FaraPy: An Augmented Reality Feedback System for Facial Paralysis using Action Unit Intensity Estimation



Figure 1: Overview FaraPy System Architecture: (a) Camera data is processed in real-time by the proposed (b) LW-FAU network. (c) Left- and right-side facial muscle activation predictions are processed in (d) JavaScript, where (2) the paralyzed side of the face is identified as the one with overall lower activations. (3) Ratios per AU are calculated to provide (4) feedback on the symmetry of activation. (5) The muscle education and mirror therapy filter is activated on the identified paralyzed side, and (5) facial expressions are scored for symmetry. (e) Activations and the facial filter are visualized in AR.

ABSTRACT

Facial paralysis is the most common facial nerve disorder. It causes functional and aesthetic deficits that often lead to emotional distress and affect psychosocial well-being. One form of treatment is mirror therapy, which has shown potential but has several mirror-related drawbacks that limit its effectiveness. We propose FaraPy, the first mobile augmented reality mirror therapy system for facial paralysis that provides real-time feedback and tracks user progress over time. We developed an efficient convolutional neural network to detect muscle activations and intensities as users perform facial exercises in front of a mobile device camera. Our model outperforms the stateof-the-art model on benchmark data for detecting facial action unit intensity. Our user study (n=20) shows high user satisfaction and greater preference for our interactive system over traditional mirror therapy.

This work is licensed under a Creative Commons Attribution International 4.0 License.

UIST '21, October 10–14, 2021, Virtual Event, USA © 2021 Copyright held by the owner/author(s). ACM ISBN 978-1-4503-8635-7/21/10. https://doi.org/10.1145/3472749.3474803

CCS CONCEPTS

Human-centered computing → Human computer interaction (HCI); Interaction paradigms; Mixed / augmented reality;
Applied computing → Life and medical sciences; Consumer health;

KEYWORDS

Facial palsy, Facial paralysis, Augmented Reality, Lightweight models, Facial action unit detection, Action unit intensity

ACM Reference Format:

Giuliana Barrios Dell'Olio and Misha Sra. 2021. 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 (UIST '21), October 10–14, 2021, Virtual Event, USA.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3472749.3474803

1 INTRODUCTION

Daily life functions such as eating, speaking, and expressing emotions are supported by the facial nerve (CN VII) [35, 66]. Facial paralysis (FP) results from damage or injury to the facial nerve [56, 64, 74]. FP causes functional and aesthetic deficits for those affected, characterized by reduced ability to chew food, changes in clarity of speech and facial symmetry [34, 59]. These individuals exhibit significant emotional distress with negative impact on quality of life and psychosocial wellbeing [15, 27, 29, 38, 74].

Physical therapy in the form of facial exercises is a common treatment for FP [64, 73]. Tailored facial exercises have been shown to improve muscle and nerve function and facial symmetry [2],

shorten recovery time, and prevent long-term paralysis [67]. While the therapeutic effect of facial exercises alone is of moderate quality [2, 67], several studies show that combining facial exercises with mirror therapy (MT) leads to a significant positive impact on FP treatment [8, 16, 32]. It may also help prevent FP sequelae, such as synkinesis [50] which causes involuntary associated facial movements, for example, eye muscle contraction leads to squinting during smiling [50].

MT is traditionally conducted by simply watching the mirror while performing facial exercises or creating the visual illusion of a non-paralyzed face [50]. For the latter, a mirror is placed in the individual's mid-sagittal plane such that the paretic side of the face is covered and the non-paretic side is reflected in the mirror [32]. Unfortunately, previous MT approaches for facial exercising have several limitations including: (a) reduced user motivation, (b) lack of meaningful feedback and muscle education, (d) mirror exposure anxiety, and (e) inability to track progress over time.

To address these limitations, we propose FaraPy, a mobile augmented reality (AR) system with a novel deep learning model in the back end. Our proposed Light Weight Facial Activation Unit (LW-FAU) model detects facial muscle activations and intensities based on input from the user's camera. The AR interface provides real-time feedback to the user on the detected facial muscle activations. FaraPy allows individuals to receive feedback on exercise performance at home, enabling them to continue their recovery outside of therapy sessions, which is an important contributor to successful recovery [20].

Our primary design goal was to make the system run on any smartphone or tablet and be accessible anywhere, anytime. In pursuing this goal, we faced two significant challenges. First, there is a lack of FP benchmark datasets and models labeled and trained with our values of interest. To address this we labeled benchmark data and also created our own user dataset of individuals with FP. Second, we needed an efficient model that would run in real-time on devices with limited computational capacity. Using the studentteacher learning approach, we integrated the knowledge from a large and complex but accurate model [21] for building an efficient and lightweight model (LW-FAU). Our key contributions are as follows:

- The first AR-based real-time physical therapy feedback system for individuals with FP.
- A novel lightweight deep learning model that runs on smartphones in real-time and outperforms state-of-the-art models.
- Prediction of unilateral facial action unit intensities (AUI) independently for each side of the face, as opposed to traditional bilateral predictions of benchmark models.
- A new FP dataset with unilateral action unit intensity labels.
- An end-to-end pipeline with a fully working system that can work on mobile and other devices with front facing cameras.

Our user study (n=20 users) results indicate that participants were very satisfied with FaraPy and would prefer to use our feedback system over traditional mirror therapy. A technical evaluation of our LW-FAU model on two real-world data sets (benchmark data: DISFA, Denver Intensity of Spontaneous Facial Action Database [45], and individuals with FP dataset: FIFA, Facial palsy Intensity of Facial Action Database), shows our model outperforms the state-ofthe-art model for activation unit intensity detection on benchmark data by Fan et al. [21], hereon referred to as FAU-Net.

While we acknowledge that face-to-face interaction with a therapist is essential in physical therapy [24], we believe that virtual applications like ours can fill gaps (accessibility, affordability, selfmonitoring, etc.) in the current treatment process by supplementing human expertise [10, 25, 47] only available during in-person therapy session.

2 RELATED WORK

The work presented in this paper builds primarily on tracking and visualizing facial activation for physical therapy.

2.1 Mirror Therapy for Facial Paralysis

Central FP results from damage to the central segment of the facial nerve [69] while peripheral FP results from injury or damage to extratemporal segments of the facial nerve [56, 64, 74]. Bell's Palsy, the idiopathic form of peripheral FP accounts for 60-75% of its cases. MT is an effective adjunctive treatment for FP [8, 16, 32]. Approaches range from real to virtual and advanced computer vision based mirror setups. One real mirror-based setup is Mirror Book Therapy [54], in which those with FP use a dual-folding mirror so that they see their full, unaffected face while performing a series of facial expression exercises [9]. In a simple virtual setup, a tablet PC was used for MT in FP patients after a stroke, resulting in significantly greater improvement in facial movement [32]. Recently, using computer vision technology, a novel face mirroring system (FMS) was engineered to create a healthy facial illusion, resulting in a better user experience and perception of facial embodiment [17]. However, mirror-based setups limit treatment efficacy [17]. Covering the affected side prevents mirror exposure anxiety [8] but people lose the opportunity to receive biofeedback on the paralyzed facial muscles. During MT, users tend to lean towards the affected side and are not often not aligned with the mirror axis [33, 46]. Thus, users do not see a well formed reconstruction of their non-paralyzed face and symmetrical face in the mirror but instead see deformed reflections along the mirror book axis.

None of these approaches provide muscle education or automatically detect the affected side of the face or provide extended quantitative and qualitative feedback. FaraPy combines these three elements, missing in previous approaches to MT, into a single system that can provide users with real-time feedback and muscle education while performing facial exercises at home.

2.2 Augmented Reality in Healthcare

AR adds virtual objects and information into the real world in ways that make the virtual elements appear part of the physical environment [7]. One widespread use of AR is entertainment in social media applications like Snapchat and others [5, 55, 63]. Snapchat offers facial and world lenses viewed through the front or the rear camera of a mobile device. These lenses are virtual effects and transformations to a user's face or 3D object and text additions to the physical environment, often shared as photos or videos [63].

