MASK SEGMENTATION AND CLASSIFICATION WITH ENHANCED GRASSHOPPER OPTIMIZATION OF 3D HAND GESTURES

FAWAD SALAM KHAN

A thesis submitted in
fulfilment of the requirement for the award of the
Doctor of Philosophy in Electrical Engineering

Faculty of Electrical and Electronic Engineering
Universiti Tun Hussein Onn Malaysia

APRIL 2023

I dedicate this work to Almighty Allah and the last messenger Muhammad (PBUH) after that to my wife Mrs. Nadia Fawad Khan, my daughter Minhal Fawad Khan and my son Zohair Fawad Khan

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my supervisor, Assoc. Prof. Ts. Dr. Mohd Norzali Bin Hj. Mohd. for all his support as and when required during my Ph.D. study. His help was not just limited to his guidance so that I can reach my Ph.D. destination, but he has also offered me several opportunities to excel in the field of education. His vision, sincerity, and knowledge inspired me a lot and I found myself privileged to be a part of his team. I would also like to thank my co-supervisor, Assoc. Prof. Dr. Dur Muhammad Soomro, for being my mentor and guiding me throughout this journey.

I should not miss out on my mother who is a pillar for being who I am today. She gave me all the freedom to choose my career path and helped me in all ups and downs of my life to reach up to this. I don’t dare to thank her because it cannot express how blessed I am to be her child. I also thank my wife and my children to their support to achieve this milestone.

I am also grateful to the almighty Allah who gave me this opportunity and showed me light whenever I feel lost. Last but not the least, I am also grateful to my dear wife and mother-in-law who encouraged me to pursue this journey.

ABSTRACT

The difficulties associated with extracting 3D hand meshes from depth image utilizing 2D convolutional neural networks. The precision of such estimations is frequently hampered by visual distortions caused by nonrigidity, complex backdrops, and shadows. This research provides a unique methodology that combines the enhanced grasshopper optimization method for feature optimization with MASK-RCNN and FCN for segmenting and classifying 3D hand gestures to address these problems. In order to evaluate the proposed method, a 3D gesture data set is generated. In addition, a skeleton model for RGB hand gestures is constructed by estimating the degree of freedom (DoF) using human kinematics. The segmentation of 3D hand gestures is computed using the ResNet50 backbone network, and the Overlap Coefficient (OC) is employed as an evaluation metric. On the other hand, the ResNet101 backbone network is used to calculate the classification of 3D hand gestures. Experimental results reveal that the proposed method achieves greater accuracy in segmenting and classifying 3D hand gestures than existing methods. The study also emphasizes the significance of using feature optimization approaches and developing skeletal models to estimate (DoF) in order to improve the precision of 3D hand gesture analysis. This study provides a promising approach for robust and precise 3D hand gesture recognition, with potential applications in disciplines such as human-computer interaction and virtual reality. The test results show best accuracy for 3D hand gesture classification and segmentation. On the private dataset, the classification accuracy is 99.05 %, whereas 99.29 % on the Kinect dataset, 99.39 % and 99.29% using SKIG and ChaLearn dataset during validation. The OC is 88.16 % and 88.19 %, respectively which is the highest available accuracy compared with other methods. The mAP of ChaLearn 93.26%, private 81.48%, SKIG 75.21% and Kinect 66.74%.

ABSTRAK

Kesukaran yang berkaitan dengan mengekstrak jerat tangan 3D daripada imej kedalaman menggunakan rangkaian saraf konvolusi 2D. Ketepatan anggaran sedemikian sering dihalang oleh herotan visual yang disebabkan oleh ketidaktegaran, latar belakang yang kompleks dan bayang-bayang. Penyelidikan ini menyediakan metodologi unik yang menggabungkan kaedah pengoptimuman belalang yang dipertingkatkan untuk pengoptimuman ciri dengan MASK-RCNN dan FCN untuk membahagikan dan mengklasifikasikan gerak isyarat tangan 3D untuk menangani masalah ini. Untuk menilai kaedah yang dicadangkan, set data gerak isyarat 3D dijana. Selain itu, model rangka untuk gerak isyarat tangan RGB dibina dengan menganggar tahap kebebasan dengan kinematik manusia (DoF). Pembahagian gerak isyarat tangan 3D dikira menggunakan rangkaian tulang belakang ResNet50, dan Pekali Pertindihan (OC) digunakan sebagai metrik penilaian. Sebaliknya, rangkaian tulang belakang ResNet101 digunakan untuk mengira klasifikasi gerak isyarat tangan 3D. Keputusan eksperimen mendedahkan bahawa kaedah yang dicadangkan mencapai ketepatan yang lebih besar dalam membahagikan dan mengklasifikasikan gerak isyarat tangan 3D daripada kaedah sedia ada. Kajian ini juga menekankan kepentingan menggunakan pendekatan pengoptimuman ciri dan membangunkan model rangka untuk menganggarkan DoF bagi meningkatkan ketepatan analisis isyarat tangan 3D. Kajian ini menyediakan pendekatan yang menjanjikan untuk pengecaman gerak isyarat tangan 3D yang mantap dan tepat, dengan aplikasi yang berpotensi dalam disiplin seperti interaksi manusia-komputer dan realiti maya. Keputusan ujian menunjukkan ketepatan terbaik untuk klasifikasi dan pembahagian gerak isyarat tangan 3D. Pada menetapkan data peribadi, ketepatan klasifikasi ialah 99.05%, manakala 99.29% pada menetapkan data Kinect, 99.39% dan 99.29% pada set data SKIG dan ChaLearn semasa pengesahan. OC ialah 88.16% dan 88.19%, masing-masing yang

menjadi ketepatan tertinggi yang tersedia berbanding dengan kaedah lain. Peta ChaLearn 93.26%, persendirian 81.48%, SKIG 75.21% dan Kinect 66.74%.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

