without limitations. The evaluation of our method is limited to a qualitative examination of the realism of the generated images, comparative visual inspection against an off-the-shelf method, and the comparison of the generated asymmetrical facial expressions against real images. For a more comprehensive evaluation, in the future, we will train a FER model using the data generated by our method, and compare the model against the state-of-the-art.

Ethical Impact Statement

We assessed the ethical implications of this work by aligning with established guidelines from the affective computing community and through thematic analysis of recent literature [34].

This paper addresses the representational bias in FER by generating synthetic images of individuals with facial palsy. Our work does not involve human subjects. We used a scientific facial palsy image dataset, i.e., MEEI [21] and CelebA dataset [31] to evaluate our method and illustrate the results.

While our primary objective is to improve fairness in FER by increasing the representation of underrepresented conditions, the generated data could be repurposed in inappropriate ways. To mitigate this, we recommend that future deployments of our method, and any data generated with it, be accompanied by clear ethical use policies.

Our method explicitly aims to reduce bias by generating data representing a marginalized and often overlooked group: individuals with facial palsy. Since if our method is used solely on existing facial image datasets, there is a risk to inherit the demographic biases present in those datasets. The synthesized images are not meant to be used for medical diagnosis or clinical assessment. The scope of our claims is restricted to their use in FER model training and evaluation under research settings, particularly for improving model robustness and fairness. We encourage future research and development building upon this work to align with relevant regulatory frameworks and ethical standards. In the spirit of open science, we will release the code and generated dataset following the completion of the peer review process.

Acknowledgements

This research project was financially supported by NWO with grant number OCENW.XS24.3.041.

References

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

[2] T. Kopalidis, V. Solachidis, N. Vretos, and P. Daras, "Advances in facial expression recognition: A survey of methods, benchmarks, models, and datasets," *Information*, vol. 15, no. 3, p. 135, 2024.

[3] T. Xu, J. White, S. Kalkan, and H. Gunes, "Investigating bias and fairness in facial expression recognition," in *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16*. Springer, 2020, pp. 506–523.

[4] S. Yucer, F. Tektas, N. Al Moubayed, and T. Breckon, "Racial bias within face recognition: A survey," *ACM Computing Surveys*, vol. 57, no. 4, pp. 1–39, 2024.

[5] A. Domnich and G. Anbarjafari, "Responsible ai: Gender bias assessment in emotion recognition," *arXiv preprint arXiv:2103.11436*, 2021.

[6] J. Deuschel, B. Finzel, and I. Rieger, "Uncovering the bias in facial expressions," *arXiv preprint arXiv:2011.11311*, 2020.

[7] S. R. Jannat and S. Canavan, "Expression recognition across age," in *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*. IEEE, 2021, pp. 1–5.

[8] R. Poyiadzi, J. Shen, S. Petridis, Y. Wang, and M. Pantic, "Domain generalisation for apparent emotional facial expression recognition across age-groups," *arXiv preprint arXiv:2110.09168*, 2021.

[9] Y. Xia, C. Nduka, R. Y. Kannan, E. Pescarini, J. E. Berner, and H. Yu, "Aflfp: A database with annotated facial landmarks for facial palsy," *IEEE Transactions on Computational Social Systems*, vol. 10, no. 4, pp. 1975–1985, 2022.

[10] Y. Zhang, W. Gao, H. Yu, J. Dong, and Y. Xia, "Artificial intelligence-based facial palsy evaluation: a survey," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2024.

[11] "Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation) (text with eea relevance)," pp. 1–88, May 2016. [Online]. Available: http://data.europa.eu/eli/reg/2016/679/oj

[12] "Regulation (eu) 2024/1689 of the european parliament and of the council of 13 june 2024 laying down harmonised rules on artificial intelligence and amending regulations (ec) no 300/2008, (eu) no 167/2013, (eu) no 168/2013, (eu) 2018/858, (eu) 2018/1139 and (eu) 2019/2144 and directives 2014/90/eu, (eu) 2016/797 and (eu) 2020/1828 (artificial intelligence act) (text with eea relevance)," Jul 2024. [Online]. Available: http://data.europa.eu/eli/reg/2024/1689/oj

[13] G. Wolberg, *Digital image warping*. IEEE computer society press Los Alamitos, CA, 1990, vol. 10662.

