Emotion Recognition Using Pretrained Deep Neural Networks

Marek Dobeš*, Natália Sabolová**

*CSPV SAV, Karpatská 5, 04001 Košice, Slovak Republic, dobes@saske.sk
**FF UPJŠ, Moyzesova 9, 04059 Košice, Slovak Republic,
natalia.sabolova@student.upjs.sk

Abstract: We use pretrained deep neural networks additionally trained using transfer learning to recognise human emotions from photographs. Furthermore, we want to establish what face regions seem to provide the most information for the network. We find that the network performed the best jn the photographs with no occlusions (accuracy of 74.20%) followed by the photographs with occluded eye region (72.01%), faces with covered noses (70.21%) and faces with covered mouths (52.10%).

Keywords: emotion recognition; pretrained networks; deep learning

1 Introduction

Face recognition and facial expressions for emotion detection has been widely studied across different scientific disciplines (including social and behavioural sciences, neuroscience, cognitive psychology, etc.). For example, its profound application can be found in a clinical and neuropsychological research of prosopagnosia that is also called "face blindness" [1], alexithymia – impaired emotion recognition, often in association with autism [2], or social anxiety disorder [3]. Research in this field has been to a great extent influenced by technological growth [4]. Computational advances in programming, computer vision and machine learning allowed researchers to use novel techniques to be able to detect faces and facial emotions from still images or video sequences [5] [6]. There are attempts to build machine learning-based models to simulate cognitive functions produced by the human brain such as face perception and recognition [7]. However, it is essential to note that neurocomputational modeling approach does not completely reflect those phenomena, although it tries to better understand them by studying brain-inspired principles. The core of computer vision research focuses on designing intelligent computer systems that are similarly efficient and accurate (if not more), as our brains [8]. For instance, there is evidence for use of computer vision-based approaches for security purposes [9], such as access control or immigration verification. In forensic settings it is criminal identification or lie detection systems, where facial micro-expression recognition [10] is processed. Furthermore, computational analysis of facial information presents great potential in robotics and internet of things [11] [12] and as it was mentioned at the beginning, healthcare – for computational diagnosis and assessment of mental or facial diseases and/or machine-assisted interventions [13], etc.

The process of facial expression recognition (FER) can be for a neurotypical human easy and automatic transfer of information about the person that is being observed [14] [15]. There are three main categories of models in the literature to quantify affective facial behaviours: categorical model – according to Ekman and Friesen [16] six prototypical facial emotions: happy, sad, anger, fear, surprise, disgust with a neutral composition of facial muscles are prominent non-verbal communication tools. The second is the dimensional model, where affective dimensions like valence and arousal are central parts of an emotional continuum [17]. The third is Facial Action Coding System (FACS) that reflects facial muscle movement specifics, where action units - AUs emotion-specified facial expressions and their combinations represent various emotions. The computational approach highlights the complexity and challenges of the FER process. Computer model requires large datasets that are divided into training data (reference data for learning), data for testing, and for validation. Firstly, a face needs to be identified and distinguished from the background by pre-processing [18]. Then features extraction and classification can be performed [19]. Depending on how the features are extracted, several extraction techniques can be recognized, such as the holistic approach, one of the most common - appearance-based approach (use of filters) and geometry-based method (shapes and locations of important facial features including eyes, mouth, and nose), colour-based technique, and template-based technique, etc., with its numerous hybrid variations [20].

What makes FER challenging is the nature of databases, annotation and labels objectivity/subjectivity, and computational efficiency [21]. Variability – interpersonal and intrapersonal is important if one aims to work with a computational model in real-life conditions [22]. While laboratory-based controlled images might present consistently taken frontal posed mugshots that are easier for computer to recognise, generalisation and effectiveness of training dataset can be better achieved with more robust, spontaneous real-world scene images from in-the-wild datasets [23]. For instance, differently posed faces (changed angle and lighting of facial features), including various ethnicities and age of people may also present difficulties in accuracy of the FER. Therefore, evolution in datasets was needed and there has been a simultaneous progress in methods and particularly algorithms that have been used: from the handcrafted and shallow learning algorithms to deep learning techniques that are more capable of overcoming these challenges [5].