3D	–	Three dimensional
DoF	–	Degree of freedom
ANN	–	Artificial neural network
RGB	–	Red Green Blue
RL	–	Reinforcement learning
IMU	–	Inertial Measurement Unit
BLDC	–	Brush less direct current
LQR	–	Linear quadratic regulator
NN	–	Neural network
CNN	–	Convolutional neural network
DQN	–	Deep quality-network
GOA	–	Grasshopper optimization algorithm
DL	–	Deep learning
US	–	Ultra sonic
LSTM	–	Long short-term memory
GPU	–	Graphical processing unit
GUI	–	Graphical user interface
RCNN	–	Reginal convolutional neural network
IOU	–	Intersection over union
mAP	–	Mean average precision
PID	–	Proportional integral derivative
HCI	–	Human-computer interaction
ROS	–	Robot operating system
ML	–	Machine learning
ANFIS	–	Adaptive network-based fuzzy inference system

LMI	–	Linear matrix inequalities
FD	–	Full Duplex
EE	–	Energy efficiency
ISMC	–	Integral sliding mode control
MAS	–	Multi agent system
ResNet	–	Residual Network
ROC	–	Receiver operating characteristic
ROI	–	a region of interest
R-CNN	–	Region-based CNN
DDPG	–	Deep deterministic policy gradient
TPR	–	True positive rate
IR	–	Infra red
RT3D	–	Real-time three dimensional
POFF	–	Predicted obstacle force field
HAAR	–	Hate African American recognition
PPO	–	Proximal policy optimization
TRPO	–	Trust region policy optimization
FPN	–	Feature pyramid network
OC	–	Overlap coefficient
RPN	–	Region proposal network
NMS	–	Non-max suppression
DIP	–	Digital image processing
RoI	–	Region of align
LWG	–	Light weight graph

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

Hand gesture detection is a contentious issue in computer vision and human-computer interface research. Hand gesture recognition and interpretation can be used to control devices like smartphones, tablets, and computers without the use of physical buttons or touchscreens. Furthermore, hand gesture detection can be used to give a more natural and immersive experience in virtual and augmented reality situations [1], [2]. 3D hand gestures typically employ sensors such as cameras or depth sensors to capture the position and movement of the hands in three-dimensional space and to store the depth information of the hands [3]. Microsoft Kinect or Intel RealSense cameras are used to capture the depth information of the hands. This enables the measurement of the distance between the hands and the camera, which may subsequently be used to identify particular hand gestures. This enables more natural and intuitive interaction with virtual objects and interfaces.

Recent advancements in depth sensing technology have allowed for the creation of 3D hand gesture detection systems. These devices collect 3D point clouds of a person's hand using depth cameras such as the Microsoft Kinect. These point clouds can then be processed to obtain determine the competitiveness in hand gesture recognition and interpretation [4].

The difficulty of 3D hand gesture identification, on the other hand, is not insignificant and the variety of hand motions is a significant difficulty. People can perform the same gesture in a diversity of ways, such as different fingers or hand orientations. Furthermore, hand gestures can be performed at various rates and with varying degrees of precision. These inconsistencies make it difficult for a gesture recognition system to identify a gesture effectively, especially with depth map images [5].

The use of virtual reality (VR) and augmented reality (AR) software has seen explosive growth in recent years. A crucial task in this area is the estimation of 3D key points, which entails deducing the 3D coordinates of the joints in a person's hand from a single RGB camera image. Creating AR/VR experiences that are accessible to a wide audience is a particularly appealing task because RGB cameras are low-cost and commonly found on the vast majority of mobile devices [6]. To enhance the precision of 3D hand pose estimation, however, researchers will need to work around two major roadblocks. Predicting the 3D coordinates of the hand joints from a 2D image is a complex algorithmic task. Second, there aren't enough high-quality datasets to train and test with.

Despite the challenging problem of background and shadows, researchers have developed some techniques to address the issue of variability and complex backgrounds to improve the performance of hand gesture-based systems [7].

Figure 1.1: 3D hand recognition with Kinect dataset [1]

New approaches to deep learning could make it possible to circumvent the limitations imposed by the technologies that are now in use. It is currently only possible to determine the 3D hand gestures with hand key points when performing 3D hand analysis using monocular RGB images. It is not possible to depict the three-dimensional contour of a hand in this manner [8].

Recent breakthroughs in deep learning and a drop in the price of depth sensors may be responsible for these gains in gesture estimation. Find a solution to the described problem, there are still a lot of challenges to conquer. Large differences in hand-shape viewpoint shifts, a large number of degrees of freedom (DoFs), constrained parameter space, self-similarity, and occlusions are some of the characteristics that contribute to the difficulty of making realistic hand-gesture estimation [9].

1.1 Background of the problem

Deep learning has improved 3D hand gesture detection by improving hand shape and movement understanding. CNNs can improve hand gesture classification and segmentation. Algorithms that can handle many hand configurations have challenges because people have different hand shapes and sizes. Changing perspectives can make the hand look different [10]. Hand movement has many DoFs, making gesture recognition difficult. Despite these challenges, deep learning research and sensor technology improvements will likely improve 3D hand gesture detection. As this technology improves, it could greatly improve the usability and accessibility of electronic devices, especially for those with physical disabilities.

The presence of image shadows and background clutter is another significant issue that needs to be addressed for the classification and segmentation of 3D hand gestures. In natural settings with a lot of shadows and background clutter, such as living rooms or offices, hand gesture detection systems are frequently used. Because of this, the system may produce false positives and negatives, which makes it difficult to recognize and correctly interpret hand gestures [11].

