IMPROVED CNN-BASED MOUTH POSITION AND STATUS DETECTION

CHOK YONG SHENG

UNIVERSITI TEKNOLOGI MALAYSIA

IMPROVED CNN-BASED MOUTH POSITION AND STATUS DETECTION

CHOK YONG SHENG

A project report submitted in partial fulfilment of the
requirements for the award of the degree of
Master of Engineering (Computer and Microelectronic Systems)

School of Electrical Engineering
Faculty of Engineering
Universiti Teknologi Malaysia

JULY 2022

DEDICATION

This thesis is dedicated to my father, who taught me that the best kind of knowledge to have is that which is learned for its own sake. It is also dedicated to my mother, who taught me that even the largest task can be accomplished if it is done one step at a time.

ACKNOWLEDGEMENT

First, I would like to convey my deepest thanks and appreciation to supervisor for this project, Dr Abdul-Malik Haider Yusef, for all his supervision and encouragement given throughout the completion of this project. He helped me to solve the problems that I faced while completing the project and provided several opportunities for me to explore new knowledge and to make decisions. Without his useful guidance and countless support, it is impossible for me to complete this report successfully.

Next, I would like to take this opportunity to thank all my lecturers who have taught and guided me throughout the years of studying life in UTM. They were willing to help and guide their students in every project and subject. Without the guidance and knowledge given, I believe that I would not be able to complete this final year project.

Not forgetting my deepest gratitude to my beloved family, especially my parents for their encouragement, constructive suggestions, and motivational support throughout the project completion, from the beginning till the end.

Last but not least, thanks to all my friends and course mates who have supported and helped me directly and indirectly throughout the completion of this report. Their help, ideas, advice, and moral support really help me in getting through the hardships while completing this project.

ABSTRACT

Mouth position and status detection system plays an important role in the auto-feeding system for paralyzed people. Through identifying the mouth status, whether it is open or close, and obtain the location of the open mouth, the system will be able to pick the correct timing to feed patients with robotic arms. There are two major problems that urge the proposal of this project. First, the existing mouth status recognition networks are built and executed on high-end and costly hardware. Second, the existing CNN mouth status related detection systems are less accurate, the highest accuracy in the researched work is only 86.8% for 3 states mouth status detection. Based on the problems, there are two research objectives that are strived to be achieved. First, to develop a high-accuracy and light CNN-based model for mouth status detection on Python platform. Second, to shorten the inference time of the CNN-based model by resizing some of the convolution layers. For methodology, the primary task is to train a mouth status detection CNN model with high accuracy. The face picture datasets fed to the model during CNN model training are diverse, covering different human races and shooting angles. YOLOv5 is chosen to be the pre-trained network due to its outstanding performance. The YOLOv5 backbone convolution layers are resized to shorten the inference time and reduce the model size. The developed CNN-based model achieved the targeted performance which is 96.8%, successfully improved inference time by 21.90% and model size by 13.20% as compared to the original model before enhancement.