This article specifically focuses on occlusions, one of the challenges that impair the recognition. It is nothing unfamiliar to speak about facial occlusions during the covid-19 pandemics, when wearing masks (lower part of the faces including mouth and nose are covered) has become a regular part of our life that may have besides its health benefits social and personal impact on individual's ability to communicate [24]. Also, machine learning can be used to analyse the emotions of people wearing head-mounted displays while playing computer games [25]. There are already datasets consisting of images with facemasks occlusions (MaskedFace-Net, Correctly Masked Face Dataset, Incorrectly Masked Face Dataset) [26]. The present paper focuses on how computational models deal with similar occlusions by using artificially covered areas of faces (mouth, nose area and eyes). Interestingly, a recent study investigated the effect of sunglasses and face masks on perceiver ability to recognize facial expression [27]. They found a support for diagnostic potential of mouth region, with the largest impact of the masks when emotions happy, surprise and disgust prevailed. The eye region, typical for expressions angry and fear was influenced by both, sunglasses, and masks while the identification of fear was the most affected by masks. Eyes and mouth are particularly known diagnostic region for emotions sad and neutral expressions, however, in that study, recognition of neutral expression was better when eyes were covered (assuming overrated importance of eye region information). On the other hand, sad expressions were affected only by the sunglasses (suggesting a critical role of the eye region). In accordance with the literature on occlusion aware facial expression recognition using convolutional neural network (CNN), CNN with attention mechanism (ACNN) inspired by the intuition, automatically identifies the blocked facial patches and diverts attention to the distinctive as well as unobstructed regions in facial image [28]. In addition, evidence from the previous literature listed recognition rates for near frontal faces: 72.6% with eyes occluded, 69.8% for the recognition with mouth-occluded region and with no occlusions 74.28% recognition rate of facial expressions, which points out the significant role of mouth region [29]. Moreover, when the nose area was covered, Vyas and Hablani [30] reported a similar trend in their results: without occlusion 94.82% recognition rate, following eyes occlusion 93.10%, nose 91.37% and mouth in occlusion reached 89.67%.

2 Methods

We used photographs of human faces from the Face Expression Recognition database from Kaggle. Photographs were divided into three categories – happy emotional expression (8989), angry emotional expression (4953 photographs) and neutral emotional expression (6198 photographs). All the photographs were gray scale with the size 48 x 48 pixels.

For later processing we created three sub-categories from each category. We covered with white noise either eyes, nose and mouth of every photograph. We did this using computer processing rather than human processing to avoid possible bias in further learning. For every photograph we created a rectangle of white noise with the size of 10 x 48 pixels and inserted it into the photograph at the position of the tenth pixel (eyes), 20th pixel (nose) and 30th pixel (mouth). In some cases the rectangle did not fit precisely, but on the average the desired areas of the face were covered. Thus we obtained nine subcategories of photographs. Examples of photographs are displayed in the Figure 1.

Figure 1
Examples of input photographs

To save time and improve efficiency we used transfer learning to create a set of deep neural networks for emotion recognition. We used googlenet network [31]. It is one of the standard pretrained networks used for image recognition and is readily available for application. We removed the last fully connected layer and softmax/output layer. We replaced them with a blank fully connected layer and a softmax/output layer to match the number of classes that we use. We froze the original googlenet layers and trained the network on our data of photographs of faces with emotional expressions. We randomly split our sample so that 70% of the data were used for training and the remaining data for validation. We introduced random augmentation of inputs and trained the network. After training we achieved the validation accuracy of 74.2% with which the network recognised individual emotions.

To investigate on what parts of face does the network concentrate, we applied the grad-cam method [32]. The algorithm will analyse what parts of the picture is the network most sensitive to. In Figure 2 we provide an example of the attentional heat map overlayed on a photograph.