In addition, the detection of hand gestures in virtual reality and augmented reality applications can result in an experience that is both more immersive and more natural. Recent developments in depth map recognition have made it possible to create 3D hand gesture detection systems that can collect 3D point clouds of a person's hand using depth cameras. These systems can be used by themselves or in conjunction with other depth-sensing devices. However, despite recent advancements, recognizing 3D hand gestures is still a difficult problem to solve because of the wide variety of hand gestures, and also has some transformation problems due to similar types of hand gestures.

The variability of hand motions is one of the most significant challenges associated with recognizing 3D hand gestures. Because different people can execute the same gesture in a variety of different ways, it can be challenging for gesture recognition systems to correctly identify and interpret hand motions. In addition, the same hand gesture can be performed at a variety of speeds and with a wide range of degrees of precision, further adding to the difficulty of the issue [12].

Another way is to extract the hand from the background using segmentation algorithms. This can be accomplished through the use of techniques such as skin color segmentation or depth-based segmentation. Once the hand has been separated from the background, it can be utilized to extract features for gesture identification [13].

Researchers have developed a variety of solutions to address these challenges. These solutions include the training of classifiers through the use of machine learning techniques and the extraction of hands from the background through the use of segmentation algorithms. When a classifier is trained on a dataset of 3D point clouds containing hand gestures, the classifier can differentiate between various hand motions when applied to new point clouds. By using segmentation algorithms to isolate hands from their backgrounds, it is possible to reduce or eliminate background clutter and make it simpler to isolate features that can be used for gesture recognition [14].

Most researchers have developed strategies to improve the performance of hand gesture-based systems despite the challenges presented by variability and complex backgrounds. Enhancing the user experience in virtual and augmented reality applications is possible with the help of these techniques, which can help to enable more natural and immersive interactions with devices [15] . Figure 1.2 shows a simple camera vision-based sensor for hand gesture extraction and identification.

REFERENCES

[1] C. Zheng *et al.*, “Deep Learning-Based Human Pose Estimation: A Survey,” Dec. 2020, [Online]. Available: http://arxiv.org/abs/2012.13392

[2] M. Yasen and S. Jusoh, “A systematic review on hand gesture recognition techniques, challenges and applications,” *PeerJ Comput Sci*, vol. 2019, no. 9, 2019, doi: 10.7717/peerj-cs.218.

[3] P. Molchanov Xiaodong Yang Shalini Gupta Kihwan Kim Stephen Tyree Jan Kautz NVIDIA, “Online Detection and Classification of Dynamic Hand Gestures with Recurrent 3D Convolutional Neural Networks.” [Online]. Available: https://research.nvidia.com/publication/

[4] C. Miron, A. Păsărică, D. Arotăritei, H. Costin, R. G. Bozomitu, and C. Rotariu, “Hand gesture detection using a stereo camera system and simulation of movement,” in *2017 10th International Symposium on Advanced Topics in Electrical Engineering, ATEE 2017*, Apr. 2017, pp. 297–300. doi: 10.1109/ATEE.2017.7905134.

[5] X. Sun, Y. Wei, S. Liang, and J. Sun, “Cascaded Hand Pose Regression.”

[6] M. Oudah, A. Al-Naji, and J. Chahl, “Hand Gesture Recognition Based on Computer Vision: A Review of Techniques,” *Journal of Imaging*, vol. 6, no. 8. MDPI, Jul. 01, 2020. doi: 10.3390/JIMAGING6080073.

[7] L. Chen, J. Fu, Y. Wu, H. Li, and B. Zheng, “Hand gesture recognition using compact CNN via surface electromyography signals,” *Sensors (Switzerland)*, vol. 20, no. 3, Feb. 2020, doi: 10.3390/s20030672.

[8] V. K. Nguyen and R. Alba-Flores, “UAVs Control Using 3D Hand Keypoint Gestures,” *Conference Proceedings - IEEE SOUTHEASTCON*, vol. 2022-March, pp. 140–144, 2022, doi: 10.1109/SoutheastCon48659.2022.9764030.

[9] N. H. Nguyen, T. D. T. Phan, S. H. Kim, H. J. Yang, and G. S. Lee, “3D skeletal joints-based hand gesture spotting and classification,” *Applied Sciences (Switzerland)*, vol. 11, no. 10, 2021, doi: 10.3390/app11104689.

[10] T. Chatzis, A. Stergioulas, D. Konstantinidis, K. Dimitropoulos, and P. Daras, "A comprehensive study on deep learning-based 3d hand pose estimation methods," *Applied Sciences (Switzerland)*, vol. 10, no. 19, pp. 1–27, Oct. 2020, doi: 10.3390/app10196850.

[11] A. Memo and P. Zanuttigh, "Head-Mounted Gesture Controlled Interface for Human-Computer Interaction."

[12] W. J. Chang, C. H. Hsu, and L. B. Chen, "A Pose Estimation-Based Fall Detection Methodology Using Artificial Intelligence Edge Computing," *IEEE Access*, vol. 9, pp. 129965–129976, 2021, doi: 10.1109/ACCESS.2021.3113824.

[13] J. H. R. Isaac, M. Manivannan, and B. Ravindran, "Corrective Filter Based on Kinematics of Human Hand for Pose Estimation," *Front Virtual Real*, vol. 2, Jul. 2021, doi: 10.3389/frvir.2021.663618.

[14] A. Malik Mohd Ali *et al.*, "Preliminary Design of a Robotic Exoskeleton for Arm Rehabilitation," in *Journal of Physics: Conference Series*, Jul. 2018, vol. 1049, no. 1. doi: 10.1088/1742-6596/1049/1/012067.

