Lightweight Mood Estimation Algorithm For Faces Under Partial Occlusion

Nikolas Petrou
Catalink Limited
Nicosia, Cyprus
nick.petrou.lim@gmail.com

Georgia Christodoulou
Catalink Limited
Nicosia, Cyprus
georgia.christodoulou@gmail.com

Konstantinos Avgerinakis
Catalink Limited
Nicosia, Cyprus
koafgeri@catalink.eu

Pavlos Kosmides
Catalink Limited
Nicosia, Cyprus
pkosmidis@catlink.eu

ABSTRACT

The latest advancements in Machine Learning have led to impressive capabilities in distinguishing emotions from facial expressions, allowing computers and smart devices to accurately detect and interpret human emotions through computer vision. While a lot of work has been conducted on understanding human expressions by utilizing visual information, most of them assume that the faces are fully exposed. In this work, we present the implementation of a lightweight mood estimation deep learning model in the presence of partial occlusion where the user is wearing eyewear equipment that completely covers the area around their eyes. Examples of such eyewear are glasses for visually impaired people or a head-mounted display in a virtual reality setting. Rather than collecting a new dataset of images illustrating individuals wearing such eyewear or virtual reality equipment, we utilized a dataset based on a previous work of ours, where the occlusion arising from such headsets was obtained through simulation. That way, we were able to make use of the transfer learning approach by fine-tuning an efficient model that was pre-trained on a typical Facial Expression Recognition task.

CCS CONCEPTS

• **Computing methodologies** → **Computer vision tasks**; **Neural networks**; • **Applied computing** → *Health informatics*; *Consumer health*; • **Human-centered computing** → **Virtual reality**; *Ubiquitous and mobile devices*; • **Social and professional topics** → **Assistive technologies**.

KEYWORDS

Deep Learning, Computer Vision, Facial Emotion Recognition, Virtual Reality

ACM Reference Format:
Nikolas Petrou, Georgia Christodoulou, Konstantinos Avgerinakis, and Pavlos Kosmides. 2023. Lightweight Mood Estimation Algorithm For Faces Under Partial Occlusion. In *Proceedings of the 16th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '23), July 05–07, 2023, Corfu, Greece.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3594806.3596553

1 INTRODUCTION

Facial emotion Recognition (FER) technologies are becoming more and more significant in various fields and industries. The emotional state of individuals is a significant indicator of their well-being, especially in cases of elder individuals or people suffering from chronic health conditions or impairments. In addition, monitoring an individual's mood over a period of time can reveal shifts and irregularities in their actions and behaviour. These variations could be indicative of a substantial decline in their functioning before the onset of clinical symptoms. Therefore, it is important to monitor the emotional status to identify any potential problems early on. Furthermore, emotional tracking of patients can play a significant role in healthcare, since it could provide assistance to clinicians, to easily monitor the progress of their patients, even remotely. With the use of low-cost smart devices, sensors, and data analysis, meaningful insights related to the users' health conditions can be collected and sent to doctors for further investigation. [21].

While commonly linked to gaming, the cutting-edge technology of Virtual Reality (VR) has the potential to transform various industries. Specifically, the healthcare sector is currently exploring promising ways in which VR can aid patients and healthcare providers in achieving improved treatments and outcomes. These applications include surgical procedures [20], pain management [1, 26], physical [8] and cognitive rehabilitation [28], mental health [9], and other areas.

The latest developments in Machine Learning (ML) have led to formidable FER capabilities, allowing computers and smart devices to accurately detect and interpret human emotions from facial expressions, and therefore revolutionizing the way we benefit from technology. These sophisticated algorithms have immense potential for improving human-computer interaction in fields such as serious games [3], health, and well-being in general. While lots of applications and studies have been conducted on automatically inferring human expressions by utilizing visual information, most of them assume that the users' faces are uncovered. Moreover, although

PETRA '23, July 05–07, 2023, Corfu, Greece

ACM ISBN 979-8-4007-0069-9/23/07…$15.00
https://doi.org/10.1145/3594806.3596553

state-of-the-art FER methods are highly effective for controlled laboratory environments, the existing approaches do not achieve similar performances when applied to groups with visual impairments (e.g. people wearing smart glasses for virtual assistance) or applications like VR environments in which severe occlusion conditions exist. Recently, the research community has identified this issue and proposed promising methods for handling severe systematic occlusion like VR setting where features of the upper half of the face are completely missing [17, 23]. However, the necessity of lightweight models is critical for FER applications under partial occlusion, in order to be employed smoothly on mobile devices and other resource-constrained platforms, without compromising accuracy and performance.