ABSTRAK

Sistem pengesanan kedudukan mulut dan status memainkan peranan penting dalam sistem penyusuan automatik untuk orang lumpuh. Melalui mengenal pasti status mulut, sama ada ia terbuka atau tertutup, dan mendapatkan lokasi mulut terbuka, sistem akan dapat memilih masa yang betul untuk memberi makan kepada pesakit dengan lengan robot. Terdapat dua masalah besar yang mendesak cadangan projek ini. Pertama, rangkaian pengecaman status mulut sedia ada dibina dan dilaksanakan pada perkakasan mewah dan mahal. Kedua, sistem pengesanan berkaitan dengan status mulut CNN sedia ada adalah kurang tepat, ketepatan tertinggi dalam kerja yang dikaji hanyalah 86.8% untuk pengesanan status mulut untuk 3 status. Berdasarkan permasalahan tersebut, terdapat dua objektif kajian yang diusahakan untuk dicapai. Pertama, untuk membangunkan model berasaskan CNN berketepatan tinggi dan ringan untuk pengesanan status mulut pada platform Python. Kedua, untuk memendekkan masa inferens model berasaskan CNN dengan mengubah saiz beberapa lapisan konvolusi. Untuk metodologi, tugas utama adalah untuk melatih model CNN pengesanan status mulut dengan ketepatan yang tinggi. Set data gambar muka yang diberikan kepada model semasa latihan model CNN adalah pelbagai, meliputi kaum manusia yang berbeza dan sudut penangkapan. YOLOv5 dipilih untuk menjadi rangkaian pra-latihan kerana prestasinya yang cemerlang. Lapisan lilitan tulang belakang YOLOv5 diubah saiz untuk memendekkan masa inferens dan mengurangkan saiz model. Pada akhir penyelidikan ini, model berasaskan CNN yang dibangunkan mencapai prestasi yang disasarkan iaitu 96.8%, berjaya meningkatkan masa inferens sebanyak 21.90% dan saiz model sebanyak 13.20% berbanding model asal sebelum peningkatan.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANN	-	Artificial Neural Network
AI	-	artificial intelligent
CNN	-	Convolutional Neural network
FN	-	False Negative
FP	-	False Positive
MATLAB	-	Matrix Laboratory
YOLO	-	You Only Look Once

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Background

Mouth status detection is one of key components in human face detection and is crucial for recognizing mouth shapes, reading lips, and authenticating identities. The application of mouth status detection is wide in safety and health sector, like driver's fatigue level detection and automated self-feeding machine for persons with physical disability. The self-feeding system with light model size yet accurate CNN mouth detection system can be marketized at a more affordable price, thus contributing to paralyzed people which are having difficulties in taking meal.

Paralyzed people can be defined as people who is suffering from losing part of body's muscle function[1]. Based on Department of Statistical Malaysia, in year 2017, Malaysia has around 453,258 Person with Disablities (PWD) and around 35.2 percent of them are suffering from physical disability which having difficulty in moving themselves[2]. Paralysis mainly can be classified into 4 categories which are localized, generalized, partial and complete paralysis, each possessing different level severity. People suffering with partial paralysis faced muscle weakness, but they still can move their affected body part with a small degree of control. There are 4 main types of paralysis[3]. Monoplegia affecting only one area, such as one arm or leg. Hemiplegia affecting 2 area from the same side of body, like left arm and left leg. Paraplegia which affects the lower body, such as paralysis of both legs. Quadriplegia affecting both arms and legs, could be also affecting functionality of organs. Paralysis can be caused by various kind of reason, for example, spinal cord injury, cerebral palsy, stroke, multiple sclerosis and others[4]. Paralysis has bring challenges to the patient like loss of voluntary movement, unwanted motions such as spasticity and spasm, loss of feeling leading to skin disintegration and loss of awareness, and discomfort[5]. The treatment for paralysis could be therapy and medical treatments, as well as the mobility assistance like wheelchair and urine collection devices.

Restorative treatments, such as body weight assisted treadmill training, have recently been studied to repair dormant circuits in the spinal cord. Tendon transfers and, in certain circumstances, entire muscle transfers are among the surgical treatments to cure paralysis. For some people, these therapy choices can result in a considerable functional improvement. However, these options leave considerable gaps in the functional restoration requirements. Restoration is almost invariably incomplete, leaving the person with significant functional loss and often requiring human help for important daily activities[5]. The goal is to provide meaningful solutions that fit into the user's daily routine and improve their independence and quality of life. There are various kind of adaptive equipment which are used to support the patients' daily activity. Mobility device like wheelchairs, canes and adapted shoes has a vast space of enhancement with creativity ideas[6]. Other than that, self-care tools, environmental control devices, positioning devices and others also have rooms for improvement to cope better with the paralyzed patients' daily lives.