[15] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," Sep. 2014, [Online]. Available: http://arxiv.org/abs/1409.1556

[16] C. Zimmermann and T. Brox, "Learning to Estimate 3D Hand Pose from Single RGB Images," May 2017, [Online]. Available: http://arxiv.org/abs/1705.01389

[17] T. Simon, H. Joo, I. Matthews, and Y. Sheikh, "Hand Keypoint Detection in Single Images using Multiview Bootstrapping," Apr. 2017, [Online]. Available: http://arxiv.org/abs/1704.07809

[18] T. Chatzis, A. Stergioulas, D. Konstantinidis, K. Dimitropoulos, and P. Daras, "A comprehensive study on deep learning-based 3d hand pose estimation methods," *Applied Sciences (Switzerland)*, vol. 10, no. 19, pp. 1–27, Oct. 2020, doi: 10.3390/app10196850.

[19] Z. Ren, J. Yuan, J. Meng, and Z. Zhang, "Robust part-based hand gesture recognition using kinect sensor," *IEEE Trans Multimedia*, vol. 15, no. 5, pp. 1110–1120, 2013, doi: 10.1109/TMM.2013.2246148.

[20] H. Zhang, C. Shen, Y. Li, Y. Cao, Y. Liu, and Y. Yan, "Exploiting temporal consistency for real-time video depth estimation *." [Online]. Available: https://tinyurl.com/STCLSTM

[21] R. P. Sharma and G. K. Verma, "Human Computer Interaction using Hand Gesture," in *Procedia Computer Science*, 2015, vol. 54, pp. 721–727. doi: 10.1016/j.procs.2015.06.085.

[22] A. Memo and P. Zanuttigh, "Head-mounted gesture controlled interface for human-computer interaction," *Multimed Tools Appl*, vol. 77, no. 1, pp. 27–53, Jan. 2018, doi: 10.1007/s11042-016-4223-3.

[23] H. Cheng, L. Yang, and Z. Liu, "Survey on 3D Hand Gesture Recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 9, pp. 1659–1673, Sep. 2016, doi: 10.1109/TCSVT.2015.2469551.

[24] S. S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: a survey," *Artif Intell Rev*, vol. 43, no. 1, pp. 1–54, Jan. 2015, doi: 10.1007/s10462-012-9356-9.

[25] J. Malik, A. Elhayek, F. Nunnari, and D. Stricker, "Simple and effective deep hand shape and pose regression from a single depth image," *Computers and Graphics (Pergamon)*, vol. 85, pp. 85–91, Dec. 2019, doi: 10.1016/j.cag.2019.10.002.

[26] K. Abdallah AHMAD, D. Christy SILPANI, and K. Yoshida, "Hand Gesture Recognition by Hand Landmark Classification."

[27] N. Baranwal, N. Singh, and G. C. Nandi, "Implementation of MFCC based hand gesture recognition on HOAP-2 using Webots platform," in *Proceedings of the 2014 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2014*, Nov. 2014, pp. 1897–1902. doi: 10.1109/ICACCI.2014.6968332.

[28] R. Herrera-Acuña, V. Argyriou, and S. A. Velastin, "A Kinect-based 3D hand-gesture interface for 3D databases," *Journal on Multimodal User Interfaces*, vol. 9, no. 2, pp. 121–139, Jun. 2015, doi: 10.1007/s12193-014-0173-0.

[29] S. Escalera *et al.*, "ChaLearn Looking at People Challenge 2014: Dataset and Results." [Online]. Available: https://www.codalab.org/competitions/

[30] R. Li, H. Wang, and Z. Liu, "Survey on Mapping Human Hand Motion to Robotic Hands for Teleoperation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 5, pp. 2647–2665, May 2022, doi: 10.1109/TCSVT.2021.3057992.

[31] B. Doosti, S. Naha, M. Mirbagheri, and D. Crandall, "HOPE-Net: A Graph-based Model for Hand-Object Pose Estimation," Mar. 2020, [Online]. Available: http://arxiv.org/abs/2004.00060

[32] L. Huang, J. Tan, J. Liu, and J. Yuan, "Hand-Transformer: Non-Autoregressive Structured Modeling for 3D Hand Pose Estimation."

[33] F. Mueller *et al.*, "GANerated Hands for Real-Time 3D Hand Tracking from Monocular RGB."

[34] D. N. Tai, I. S. Na, and S. H. Kim, "HSFE network and fusion model based dynamic hand gesture recognition," *KSII Transactions on Internet and Information Systems*, vol. 14, no. 9, pp. 3924–3940, 2020, doi: 10.3837/tiis.2020.09.020.

[35] J. Malik, A. Elhayek, F. Nunnari, and D. Stricker, "Simple and effective deep hand shape and pose regression from a single depth image," *Computers and Graphics (Pergamon)*, vol. 85, no. xxxx, pp. 85–91, 2019, doi: 10.1016/j.cag.2019.10.002.

[36] S. S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: a survey," *Artif Intell Rev*, vol. 43, no. 1, pp. 1–54, 2015, doi: 10.1007/s10462-012-9356-9.

[37] P. Molchanov, S. Gupta, K. Kim, and J. Kautz, "Hand Gesture Recognition with 3D Convolutional Neural Networks."

[38] F. Mueller *et al.*, "GANerated Hands for Real-Time 3D Hand Tracking from Monocular RGB," *ArXiv*, 2017.

[39] F. Mueller *et al.*, "Real-time pose and shape reconstruction of two interacting hands with a single depth camera," *ACM Trans Graph*, vol. 38, no. 4, 2019, doi: 10.1145/3306346.3322958.

[40] P. Molchanov, S. Gupta, K. Kim, and J. Kautz, "Hand Gesture Recognition with 3D Convolutional Neural Networks."

[41] A. Erol, G. Bebis, M. Nicolescu, R. D. Boyle, and X. Twombly, "Vision-based hand pose estimation: A review," *Computer Vision and Image Understanding*, vol. 108, no. 1–2, pp. 52–73, 2007, doi: 10.1016/j.cviu.2006.10.012.