Beyond entertainment, there is an increasing demand for AR solutions in the medical field focused on surgery, therapy, rehabilitation, and training [19, 44]. In these areas, AR provides several benefits including "visualization, directing attention, intrinsic benefits of motivation, physical interaction activating kinesthetic schemes, patient safety, skill retention, simulation confidence related to transferability, mobile learning and using oneself as a learning object" [36, 49, 52, 61]. There is a growing body of evidence evaluating the use of AR for upper and lower limb stroke rehab [25]. Alamri et al. [6] proposed an AR framework for rehabilitation after stroke with a webcam and fidicial markers making it cumbersome to setup and use. Other work has investigated grasp-and-release tasks [40] and hand opening tasks [41] for post-stroke rehab. GenVirtual [13] is a musical AR game designed to improve muscle coordination in finger and toes though individuals with low muscle movements are unable to play. Luo et al. [65] presented a virtual training environment for post-stroke rehab of hand-opening though the wearable equipment used for therapy was difficult to put on without help.

Compared to the above platforms, our system does not require the individual to wear any equipment or markers. FaraPy works on mobile devices that people most likely already own making it easy to use without the need for extra help. The camera based facial muscle detection supports a variety of facial exercises commonly prescribed for individuals with an FP condition while providing feedback in real-time on how much and how well a muscle is being activated. To our knowledge, ours is the first AR system explicitly tailored for FP rehabilitation.

2.3 Facial Action Unit Intensity Estimation

Facial expressions are generated by contractions of facial muscles, which results in temporary changes to eye lids, eye brows, nose, lips, and skin texture [22]. The facial action coding system (FACS), developed by Ekman and Friesen [23] uses 44 AUs to describe facial actions with regard to their location and intensity (AUI) which is coded with five levels of magnitude [22]. Facial expressions may be modeled by single AUs or combinations to reflect emotions, social interaction, and physiological signals. Over the last 30 years, there has been extensive research on facial expression analysis using FACS [22, 31, 43, 53, 76].

Deep neural networks have recently been used for AU detection [12, 26, 30]. However, estimating different levels of AUIs remains a far more challenging task. Most existing approaches use supervised techniques for AU intensity estimation such as [48, 51, 72] while other methods exploit spatial relationships among the intensities of multiple AUs through probabilistic graphical models [57, 70]. Prior works have also used weakly supervised convolutional neural networks [78], hybrid Bayesian Networks to capture global dependencies among AUs [72] or presented a general framework for AU intensity estimation [79]. A recent heatmap regression framework preserves semantic information of AU intensities (AUIs) as well as their locations [21]. The authors demonstrated the effectiveness of their framework on two benchmark datasets BP4D [77] and DISFA [45], outperforming state-of-the-art models.

While these approaches produce accurate results, they are highly complex and cannot run in real-time on a device with limited computational power, such as a smartphone. Moreover, none of these systems consider each side of the face independently, which means that AUIs are decoded and estimated at a bilateral level. In contrast, FaraPy presents a light weight architecture that enables real-time processing on a mobile device while providing unilateral AUI estimation capable of detecting asymmetric facial muscle activation.

3 SYSTEM DESIGN

FaraPy is a real-time interactive visual feedback system that supports at-home physical therapy to help individuals with FP drive their own recovery outside of in-person therapy [20]. Our user experience design goals focus on building an easy to use, informative, efficient, and novel interface. Figure 1 shows our system pipeline. Input facial video from the mobile device camera is processed by our LW-FAU model to generate real-time feedback which is visualized for the user in AR. The computed feedback is shown as an augmentation to the user's face and as a data visualization of activated facial muscles (Figure 2).

To help readers replicate FaraPy, we now provide the necessary design and technical details. The system has two main components: (1) a back-end LW-FAU deep learning model, and (2) a front-end AR interface.



Figure 2: FaraPy AR Interface: Muscle Activations and Symmetry Feedback for (a) Rest Pose and (b) Asymmetrical rightsided Teeth Clenching. The fifth AU in our system is not activated with these facial expressions and is therefore not visible in the color bars.

3.1 Light Weight Facial Activation Unit model (LW-FAU)

The feedback provided in AR is based on estimated real-time facial muscle activations. Most AUs in FACS are defined at the bilateral level, i.e., muscles with visible activations on both sides of the face are considered a single AU. For example, Lip Corner Puller (AU12) is a smile with the same muscle activated on both sides of the face (Figure 3a). Together with Cheek Raiser (AU6) it represents the facial expression corresponding to Happiness.

To our knowledge, there is currently no model that predicts unilateral labels nor dataset with unilateral AUI labels. Our model is the first to explicitly provide one-sided AUIs as final values in realtime on a mobile device. The model design presented two unique UIST '21, October 10-14, 2021, Virtual Event, USA



Figure 3: (a) Facial expression examples represented as AU combinations [37]. (b) Student-teacher learning setup where our LW-FAU (student model) learns from FAU-Net [21] (teacher model) using the DISFA training dataset [4].

challenges. The first was to detect AUs and AUIs independently for muscles on each side of the face. This is critical for individuals with FP as we need to track muscles on the paralyzed side of the face separately from the same muscles on the non-paralyzed or healthy side of the individual's face. The second challenge was for the model to run in real-time on a mobile device. To address these challenges we developed a new Light Weight Facial Activation Unit model (LW-FAU) that can estimate facial muscle movement intensities at the unilateral level. In contrast with FAU-Net by Fan et al. [21], our work differs in three main aspects: (1) the method of deriving the value (FAU-Net: from a heat map; LW-FAU: final predictions of a neural network), (2) the architecture of the model (FAU-Net: Complex learning framework using Semantic Correspondence Learning with Dynamic Graph Convolution; LW-FAU: A single Convolutional Neural Network with Depth-wise Separable Convolutions), (3) the model size and efficiency of prediction (ONNX file size FAU-Net: 132,720 KB; LW-FAU: 381 KB).

In particular, the LW-FAU model is built using the concepts of depthwise separable convolution, multi-task learning, and knowledge distillation.

3.1.1 Depthwise Separable Convolution. Depthwise separable convolution was first introduced by Sifre et al. [62] and subsequently used in MobileNets [28], a class of computationally efficient models used for mobile and embedded vision applications. The MobileNet lightweight architecture consists of depthwise separable filters, a form of factorized convolution resulting in reduced computation time and model size. A kernel is split into two separate convolutions (a) a depthwise convolution, and (b) a point-wise convolution [28]. The depthwise convolution applies separate kernels to each channel giving an output image with the same number of channels as the input image. It is highly efficient with far fewer computations compared to standard convolution [28]. The point-wise convolution applies a 1x1 kernel, essentially a kernel that iterates through every single point in the image to create outputs with as many channels as desired [28]. For faster computation with fewer parameters, we incorporate depth-wise separable convolution with batch normalization from the MobileNet V1 [28] architecture in LW-FAU.

Barrios Dell'Olio and Sra

Table 1: LW-FAU Architecture.

Туре	#Block	Layer/Stride	Input Size	Output
		Conv2d/s1	64x64x3	3
1x		BatchNorm2d	64x64x3	-
Shared	1	ReLU	64x64x3	-
		Downsample	64x64x3	-
		SeparableConvBN/s1	32x32x3	16
	2	ReLU 32x32x3		-
		Downsample	32x32x3	-
	3	SeparableConvB/s1	16x16x16	32
		ReLU	16x16x16	-
		Downsample	16x16x16	-
	4	SeparableConvBN/s1	8x8x32	32
10x		ReLU	8x8x32	-
Branch		Downsample	8x8x32	-
		SeparableConvBN/s1	4x4x32	64
	5	ReLU	4x4x32	-
		Downsample	4x4 x32	-
	6	AdaptiveAvgPool2d	1x1x64	-
		Conv2d/s1	1x1x64	64
		Flatten	1x1x64	-
		Linear	1x64	-

3.1.2 Multi-Task-Learning. While MobileNets enable efficient computation for vision tasks, the base architecture only provides for solving one task at a time. However, our goal is to simultaneously predict the AUI of multiple facial muscles, where each unilateral AUI estimation represents a unique task. Multi-task learning (MTL) is a machine learning training paradigm defined as learning or solving multiple tasks simultaneously through a common model by exploiting similarities and differences between the tasks [14]. Traditionally, many multi-task architectures in computer vision follow the principle of a global feature extractor consisting of convolutional layers shared by all tasks, referred to as the "shared trunk." We extend this standard shared model by adding a separate branch for each task in the output layer of the network (i.e., *n* branches for *n* tasks) [14]. Since shared trunks are particularly effective for face recognition tasks [14], we incorporate this approach into our model by adding shared layers for high-level learning and splitting into multiple branches for low-level learning, with each branch specializing in a unilateral AU (Table 1).