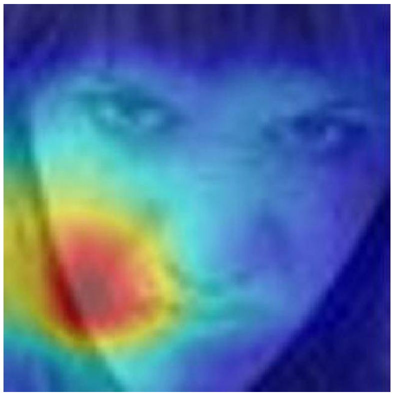

Figure 2
An example of the grad-cam method overlayed on a photograph with angry expression

3 Results

To see how the network will perform on the set of photographs with covered parts of faces, we repeated the transfer learning process with the datasets with covered face parts and achieved the validation accuracy of 72.01 for faces with covered eyes, 70.21 for faces with covered noses and 52.10 for faces with covered mouths.

To analyse how does a network focus when a part of a face is occluded we applied the grad-cam method to the networks trained on occluded data. The results are summarised in Figure 3.

We may notice the trend that presents itself over all three emotional expressions. If the area of eyes and the area of mouth is occluded the network focuses on this area. If the area of nose is occluded, the focus of the network shifts upwards to the area of eyes. This may explain why the network that was trained in the photographs with occluded noses had significantly higher accuracy than the network that was trained on the region of mouth. However, it does not explain the relatively high accuracy of the network that has been trained in the photographs with occluded eyes.

Conclusions

Our findings are in accord with previous research by Goodarzi et al. [29] and Vyas and Hablani [30]. We found that the deep neural network performed best on faces without occlusions and decreased its performance as eyes, nose and mouth were covered.

The later analysis using the grad-cam algorithm may answer why the network that used the photographs with occluded mouth region performed poorly as it focused on the occluded face region that provided little information. However, we were not able to identify why the network trained on the photographs with occluded eyes performed relatively well, although it also focused on the occluded region.

It would be interesting for further research to establish why the eye region seems to provide more information for the computer network even if the substantial part of it is covered. Also, possible differences between genders could be investigated Bilalpur et al. [33].

We understand that our study has its limitations, such as the use of specific database of training photographs or use of only one pre-trained network. However, we assume that both database and network are relatively robust and should not significantly alter the outcomes should they be trained on different data.

Further research could shed more light on more similarities and differences between how deep neural networks analyse the real-world data [34] and how do the brains of mammals perform similar tasks [35].

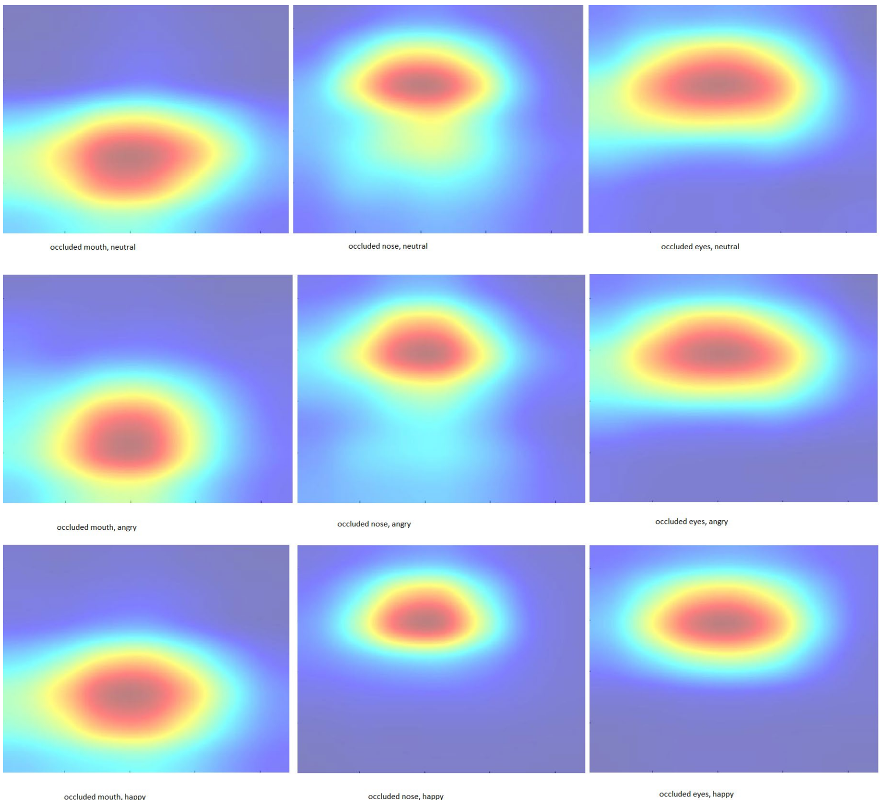

Figure 3
Grad-cam algorithm averaged over 1000 randomly selected samples. Below each sub-figure it is included what part of face is covered and what emotional expression is analysed. Red means the biggest focus, blue the lowest focus.