In this work, our key contributions are the following:

- We investigate the task of FER on facial images that have the upper face region covered, and show how occlusion causes certain emotions to be less distinguishable.
- We re-purpose our previous computationally efficient mood estimation model [5] on images that do not include the upper face region, exploring in that way the possibility of employing such application on users that wear VR headsets or smart glasses.
- We demonstrate that with the transfer learning technique and appropriate fine-tuning, a lightweight baseline model is obtained, which achieves a sufficient performance for FER tasks under occlusion. Due to its low-overhead, the developed model can be easily deployed to many applications and devices.

2 RELATED WORK

Numerous studies in the academic literature have attempted to recognize emotions through analysing image data, utilizing both traditional ML methods and advanced deep learning techniques. In order to accurately estimate emotions, researchers have focused on extracting features based on facial landmarks and using training models like Support Vector Machines (SVM) and Gradient Boosting Trees as classifiers, based on extracted features [25]. Since the recent advances in deep learning, Convolutional Neural Networks (CNNs) have been proven to achieve superior performance when compared with conventional Machine Learning models and hand-extracted features. Some of the recently hypertrophied and commonly exploited techniques for mood estimation are based on Convolutional Neural Networks (CNN). In [7], a video-based emotion recognition approach is proposed, where a MobileNet feature extractor is used in combination with an SVM classifier. Regarding the Healthcare domain, authors have previously implemented a 3D CNN model for capturing and analysing video frames obtained remotely from the house of patients and reporting back to the clinicians their detected emotional status [16].

Previous studies have shown the significance of stimulating user's emotional stage through VR environments. Particularly, a recent study has shown that VR was seen to have the highest common usage for emotion classification among other stimuli used [24]. Despite its significance, there is a lack of available image datasets that could be directly utilized in FER under partial occlusion. Hickson et al. proposed an algorithm that can detect facial expressions by utilizing information obtained only from a person's eyes, which were captured through an infrared gaze-tracking camera in a VR head-mounted display [15].

As not all VR devices have embedded eye-tracking sensors, a widely applicable approach, which was also proven to be less costly, was followed by Georgescu and Ionescu [10]. In their research, Georgescu and Ionescu introduced a method for detecting the facial expressions of individuals who wear a VR headset while exploiting the use of an external camera. To do so, Georgescu and Ionescu focused on training VGG-like models, based on modified training images in which the upper half of the face was completely occluded. This has proven that the neural network was forced to find distinguishing patterns in the lower half of the face. In their work, they followed a fine-tuning procedure that consisted of two phases. During the first stage, the model was fine-tuned based on the original images that included full-face images, while in the second phase, the model was further tuned based on images in which the area around the eyes was occluded. In another deep learning approach for occluded targets, Cheng et al. simulated the occlusion by drawling graphic masks on the images of FER datasets, since datasets with natural partial occlusion facial images were not available [4]. However, that approach had a considerable limitation, since it does not account for realistic occlusion resulting from particular devices with specific dimensions. Recent studies aimed to address this issue by defining certain algorithms for simulating under partial occlusion [17, 23].