One of the most popular adaptive equipment in the market is the adaptive feeding device for paralyzed people[7]. For example, Hand cuffs are manual devices that are simple to operate and effective for those who have limited hand dexterity. Another advantage of these aids is that you may quickly replace your dining utensils with a toothbrush or another tiny object. There are also some robotic devices invented for patients suffering with Quadriplegics. For example, meal buddy system which is using a robotic arm and bowls built on a based, which are linked to each other through magnetic connectors. The user has complete control on the feeding pace and bowl choice. Deep learning concept appeared firstly during 2006 as a new field of research within machine learning[8]. It has been widely implemented in many research fields related to pattern recognition.

Deep learning model performs feature extraction and classification using cascaded and multilayer processing units. The learning process of deep learning could be supervised or unsupervised. Supervised learning refers to learning process with labelled target classes while unsupervised learning refers to learning process without labelled target classes [9]. One of the advantages of deep learning is its ability to perform automatic feature extraction instead of classifying handpicked feature based on domain-specific knowledge. This helps in performing detection and classification on certain patterns effectively [10]. Many deep learning algorithms are task-specific, which are trained to carry out the intended function and purpose. If the feature changes, the models must be rebuilt from scratch. Transfer learning is a method that resolves such tradeoff by using the knowledge learned for old task to seek for solution for another new task. Transfer learning technique gives better performance result with smaller sample size data due to its pre-trained weights and improved efficiency. Pre-trained model is defined as a model that was trained with a huge benchmark dataset like ImageNet to seek solutions a general issues [11]. The example for pre-trained model includes VGG-16, ResNet, AlexNet, EfficientNet, GoogleNet, and others.

1.2 Problem Statement

First, the existing mouth status recognition networks are built and executed on high-end and costly hardware, for example, the mouth status classification with high-spec GPU[12] and Lips state identification using Kinect2[13]. The main reason is because the existing networks trained for mouth status detection are complex and huge. The hardware requirement to adapt such complex network is relatively high, like relatively large memory to contain the big CNN model and the high-performance processor to obtain inference time which is fast enough for real-time-cam detection.

Second, the existing CNN mouth status related detection systems proposed have only moderate accuracy, which is less than 90% for 3 states mouth classification. Although the LSTM(Long-Short Term Memory) network implemented by Pinzon-Arenas in research in year 2019[12] and research in year 2018[14] have accuracy up to 97.9% and 99.3% respectively, but the 3-mouth states recognitions have only 84.8% and 86.8%. The driver's mouth-and-eye-based fatigue detection system proposed by Deng Wanghua[15] is also having only 92% accuracy, and it would drop slightly if the driver is wearing glasses. The other mouth-centric emotion recognition system using CNN proposed by Valentina Franzoni[16] is also having around 79.5% accuracy only.

1.3 Research Objectives

(a) To prepare mouth states datasets with open, close and intermediate open-close states for model training.

(b) To develop a high-accuracy CNN-based model for mouth status detection on Python platform.

(c) To shorten the inference time of the CNN-based model by resizing the convolution layers.

1.4 Scope of Work

(a) The mouth position and status detection system will utilize the face samples from public datasets.

(b) The neural network is trained using public images dataset that contains multi-racial and multi-angle-shot faces with open, close and intermediate open-close mouth.

(c) YOLOv5 was selected as the CNN model to be used in mouth status detection process.

(d) YOLO architecture was analysed and enhanced to shorten inference time and reduce model size.

1.5 Report Organization

The purpose of this report is to develop a mouth position and status detection system using a transfer learning approach. The overall five chapters in the report are organized in the following sequences: introduction, literature review, research methodology, preliminary results and lastly the conclusion and recommendation.

The first chapter introduces and give brief information about the purpose of the study. The introduction concludes the overview of the entire topic of study. The problem statements, research objectives and the related scopes on each objective are all included in this chapter.

Chapter 2 is a study of journal articles and other materials from other researchers, and it falls under the literature review section. The study focuses on specific and pertinent subjects, summarises the researcher's findings and provide personal perspectives and comments on the research paper.

Chapter 3 is focusing on the research methodology. This chapter's topic demonstrates the methodology and framework to implement the technique to meet the specified objectives. This chapter also clarifies the basic methods and methodology for doing research.