[42] J. S. Supančič, G. Rogez, Y. Yang, J. Shotton, and D. Ramanan, "Depth-Based Hand Pose Estimation: Methods, Data, and Challenges," *Int J Comput Vis*, vol. 126, no. 11, pp. 1180–1198, 2018, doi: 10.1007/s11263-018-1081-7.

[43] J. S. Supančič, G. Rogez, Y. Yang, J. Shotton, and D. Ramanan, "Depth-Based Hand Pose Estimation: Methods, Data, and Challenges," *Int J Comput Vis*, vol. 126, no. 11, pp. 1180–1198, 2018, doi: 10.1007/s11263-018-1081-7.

[44] S. Yuan *et al.*, "Depth-Based 3D Hand Pose Estimation: From Current Achievements to Future Goals," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2636–2645, 2018, doi: 10.1109/CVPR.2018.00279.

[45] B. Kong, J. Supančič, D. Ramanan, and C. C. Fowlkes, "Cross-Domain Image Matching with Deep Feature Maps," *Int J Comput Vis*, vol. 127, no. 11–12, pp. 1738–1750, 2019, doi: 10.1007/s11263-018-01143-3.

[46] M. Oberweger and V. Lepetit, "DeepPrior++: Improving Fast and Accurate 3D Hand Pose Estimation," *Proceedings - 2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017*, vol. 2018-Janua, pp. 585–594, 2017, doi: 10.1109/ICCVW.2017.75.

[47] Y. Chen, Z. Tu, L. Ge, D. Zhang, R. Chen, and J. Yuan, "SO-handNet: Self-organizing network for 3D hand pose estimation with semi-supervised learning," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2019-Octob, pp. 6960–6969, 2019, doi: 10.1109/ICCV.2019.00706.

[48] H. Cheng, L. Yang, and Z. Liu, "Survey on 3D Hand Gesture Recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 9, pp. 1659–1673, 2016, doi: 10.1109/TCSVT.2015.2469551.

[49] H. Tan *et al.*, "A novel preoperative planning technique using a combination of CT angiography and three-dimensional printing for complex toe-to-hand reconstruction," *J Reconstr Microsurg*, vol. 31, no. 5, pp. 369–377, 2015, doi: 10.1055/s-0035-1546419.

[50] C. Qian, X. Sun, Y. Wei, X. Tang, and J. Sun, “Realtime and robust hand tracking from depth,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1106–1113, 2014, doi: 10.1109/CVPR.2014.145.

[51] C. Y. Fan, M. H. Lin, T. F. Su, S. H. Lai, and C. H. Yu, “3D hand skeleton model estimation from a depth image,” *Proceedings of the 14th IAPR International Conference on Machine Vision Applications, MVA 2015*, pp. 489–492, 2015, doi: 10.1109/MVA.2015.7153237.

[52] Z. A. Kahar, P. S. Sulaiman, F. Khalid, and A. Azman, “Skeleton Joints Moment (SJM): A Hand Gesture Dimensionality Reduction for Central Nervous System Interaction,” *IEEE Access*, vol. 9, pp. 146640–146652, 2021, doi: 10.1109/ACCESS.2021.3123570.

[53] A. Daffertshofer, C. J. C. Lamoth, O. G. Meijer, and P. J. Beek, “PCA in studying coordination and variability: A tutorial,” *Clinical Biomechanics*, vol. 19, no. 4, pp. 415–428, 2004, doi: 10.1016/j.clinbiomech.2004.01.005.

[54] M. Santello *et al.*, *Hand synergies: Integration of robotics and neuroscience for understanding the control of biological and artificial hands*, vol. 17. 2016. doi: 10.1016/j.plrev.2016.02.001.

[55] E. Remelli, A. Tkach, A. Tagliasacchi, and M. Pauly, “Low-Dimensionality Calibration through Local Anisotropic Scaling for Robust Hand Model Personalization,” *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2017-Octob, pp. 2554–2562, 2017, doi: 10.1109/ICCV.2017.277.

[56] P. Ren, H. Sun, Q. Qi, J. Wang, and W. Huang, “SRN: Stacked regression network for real-time 3D hand pose estimation,” *30th British Machine Vision Conference 2019, BMVC 2019*, 2020.

[57] L. Ge, H. Liang, J. Yuan, and D. Thalmann, “Robust 3D Hand Pose Estimation from Single Depth Images Using Multi-View CNNs,” *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4422–4436, 2018, doi: 10.1109/TIP.2018.2834824.

[58] R. Mirsu, G. Simion, C. D. Caleanu, and I. M. Pop-Calimanu, “A pointnet-based solution for 3D hand gesture recognition,” *Sensors (Switzerland)*, vol. 20, no. 11, pp. 1–13, 2020, doi: 10.3390/s20113226.

[59] S. Alashhab, A. J. Gallego, and M. Á. Lozano, “Efficient Gesture Recognition for the Assistance of Visually Impaired People using Multi-Head Neural Networks,” May 2022, doi: 10.1016/j.engappai.2022.105188.

[60] Q. Gao, Y. Chen, Z. Ju, and Y. Liang, “Dynamic Hand Gesture Recognition Based on 3D Hand Pose Estimation for Human-Robot Interaction,” *IEEE Sens J*, 2021, doi: 10.1109/JSEN.2021.3059685.

[61] P. Molchanov, X. Yang, S. Gupta, K. Kim, S. Tyree, and J. Kautz, “Online Detection and Classification of Dynamic Hand Gestures with Recurrent 3D Convolutional Neural Networks,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-Decem, pp. 4207–4215, 2016, doi: 10.1109/CVPR.2016.456.