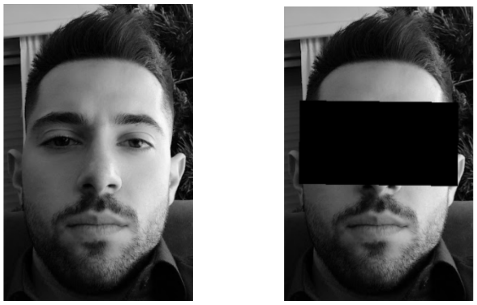
The work presented in this paper expands on previous studies that investigate how occlusion resulting from VR goggles affects the recognition of facial expressions. We concentrated on exploiting our previously constructed dataset and pre-trained model, which was trained for unoccluded patients' mood estimation (initially presented in [5]). Grounded on the data collected during our previous work, a simulation of the occlusion is performed, by generating rectangles that are applied on the upper region of the faces based on the methodology proposed in [23]. Then, a transfer learning approach was deployed for fine-tuning the pre-trained model. We provide a comparison regarding different transfer learning methodologies we applied and we present the results of our experiments. Concluding, we briefly describe the future directions that we plan to follow in order to expand and further improve our work.

3 METHODOLOGY

3.1 Dataset & Simulated Occlusion

Based on the task's requirements, a model for facial expression classification should be developed, trained, and validated, on faces occluded by VR headsets. Specifically, the proposed model should be lightweight, fast during inference, and demonstrate good performance on classification tasks for partially occluded faces.

However, for deep learning networks training, a large amount of training data (in the order of thousands) is required so that the model learns to conduct a classification task successfully. In our dataset, which was firstly collected and employed for [5], we included multiple and various facial expressions within each emotion category, in order to create a dataset that well-represents the different human facial expressions. Facial images are hard to be found



(a) Original image (b) Image with applied occlusion

Figure 1: Example of the simulated occlusion

available online, due to the strict copyright licenses. For that reason, our images were gathered from online resources that provided copyright-free images, such as Kaggle (FER 2013 dataset[1], Jafar Hussain Human emotions dataset[2]) and other open source databases such as Unsplash[3] , Pexels[4] and Pixabay[5].

Our original image collection consists of roughly 50,000 images of facial expressions, which are categorized into seven emotion classes ('angry', 'disgusted', 'scared', 'happy', 'sad', 'surprised', 'neutral'). The categories have unequal amounts of instances, making the dataset imbalanced. Due to this work's and MuseIT's [6] purposes, we decided to focus on the three most basic emotions, i.e. 'happy', 'sad', and 'neutral', and for that reason, we grouped the rest categories into a fourth class, named 'other'.

In order to obtain representative image instances which are identical to occluded faces, a preprocessing procedure was performed. Analytically, the collected images were adjusted to our new task, by occluding the upper part of the face (i.e. the eyes and parts of the forehead and nose), inspired by the methodology originally proposed in [23]. Initially, the preprocessing algorithm uses a Multi-task Cascade Convolutional Neural Network (MTCNN) [27] to detect five facial landmarks (two for the center of each eye, one for the nose center and two for the right and left side of the mouth). Based on the detected eye and nose landmarks, as well as the distances specified by the algorithm suggested by Rodrigues et al. [23] a rectangle is drawn on top of each image. Therefore, the upper part of the faces is hidden, simulating in such way the inclusion of VR headsets. An example of a pair that consists of an image and its occluded version is elucidated in Figure 1.

As it was also indicated in [23], the MTCNN is not always able to identify the facial landmarks, and therefore it is not feasible to draw the artificial rectangle for some instances based on the proposed methodology. In their work, Rodrigues et al. included the original images for the cases where the facial landmarks were not detected [23]. Instead, in our work, we did not include such instances in our occluded dataset, as we wanted to directly compare and study the impact of occlusion. Therefore, it was decided to avoid including not representative observations, and due to that choice, the size of the resultant dataset decreased by approximately 10%. Ultimately, the class distribution and number of classes are shown in Table 1.

Table 1: Experimental settings & Results for different model

CLASS	NO. TRAINING INSTANCES	NO. VALIDATION INSTANCES	NO. TEST INSTANCES
Sad	8234	680	935
Neutral	9249	985	749
Happy	12196	1247	830
Other	15065	1630	1378
TOTAL	**44744**	**4542**	**3892**

3.2 Model Architecture

As already mentioned, our application demands a lightweight model, that would be able to run fast the inference process (ideally in real-time) and demonstrate good performance on the task of FER. Moreover, the model must be easily deployed on smart devices, without the need to sacrifice much of the device's memory and computational resources.