Chapter 4 presents the preliminary results. The overall experimental and simulation results are interpreted and analysed in this chapter. By referring to any plotted graph or table, the discussion examines whether the resulting result is valid or invalid. Make comparison of any modifications in different parameter and factor as they will have varied outcomes.

Chapter 5 is the summary and conclusion of the entire research paper. It summarises the essential elements of the main study topic as well as concluding the final obtained result to justify the result is acceptable or not. Any recommendations for further research as well as the limitations encountered during the study are explored

REFERENCES

[1] H. Stubblefield, “Paralysis,” 2018. https://www.healthline.com/health/paralysis (accessed Dec. 19, 2021).

[2] M. Department of Statistics, “Social Statistics Bulletin, Malaysia,” *Department of Statistics, Malaysia*, 2012. https://www.dosm.gov.my/v1/index.php?r=column/cthemeByCat&cat=152&bul_id=NU5hZTRkOSs0RVZwRytTRE5zSitLUT09&menu_id=U3VPMldoYUxzVzFaYmNkWXZteGduZz09 (accessed Dec. 19, 2021).

[3] J. Eske, “Paralysis: What is it?,” *MedicalNewsToday,* 2020. https://www.medicalnewstoday.com/articles/paralysis (accessed Dec. 19, 2021).

[4] A. Eske, “What are the differences between paraplegia and quadriplegia?,” *MedicalNewsToday*, 2020. https://www.medicalnewstoday.com/articles/paraplegia-vs-quadriplegia (accessed Dec. 19, 2021).

[5] P. H. Peckham and K. L. Kilgore, “Challenges and opportunities in restoring function after paralysis,” *IEEE Trans. Biomed. Eng.*, vol. 60, no. 3, pp. 602–609, 2013, doi: 10.1109/TBME.2013.2245128.

[6] K. StayWell, “Spinal Cord Injury (SCI): Adaptive Equipment | MHealth.org.” https://www.mhealth.org/Patient-Education/Articles/English/s/p/i/n/a/Spinal_Cord_Injury_SCI_Adaptive_Equipment_41175.

[7] S. co. Team, “4 Adaptive Feeding Devices for Quadriplegics,” *SpinalCord.com*, 2020. https://www.spinalcord.com/blog/4-adaptive-feeding-devices-for-quadriplegics (accessed Dec. 12, 2021).

[8] V. M. Kota, V. Manoj Kumar, and C. Bharatiraja, “Deep Learning - A Review,” *IOP Conf. Ser. Mater. Sci. Eng.,* vol. 912, no. 3, 2020, doi: 10.1088/1757-899X/912/3/032068.

[9] R. Sathya and A. Abraham, “Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification,” *Int. J. Adv. Res. Artif. Intell.,* vol. 2, no. 2, pp. 34–38, 2013, doi: 10.14569/ijarai.2013.020206.

[10] S. T. Hwa Kieu, A. Bade, M. H. Ahmad Hijazi, and H. Kolivand, “A survey of deep learning for lung disease detection on medical images: State-of-the-art, taxonomy, issues and future directions,” *J. Imaging*, vol. 6, no. 12, 2020, doi: 10.3390/jimaging6120131.

[11] M. C. Phillips, R. Stein, and T. Park, “Analyzing Pre-Trained Neural Network Behavior with Layer Activation Optimization,” *2020 Syst. Inf. Eng. Des. Symp. SIEDS 2020*, 2020, doi: 10.1109/SIEDS49339.2020.9106628.

[12] J. O. Pinzon-Arenas, R. Jimenez-Moreno, and A. Rubiano-Fonseca, “Mouth states using LSTM neural network,” *IEEE ICA-ACCA 2018 - IEEE Int. Conf. Autom. Congr. Chil. Assoc. Autom. Control Towar. an Ind. 4.0 - Proc.*, pp. 17–21, 2019, doi: 10.1109/ICA-ACCA.2018.8609837.