3.1.3 Knowledge Distillation. Neural models have shown success in a large variety of tasks. However, these models can have millions of parameters making them too large to be deployed on mobile devices. Knowledge Distillation (KD) is the idea of model compression by teaching a smaller network using a bigger trained network. It is often characterized by a Student-Teacher-Learning (STL) framework [71]. STL is an effective technique to transfer knowledge from a complex model (teacher) to a lightweight, efficient model (student). Since our goal is to perform the challenging task of AUI detection in real-time on a mobile device, we required a lightweight but powerful model. We use FAU-Net as the teacher for our learning our efficient LW-FAU model. Specifically, we used the "model compression and knowledge transfer" approach [71]. The student model learns from its teacher's predictions, which serve as the gold labels. Thus for us the predictions of the pre-trained FAU-Net served as the gold standard. During the learning process (Figure 3b), the mean absolute error (MAE) between the predictions of LW-FAU and FAU-Net was calculated to evaluate LW-FAU's performance.

Model Target Values: While FACS assumes symmetric faces and decodes AUs at the bilateral level, we are specifically interested in detecting AUI divergence between the affected and unaffected sides of the face in individuals with FP. The FAU-Net model creates heatmaps as an intermediate step to visualize AU activations on the left and right sides of the face [21]. They calculate final bilateral AUIs AUI_{Final} by averaging the unilateral activations, as follows:

$$AUI_{Final} = \frac{AUI_{LHeatmap} + AUI_{RHeatmap}}{2} \tag{1}$$

While unilateral AUI heat maps are an intermediate step in their approach, they serve as our target values $AUI06_L$, $AUI06_R$, $AUI10_L$, $AUI10_R$, $AUI12_L$, $AUI12_R$, $AUI14_L$, $AUI14_R$, $AUI17_L$, and $AUI17_R$, which we use in the next step to label the benchmark and our new dataset of individuals with FP.

3.2 LW-FAU Architecture

The base LW-FAU architecture (Table 1) is composed of six stacked building blocks, each with (1) a depthwise separable convolution, followed by (2) a batchnorm, (3) a ReLu nonlinearity, and (4) a downsampling layer. The downsampling layers were introduced by Zhang et al. [79] to deconvolve the network with classical signal processing techniques that help make the network shift-invariant, leading to increased accuracy and better generalization. The first layer is a standard convolution, and the last fully connected layer comes without the ReLu activation layer. A final average pooling reduces the spatial resolution to one. Downsampling is addressed in the first layer and within each building block after the separable convolution.

For MTL, the base architecture was split into a sequence of one shared (first building block) and 10×5 unshared layers (five remaining building blocks). We designed for high-level features to be shared by all tasks and low-level features to be learned by specialized tasks. We determined the network's splitting point through experimental testing. Five different versions of the multi-task model were created by splitting the model after each of the six building blocks. The model was validated after 50 epochs of training, with the best results obtained with a total loss of $L_total = 136.88$ and a learning rate of LR = 1,000e - 07 for a split after the first building block. Building on the findings of [28], we applied a weight decay (*Weight_Dec* = 1e - 4) to the depthwise filters to prevent overfitting and to penalize complexity, albeit with small magnitude since the convolutions contain few parameters.

Table 2 shows all hyper-parameters used in training.

3.3 Training

Our model is trained in PyTorch using the STL approach, with the FAU-Net serving as a teacher. Figure 3b depicts the training process, where the MAE between our model's predictions and FAU's intermediate heatmap values is calculated and propagated back

Table 2: Hyper-parameters for LW-FAU Model Training.

Parameter Value Description			
Batch_Size	64	# Training samples per iteration	
Epochs	100	# Training epochs	
LR	1e-2	Learning Rate	
LR_Dec_Step	7	# Epochs before LR decrease	
LR_Dec_Gamma	0.1	Decaying coefficient	
Weight_Dec	1e-4	Weight dec. coeff., overfit/complex	
Cons_Weight	0.1	Coeff. for consist. part of loss fct.	
Num_Workers	4	# Parallel image loading	

through LW-FAU. Precisely, the multi-task model's loss is computed by accumulating the individual MAE losses $loss_n$ for each of the ten branches of the model.

Thus, for $loss_n = MAE(X_{pred}, X_{true})$, the loss function L_{Total} can be defined as follows:

$$L_{Total} = \sum_{n=1}^{10} loss_n \tag{2}$$

3.3.1 Datasets. For training and validation, we used the benchmark facial expression dataset DISFA (Denver Intensity of Spontaneous Facial Action Database), a non-posed facial expression database with videos from 27 adults [4]). We labeled DISFA with the unilateral target values described in section 3.1.3, using the intermediate predictions of the FAU-Net model on the data to allow direct comparison between FAU and LW-FAU performance. In addition to labeling DISFA, for testing, we created a new dataset FIFA with video data from FP individuals, which was also labeled with FAU-Net using the same procedure described above used to label DISFA. FIFA is a posed facial expression dataset for researchers interested in developing computer algorithms for automatic detection of AUs and their intensities in individuals with FP. This dataset contains 5-minute videos of the 20 study participants (16 women and 4 men from different ethnicities) sequentially performing the five facial exercises in our user study. Videos were captured using Zoom's local capture option during the remote user study and cropped to 256x256 resolution. To our knowledge, ours is the first image FP dataset labeled with unilateral AUIs.

We would like to emphasize our high priority to protect user privacy, especially given the sensitive nature of the FP condition and the need for including videos with full facial information in the released dataset. That said, we are committed to enable future researchers to develop new technologies that can advance management and care for individuals with FP and related conditions. We commit to releasing the dataset for research purposes in a controlled manner, with access available after submission of an agreement form (similar to the form required to download the DISFA dataset). We will make our trained model and code publicly available on Github.

3.4 Augmented Reality Interface

The interactive feedback system's architecture consists of six components, illustrated in Figure 1. The LW-FAU model runs in the back end of the AR application and uses the input facial video data



Figure 4: Components of the AR Muscle Education / MT Filter: (a) Face Mesh UV file, Image from [11] (b) 3D model of the Face Mesh, Image from [11] (c) 2D Muscle Education Texture, Right-Sided, Image from [18]

from the user's mobile device camera to independently estimate AUIs for each side of the face. Specifically, 10 unilateral AUI float values of the following five AUs (with corresponding facial muscles) are computed (Figure 3a): AU06 ("Cheek Raiser": Orbicularis oculi, pars orbitalis), AU10 ("Upper Lip Raiser": Levator Labii Superioris, Caput infraorbitalis), AU12 ("Lip Corner Puller": Zygomatic Major), AU14 ("Dimpler": Buccinator), and AU17 ("Chin Raiser": Mentalis) [23].

After retrieving the model's predictions, we determine which side of the face is paralyzed. Since unilateral FP is characterized by weakness on one side, we identify the affected side based on overall lower muscle activations as shown in (3). Let \mathcal{AU} be the set of all estimated AUIs and $AU_{Ln} \cup AU_{Rm} \in \mathcal{AU}$ holds for all left (AU_L) and right sided (AU_R) AUS. The affected side S is defined as:

$$S = \begin{cases} Left, & \text{if } \sum_{n=1}^{5} AU_{Ln} < \sum_{m=1}^{5} AU_{Rm} \\ Right, & \text{otherwise.} \end{cases}$$
(3)

The goal of facial exercises is to produce visually discernible control of small facial muscles for creating a symmetrical facial expression [39]. Using the AUI outputs from our model, we calculate the ratio between the activations of the healthy and unhealthy sides of the face. This difference *dRatio* is visualized for the user in real-time per muscle group (Figure 2). The ratio provides information about how evenly, and thus symmetrically, a particular facial muscle group was activated and can serve as an indicator of performance aimed at visual symmetry [39].