After studying the state-of-the-Art works, we found out that the most recently developed model which fits the requirements and is appropriate for this task is the mini-Xception deep learning model [2]. Mini-Xception is the successor of the original Xception model [6], and it is proved that it demonstrates high accuracy rates on emotion recognition tasks. The success of the mini-Xception architecture lies mostly in the fact that they use residual modules [14], and depth-wise separable convolutions [18]. Residual connections are speeding up the convergence of the model, both in terms of speed and final classification performance [6]. Also, depth-wise separable convolutions demand significantly fewer computations compared to normal convolutions, resulting in a network that demands less computational power for both training and inference.

Another characteristic of the mini-Xception is the elimination of fully-connected layers and their replacement by the Global Average Pooling operation. In that way, mini-Xception's number of parameters is significantly reduced, ending up with an overall of 58,000 parameters. Lastly, the final model's size is less than 1MB, so it can seamlessly be deployed and run on hardware-constrained devices.

The architecture of mini-Xception includes two Convolution layers (which are followed by Batch Normalization [19] and ReLU [12]) that are followed by four residual blocks. Each block contains a convolution layer on the skip connection side, and the other side consists of two separable convolutions followed by a Max Pooling layer. Lastly, the Global Average Pooling operation takes place, and the results go through the softmax function, which gives the final result.

3.3 Experimental settings

In general, convolutional networks extract low-level features that are common for various datasets during the convolution process. Training a CNN from scratch can be expensive, particularly for large datasets. To address this, a different approach is to transfer the parameters from pre-existing models and fine-tune them based

[1] https://www.kaggle.com/datasets/msambare/fer2013
[2] https://www.kaggle.com/jafarhussain786/datasets
[3] https://unsplash.com/
[4] https://www.pexels.com/search/face/
[5] https://pixabay.com/vectors/
[6] https://www.muse-it.eu/

Table 2: Experimental settings & Results

ID	MINI-XCEPTION MODEL SETTINGS	TRAINABLE PARAMETERS	DESCRIPTION	TEST ACCURACY	TEST F1-MACRO	TEST F1-WEIGHTED
MODEL_1	Pre-trained for [5]	0	No additional training involved	0.49	0.46	0.49
MODEL_2	Pre-trained for [5] & Unfreeze Last Layer	4,612	All parameters apart from the last convolutional layer were frozen during training on the occluded dataset	0.63	0.61	0.63
MODEL_3	Pre-trained for [5] & Unfreeze All Layers	53,636	Parameters initialized based on [5] and continued training on the occluded dataset	**0.69**	**0.68**	**0.69**
MODEL_4	Trained from Scratch	53,636	Parameters were reinitialized	0.68	0.67	0.68

on the new dataset. Hence, for classifying facial expressions under occlusion we chose to utilize the mini-Xception model of our previous work, pre-trained for a non-occluded FER task. Therefore, the process aimed in utilizing the already learned knowledge of the pre-trained network, in order to reduce the training time as well as to improve the overall classification performance for the occluded scenario.

In order to provide a fruitful comparison and applicable empirical results during our experiments, we focused on the experimentation and evaluation of four different settings for the occluded dataset: 1) For the first scenario, the mini-Xception network pre-trained on the non-occluded dataset of [5] was evaluated for baseline results. Next, the values of the parameters of that same network were used in two different transfer learning settings. 2) In order to use the first convolutional layers of the pre-trained network as feature extractors, the network was trained on the occluded dataset by freezing all the parameters except the ones which belong to the last convolutional layer. 3) In the third setting, the values of the network's parameters were initialized based on the first setting, but this time none of them were left frozen during the training on the occluded dataset. 4) Lastly, for the final setting, the mini-Xception architecture was trained from scratch on the occluded dataset, and the parameters of the network were initialized based on Xavier uniform initializer [11].