[13] M. Z. Yazdi, “Depth-Based Lip Localization and Identification of Open or Closed Mouth, Using Kinect 2,” *Proceedings*, vol. 27, no. 1, p. 22, 2019, doi: 10.3390/proceedings2019027022.

[14] J. O. Pinzón-arenas, R. Jiménez-moreno, and A. Rubiano-fonseca, “Mouth States Recognition System Focused on Feeding task,” vol. 13, no. 22, pp. 15872–15877, 2018.

[15] W. Deng and R. Wu, “Real-Time Driver-Drowsiness Detection System Using Facial Features,” *IEEE Access*, vol. 7, pp. 118727–118738, 2019, doi: 10.1109/ACCESS.2019.2936663.

[16] V. Franzoni, G. Biondi, D. Perri, and O. Gervasi, “Enhancing mouth-based emotion recognition using transfer learning,” *Sensors (Switzerland)*, vol. 20, no. 18, pp. 1–15, 2020, doi: 10.3390/s20185222.

[17] M. DIxit, A. Tiwari, H. Pathak, and R. Astya, “An overview of deep learning architectures, libraries and its applications areas,” *Proc. - IEEE 2018 Int. Conf. Adv. Comput. Commun. Control Networking, ICACCCN 2018*, pp. 293–297, 2018, doi: 10.1109/ICACCCN.2018.8748442.

[18] A. Abubakar, M. Ajuji, and I. U. Yahya, “Comparison of deep transfer learning techniques in human skin burns discrimination,” *Appl. Syst. Innov.*, vol. 3, no. 2, pp. 1–15, 2020, doi: 10.3390/asi3020020.

[19] X. Kang, B. Song, and F. Sun, “A deep similarity metric method based on incomplete data for traffic anomaly detection in IoT,” *Appl. Sci.*, vol. 9, no. 1, 2019, doi: 10.3390/app9010135.

[20] A. Nasiri, A. Taheri-Garavand, and Y. D. Zhang, “Image-based deep learning

automated sorting of date fruit," *Postharvest Biol. Technol.*, vol. 153, pp. 133–141, 2019, doi: 10.1016/j.postharvbio.2019.04.003.

[21] G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, and N. Srivastava, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[22] J. Redmon and A. Farhadi, "Yolo V2.0," *Cvpr2017*, no. April, pp. 187–213, 2017, [Online]. Available: http://www.worldscientific.com/doi/abs/10.1142/9789812771728_0012.

[23] P. S. Raskar and S. K. Shah, "Real time object-based video forgery detection using YOLO (V2)," *Forensic Sci. Int.*, vol. 327, 2021, doi: 10.1016/j.forsciint.2021.110979.

[24] L. Jianwei and G. Kai, "Research on Monocular Visual Gesture Positioning based on YOLO v2," *Chinese Control Conf. CCC*, vol. 2018-July, pp. 9101–9106, 2018, doi: 10.23919/ChiCC.2018.8483096.

[25] F. Bi and J. Yang, "Target Detection System Design and FPGA Implementation Based on YOLO v2 Algorithm," *2019 3rd Int. Conf. Imaging, Signal Process. Commun. ICISPC 2019*, pp. 10–14, 2019, doi: 10.1109/ICISPC.2019.8935783.

[26] K. Jo, J. Im, J. Kim, and D. S. Kim, "A real-Time multi-class multi-object tracker using YOLOv2," *Proc. 2017 IEEE Int. Conf. Signal Image Process. Appl. ICSIPA 2017*, pp. 507–511, 2017, doi: 10.1109/ICSIPA.2017.8120665.

[27] Y. Tian, G. Yang, Z. Wang, H. Wang, E. Li, and Z. Liang, "Apple detection during different growth stages in orchards using the improved YOLO-V3 model," *Comput. Electron. Agric.*, vol. 157, no. October 2018, pp. 417–426, 2019, doi: 10.1016/j.compag.2019.01.012.

[28] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," 2018, [Online]. Available: http://arxiv.org/abs/1804.02767.