Let AU_H be the set of all AUIs of the determined healthy face side and AU_U be the set of all AUs of the unhealthy face side. Furthermore, $AU_h \in AU_H$ and $AU_u \in AU_U$. The difference ratio dRatio is then defined as:

$$dRatio = \frac{AU_h}{AU_u} \tag{4}$$

Since an AUI can take a positive value in range 0-5, reflecting the strength of activation, the *dRatio* can never become negative, but it can take a small positive value between 0-1. We generate an overall symmetry Ratio score *symmRatio* for the entire face and define it as:

$$symmRatio = \frac{\sum_{n=1}^{5} AU_{hn}}{\sum_{m=1}^{5} AU_{um}}$$
(5)

AR feedback provides information on facial muscle movements and AUIs. The information is quantified and visualized on the user's

Barrios Dell'Olio and Sra

Table 3: Scoring System: Categories and Intervals of dRatio.

Category	Lower	Upper	Color	Score
Symmetrical Over	-0.1	0.2	Green Purple	1 0
Under	0.2	1.0	Blue	0

face in the form of a muscle filter and on the device screen as bars that dynamically depict the degree of activation as seen in Figure 2. We use a color coding system to indicate symmetry of facial muscle activations and provide feedback on how much (more or less) do the muscles on the paralyzed side need to be activated to match the movements of the corresponding muscles on the healthy side of the face (numerically displayed in the "difference" column, see Figure 2).

For providing muscle education, we created a 3D AR facial muscle filter as shown in Figure 4. This filter maps the 2D texture of a human facial muscle image onto a 3D face mesh [1]. The filter is displayed in real-time on the user's face as a half mask, covering their paralyzed side to reduce anxiety while providing in-situ information on which muscles are currently being activated.

Intrinsic motivation is a key factor in rehab therapy and is often used as a determinant of the outcome [42]. While the AR technology itself provides intrinsic motivation [49], we also integrated a gamification element [58] that would track each individual's progress over time, something a mirror is unable to do. To that end we added a scoring system that rewards users for more symmetrical facial movements, indicated via color bars showing each muscle's activation during exercise performance.

The scoring system works as follows:

- If a muscle is *symmetrically* activated (i.e., *dRatio* is within the range] – 0.1, 0.2[, see Table 3), the muscle's activation bar turns green, and the user receives a point. This increases the current "score" by one.
- If a muscle is *over* activated (i.e., *dRatio* is in the range [-1.0, -0.1]), the muscle's activation bar turns purple and the user does not receive a point.
- If a muscle is *under* activated (i.e., *dRatio* is in the range [0.2, 1.0]), the muscle's activation bar turns blue, and the user receives no point.

Thresholds for determining whether muscles were symmetrical, over-, or under-activated were defined based on video data (4x6) of six healthy subjects (built in test videos in Lens Studio [63] of 3 female, 3 male individuals of different ethnicities; facial expressions performed: (1) neutral, (2) open mouth, (3) raised eyebrows, and (4) smile). The ratios obtained from these healthy subjects were used to determine the range of acceptable inter-subject variation for "symmetric" muscle motion. Values outside this range were defined as "asymmetric" according.

Points are collected and accumulated from the moment the user presses the start button. Users can track their current score $S_{Current}$ under "Score" in the application header (Figure 2). The best, highest score to date S_{High} is stored under "Best" in the menu. This logic can be formulated as follows:

FaraPy: An Augmented Reality Feedback System for Facial Paralysis using Action Unit Intensity Estimation

$$S_{High} = \begin{cases} S_{Current}, & \text{if } S_{Current} > S_{High} \\ S_{High}, & \text{otherwise.} \end{cases}$$
(6)

The system was implemented using Lens Studio 3.3 [3], a desktop tool for creating AR experiences for mobile devices. By building in Lens Studio, we were able to make our application accessible to study participants via SnapChat [63] by simply sharing a SnapCode [3] Our system stores neither the personal input data nor the processed output for data protection and user privacy. Only the high score is stored which is visible to the user. We did not access any local data from the user's mobile device.

4 TECHNICAL EVALUATION

The model was trained for a total of 100 epochs.

The best results were obtained at *epoch* = 85 with a validation loss of L_valid = 177.688, a total loss of L_total = 211.564 and a learning rate LR = 1.000e - 14.

4.1 Benchmark Data

LW-FAU outperforms the current state-of-the-art FAU-Net [21] by achieving a lower mean MAE of $MAE_{\mu} = 0.14$ and a higher ICC $ICC_{\mu} = 0.77$. These results are only indirectly comparable with benchmark models because they provide two-sided AUIs. However, comparing with FAU-Net we found no significant differences in AUI μ values for p < 0.05, across all AUs (Figure 5) and participants. Since our model was trained using FAU-Net as the teacher model, this result indicates that we can get comparable performance with a much smaller footprint making it is well suited for real-time use.

4.2 User Data

Testing our proposed model on our created FIFA dataset yielded a mean MAE of $MAE_{\mu} = 0.41$, which is within the range of FAU-Net performance on benchmark data with $MAE_{\mu} = 0.58$ for the BP4D and $MAE_{\mu} = 0.20$ for the DISFA datasets. While the mean ICC was moderate at $ICC_{\mu} = 0.33$, high ICC values were seen for certain participants (e.g., $ICC_{FPS002} = 0.68$, $ICC_{FPS003} = 0.60$). 80% (160 of 200) of the ICC measures were significant for p < 0.05, indicating that in almost all cases there was a positive correlation between the student and teacher predictions. Looking at μ AUI values across all patients, significant differences were found between student and teacher predictions in 50% of cases. Measuring AUI μ across all AUs, there were significant differences 20% of the time (Figure 6). We speculate on two possible reasons for this performance drop on user data, particularly for certain individuals. First, there may be limited generalization of the model from the training data (benchmark DISFA dataset of healthy faces) to real-world test data (FIFA dataset of individuals with FP). Second, the quality of the FIFA dataset might be causing the performance drop for some individuals. The dataset is built from Zoom recordings which have dependencies on local conditions of each study participant (webcam quality, lightning, motion, obstacles, positioning).

4.2.1 *Degree of Facial Paralysis.* While the model showed differential performance on some user data, the degree of FP does not appear to be the reason. For p < 0.05, non-significant Pearson correlations were found between model performance (MAE and ICC)

Table 4: Pearson r: FP Degree & Model Performance, for p < 0.05.

Metric	Pearson r	p-value	power	CI95%
MAE	-0.315	0.177	0.278	[-0.66, 0.15]
ICC	0.072	0.764	0.060	[-0.38, 0.50]

and degree of FP (Table 4). However, the tests were of low statistical power with *power* < 0.80 and would need to be repeated with a larger sample size to increase the likelihood of finding no effect between FP grade and model performance, if there is none. Degree of FP determination was inspired by the facial reanimation measurement system [68]. This system measures angles and distances between landmarks in resting and smiling positions. We introduced a new assessment technique by taking the ratios between the 2sided measurements of the right and left commissure (RC, LC), right and left middle upper lip (RMUL, LMUL), and right and left middle lower lip (RMLL, LMLL) in the frontal view of the face, as shown in Figure 7. We then took the mean ratio R_{Avg} , which corresponds to the continuous FP degree over all three ratios R_C , R_{MUL} , and R_{MLL} , defined by:

$$R_{Avg} = Mean\left(\frac{min(RC,LC)}{max(RC,LC)}, \frac{min(RMUL,LMUL)}{max(RMUL,LMUL)}, \frac{min(RMLL,LMLL)}{max(RMLL,LMLL)}\right)$$
(7)

Our technique quantifies how symmetric a smile activation is compared to the resting pose. A higher R_{Avg} value represents higher facial muscle functionality.

5 USER EVALUATION

We performed a study with 20 remotely located participants to better understand the subjective user experience when interacting with FaraPy (Figure 8). The study protocol was approved by *anonymized*. FaraPy was integrated with the Snapchat [63] application for easy access and use. Participant interaction with FaraPy was video recorded (Zoom recording) and the data used to build an FP individuals dataset as well as evaluate our model on user data.