A brief summarization of the above-mentioned model settings is available in Table 2. Our input images were 3-channel RGB images of size 64×64. Regarding the choice of hyperparameters and other training options, a procedure similar to work [5] was followed. Specifically, for all the trained models, the *Adam* optimizer was used, with the initial learning rate hyperparameter tuned individually for each scenario based on a validation set. In addition, the learning rate was gradually reduced based on the *Reduce Learning Rate on Plateau* technique. Furthermore, for regularization, we applied the L2 regularization method with $\lambda = 0.01$. A batch size of 64 was used for all the models, as it was also used in our previous work. Finally, for the class imbalance problem handling, we applied class weights to the loss function, and during the training process, the best model was saved based on the F1-macro average score [13] of the validation dataset.

4 EXPERIMENTAL RESULTS

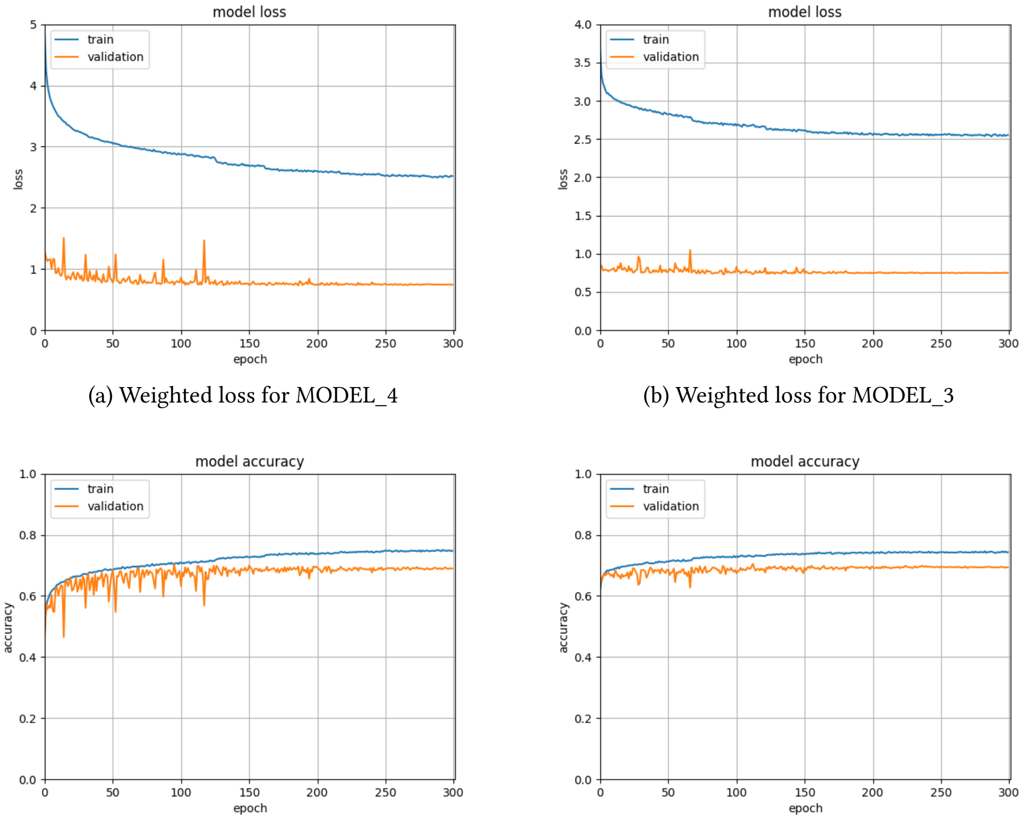
In order to demonstrate the mini-Xception's performance for different settings for the specific application, Table 2 provides a comparison between the best version of each model for the occluded test set, that makes up roughly 7% of the observations. As expected, the classification results of the model that was pre-trained solely on non-occluded images (MODEL_1) is inferior compared to the rest. The model which performed the best with respect to all three evaluation metrics was the pretrained model that was further trained and fine-tuned for the occluded task (MODEL_3).