Acknowledgement

This work was supported by VEGA 2/0043/17 grant.

References

[1] Robotham, R. J., & Starrfelt, R. (2018) Tests of whole upright face processing in prosopagnosia: A literature review. *Neuropsychologia, 121,*106-121, doi:10.1016/j.neuropsychologia.2018.10.018

[2] Cuve, H. C., Castiello, S., Shiferaw, B., Ichijo, E., Catmur, C., & Bird, G. (2021) Alexithymia explains atypical spatio-temporal dynamics of eye gaze in autism. *Cognition, 212,* doi.org/10.31234/osf.io/gzubp

[3] Claudino, R. G.e., de Lima, L. K. S., de Assis, E. D. B., & Torro, N. (2019) Facial expressions and eye tracking in individuals with social anxiety disorder: A systematic review. Psicologia: Reflexão e Crítica, 32, 9, doi.org/10.1186/s41155-019-0121-8

[4] Dodig-Crnkovic, G. (2013) The development of models of computation with advances in technology and natural sciences. *The 6th AISB Symposium on Computing and Philosophy: The Scandal of Computation - What is Computation?* The Society for the Study of Artificial Intelligence and the Simulation of Behaviour

[5] Li, S., & Deng, W. (2020) Deep facial expression recognition: A survey. *IEEE Transactions on Affective Computin.* doi: 10.1109/TAFFC.2020.2981446

[6] Mehendale, N. (2020) Facial emotion recognition using convolutional neural networks (FERC) *SN Applied Sciences, 2*(3) doi:10.1007/s42452-020-2234-1

[7] Martinez, A. M. (2017) Computational models of face perception. *Current Directions in Psychological Science, 26*(3) 263-269, doi:10.1177/0963721417698535

[8] Jonathan, Lim, A., P., Paoline, Kusuma, G. P., & Zahra, A. (2018) Facial Emotion Recognition Using Computer Vision. *IEEE 2018 Indonesian Association for Pattern Recognition International Conference - Jakarta, Indonesia*, 46-50, doi:10.1109/INAPR.2018.8626999

[9] Saad, F. S. A., Lee, M. V., & Basah, S. (2019) Face recognition using machine vision for Security system. *IOP Conference Series: Materials Science and Engineering, 705*, doi:10.1088/1757-899X/705/1/012005

[10] Tamuli, R., Samantara, S., Sarkar, S., & Adhikari, S. (2020) Lie detection using facial micro-expresions. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 4,* (11), 25-30, Retrieved from: https://ijsrcseit.com/paper/CSEIT203115.pdf

[11] Berns, K., & Zafar, Z. (2018) Emotion based human-robot interaction. *MATEC Web of Conferences, 161*, 01001, doi:10.1051/matecconf/201816101001

[12] Balogh, Z., Magdin, M., & Molnár, G. (2019) Motion Detection and Face Recognition using Raspberry Pi, as a Part of, the Internet of Things. *Acta Polytechnica Hungarica*, *16(3)*, 167-185

[13] Leo, M., Carcagni, P., Mazzeo, P. L., Spagnolo, P., Cazzato, D., & Distante, C. (2020) Analysis of facial information for healthcare applications: A survey on computer vision-based approaches. *Information, 11*(3), 128, doi:10.3390/info11030128

[14] Bruce, V., & Young, A. (1986) Understanding face recognition. *British Journal of Psychiatry, 77*, 305-327, doi:10.1111/j.2044-8295.1986.tb02199.x

[15] Darwin, C. & Prodger, P. (1998) *The Expression of the Emotions in Man and Animals*. Oxford University Press

[16] Ekman, P., & Friesen, W. V. (1978) *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto, CA: Consulting Psychologists Press

[17] Russell, J. A. (1980) A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161-1178, doi: 10.1037/h0077714

