INFORMATION CONTENT DETECTION IN FACE PARTS USING DEEP LEARNING

Marek DOBEŠ*
CSPV SAV, Karpatská 5, 04001 Košice, Slovak Republic
Natália SABOLOVÁ
FF UPJŠ, Moyzesova 9, 040 59 Košice, Slovak Republic
*Corresponding author. E-mail: dobes@saske.sk

   Abstract. The present paper introduces use of deep neural network for classification of three different categories of emotions - angry, happy and neutral. The database consisted of 48x48 pixel grayscale images of faces from the Face expression recognition dataset from Kaggle. Separate parts of faces such as eyes, nose, or mouths were occluded by a manually inserted 48x15 pixel black rectangle to see what part of the face carries the most significant information about the expressed emotions. By applying pretrained Inception network provided as a part of Keras/Tensorflow environment, we found that, to our surprise, faces with eyes covered were more easily identified. Results were replicated using augmented data.

   Keywords: emotions; deep learning; information content

1. INTRODUCTION

   It is easy to imagine the brain as a super-efficient computer running algorithms, for example, dedicated to the interpretation of people's facial expressions and emotions. This can serve as a metaphor for understanding what computational modelling is attempting to, to decode and understand these algorithms [1]. Importantly, computational models do not reproduce actual brain architecture and cannot explain insights of complex mechanisms used by the brain, however, researchers are able to design computer algorithms that lead to face and emotion recognition. For example, algorithms that automatically code facial actions in still images and video sequences present a great potential in psychology, neuroscience [2], security or forensic setting, as a polygraph aid during interrogation or interviews [3,4], robotics [5] and retail or consumer science etc. [6].
   Facial expression recognition is a successfully growing sub-field of image processing [7]. A great number of computational models of emotion recognition have been compared and validated by looking at its results, correctness, and accuracy [8, 9]. Researchers' aim is to develop the most efficient machine learning or computer vision methods and algorithms by further fine-tuning, editing and optimization of parameters [10, 11]. The process requires making design choices that reflect the nature of data. Techniques that have been used include repeating what worked in other settings and model selection by trial and error search [12]. Firstly, the accuracy and characteristics of databases (background, distance from the camera, resolution of pictures, number of expressions used etc.) play an important role in facial expression recognition research. While using the intra-database protocol (training and testing data are from the same database), well-established databases such as Extended Cohn-Kanade Database CK+ [13], MMI [14], Radboud Faces Database RaFD [15] or Karolinska Directed Emotional Faces KDEF [16] models perform with high accuracies. However, when a cross-database protocol is adopted (one or more databases are used for training and other databases are utilized for evaluating) models achieve notably lower accuracies [11]. Furthermore, even when choice of databases is considered, researchers must cope with an intra-class variation, clutter, occlusion, face pose changes and excessive make-up of participants etc. [17, 18].

For over two millennia, research has been built on the hypothesis that certain discrete and discernible facial muscle and muscle groups changes are linked to specific emotions [19, 20, 21]. Thus, Action Units (AUs), discriminable by facial muscle changes are considered to be expression markers [22] for the most widely used and validated method of measuring and describing facial behaviours - Facial Action Coding System (FACS), developed by Paul Ekman and Wallace V. Friesen [23]. Action Units detail facial movement specifics. For example, AU 12 represents a lip corner puller that allows one to smile (zygomatic major muscle), AU 6 represents a cheek riser displayed also as wrinkling lateral to the eyes (orbicularis oculi contraction and pars orbitalis) and the combinations of those AUs refer to emotion-specified facial expression - happy. Therefore, by allocating AUs such as brow rising and lowering, nostrils dilating etc. and its combination, FACS defines seven emotions: happy, sad, anger, fear, surprise, disgust, and neutrality that might be observed at different intensities.

The problem of emotion recognition (and object recognition in general) has typically been approached using parts-and-shape models that represent not only the appearance of individual object components, but also the spatial relations between them [1, 24, 25]. Similarly, it has been proposed that the brain or computer needs to identify the shape and shading features that are invariant to identity, pose and illumination, and incorporate it into algorithms that identify which of these features best discriminate the muscles involved in each expression [26]. On the other hand, Benitez-Quiroz with colleagues argue that facial expressions of emotion and neutral faces can be discriminated solely depending on colour features without information on shading provided by facial muscle movements.