[62] A. Caputo *et al.*, “SFINGE 3D: A novel benchmark for online detection and recognition of heterogeneous hand gestures from 3D fingers’ trajectories,” *Computers and Graphics (Pergamon)*, vol. 91, pp. 232–242, 2020, doi: 10.1016/j.cag.2020.07.014.

[63] Y. Hasson *et al.*, “Learning joint reconstruction of hands and manipulated objects,” *ArXiv*, pp. 11807–11816, 2019.

[64] F. S. Khan, M. N. H. Mohd, D. M. Soomro, S. Bagchi, and M. D. Khan, “3D Hand Gestures Segmentation and Optimized Classification Using Deep Learning,” *IEEE Access*, vol. 9, pp. 131614–131624, 2021, doi: 10.1109/ACCESS.2021.3114871.

[65] C. Qian, X. Sun, Y. Wei, X. Tang, and J. Sun, “Realtime and robust hand tracking from depth,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Sep. 2014, pp. 1106–1113. doi: 10.1109/CVPR.2014.145.

[66] R. P. Sharma and G. K. Verma, “Human Computer Interaction using Hand Gesture,” *Procedia Comput Sci*, vol. 54, pp. 721–727, 2015, doi: 10.1016/j.procs.2015.06.085.

[67] A. Memo and P. Zanuttigh, “Head-mounted gesture controlled interface for human-computer interaction,” *Multimed Tools Appl*, vol. 77, no. 1, pp. 27–53, 2018, doi: 10.1007/s11042-016-4223-3.

[68] L. Zhang, G. Zhu, L. Mei, P. Shen, S. A. A. Shah, and M. Bennamoun, "Attention in convolutional LSTM for gesture recognition," *Adv Neural Inf Process Syst*, vol. 2018-Decem, no. NeurIPS, pp. 1953–1962, 2018.

[69] F. Mueller, D. Mehta, O. Sotnychenko, S. Sridhar, D. Casas, and C. Theobalt, "Real-Time Hand Tracking under Occlusion from an Egocentric RGB-D Sensor," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2017-Octob, pp. 1163–1172, 2017, doi: 10.1109/ICCV.2017.131.

[70] X. Ma and J. Peng, "Kinect sensor-based long-distance hand gesture recognition and fingertip detection with depth information," *J Sens*, vol. 2018, 2018, doi: 10.1155/2018/5809769.

[71] J. Zhang, J. Jiao, M. Chen, L. Qu, X. Xu, and Q. Yang, "3D Hand Pose Tracking and Estimation Using Stereo Matching," Oct. 2016, [Online]. Available: http://arxiv.org/abs/1610.07214

[72] H. Sahbi, "Skeleton-based Hand-Gesture Recognition with Lightweight Graph Convolutional Networks," Apr. 2021, [Online]. Available: http://arxiv.org/abs/2104.04255

[73] C. Y. Fan, M. H. Lin, T. F. Su, S. H. Lai, and C. H. Yu, "3D hand skeleton model estimation from a depth image," in *Proceedings of the 14th IAPR International Conference on Machine Vision Applications, MVA 2015*, Jul. 2015, pp. 489–492. doi: 10.1109/MVA.2015.7153237.

[74] Y. Hasson *et al.*, "Learning joint reconstruction of hands and manipulated objects," Apr. 2019, [Online]. Available: http://arxiv.org/abs/1904.05767

[75] J. Taylor *et al.*, "Efficient and precise interactive hand tracking through joint, continuous optimization of pose and correspondences," in *ACM Transactions on Graphics*, Jul. 2016, vol. 35, no. 4. doi: 10.1145/2897824.2925965.

[76] J. Li *et al.*, "Attentive 3D-Ghost Module for Dynamic Hand Gesture Recognition with Positive Knowledge Transfer," *Comput Intell Neurosci*, vol. 2021, 2021, doi: 10.1155/2021/5044916.

[77] D. Rukhovich, A. Vorontsova, and A. Konushin, "FCAF3D: Fully Convolutional Anchor-Free 3D Object Detection." [Online]. Available: https://github.com/samsunglabs/fcaf3d.

[78] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," Dec. 2018, [Online]. Available: http://arxiv.org/abs/1812.08008

[79] Q. Gao, Y. Chen, Z. Ju, S. Member, and Y. Liang, "Dynamic Hand Gesture Recognition Based on 3D Hand Pose Estimation for Human-Robot Interaction."

[80] P. Shenoy, V. Sompur, and S. K. M. Varadhan, "Methods for Measurement and Analysis of Full Hand Angular Kinematics Using Electromagnetic Tracking Sensors," *IEEE Access*, vol. 10, pp. 42673–42689, 2022, doi: 10.1109/ACCESS.2022.3168674.

[81] D. N. Tai, I. S. Na, and S. H. Kim, "HSFE network and fusion model based dynamic hand gesture recognition," *KSII Transactions on Internet and Information Systems*, vol. 14, no. 9, pp. 3924–3940, Sep. 2020, doi: 10.3837/tiis.2020.09.020.

[82] G. K. O, "Text classification 방법을 사용한 행동 인식 , 손동작 인식 및 감정 인식 Action recognition , hand gesture recognition , and emotion recognition using text classification method," pp. 213–216, 2021.

[83] Y. Cai, M. Landis, D. T. Laidley, A. Kornecki, A. Lum, and S. Li, "Multi-modal vertebrae recognition using Transformed Deep Convolution Network," *Computerized Medical Imaging and Graphics*, vol. 51, pp. 11–19, Jul. 2016, doi: 10.1016/j.compmedimag.2016.02.002.

[84] E. Ebeid, M. Skriver, and J. Jin, "A Survey on Open-Source Flight Control Platforms of Unmanned Aerial Vehicle," *Proceedings - 20th Euromicro Conference on Digital System Design, DSD 2017*, pp. 396–402, 2017, doi: 10.1109/DSD.2017.30.