[14] J. Guo, D. Zhang, X. Liu, Z. Zhong, Y. Zhang, P. Wan, and D. Zhang, "Liveportrait: Efficient portrait animation with stitching and retargeting control," *arXiv preprint arXiv:2407.03168*, 2024.

[15] M. M. Hosseini, A. P. Fard, and M. H. Mahoor, "Faces of fairness: Examining bias in facial expression recognition datasets and models," *arXiv preprint arXiv:2502.11049*, 2025.

[16] S. Siddique, M. A. Haque, R. George, K. D. Gupta, D. Gupta, and M. J. H. Faruk, "Survey on machine learning biases and mitigation techniques," *Digital*, vol. 4, no. 1, pp. 1–68, 2023.

[17] M. Hort, Z. Chen, J. M. Zhang, M. Harman, and F. Sarro, "Bias mitigation for machine learning classifiers: A comprehensive survey," *ACM Journal on Responsible Computing*, vol. 1, no. 2, pp. 1–52, 2024.

[18] S. Dehdashtian, R. He, Y. Li, G. Balakrishnan, N. Vasconcelos, V. Ordonez, and V. N. Boddeti, "Fairness and bias mitigation in computer vision: A survey," *arXiv preprint arXiv:2408.02464*, 2024.

[19] M. Mattioli and F. Cabitza, "Not in my face: Challenges and ethical considerations in automatic face emotion recognition technology," *Machine Learning and Knowledge Extraction*, vol. 6, no. 4, pp. 2201–2231, 2024.

[20] G.-S. Jison Hsu, W.-F. Huang, and J.-H. Kang, "Hierarchical network for facial palsy detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 580–586.

[21] 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.

[22] 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.

[23] G. Retsinas, P. P. Filntisis, R. Danecek, V. F. Abrevaya, A. Roussos, T. Bolkart, and P. Maragos, "3d facial expressions through analysis-by-neural-synthesis," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 2490–2501.

[24] J. H. Soeseno, D. S. Tan, W.-Y. Chen, and K.-L. Hua, "Faster, smaller, and simpler model for multiple facial attributes transformation," *IEEE Access*, vol. 7, pp. 36 400–36 412, 2019.

[25] D. S. Tan, J. H. Soeseno, and K.-L. Hua, "Controllable and identity-aware facial attribute transformation," *IEEE Transactions on Cybernetics*, vol. 52, no. 6, pp. 4825–4836, 2021.

[26] D. S. Tan, Y.-X. Lin, and K.-L. Hua, "Incremental learning of multi-domain image-to-image translations," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 4, pp. 1526–1539, 2020.

[27] T. Kirschstein, S. Giebenhain, and M. Nießner, "Diffusionavatars: Deferred diffusion for high-fidelity 3d head avatars," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 5481–5492.

[28] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. Yong, J. Lee *et al.*, "Mediapipe: A framework for perceiving and processing reality," in *Third workshop on computer vision for AR/VR at IEEE computer vision and pattern recognition (CVPR)*, vol. 2019, 2019.

[29] W. S. Cleveland and S. J. Devlin, "Locally weighted regression: an approach to regression analysis by local fitting," *Journal of the American statistical association*, vol. 83, no. 403, pp. 596–610, 1988.

[30] G. Wolberg, "Image morphing: a survey," *The visual computer*, vol. 14, no. 8-9, pp. 360–372, 1998.

[31] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proceedings of International Conference on Computer Vision (ICCV)*, December 2015.

[32] D. Iren and D. S. Tan, "Unilateral facial action unit detection: Revealing nuanced facial expressions," in *IEEE Affective Computing and Intelligent Interfaces*, 2024.

[33] T. Gajarsky, "Facetorch: A python library for analyzing faces using pytorch," https://github.com/tomas-gajarsky/facetorch, 2024.

[34] 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.