5.1 Participants

We recruited 20 adults with FP from FP support groups on Facebook (16 self-identified as female; 4 as male; mean age = 45 years old, SD = 11.18). We created a non-sponsored social media post that described the feedback system and provided our academic email address so that interested individuals could contact us on a voluntary and self-determined basis. Personal information available to Facebook group members was not used for outreach or recruitment. Moderators reviewed and approved our post before it was shared with the communities.

We excluded participants with complete or early-stage FP since we needed individuals who could perform the facial exercises without risk of harm [67]. 65% of the participants reported Bell's palsy as their FP condition, 25% were affected by Synkinesis, and 10% by the Ramsay Hunt syndrome. The largest group of participants (45%) had been affected by the FP disease for more than two years, 40% from 0.5 to 2 years, and 15% for less than 0.5 years. Half of the study



Figure 5: AUI μ over all AUs, on DISFA benchmark data: LW-FAU (pred) vs. FAU-Net (true). No significant differences were found, for *p* < 0.05.



Figure 6: AUI μ over all AUs on FIFA user data: LW-FAU (pred) vs. FAU-Net (true). Significant differences are marked with *, for p < 0.05.

participants accessed FaraPy from an iPhone mobile device, while the rest tested FaraPy from Android devices. Participants received a \$20 gift card for 30 minutes of their time.

5.2 Procedure

The study took place over a Zoom videoconferencing call. Following consent, participants were informed about the technology that was being studied: a real-time MT system in AR. They were then guided through the download, installation and setup process of our AR interface on their smartphone device. They were instructed on color changes in the muscle activation indicator bars based on their facial expressions, with green being the target color and blue and purple indicating under and over activation of a muscle requiring correction (Figure 2 and Table 3). They were told a facial filter showing muscles would display on one half of their face with labels for each muscle as an aid to memory and learning. The



Figure 7: FP Degree determination based on three twosided ratio measurements RC, LC; RMUL, LMUL; and RMLL, LMLL. (a) Left-sided paralysis and (b) right-sided paralysis example.



Figure 8: Virtual study setup: (1) Researcher instructs participants and records Zoom session. (2) Participants perform facial exercises with FaraPy using their mobile device camera to retrieve feedback. Their laptop webcam records their facial expressions as they perform the facial exercises.

muscles shown on the face were also visualized in the color bars to help the participants consciously attempt to activate specific facial muscles to turn the bar green, indicating correct exercise performance. Following these instructions, participants were asked to perform five facial exercises 1) smiling, 2) opening and closing the jaw, 3) saying A-O-I, 4) clenching teeth, 5) making a big yawn. Each exercise was repeated five times resulting in a total of 25 facial exercises per participant. These exercises were taken from an online physical therapy video resource [75]. Each facial exercise took approximately 1 minute for a total of 5 minutes of exercise per participant.

After completing the exercises, participants were asked about their overall impression and if FaraPy improved their MT experience in a semi-structured interview. They also filled out a post-study UX questionnaire by Schrepp et al. [60] that included a series of questions about their perception of the application's (1) Attractiveness ("Overall impression of the product. Do you like it or dislike our system? Is it attractive, enjoyable, or pleasing?"), (2) Perspicuity ("Is it easy to get familiar with the system? Is it easy to learn? Is the system easy to understand and clear?"), (3) Efficiency ("Can you solve your tasks without unnecessary effort? Is the interaction efficient and fast? Does the system react fast to your input?"), (4) Dependability ("Do you feel in control of the interaction? Can you predict the system behavior? Do you feel safe when working with the system?"), (5) Stimulation ("Is it exciting, motivating or fun to use the system?"), and (6) Novelty ("Is the system innovative and creative? Does it capture your attention?"). Participants indicated their level of agreement on a Likert scale where from 1 (strongly disagree) to 7 (strongly agree). The questionnaire also included six open-ended questions asking about their experience and feedback.

From the UX questionnaire and the semi-structured interviews, we inferred that none of the participants had previously engaged in mirror therapy, but that home exercises in front of a conventional mirror were popular. This suggests that FaraPy is not only a realtime interactive evolution of mirror therapy (central-transverse positioning of the mirror to reflect the healthy body half onto the unhealthy half), but also of exercising in front of a mirror in general (frontal positioning of the mirror).

5.3 Results and Discussion

Overall, we found that participants were very satisfied with the AR feedback system, with a μ rating of 6.4 (SD = 0.8) out of 7 across all items in the questionnaire. From the written responses to the open-ended questions and the semi-structured interview data, we inferred that most participants preferred FaraPy to a mirror for performing facial exercises (85%) and would like to continue using FaraPy (95%).

Figure 9 shows participants' ratings for FaraPy, measured across the six UX categories in the questionnaire. In particular, participants rated efficiency, dependability, stimulation, and novelty high with a mean score of $\mu >= 6.5$ and low standard deviation SD <= 0.8. Within this group, stimulation was rated the highest. 70% of participants rated stimulation ($\mu = 6.9$, SD = 0.3) with the highest score of 7 across all four items operationalizing the constructs: valuable, exciting, interesting, and motivating. No participant assessed any of these four items with less than a 6.



Figure 9: Quantitative results on six different UX categories. Highest μ ratings were measured for stimulation (6.9) and novelty (6.8).

In contrast, attractiveness and perspicuity received a wider range of evaluation scores with a mean $\mu >= 5.9$ and SD <= 1.4. While most participants rated attractiveness ($\mu = 6.1$,SD = 1.4) high (70%) assessed attractiveness higher than 6.3), two (10%) participants gave less than neutral ratings, with the lowest possible score of 1 on the items good, attractive, and friendly. Quotes from these individuals in response to the open-ended question ("Please rate whether (and why) you like using our system and would like to use it in the future.") in the Attractiveness section are: "Yes" (FPS010) and "I would definitely use this product" (FPS012). These positive qualitative responses contradict the quantitative ratings these participants provided. Both these participants rated the other five categories positively with $\mu = 6.7$, SD = 0.8. This could indicate a misunderstanding of the item rating system because of the included control items in the original questionnaire to maintain participant attention and engagement (e.g., positive descriptors were not always on the right side of the Likert scale [60]).

In its current state, our system seems to satisfy the needs of the participants as evident in the questionnaire results and feedback. Regarding the real-time feedback provided by FaraPy, FPS001 said, "I like how the muscles were identified/separated to focus on each movement individually." FSP016 similarly remarked, "I've never received direct feed back before." Participants were excited about the scoring system with FPS006 saying, 'The scoring and gamification aspects make it more interesting to me and at this point." FPS020 found FaraPy motivating, which leads us to believe there is potential in digital therapies to engage individuals in the process of their recovery, outside of therapy sessions. They said, "MORE MOTIVATION I LIKE THE COLORS AND THE GAME," and "Motivating I felt like I was in a video game." Participants appreciated being able to perform exercises at home. FPS008 said, "I think it's really great to see the feedback at home." FPS011 was excited about FaraPy running on their phone saying, "Because I can use the phone and be with it everywhere, it's a good innovation." FPS013 was glad for access to Farapy and said, "Since there is no medical professional physically monitoring the exercises, it's the only way for patients with facial paralysis to follow their own process and notice how their face moves."

5.4 Design Considerations

Study participants offered suggestions and feedback which could be included in future versions of the system, though some suggestions may require therapist input before being integrated. FPS008 expressed a desire to slow-down the feedback saying, "It's great that it's in real-time, but that makes it almost too difficult to record the feedback since the exercises are continuous." FPS007 indicated need for instructions and help built into the system, not only for using the application but also for the exercises. They said, "first time without supervision would be a bit difficult, but it's easy to get the hang of. I need clarity on the feedback and how to exercise certain muscles more." FPS013 remarked upon the disconnect between the feedback showing on screen vs on the face. They said, "If the colors of the muscles changed instead of the bars at the bottom, that would be helpful, and there would be less information to figure out which muscle is being exercised too much/too little." Study participants also expressed a desire for eye-tracking that would

Table 5: Pearson r: User Satisfaction & Model Performance, for p < 0.05.