[85] A. Kratimenos, G. Pavlakos, and P. Maragos, "Independent sign language recognition with 3d body, hands, and face reconstruction," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. 2021-June, pp. 4270–4274, 2021, doi: 10.1109/ICASSP39728.2021.9414278.

[86] Q. Xu, G. Qin, M. Sun, J. Yan, H. Jiang, and Z. Zhang, "Feature fusion based hand gesture recognition method for automotive interfaces," *Chinese Journal of Electronics*, vol. 29, no. 6, pp. 1153–1164, 2020, doi: 10.1049/cje.2020.06.008.

[87] T. Chatzis, A. Stergioulas, D. Konstantinidis, K. Dimitropoulos, and P. Daras, "A comprehensive study on deep learning-based 3d hand pose estimation methods," *Applied Sciences (Switzerland)*, vol. 10, no. 19, pp. 1–27, 2020, doi: 10.3390/app10196850.

[88] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Trans Pattern Anal Mach Intell*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: 10.1109/TPAMI.2016.2577031.

[89] X. Yu, "Hand Gesture Recognition Based on Faster-RCNN Deep Learning," *J Comput (Taipei)*, vol. 14, no. 2, pp. 101–110, 2019, doi: 10.17706/jcp.14.2.101-110.

[90] O. Mazhar, B. Navarro, S. Ramdani, R. Passama, and A. Cherubini, *A real-time human-robot interaction framework with robust background invariant hand gesture detection*, vol. 60. 2019. doi: 10.1016/j.rcim.2019.05.008.

[91] A. P. YUNUS, N. C. SHIRAI, K. MORITA, and T. WAKABAYASHI, "Time Series Human Motion Prediction Using RGB Camera and OpenPose," *International Symposium on Affective Science and Engineering*, vol. ISASE2020, no. 0, pp. 1–4, 2020, doi: 10.5057/isase.2020-c000037.

[92] N. Aloysius and M. Geetha, "A review on deep convolutional neural networks," in *Proceedings of the 2017 IEEE International Conference on Communication and Signal Processing, ICCSP 2017*, Feb. 2018, vol. 2018-January, pp. 588–592. doi: 10.1109/ICCSP.2017.8286426.

[93] L. Jia, X. Zhou, H. Qin, R. Bai, L. Wang, and C. Xue, "Research on discrete semantics in continuous hand joint movement based on perception and expression," *Sensors*, vol. 21, no. 11, Jun. 2021, doi: 10.3390/s21113735.

[94] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," Dec. 2014, [Online]. Available: http://arxiv.org/abs/1412.6980

[95] L. Liu and L. Shao, "Learning Discriminative Representations from RGB-D Video Data."

[96] K. Rahul and A. K. Tiwari, "FQI: Feature-based reduced-reference image quality assessment method for screen content images," *IET Image Process*, vol. 13, no. 7, pp. 1170–1180, 2019, doi: 10.1049/iet-ipr.2018.5496.

[97] Y. Yuan, S.-E. Wei, T. Simon, K. Kitani, and J. Saragih, "SimPoE: Simulated Character Control for 3D Human Pose Estimation." [Online]. Available: https://www.ye-yuan.com/simpoe

[98] F. Mueller, D. Mehta, O. Sotnychenko, S. Sridhar, D. Casas, and C. Theobalt, "Real-time Hand Tracking under Occlusion from an Egocentric RGB-D Sensor," Apr. 2017, doi: 10.1109/ICCV.2017.131.

[99] J. J. Bird, "Statistical and Spatio-temporal Hand Gesture Features for Sign Language Recognition using the Leap Motion Sensor," Feb. 2022, [Online]. Available: http://arxiv.org/abs/2202.11005

[100] O. Chernytska, "3D Hand Pose Estimation from Single RGB Camera."

[101] J. S. Supancic, G. Rogez, Y. Yang, J. Shotton, and D. Ramanan, "Depth-based hand pose estimation: methods, data, and challenges," Apr. 2015, [Online]. Available: http://arxiv.org/abs/1504.06378

[102] M. A. Rahim, M. R. Islam, and J. Shin, "Non-touch sign word recognition based on dynamic hand gesture using hybrid segmentation and CNN feature fusion," *Applied Sciences (Switzerland)*, vol. 9, no. 18, Sep. 2019, doi: 10.3390/app9183790.

[103] J. Suarez and R. R. Murphy, "Hand gesture recognition with depth images: A review," in *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, 2012, pp. 411–417. doi: 10.1109/ROMAN.2012.6343787.

[104] Y. Fang, K. Wang, J. Cheng, and H. Lu, "A REAL-TIME HAND GESTURE RECOGNITION METHOD."

[105] M. Rezaei, R. Rastgoo, and V. Athitsos, "TriHorn-Net: A Model for Accurate Depth-Based 3D Hand Pose Estimation," Jun. 2022, [Online]. Available: http://arxiv.org/abs/2206.07117

[106] L. Ge, Y. Cai, J. Weng, and J. Yuan, "Hand PointNet: 3D Hand Pose Estimation using Point Sets."

[107] J. Supancic *et al.*, "Depth-based hand pose estimation: data, methods, and challenges," pp. 1868–1876, 2015, doi: 10.1109/ICCV.2015.217ï.

[108] R. Mirsu, G. Simion, C. D. Caleanu, and I. M. Pop-Calimanu, "A pointnet-based solution for 3D hand gesture recognition," *Sensors (Switzerland)*, vol. 20, no. 11, pp. 1–13, Jun. 2020, doi: 10.3390/s20113226.

[109] D. Wu *et al.*, "Deep Dynamic Neural Networks for Multimodal Gesture Segmentation and Recognition," *IEEE Trans Pattern Anal Mach Intell*, vol. 38, no. 8, pp. 1583–1597, Aug. 2016, doi: 10.1109/TPAMI.2016.2537340.