Even though the difference in performance between MODEL_3 and the version which was trained from scratch for the occluded dataset (MODEL_4) was approximately 1%, there are more advantages in utilizing the MODEL_3's training scheme. Specifically, as Figure 2 illustrates, MODEL_3's initialization was more ideal since the loss and accuracy were satisfactory even after only a few epochs of training. On the contrary, it was required for MODEL_4 to be trained approximately for 40 to 50 epochs in order to achieve a performance similar to the one that MODEL_3 achieved after 5 to 10 epochs.

To obtain an assessment of the diminishment in performance by using occlusion, we also conducted an evaluation of the classification capabilities of MODEL_1 for the non-occluded version of the test-set. The F1-weighted average score for the non-occluded test-set was 0.73, while the F1-macro average score was 0.72 (Fig 3). By examining the evaluation results of the non-occluded and occluded scenarios for the pre-trained MODEL_1, the diminishment in performance is immense, as in the non-occluded case, the model achieved a F1-macro average score of 0.46. However, comparing the performance between the results of MODEL_1 for the non-occluded version of the test-set as well as the results of our best model (MODEL_3) for the occluded version of the test-set, it is noticed that the overall performance is only decreased by a small amount of 4%, when occlusion is introduced. Furthermore, by comparing the performance diminishment between the two above-mentioned scenarios for the different classes, it was observed that a large amount of misclassifications has risen for the classes "Sad" and "Neutral". It is believed that this is due to the fact that apart from having the lip corners pulled down, people often express their sadness by crying or by raising their inner corners of eyebrows raised and eyelids loose [22]. Therefore, this information is hard to be utilized under partial or severe occlusion. The confusion matrix which shows the inference results of MODEL_1 for the non-occluded version of the test-set as well as the results of our best model (MODEL_3) for the occluded version of the test-set are illustrated in Fig 3.

5 CONCLUSIONS & FUTURE WORK

In this work, we developed a computationally efficient transfer learning-based model for addressing facial expression recognition in the presence of occlusion where the user is assumed to be wearing head-mounted equipment that covers a large part of the area around



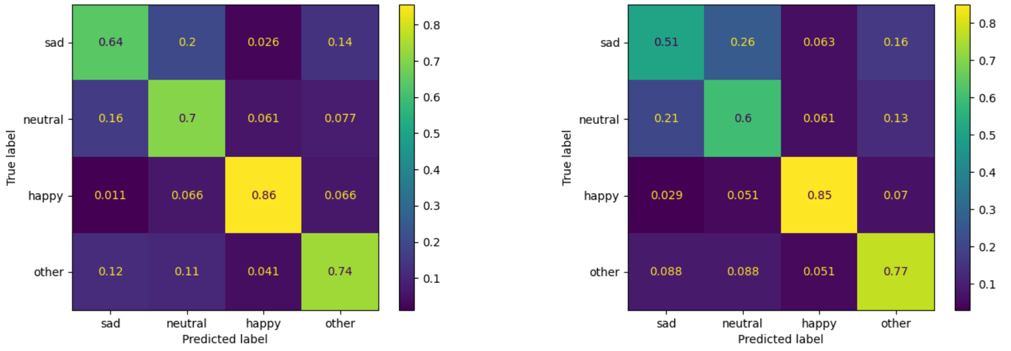
(a) Weighted loss for MODEL_4

(b) Weighted loss for MODEL_3

(c) Accuracy for MODEL_4

(d) Accuracy for MODEL_3

Figure 2: Learning curves for MODEL_4 and MODEL_3



(a) MODEL_1 results for non-occluded test set

(b) MODEL_3 results for occluded test set

Figure 3: Confusion Matrices for non-occluded and occluded test set

their eyes. We altered our existing dataset, which consists of roughly 50,000 images, to artificially mimic a VR occlusion by utilizing a geometric simulation method based on the work of Rodrigues et al. [23]. We employed in different settings our pre-trained model which was trained on our original non-occluded database and further fine-tuned the parameters on the occluded images. Comparing the performance between our best models for the non-occluded and occluded cases, it was noticed that the overall performance was only reduced by a small amount of 4% when occlusion was introduced. This fact indicates that FER under occlusion is still possible. Furthermore, the results confirm that the exploitation of transfer learning as well as the simulation techniques for synthetic occlusion can lead to a respectable model that produces results that keep pace with frameworks that utilize information from the periocular area and eyes.