[29] W. Jia *et al.*, "Real-time automatic helmet detection of motorcyclists in urban traffic using improved YOLOv5 detector," *IET Image Process.*, vol. 15, no. 14, pp. 3623–3637, 2021, doi: 10.1049/ipr2.12295.

[30] "ultralytics/yolov5." https://github.com/ultralytics/yolov5.

[31] J. Yao, J. Qi, J. Zhang, H. Shao, J. Yang, and X. Li, "A real-time detection algorithm for kiwifruit defects based on yolov5," *Electron.*, vol. 10, no. 14, 2021, doi: 10.3390/electronics10141711.

[32] "SSD vs. YOLO for Detection of Outdoor UrbanAdvertising Panels under Multiple Variabilities.pdf." .

[33] W. Liu *et al.*, "SSD: Single shot multibox detector," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9905 LNCS, pp. 21–37, 2016, doi: 10.1007/978-3-319-46448-0_2.

[34] Z. Li and F. Zhou, "FSSD: Feature Fusion Single Shot Multibox Detector," vol. 1, 2017, [Online]. Available: http://arxiv.org/abs/1712.00960.

[35] C. Chen, M. Y. Liu, O. Tuzel, and J. Xiao, "R-CNN for small object detection," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10115 LNCS, pp. 214–230, 2017, doi: 10.1007/978-3-319-54193-8_14.

[36] X. Lu, X. Kang, S. Nishide, and F. Ren, "Object detection based on SSD-ResNet," *Proc. 2019 6th IEEE Int. Conf. Cloud Comput. Intell. Syst. CCIS 2019*, pp. 89–92, 2019, doi: 10.1109/CCIS48116.2019.9073753.

[37] L. Posilovic, D. Medak, M. Subasic, T. Petkovic, M. Budimir, and S. Loncaric, "Flaw detection from ultrasonic images using YOLO and SSD," *Int. Symp. Image Signal Process. Anal. ISPA*, vol. 2019-Septe, pp. 163–168, 2019, doi: 10.1109/ISPA.2019.8868929.

[38] J. A. Kim, J. Y. Sung, and S. H. Park, "Comparison of Faster-RCNN, YOLO, and SSD for Real-Time Vehicle Type Recognition," *2020 IEEE Int. Conf. Consum. Electron. - Asia, ICCE-Asia 2020*, 2020, doi: 10.1109/ICCE-Asia49877.2020.9277040.

[39] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *Adv. Neural Inf. Process. Syst.*, vol. 2015-Janua, pp. 91–99, 2015.

[40] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal Loss for Dense Object Detection," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2017-Octob, pp. 2999–3007, 2017, doi: 10.1109/ICCV.2017.324.

[41] C. Bouvier, A. Benoit, A. Caplier, and P. Y. Coulon, "Open or closed mouth state detection: Static supervised classification based on log-polar signature," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5259 LNCS, pp. 1093–1102, 2008, doi: 10.1007/978-3-540-88458-3_99.

[42] F. B. Bryant and P. R. Yarnold, "Principal-Components Analysis and

Exploratory and Confirmatory Factor Analysis," *Read. Underst. Multivar. Stat.*, pp. 99–136, 1995, [Online]. Available: http://psycnet.apa.org/psycinfo/1995-97110-000.

[43] Y. Yang, J. Li, and Y. Yang, "The research of the fast SVM classifier method," *2015 12th Int. Comput. Conf. Wavelet Act. Media Technol. Inf. Process. ICCWAMTIP 2015*, pp. 121–124, 2015, doi: 10.1109/ICCWAMTIP.2015.7493959.

[44] a M. Martinez and R. Benavente, "The AR face database," *CVC Tech. Rep. 24*, vol. %6, p. %&, 1998.

[45] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," *IEE Conf. Publ.*, vol. 2, no. 470, pp. 850–855, 1999, doi: 10.1049/cp:19991218.