[18] Revina, I. M., & Emmanuel, W. R. S. (2021) A survey on human face expression recognition techniques. *Journal of King Saud University - Computer and Information Sciences, 3*(6), 619-628, doi: 10.1016/j.jksuci.2018.09.002

[19] Mlakar, U., & Potočnik, B. (2015) Automated facial expression recognition based on histograms of oriented gradient feature vector differences. *Signal, Image and Video Processing, 9* (1), 245-253, doi: 10.1007/s11760-015-0810-4

[20] Fuad, M. T. H., Fime, A. A., Sikder, D., Iftee, M. A. R., Rabbi, J., Al-rakhami, M. S., Gumae, A., Sen, O., Fuad, M., & Islam, M. N. (2021) *Recent advances in deep learning techniques for face recognition*. PsyArXiv. doi: 10.13140/RG.2.2.13444.86402

[21] Martinez, B., & Valstar, M. F. (2016) Advances, challenges, and opportunities in automatic facial expression recognition. In M. Kawulok, E. M. Celebi, & B. Smolka (Eds.), *Advances in Face Detection and Facial Image Analysis*, (pp. 63-100) Springer. doi: 10.1007/978-3-319-25958-1_4

[22] Weber, R., Soladié, C., & Séguier, R. A. (2018) Survey on databases for facial expression analysis. *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, 5,* 73-84 doi: 10.5220/0006553900730084

[23] Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2017) *AffectNet: A database for facial expression, valence, and arousal computing in the wild.* PsyArXiv. doi: 10.1109/TAFFC.2017.2740923

[24] Marini, M., Ansani, A., Paglieri, F., Caruana, F., & Viola, M. (2021) The impact of facemasks on emotion recognition, trust attribution and re-identification. *Scientific Report, 11*, 5577, doi: 10.1038/s41598-021-84806-5

[25] Yong, H., Lee, J., & Choi, J. (2019 March) Emotion recognition in gamers wearing head-mounted display. In 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR) (pp. 1251-1252) IEEE

[26] Cabani, A., Hammoudi, K., Benhabiles, H., & Melkemi, M. (2021) MaskedFace-Net - A dataset of correctly/incorrectly masked face images in the context of COVID-19. *Smart health (Amsterdam, Netherlands), 19*, 100144, doi: 10.1016/j.smhl.2020.100144

[27] Noyes, E., Davis, J. P., Petrov, N., Gray, K. L. H., & Ritchie, K. L. (2021) The effect of face masks and sunglasses on identity and expression recognition with super-recognizers and typical observers. *Royal Society Open Science, 8,* 201169, doi: 10.1098/rsos.201169

[28] Li, Y., Zeng, J., Shan, S., Chen, X. (2018) Occlusion aware facial expression recognition using CNN with attention mechanism. *IEEE Transactions on Image Processing,* 1–1. doi:10.1109/TIP.2018.2886767

[29] Goodarzi, F., Saripan, M. I., Marhaban, M. H., & Rokhani, F. Z. (2018) Real time facial expression recognition in the presence of rotation and partial occlusions. *Journal of Theoretical and Applied Information Technology, 96* (4), 854-864

[30] Vyas, A., & Hablani, R. (2014) Effect of different occlusion on facial expressions recognition*. International Journal of Engineering Research and Applications, 4* (10), 40-44

[31] Anand, R., Shanthi, T., Nithish, M. S., & Lakshman, S. (2020) Face recognition and classification using GoogleNET architecture. In Soft Computing for Problem Solving (pp. 261-269) Springer, Singapore

[32] Selvaraju, R. R. et al. (2017) Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. In: Proceedings of the IEEE international conference on computer vision (pp. 618-626)

[33] Bilalpur, M., Kia, S. M., Chua, T. S., & Subramanian, R. (2017, October) Discovering gender differences in facial emotion recognition via implicit behavioral cues. In 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII) (pp. 119-124) IEEE

[34] Majaj, N. J., & Pelli, D. G. (2018) Deep learning—Using machine learning to study biological vision. *Journal of vision*, *18*(13), 2-2

[35] Storrs, K. R., & Kriegeskorte, N. (2019) Deep learning for cognitive neuroscience. *arXiv preprint arXiv:1903.01458*