As part of the European-funded project, MuseIT, the final model will be integrated into multisensory technologies, with the purpose of extracting insights regarding the emotional state of users during their engagement with cultural assets and music. The technologies developed, and the mood predictions extracted, aim to improve the inclusion, accessibility as well as the whole experience of cultural assets for all, with a particular focus on the needs of people with disabilities.

Concluding, the developed lightweight model and formulated methodology can serve as a tool to monitor the emotional status of individuals through mobile devices. Additionally, we anticipate that both healthcare and serious-games sectors can be greatly benefited when employing such models in VR environments or for visually impaired users that wear certain optic equipment. For future work, we aim to further improve our model as well as to evaluate its effectiveness in actual real-life scenarios. Finally, we also intend to implement and make experiments for a fused model that could be efficaciously exploited for both occluded and non-occluded settings.

ACKNOWLEDGMENTS

This work has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101061441 - MuseIT: Multi-sensory, User-centred, Shared cultural Experiences through Interactive Technologies. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

REFERENCES

[1] Naseem Ahmadpour, Hayden Randall, Harsham Choksi, Antony Gao, Christopher Vaughan, and Philip Poronnik. 2019. Virtual Reality interventions for acute and chronic pain management. *The international journal of biochemistry & cell biology* 114 (2019), 105568.

[2] Octavio Arriaga, Matias Valdenegro-Toro, and Paul Plöger. 2017. Real-time Convolutional Neural Networks for Emotion and Gender Classification. arXiv:1710.07557 [cs.CV]

[3] David Checa and Andres Bustillo. 2020. A review of immersive virtual reality serious games to enhance learning and training. *Multimedia Tools and Applications* 79 (2020), 5501–5527.

[4] Yue Cheng, Bin Jiang, and Kebin Jia. 2014. A deep structure for facial expression recognition under partial occlusion. In *2014 Tenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*. IEEE, Kitakyushu, Japan, 211–214.

[5] Chloe Chira, Evangelos Mathioudis, Christina Michailidou, Pantelis Agathangelou, Georgia Christodoulou, Ioannis Katakis, Efstratios Kontopoulos, and Konstantinos Avgerinakis. 2023. An Affective Multi-modal Conversational Agent for Non Intrusive Data Collection from Patients with Brain Diseases. In *Chatbot Research and Design: 6th International Workshop, CONVERSATIONS 2022, Amsterdam, The Netherlands, November 22–23, 2022, Revised Selected Papers*. Springer, Amsterdam, The Netherlands, 134–149.

[6] François Chollet. 2017. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE Computer Society, Los Alamitos, CA, USA, 1251–1258.

[7] Polina Demochkina and Andrey V Savchenko. 2021. MobileEmotiFace: Efficient facial image representations in video-based emotion recognition on mobile devices. In *Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10–15, 2021, Proceedings, Part V*. Springer, Online, 266–274.

[8] Hao Feng, Cuiyun Li, Jiayu Liu, Liang Wang, Jing Ma, Guanglei Li, Lu Gan, Xiaoying Shang, and Zhixuan Wu. 2019. Virtual reality rehabilitation versus conventional physical therapy for improving balance and gait in Parkinson's disease patients: a randomized controlled trial. *Medical science monitor: international medical journal of experimental and clinical research* 25 (2019), 4186.

[9] Daniel Freeman, Polly Haselton, Jason Freeman, Bernhard Spanlang, Sameer Kishore, Emily Albery, Megan Denne, Poppy Brown, Mel Slater, and Alecia Nickless. 2018. Automated psychological therapy using immersive virtual reality for treatment of fear of heights: a single-blind, parallel-group, randomised controlled trial. *The Lancet Psychiatry* 5, 8 (2018), 625–632.