[46] Y. Ji, S. Wang, Y. Lu, J. Wei, and Y. Zhao, "Eye and mouth state detection algorithm based on contour feature extraction," *J. Electron. Imaging*, vol. 27, no. 05, p. 1, 2018, doi: 10.1117/1.jei.27.5.051205.

[47] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," *Image Vis. Comput.*, vol. 16, no. 5, pp. 295–306, 1998, doi: 10.1016/s0262-8856(97)00070-x.

[48] H. Kalbkhani and M. C. Amirani, "An Efficient Algorithm for Lip Segmentation in Color Face Images," vol. 1, no. 1, pp. 12–16, 2012.

[49] A. B. A. Hassanat, M. Alkasassbeh, M. Al-Awadi, and E. A. A. Alhasanat, "Colour-based lips segmentation method using artificial neural networks," *2015 6th Int. Conf. Inf. Commun. Syst. ICICS 2015*, pp. 188–193, 2015, doi: 10.1109/IACS.2015.7103225.

[50] S. C. Castellanos, I. S. Moreno, J. A. C. Ceballos, R. A. Vargas, and P. L. M. Quintal, "An Approach to Improve Mouth-State Detection to Support the ICAO Biometric Standard for Face Image Validation," *Proc. - 2015 Int. Conf. Mechatronics, Electron. Automot. Eng. ICMEAE 2015*, pp. 35–40, 2015, doi: 10.1109/ICMEAE.2015.12.

[51] Itseez, "Open Source Computer Vision," *Open CV*, 2016. http://opencv.org/.

[52] A. Kasiński, A. Florek, and A. Schmidt, "The put face database," *Image Process. Commun.*, vol. Vol. 13, n, pp. 59–64, 2008, [Online]. Available: http://yadda.icm.edu.pl/baztech/element/bwmeta1.element.baztech-article-

BAT5-0037-0007.

[53] D. D. Hromada, C. Tijus, S. Poitrenaud, and J. Nadel, “Zygomatic smile detection: The semi-supervised haar training of a fast and frugal system. A gift to OpenCV community,” *2010 IEEE-RIVF Int. Conf. Comput. Commun. Technol. Res. Innov. Vis. Futur. RIVF 2010*, 2010, doi: 10.1109/RIVF.2010.5633176.

[54] O. Gervasi, V. Franzoni, M. Riganelli, and S. Tasso, “Automating facial emotion recognition,” *Web Intell.*, vol. 17, no. 1, pp. 17–27, 2019, doi: 10.3233/WEB-190397.

[55] F. Chollet, “Keras,” 2015. .

[56] A. Mollahosseini, B. Hasani, and M. H. Mahoor, “AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild,” *IEEE Trans. Affect. Comput.*, vol. 10, no. 1, pp. 18–31, 2019, doi: 10.1109/TAFFC.2017.2740923.

[57] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, “High-speed tracking with kernelized correlation filters,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 583–596, 2015, doi: 10.1109/TPAMI.2014.2345390.

[58] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, “SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size,” 2016, [Online]. Available: http://arxiv.org/abs/1602.07360.

[59] “YAWDD: YAWNING DETECTION DATASET.” https://ieee-dataport.org/open-access/yawdd-yawning-detection-dataset (accessed Dec. 30, 2021).

[60] “Roboflow.” https://roboflow.com/ (accessed Dec. 30, 2021).

[61] “cocodataset.” https://cocodataset.org/#home (accessed Dec. 28, 2021).

[62] T. Balaiah, T. J. T. Jeyadoss, S. S. Thirumurugan, and R. C. Ravi, “A deep learning framework for automated transfer learning of neural networks,” *Proc. 11th Int. Conf. Adv. Comput. ICoAC 2019*, pp. 428–432, 2019, doi: 10.1109/ICoAC48765.2019.246880.

[63] N. Becherer, J. Pecarina, S. Nykl, and K. Hopkinson, “Improving optimization of convolutional neural networks through parameter fine-tuning,” *Neural Comput. Appl.*, vol. 31, no. 8, pp. 3469–3479, 2019, doi: 10.1007/s00521-017-3285-0.