Because of the complexity of faces [27] and emotional expressions, the aim of this paper is to look closer at stimuli present in faces that may convey important information for emotion recognition as this should be regarded as the cornerstone of the computational models and approaches leading to emotion recognition. Similarly, Happy and Routray [28] examined 'salient facial patches' as distinct pairs of expression classes in the area of inner facial features such as eyes, nose and mouth. In other words, this study wants to explore the aforementioned area of taxonomic reference setting, (computational descriptors, facial landmarks or classifiers), different feature extraction conditions and their effects on identifying targeted emotions. One may argue that perception and expression of emotions vary cross-culturally [29]. For instance, there is evidence that more expressive English-speaking western cultures focus more on the area of mouth than people from Asian countries, who pay more attention to eyes [30]. This was also reflected in analyses of cross-cultural use of emoticons on Twitter that supported differences of emoticon preferences, as easterners expressed their feelings and states with their eyes, whereas westerners did so with their mouth [31]. However, Srinivasan and Martinez [32] in their study identified 35 facial cross-cultural expression configurations (shared expressions across countries) and only 8 culture-specific (used in some, but not all, cultures).

In this study, we look at the information content that may be hidden in specific face parts when displaying emotions. We use a deep learning framework to provide an objective measure of how rich in information content these face parts are.

2. METHODOLOGY

For the data analysis we used photographs of people from the Face expression recognition dataset from Kaggle [33]. All photographs are 48x48 pixel grayscale images of faces, each image corresponds to a facial expression of one emotion category. For the simplicity purposes of this preliminary study, we chose only three categories of emotions - angry, happy and neutral. In every category, we split the data into four groups, either we left pictures intact or we covered eyes, nose, or mouths by a manually inserted 48x15 pixel black rectangle into the photograph. Figure 1 shows samples of input data.

Pretrained Inception network provided as a part of Keras/Tensorflow environment was employed. After removing the final layer, we froze the network and added a softmax layer on the top to allow for training of the data.

Figure 1 Example of input photographs

For every emotion category, we used 40 pictures as a training set, 30 as a validation set and 30 as a test set. The network was trained by applying RMSprop optimizer, for the loss we ran sparse categorical crossentropy with 50 epochs.

3. RESULTS

For every set of data, we ran 50 simulations to adjust for the stochastic nature of deep learning algorithms. We found an average accuracy of 0.54 for the sample without covering any part of the face (average loss was 1.48). For the sample with covered nose, an average accuracy reached 0.58 (average loss 1.36). For the sample with covered eyes, we found an average accuracy of 0.60 (average loss 1.24). In the sample with covered mouth, there was an average accuracy of 0.44 (average loss 1.75).

We find these results surprising as the accuracy for pictures with a covered nose and eyes respectively was higher than for a non-covered face. To see whether these results may not be caused by some artefact in the data we used augmentation of pictures. Following data augmentation parameters in Keras datagen to modify pictures were used:

```
rotation_range=20,
width_shift_range=0.02,
height_shift_range=0.02,
horizontal_flip=True
```

Again, we ran 50 simulations and found an average accuracy of 0.52 for the sample without covering any part of the face. For the sample with covered nose, we found an average accuracy of 0.50. For the sample with covered eyes, we found an average accuracy of 0.53. For the sample with covered mouth, we found an average accuracy of 0.40.

4. CONCLUSIONS

In reference to our findings and previous literature, we can conclude that the accuracy of recognition of emotions based on covering of specific face parts varies. This is not an unexpected result. However, the deep network most accurately recognised the emotions with eyes of subjects covered. This finding was replicated also with the augmented data although the differences were less pronounced. We may only hypothesize what may cause such a result. There may be an unknown artefact in the data that may cause this result. Furthermore, the network may work more efficiently when considering less complex data. In future research, we want to operate with larger samples to see whether these results will still hold.