Metric	Pearson r	p-value	power	CI95%
MAE	0.164	0.489	0.107	[-0.3, 0.57]
ICC	0.360	0.119	0.355	[-0.1, 0.69]

allow them to perform synkinesis-related exercises, if needed. We believe this information will be valuable to researchers interested in applying AI and AR technologies to future mirror therapy for FP with synkinesis.

5.4.1 Model Performance and User Satisfaction. User satisfaction does not appear to be correlated to model performance. When measuring model performance (in MAE and ICC), no significant Pearson correlation was found for p < 0.05 with the variable user satisfaction, operationalized as mean UX questionnaire score per participant (Table 5). This means that slight losses in accuracy might be acceptable in favor of efficiency in a mobile application, as they do not seem to adversely affect user satisfaction. It is possible that participants were unable to perceive or evaluate the model's performance or if they could, it did not impact their level of satisfaction with the application. Additionally, the sample size of 20 may not have been sufficient to draw firm conclusions. The tests were of low statistical power with power < 0.80 and would need to be repeated with larger sample sizes. Beyond satisfaction, the impact of any estimation inaccuracies on long term recovery needs to be explored in further studies.

6 LIMITATIONS AND FUTURE WORK

The results of the technical and user evaluations revealed several research questions that motivate future work on machine learning based patient care systems. In the following section we discuss limitations and research opportunities beyond those findings:

6.1 Lack of FP Benchmark Models and Data Sets

There are no benchmark models or datasets for one-sided AUI values. Manual labeling of datasets is time and cost intensive and in this particular case, requires expertise in detecting and labeling facial muscle activations for paralyzed faces. Therefore, we used the current state-of-the-art FAU-Net model for labeling our FIFA user dataset and the DISFA benchmark dataset. However, the FAU-Net model has never been explicitly trained on individuals with FP. Before applying the FAU-Net model to the entire user dataset, we probed its feasibility by applying it to FP data samples from individuals with FP condition retrieved through a Google image search and obtained reasonable intensity predictions. This could be explained by FAU-Net's design, which considers one-sided AUIs as an intermediate step in the heat map output. Thus, the model may capture this deviation from known standard inputs of symmetric faces better than benchmark approaches. Thus, we measured how well our LW-FAU model can reproduce FAU-Net's predictions, rather than how close the LW-FAU predictions are to the datasets' true labels.

Future work could explicitly address the lack of accessible datasets on individuals with FP. There is also a need to build a comprehensive system that describes facial muscle activations at a more granular level, e.g., by considering both halves of the face independently. Furthermore, more work is needed in building efficient and powerful ML models that can be applied to real-world scenarios and mobile devices. We hope more researchers feel encouraged to focus on problem areas that deeply affect people, even if it is only a niche user group. After all, these individuals could benefit the most from research advances in machine learning and immersive technologies.

6.2 Lack of Local User Data

We did not have access to participant device data. Due to COVID-19, we were forced to conduct our study with remote participants. Therefore, we could not test our system on a single device under controlled but randomized conditions in our laboratory. We reconstructed how the model might have performed by capturing videos in Zoom of participants interacting with our system and feeding this data as input to our LW-FAU model. Follow-up studies are needed to test the model's performance on a single device for better comparison and standardization.

6.3 User Study Sample Size

Although our sample size of $sample_{size} = 20$ was equal to or higher than that in benchmark FP user studies [8, 32, 33], the small number does impose limitations on interpreting the statistical results and generalizability. Future studies could be conducted to validate identified trends with higher statistical power.

6.4 AU Selection

FP with synkinesis requires focus on facial exercises related to eye movements. This type of FP and related exercises are not fully supported by our current system which lacks explicit eye-tracking. For our model LW-FAU, we prioritized AUs related to the facial mouth parts and those were the main focus of FP exercises performed during the user study. Future versions of our prototype would be enhanced by incorporating eye-tracking to support synkinesis-related exercises.

7 CONCLUSION

In this work we presented FaraPy, the first mobile augmented reality mirror therapy system that provides real-time feedback on facial exercises for individuals with facial paralysis. FaraPy enables individuals with FP to continue therapy at home by providing feedback on their facial expression symmetry, a target goal of facial exercise therapy. To support real-time feedback, we developed a lightweight facial activation unit intensity detection model. Unlike all existing models that detect bilateral facial muscle activation, our model detects muscle activations independently for each side of the face, a necessity for facial paralysis patients.

Our evaluation showed that our lightweight model outperforms the state-of-the-art model on benchmark data for estimating unilateral action unit intensities and achieves comparable performance FaraPy: An Augmented Reality Feedback System for Facial Paralysis using Action Unit Intensity Estimation

on user data. Results from our user study of 20 participants with facial palsy showed high levels of satisfaction and greater preference for AR over traditional mirror therapy based biofeedback.

ACKNOWLEDGMENTS

We thank Atieh Taheri and Bowen Zhang for help with testing our system.

REFERENCES

- [1] [n.d.]. Face Mesh Lens Studio by Snap Inc. https://lensstudio.snapchat.com/ guides/face-fface-effects/face-mesh/
- [2] 2017. Facial nerve palsy including Bell's palsy: Case definitions and guidelines for collection, analysis, and presentation of immunisation safety data. Vaccine 35, 15 (2017), 1972–1983. https://doi.org/10.1016/j.vaccine.2016.05.023
- [3] 2/12/2021. Lens Studio Lens Studio by Snap Inc. https://lensstudio.snapchat. com/
- [4] 2/23/2021. DISFA+ Mohammad H. Mahoor, PhD. http://mohammadmahoor. com/disfa/
- [5] Adobe Aero 2020. Augmented reality. Now a reality. Retrieved March 22, 2021 from https://www.adobe.com/products/aero.html
- [6] Atif Alamri, Jongeun Cha, and Abdulmotaleb El Saddik. 2010. AR-REHAB: An Augmented Reality Framework for Poststroke-Patient Rehabilitation. *IEEE T. Instrumentation and Measurement* 59 (10 2010), 2554–2563. https://doi.org/10. 1109/TIM.2010.2057750
- [7] Ronald Azuma, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. 2001. Recent advances in augmented reality. *IEEE computer* graphics and applications 21, 6 (2001), 34–47.
- [8] Jodi Maron Barth, Gincy L. Stezar, Gabriela C. Acierno, Thomas J. Kim, and Michael J. Reilly. 2014. Mirror book therapy for the treatment of idiopathic facial palsy. *Ear, nose, & throat journal* 93, 9 (2014), E11–4.
- [9] Jodi Maron Barth, Gincy L Stezar, Gabriela C Acierno, Thomas J Kim, and Michael J Reilly. 2014. Mirror book therapy for the treatment of idiopathic facial palsy. *Ear, Nose & Throat Journal* (2014), 0145561320913211.
- [10] Alessandra Berton, Umile Giuseppe Longo, Vincenzo Candela, Sara Fioravanti, Lucia Giannone, Valeria Arcangeli, Viviana Alciati, Claudia Berton, Gabriella Facchinetti, Anna Marchetti, et al. 2020. Virtual reality, augmented reality, gamification, and telerehabilitation: Psychological impact on orthopedic patients' rehabilitation. Journal of Clinical Medicine 9, 8 (2020), 2567.
- [11] Lens Studio by Snap Inc. [n.d.]. Face Mesh Lens Studio by Snap Inc. https: //lensstudio.snapchat.com/guides/face/face-effects/face-mesh/
- [12] Ciprian Corneanu, Meysam Madadi, and Sergio Escalera. 2018. Deep structure inference network for facial action unit recognition. In Proceedings of the European Conference on Computer Vision (ECCV). 298–313.
- [13] Ana Grasielle Dionisio Correa, Gilda Aparecida De Assis, Marilena do Nascimento, Irene Ficheman, and Roseli de Deus Lopes. 2007. Genvirtual: An augmented reality musical game for cognitive and motor rehabilitation. In 2007 Virtual Rehabilitation. IEEE, 1–6.
- [14] Michael Crawshaw. 2020. Multi-Task Learning with Deep Neural Networks: A Survey. (2020). arXiv:2009.09796 [cs.LG]
- [15] Rima M Dafer, Murali Rao, Aisha Shareef, and Aparna Sharma. 2008. Poststroke depression. *Topics in stroke rehabilitation* 15, 1 (2008), 13–21.
- [16] HILAL DENIZOĞLU KÜLLİ, ELİF DURGUT, KÜBRA ALPAY, and HÜLYA NİL-GÜN GÜRSES. 2017. A New Adapted Method for Using Mirror Therapy in a Facial Paralysis Patient. (2017).
- [17] Li Ding, Li Li, Zhimin Xu, Jing Tian, Shugeng Chen, Hewei Wang, Ming Yang, Xiao Cui, Lianying Cao, and Jie Jia. 2020. Computer vision technology-based face mirroring system providing mirror therapy for Bell's palsy patients. *Disability* and rehabilitation 42, 6 (2020), 833–840. https://doi.org/10.1080/09638288.2018. 1510551
- [18] Dreamstime. [n.d.]. Gesichtsmuskeln vektor abbildung. Illustration von gesichtsmuskeln - 107588324. https://de.dreamstime.com/gesichtsmuskelnimage107588324
- [19] Martin Eckert, Julia S Volmerg, and Christoph M Friedrich. 2019. Augmented reality in medicine: systematic and bibliographic review. *JMIR mHealth and* uHealth 7, 4 (2019), e10967.
- [20] Xue Wen Eng, Sandra G Brauer, Suzanne S Kuys, Matthew Lord, and Kathryn S Hayward. 2014. Factors affecting the ability of the stroke survivor to drive their own recovery outside of therapy during inpatient stroke rehabilitation. *Stroke research and treatment* 2014 (2014).
- [21] Yingruo Fan, Jacqueline C. K. Lam, and Victor O. K. Li. 2020. Facial Action Unit Intensity Estimation via Semantic Correspondence Learning with Dynamic Graph Convolution. (2020). https://arxiv.org/pdf/2004.09681
- [22] Beat Fasel and Juergen Luettin. 2003. Automatic facial expression analysis: a survey. Pattern recognition 36, 1 (2003), 259–275.