[10] Mariana-Iuliana Georgescu and Radu Tudor Ionescu. 2019. Recognizing facial expressions of occluded faces using convolutional neural networks. In *Neural Information Processing: 26th International Conference, ICONIP 2019, Sydney, NSW, Australia, December 12–15, 2019, Proceedings, Part IV 26*. Springer, Sydney, NSW, Australia, 645–653.

[11] Xavier Glorot and Yoshua Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, Chia Laguna Resort, Sardinia, Italy, 249–256.

[12] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. 2011. Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, Fort Lauderdale, USA, 315–323.

[13] Thamme Gowda, Weiqiu You, Constantine Lignos, and Jonathan May. 2021. Macro-Average: Rare Types Are Important Too. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, Online, 1138–1157.

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE, Las Vegas, NV, USA, 770–778.

[15] Steven Hickson, Nick Dufour, Avneesh Sud, Vivek Kwatra, and Irfan Essa. 2019. Eyemotion: Classifying facial expressions in VR using eye-tracking cameras. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, Hawaii, 1626–1635.

[16] M Shamim Hossain and Ghulam Muhammad. 2019. An audio-visual emotion recognition system using deep learning fusion for a cognitive wireless framework. *IEEE Wireless Communications* 26, 3 (2019), 62–68.

[17] Bita Houshmand and Naimul Mefraz Khan. 2020. Facial expression recognition under partial occlusion from virtual reality headsets based on transfer learning. In *2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM)*. IEEE, New Delhi, India, 70–75.

[18] 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. arXiv:1704.04861 [cs.CV]

[19] Sergey Ioffe and Christian Szegedy. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*. pmlr, Lille, France, 448–456.

[20] Mohd Javaid and Abid Haleem. 2020. Virtual reality applications toward medical field. *Clinical Epidemiology and Global Health* 8, 2 (2020), 600–605.

[21] Seena Naik and E Sudarshan. 2019. Smart healthcare monitoring system using raspberry Pi on IoT platform. *ARPN Journal of Engineering and Applied Sciences* 14, 4 (2019), 872–876.

[22] Lawrence Ian Reed and Peter DeScioli. 2017. The communicative function of sad facial expressions. *Evolutionary Psychology* 15, 1 (2017), 1474704917700418.

[23] Ana Sofia Figueiredo Rodrigues, Júlio Castro Lopes, Rui Pedro Lopes, and Luís F Teixeira. 2023. Classification of facial expressions under partial occlusion for VR games. In *Optimization, Learning Algorithms and Applications: Second International Conference, OL2A 2022, Póvoa de Varzim, Portugal, October 24-25, 2022, Proceedings*. Springer, Póvoa de Varzim, Portugal, 804–819.

[24] Nazmi Sofian Suhaimi, James Mountstephens, Jason Teo, et al. 2020. EEG-based emotion recognition: A state-of-the-art review of current trends and opportunities.

[25] J Sujanaa, S Palanivel, and M Balasubramanian. 2021. Emotion recognition using support vector machine and one-dimensional convolutional neural network. *Multimedia Tools and Applications* 80 (2021), 27171–27185.

[26] Melissa S Wong, Brennan MR Spiegel, and Kimberly D Gregory. 2021. Virtual reality reduces pain in laboring women: a randomized controlled trial. *American Journal of Perinatology* 38, S 01 (2021), e167–e172.

[27] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao. 2016. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. *IEEE Signal Processing Letters* 23, 10 (Oct 2016), 1499–1503. https://doi.org/10.1109/LSP.2016.2603342

[28] Shizhe Zhu, Youxin Sui, Ying Shen, Yi Zhu, Nawab Ali, Chuan Guo, and Tong Wang. 2021. Effects of virtual reality intervention on cognition and motor function in older adults with mild cognitive impairment or dementia: a systematic review and meta-analysis. *Frontiers in Aging Neuroscience* 13 (2021), 586999.