References

[1] Martinez, A. M. Computational models of face perception. *Current Directions in Psychological Science*. 2017. Vol. 26. No. 3. P. 263-269. DOI:10.1177/0963721417698535

[2] Martinez, A. M. The promises and perils of automated facial action coding in studying children's emotions. *Developmental Psychology*. 2019. Vol. 55. No. 9. P. 1965-1981. DOI 10.1037/dev0000728

[3] Barathi, C. S. Lie detection based on facial micro expression, body language and speech analysis. *International Journal of Engineering Research & Technology*. 2016. Vol. 5. No. 2. P. ISSN 2278-0181.

[4] Ryan, A.; et al. Automated facial expression recognition system. In: *43rd Annual 2009 International Carnahan Conference on Security Technology*. 2009. DOI 10.1109/ccst.2009.5335546

[5] Berns, K.; Zafar, Z. Emotion based human-robot interaction. *MATEC Web of Conferences*. 2018. Vol. 161. No. 01001. DOI 10.1051/matecconf/201816101001

[6] Danner, L.; Duerrschmid, K. Automatic facial expressions analysis in consumer science. In: Ares, G.; Verela, P., eds. *Methods in Consumer Research*. Kidlington: Woodhead Publishing, 2018. pp. 231-252.

[7] Mehendale, N. Facial emotion recognition using convolutional neural networks (FERC). *SN Applied Sciences*. 2020. Vol. 2. No. 446. DOI 10.1007/s42452-020-2234-1

[8] Goodfellow, I. J.; et al. Challenges in representation learning: A report on three machine learning contests. In: Lee, M.; Hirose, A.; Hou, Z. G.; Kil, R. M., eds. *Neural Information Processing ICONIP*. Berlin: Springer, 2013. ISBN 978-3-642-42051-1. DOI 10.1007/978-3-642-42051-1_16

[9] Ngoc, Q. T.; Lee, S.; Song, B. C. Facial landmark-based emotion recognition via directed graph neural network. *Electronics*. 2020. Vol. 9. No. 5. P. 764. DOI 10.3390/electronics9050764

[10] Li, L.; Yuan, Y.; Li, M.; Xu, H.; Li, R.; Lu, S. Subject independent facial expression recognition. In: *2019 International Conference on Image, Video and Signal Processing - IVSP 2019*. Shanghai, 2019. DOI 10.1145/3317640.3317662

[11] Zavarez, M. V.; Berriel, R. F.; Oliveira-Santos, T. Cross-database facial expression recognition based on fine-tuned deep convolutional network. In: *2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. Niterói, 2017. DOI 10.1109/sibgrapi.2017.60

[12] Bergstra, J.; Cox, D. D. Hyperparameter optimization and boosting for classifying facial expressions: How good can a "null" model be? In: *Workshop on Challenges in Representation Learning, ICML*. Atlanta, 2013. Available from: <https://arxiv.org/pdf/1306.3476.pdf>

[13] Lucey, P.; Cohn, J. F.; Kanade, T.; Saragih, J.; Ambadar Z.; Matthews, I. The extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition – Workshops*. San Francisco, 2010.

[14] Pantic, M.; Valstar, M.; Rademaker, R; Maat, L. Web-based database for facial expression analysis. In: *International Conference on Multimedia and Expo*. Amsterdam, 2005. DOI 10.1109/ICME.2005.

[15] Langner, O.; Dotsch, R.; Bijlstra, G.; Wigboldus, D. H. J.; Hawk, S. T.; van Knippenberg, A. Presentation and validation of the Radboud Faces Database. *Cognition & Emotion*. 2010. Vol. 24. No. 8. P. 1377-1388.