- [23] E Friesen and Paul Ekman. 1978. Facial action coding system: a technique for the measurement of facial movement. *Palo Alto* 3, 2 (1978), 5.
- [24] Mark Fuerst. 2017. Patients Prefer Face-to-Face Communications With Doctors. Oncology Times 39 (11 2017), 62. https://doi.org/10.1097/01.COT.0000527379. 44018.e1
- [25] Cassandra Gorman and Louise Gustafsson. 2020. The use of augmented reality for rehabilitation after stroke: a narrative review. *Disability and Rehabilitation: Assistive Technology* (2020), 1–9.
- [26] Shizhong Han, Zibo Meng, Ahmed Shehab Khan, and Yan Tong. 2017. Incremental boosting convolutional neural network for facial action unit recognition. arXiv preprint arXiv:1707.05395 (2017).
- [27] Adelyn L Ho, Amie M Scott, Anne F Klassen, Stefan J Cano, Andrea L Pusic, and Nancy Van Laeken. 2012. Measuring quality of life and patient satisfaction in facial paralysis patients: a systematic review of patient-reported outcome measures. *Plastic and reconstructive surgery* 130, 1 (2012), 91–99.
- [28] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. 2017. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. (2017). arXiv:1704.04861 [cs.CV]
- [29] Lisa E Ishii, Jason C Nellis, Kofi Derek Boahene, Patrick Byrne, and Masaru Ishii. 2018. The importance and psychology of facial expression. *Otolaryngol Clin North Am* 51, 6 (2018), 1011–1017.
- [30] Shashank Jaiswal and Michel Valstar. 2016. Deep learning the dynamic appearance and shape of facial action units. In 2016 IEEE winter conference on applications of computer vision (WACV). IEEE, 1–8.
- [31] Takeo Kanade, Jeffrey F Cohn, and Yingli Tian. 2000. Comprehensive database for facial expression analysis. In Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580). IEEE, 46–53.
- [32] Jung-A Kang, Min Ho Chun, Su Jin Choi, Min Cheol Chang, and You Gyoung Yi. 2017. Effects of Mirror Therapy Using a Tablet PC on Central Facial Paresis in Stroke Patients. Annals of rehabilitation medicine 41, 3 (2017), 347–353. https: //doi.org/10.5535/arm.2017.41.3.347
- [33] Jinmin Kim, Jaehoon Yi, and Chang-Ho Song. 2017. Kinematic analysis of head, trunk, and pelvic motion during mirror therapy for stroke patients. *Journal of Physical Therapy Science* 29, 10 (2017), 1793–1799. https://doi.org/10.1589/jpts. 29.1793
- [34] Petr Konecny, Milan Elfmark, and Karel Urbanek. 2011. Facial paresis after stroke and its impact on patients' facial movement and mental status. *Journal of Rehabilitation Medicine* 43, 1 (2011), 73–75.
- [35] Martin Krippl, Ahmed A Karim, and André Brechmann. 2015. Neuronal correlates of voluntary facial movements. Frontiers in human neuroscience 9 (2015), 598.
- [36] Sevda Küçük, Samet Kapakin, and Yüksel Göktaş. 2016. Learning anatomy via mobile augmented reality: Effects on achievement and cognitive load. *Anatomical* sciences education 9, 5 (2016), 411–421.
- [37] F la Torre De, WS Chu, X Xiong, F Vicente, X Ding, and J Cohn. 2015. IntraFace. In IEEE Int. Conf. on Automatic Face & Gesture Recognition, Vol. 1.
- [38] Ho Yun Lee, Myung Gu Kim, Dong Choon Park, Moon Suh Park, Jae Yong Byun, and Seung Geun Yeo. 2012. Zoster sine herpete causing facial palsy. *American journal of otolaryngology* 33, 5 (2012), 565–571. https://doi.org/10.1016/j.amjoto. 2012.02.001
- [39] Robin W. Lindsay, Mara Robinson, and Tessa A. Hadlock. 2010. Comprehensive facial rehabilitation improves function in people with facial paralysis: a 5-year experience at the Massachusetts Eye and Ear Infirmary. *Physical therapy* 90, 3 (2010), 391–397. https://doi.org/10.2522/ptj.20090176
- [40] Xun Luo, Robert V Kenyon, Tiffany Kline, Heidi C Waldinger, and Derek G Kamper. 2005. An augmented reality training environment for post-stroke finger extension rehabilitation. In 9th International Conference on Rehabilitation Robotics, 2005. ICORR 2005. IEEE, 329–332.
- [41] Xun Luo, Tiffany Kline, Heidi C Fischer, Kathy A Stubblefield, Robert V Kenyon, and Derek G Kamper. 2006. Integration of augmented reality and assistive devices for post-stroke hand opening rehabilitation. In 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference. IEEE, 6855–6858.
- [42] Niall Maclean, Pandora Pound, Charles Wolfe, and Anthony Rudd. 2000. Qualitative analysis of stroke patients' motivation for rehabilitation. BMJ (Clinical research ed.) 321 (11 2000), 1051–4. https://doi.org/10.1136/bmj.321.7268.1051
- [43] Brais Martinez, Michel F Valstar, Bihan Jiang, and Maja Pantic. 2017. Automatic analysis of facial actions: A survey. *IEEE transactions on affective computing* 10, 3 (2017), 325–347.
- [44] Anabela Marto, Henrique A Almeida, and Alexandrino Gonçalves. 2019. Using augmented reality in patients with autism: A systematic review. In ECCOMAS Thematic Conference on Computational Vision and Medical Image Processing. Springer, 454–463.
- [45] S Mohammad Mavadati, Mohammad H Mahoor, Kevin Bartlett, Philip Trinh, and Jeffrey F Cohn. 2013. Disfa: A spontaneous facial action intensity database. *IEEE Transactions on Affective Computing* 4, 2 (2013), 151–160.
- [46] Jan Mehnert, Maddalena Brunetti, Jens Steinbrink, Michael Niedeggen, and Christian Dohle. 2013. Effect of a mirror-like illusion on activation in the precuneus assessed with functional near-infrared spectroscopy. *Journal of Biomedical Optics*