[16] Goeleven, E.; De Raedt, R.; Leyman, L.; Verschuere, B. The Karolinska Directed Emotional Faces: A validation study. *Cognition and Emotion*. 2008. Vol. 22. No. 6. P. 1094-1118. doi:10.1080/02699930701626582

[17] Sajid, M.; et al. Data augmentation-asisted makeup-invariant face recognition. *Mathematical Problems in Engineering*. 2018. Vol. 2018. DOI 10.1155/2018/2850632

[18] Szirtes, G.; Szolgay, D.; Utasi, Á.; Takács, D.; Petrás, I.; Fodor, G. Facing reality: An industrial view on large scale use of facial expression analysis. In: *ACM International Conference on Multimodal Interaction (ICMI 2013)*. Sydney, 2013. DOI 10.1145/2531923.253192

[19] Aristotle. *Minor Works: On Colours. On Things Heard. Physiognomics. On Plants. On Marvellous Things Heard. Mechanical Problems. On Indivisible Lines. The Situations and Names of Winds. On Melissus, Xenophanes, Gorgias*. Cambridge, MA: Harvard University Press. 1936.

[20] Darwin, C.; Prodger, P. *The Expression of the Emotions in Man and Animals*. New York: Oxford University Press. 1998.

[21] Du, S.; Tao, Y.; Martinez, A. M. Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences of the United States of America*. 2014. Vol. 111. No. 15. P. 1454-1462. DOI 10.1073/pnas.1322355111

[22] Yan, J.; Zheng, W.; Cui, Z.; Tang, C.; Zhang, T.; Zong, Y. Multi-cue fusion for emotion recognition in the wild. *Neurocomputing*. 2018. Vol. 309. P. 27-35. DOI 10.1016/j.neucom.2018.03.068

[23] Ekman, P.; Friesen, W. V. *Facial Action Coding System: A Technique for the Measurement of facial Movemen*t. Palo Alto, CA: Consulting Psychologists Press. 1978.

[24] Martinez, A. M. Visual perception of facial expressions of emotion. *Current Opinion in Psychology*. 2017. Vol. 17. P. 27-33. DOI 10.1016/j.copsyc.2017.06.009

[25] Vesker, M.; Wilson, H. R. Face context advantage explained by Vernier and separation discrimination acuity. *Frontiers in Psychology*. 2013. Vol. 3. DOI 10.3389/fpsyg.2012.00617

[26] Benitez-Quiroz, C. F.; Srinivasan, R.; Martinez, A. M. EmotioNet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. In: *IEEE International Conference on Computer Vision & Pattern Recognition (CVPR'16)*. Las Vegas, 2016. DOI 10.1109/CVPR.2016.600.

[27] Logan, A. J.; Gordon, G. E.; Loffler, G. Contributions of individual face features to face discrimination. *Vision Research*. 2017. Vol. 137. P. 29-39. DOI 10.1016/j.visres.2017.05.011

[28] Happy, S. L.; Routray, A. Automatic facial expression recognition using features of salient facial patches. *IEEE Transactions on Affective Computing*. 2015. Vol. 6. No. 1. P. 1-12. DOI 10.1109/taffc.2014.2386334

[29] Rychlowska, M.; et al. Heterogeneity of long-history migration explains cultural differences in reports of emotional expressivity and the functions of smiles. *Proceedings of the National Academy of Sciences of the United States of America*. 2015. Vol. 112. No. 19. P. 2429-2436. DOI 10.1073/pnas.1413661112

[30] Yuki, M.; Maddux, W. W.; Masuda, T. Are the windows to the soul the same in the East and West? Cultural differences in using the eyes and mouth as cues to recognize emotions in Japan and the United States. *Journal of Experimental Social Psychology*. 2007. Vol. 43. No. 2. P. 303-311. DOI 10.1016/j.jesp.2006.02.004

[31] Park, J.; Barash, V.; Fink, C.; Cha, M. Emoticon style: Interpreting differences in emoticons across cultures. *Seventh International AAAI Conference on Weblogs and Social Media*. Cambridge, 2013.

[32] Srinivasan, R.; Martinez, A. M. Cross-cultural and cultural-specific production and perception of facial expressions of emotion in the wild. *IEEE Transactions on Affective Computing*. 2018. Vol. 12. No. 3. P. 707-721 DOI 10.1109/taffc.2018.2887267

[33] *Face expression recognition with Deep Learning*. Face expression recognition dataset. Available at: https://www.kaggle.com/jonathanoheix/face-expression-recognition-with-deep-learning/data