18, 6 (2013), 066001. https://doi.org/10.1117/1.JBO.18.6.066001

- [47] Destaw B Mekbib, Jiawei Han, Li Zhang, Shan Fang, Hongjie Jiang, Junming Zhu, Anna W Roe, and Dongrong Xu. 2020. Virtual reality therapy for upper limb rehabilitation in patients with stroke: a meta-analysis of randomized clinical trials. *Brain injury* 34, 4 (2020), 456–465.
- [48] Zuheng Ming, Aurélie Bugeau, Jean-Luc Rouas, and Takaaki Shochi. 2015. Facial action units intensity estimation by the fusion of features with multi-kernel support vector machine. In 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Vol. 6. IEEE, 1–6.
- [49] Christian Moro, Zane Štromberga, Athanasios Raikos, and Allan Stirling. 2017. The effectiveness of virtual and augmented reality in health sciences and medical anatomy. Anatomical sciences education 10, 6 (2017), 549–559.
- [50] Katsuhiko Nakamura, Naoki Toda, Koichiro Sakamaki, Kenji Kashima, and Noriaki Takeda. 2003. Biofeedback rehabilitation for prevention of synkinesis after facial palsy. Otolaryngology—Head and Neck Surgery 128, 4 (2003), 539–543.
- [51] Jeremie Nicolle, Kevin Bailly, and Mohamed Chetouani. 2015. Facial action unit intensity prediction via hard multi-task metric learning for kernel regression. In 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Vol. 6. IEEE, 1–6.
- [52] Christoph Noll, Ute von Jan, Ulrike Raap, and Urs-Vito Albrecht. 2017. Mobile augmented reality as a feature for self-oriented, blended learning in medicine: randomized controlled trial. *JMIR mHealth and uHealth* 5, 9 (2017), e139.
- [53] Maja Pantic and Leon J. M. Rothkrantz. 2000. Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on pattern analysis and machine intelligence* 22, 12 (2000), 1424–1445.
- [54] Vilayanur S Ramachandran and Diane Rogers-Ramachandran. 1996. Synaesthesia in phantom limbs induced with mirrors. Proceedings of the Royal Society of London. Series B: Biological Sciences 263, 1369 (1996), 377–386.
- [55] Reality Composer 2020. Build Augmented Reality. Retrieved March 22, 2021 from https://apps.apple.com/us/app/reality-composer/id1462358802
- [56] Gudrun Roob, Franz Fazekas, and Hans-Peter Hartung. 1999. Peripheral facial palsy: etiology, diagnosis and treatment. European neurology 41, 1 (1999), 3–9.
- [57] Georgia Sandbach, Stefanos Zafeiriou, and Maja Pantic. 2013. Markov random field structures for facial action unit intensity estimation. In Proceedings of the IEEE International Conference on Computer Vision Workshops. 738–745.
- [58] Lamyae Sardi, Ali Idri, and José Luis Fernández-Alemán. 2017. A systematic review of gamification in e-Health. *Journal of biomedical informatics* 71 (2017), 31–48.
- [59] Martin Schimmel, T Ono, OLT Lam, and F Müller. 2017. Oro-facial impairment in stroke patients. *Journal of oral rehabilitation* 44, 4 (2017), 313–326.
- [60] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. 2017. Construction of a Benchmark for the User Experience Questionnaire (UEQ). International Journal of Interactive Multimedia and Artificial Intelligence 4 (06 2017), 40–44. https://doi.org/10.9781/ijimai.2017.445
- [61] Johan N Siebert, Frederic Ehrler, Alain Gervaix, Kevin Haddad, Laurence Lacroix, Philippe Schrurs, Ayhan Sahin, Christian Lovis, and Sergio Manzano. 2017. Adherence to AHA guidelines when adapted for augmented reality glasses for assisted pediatric cardiopulmonary resuscitation: a randomized controlled trial. *Journal of medical Internet research* 19, 5 (2017), e183.
- [62] L. Sifre and S. Mallat. 2014. Rigid-Motion Scattering for Texture Classification. ArXiv abs/1403.1687 (2014).
- [63] Snapchat 2021. The fastest way to share a moment. Retrieved March 22, 2021 from https://www.snapchat.com/

- [64] Kudamo Song, Sehun Chang, Jun Lee, Sun Ae Shin, and Ho Yun Lee. 2018. Clinical Characteristics of Dizziness Associated with Acute Peripheral Facial Palsy. *Journal of audiology & otology* 22, 3 (2018), 148–153. https://doi.org/10.7874/jao. 2017.00374
- [65] Luis Enrique Sucar, Ron S Leder, David J Reinkensmeyer, Jorge Hernández, Gildardo Azcárate, Nallely Casteñeda, and Pedro Saucedo. 2008. Gesture Therapy-A Low-Cost Vision-Based System for Rehabilitation after Stroke.. In *HEALTHINF* (2). 107–111.
- [66] Kojiro Takezawa, Grant Townsend, and Mounir Ghabriel. 2018. The facial nerve: anatomy and associated disorders for oral health professionals. *Odontology* 106, 2 (2018), 103–116.
- [67] Lázaro J. Teixeira, Juliana S. Valbuza, and Gilmar F. Prado. 2011. Physical therapy for Bell's palsy (idiopathic facial paralysis). *The Cochrane database of systematic reviews* 12 (2011), CD006283. https://doi.org/10.1002/14651858.CD006283.pub3
- [68] Laura R. Tomat and Ralph T. Manktelow. 2005. Evaluation of a new measurement tool for facial paralysis reconstruction. *Plastic and reconstructive surgery* 115, 3 (2005), 696–704. https://doi.org/10.1097/01.prs.0000152431.55774.7e
- [69] Gerd Fabian Volk, Anika Steinerstauch, Annegret Lorenz, Luise Modersohn, Oliver Mothes, Joachim Denzler, Carsten M Klingner, Farsin Hamzei, and Orlando Guntinas-Lichius. 2019. Facial motor and non-motor disabilities in patients with central facial paresis: a prospective cohort study. *Journal of neurology* 266, 1 (2019), 46–56.
- [70] Robert Walecki, Ognjen (Oggi) Rudovic, Vladimir Pavlovic, Bjoern Schuller, and Maja Pantic. 2017. Deep Structured Learning for Facial Action Unit Intensity Estimation. (July 2017).
- [71] Lin Wang and Kuk-Jin Yoon. 2021. Knowledge Distillation and Student-Teacher Learning for Visual Intelligence: A Review and New Outlooks. *IEEE Transactions* on Pattern Analysis and Machine Intelligence (2021), 1–1. https://doi.org/10.1109/ tpami.2021.3055564
- [72] Shangfei Wang, Longfei Hao, and Qiang Ji. 2018. Facial action unit recognition and intensity estimation enhanced through label dependencies. *IEEE Transactions* on Image Processing 28, 3 (2018), 1428–1442.
- [73] Matthew J Warner, Julia Hutchison, and Matthew Varacallo. 2020. Bell Palsy. StatPearls [Internet] (2020).
- [74] Mark K Wax. 2014. Facial Paralysis: A Comprehensive Rehabilitative Approach. Plural Publishing.
- [75] YouTube. 3/1/2021. 21 EXERCISES FOR BELL'S PALSY DO NOT DO THESE DURING COMPLETE PALSY/INITIAL DAYS. https://www.youtube.com/watch? app=desktop&v=amonWj6_dS8
- [76] Stefanos Zafeiriou, Cha Zhang, and Zhengyou Zhang. 2015. A survey on face detection in the wild: past, present and future. Computer Vision and Image Understanding 138 (2015), 1-24.
- [77] Xing Zhang, Lijun Yin, Jeffrey F Cohn, Shaun Canavan, Michael Reale, Andy Horowitz, Peng Liu, and Jeffrey M Girard. 2014. Bp4d-spontaneous: a highresolution spontaneous 3d dynamic facial expression database. *Image and Vision Computing* 32, 10 (2014), 692–706.
- [78] Yong Zhang, Weiming Dong, Bao-Gang Hu, and Qiang Ji. 2018. Weaklysupervised deep convolutional neural network learning for facial action unit intensity estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2314–2323.
- [79] Yong Zhang, Baoyuan Wu, Weiming Dong, Zhifeng Li, Wei Liu, Bao-Gang Hu, and Qiang Ji. 2019. Joint Representation and Estimator Learning for Facial Action Unit Intensity Estimation. (June 2019).