[110] E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," May 2016, [Online]. Available: http://arxiv.org/abs/1605.06211

[111] Y. Luo, G. Cui, and D. Li, "An Improved Gesture Segmentation Method for Gesture Recognition Based on CNN and YCbCr," *Journal of Electrical and Computer Engineering*, vol. 2021, 2021, doi: 10.1155/2021/1783246.

[112] H. Chen, X. Liu, J. Shi, and G. Zhao, "Temporal Hierarchical Dictionary Guided Decoding for Online Gesture Segmentation and Recognition," *IEEE Transactions on Image Processing*, vol. 29, pp. 9689–9702, 2020, doi: 10.1109/TIP.2020.3028962.

[113] J. Malik *et al.*, "DeepHPS: End-to-end Estimation of 3D Hand Pose and Shape by Learning from Synthetic Depth," Aug. 2018, [Online]. Available: http://arxiv.org/abs/1808.09208

[114] N. Ali, K. B. Bajwa, R. Sablatnig, and Z. Mehmood, "Image retrieval by addition of spatial information based on histograms of triangular regions," *Computers and Electrical Engineering*, vol. 54, pp. 539–550, Aug. 2016, doi: 10.1016/j.compeleceng.2016.04.002.

[115] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," Nov. 2013, [Online]. Available: http://arxiv.org/abs/1311.2524

[116] G. Benitez-Garcia *et al.*, "Improving real-time hand gesture recognition with semantic segmentation," *Sensors (Switzerland)*, vol. 21, no. 2, pp. 1–16, Jan. 2021, doi: 10.3390/s21020356.

[117] S. Yang, J. Liu, S. Lu, M. Hwa Er, and A. C. Kot, "Collaborative Learning of Gesture Recognition and 3D Hand Pose Estimation with Multi-Order Feature Analysis."

[118] P. O. Pinheiro, R. Collobert, and P. Dollar, "Learning to Segment Object Candidates," Jun. 2015, [Online]. Available: http://arxiv.org/abs/1506.06204

[119] X. Wang, R. Zhang, T. Kong, L. Li, and C. Shen, "SOLOv2: Dynamic and Fast Instance Segmentation," Mar. 2020, [Online]. Available: http://arxiv.org/abs/2003.10152

[120] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation." [Online]. Available: http://arxiv.

[121] A. G. Neve, G. M. Kakandikar, and O. Kulkarni, "Application of Grasshopper Optimization Algorithm for Constrained and Unconstrained Test Functions," *International Journal of Swarm Intelligence and Evolutionary Computation*, vol. 06, no. 03, 2017, doi: 10.4172/2090-4908.1000165.

[122] K. A. Ahmad, D. C. Silpani, and K. Yoshida, "Hand Gesture Recognition by Hand Landmark Classification," pp. 2–4, 2022.

[123] B. Hekimoğlu and S. Ekinci, "Grasshopper optimization algorithm for automatic voltage regulator system," *2018 5th International Conference on Electrical and Electronics Engineering, ICEEE 2018*, pp. 152–156, 2018, doi: 10.1109/ICEEE2.2018.8391320.

[124] F. S. Problem, "Improved Binary Grasshopper Optimization Algorithm for," pp. 1–18, 2022.

[125] S. Saremi, S. Mirjalili, and A. Lewis, "Grasshopper Optimisation Algorithm: Theory and application," *Advances in Engineering Software*, vol. 105, pp. 30–47, 2017, doi: 10.1016/j.advengsoft.2017.01.004.

[126] S. Z. Mirjalili, S. Mirjalili, S. Saremi, H. Faris, and I. Aljarah, "Grasshopper optimization algorithm for multi-objective optimization problems," *Applied Intelligence*, vol. 48, no. 4, pp. 805–820, 2018, doi: 10.1007/s10489-017-1019-8.

[127] H. Zhou *et al.*, "An Improved Grasshopper Optimizer for Global Tasks," *Complexity*, vol. 2020. Hindawi Limited, 2020. doi: 10.1155/2020/4873501.

[128] P. Qin, H. Hu, and Z. Yang, "The improved grasshopper optimization algorithm and its applications," *Sci Rep*, vol. 11, no. 1, Dec. 2021, doi: 10.1038/s41598-021-03049-6.

[129] M. Mafarja *et al.*, "Evolutionary Population Dynamics and Grasshopper Optimization approaches for feature selection problems," *Knowl Based Syst*, vol. 145, pp. 25–45, Apr. 2018, doi: 10.1016/j.knosys.2017.12.037.

[130] I. Aljarah *et al.*, "Simultaneous feature selection and Support Vector Machine optimisation using the Grasshopper Optimisation Algorithm."

[131] B. Alatas and U. Can, "Performance of Grasshopper Optimization Algorithm and other swarm based methods on benchmark functions." [Online]. Available: https://www.researchgate.net/publication/330702496

[132] H. T. Ibrahim, W. J. Mazher, O. N. Ucan, and O. Bayat, "A grasshopper optimizer approach for feature selection and optimizing SVM parameters utilizing real biomedical data sets," *Neural Comput Appl*, vol. 31, no. 10, pp. 5965–5974, Oct. 2019, doi: 10.1007/s00521-018-3414-4.

[133] A. G. Neve, G. M. Kakandikar, and O. Kulkarni, "Application of Grasshopper Optimization Algorithm for Constrained and Unconstrained Test Functions," *International Journal of Swarm Intelligence and Evolutionary Computation*, vol. 06, no. 03, 2017, doi: 10.4172/2090-4908.1000165.

[134] G. L. Wang, S. C. Chu, A. Q. Tian, T. Liu, and J. S. Pan, "Improved Binary Grasshopper Optimization Algorithm for Feature Selection Problem," *Entropy*, vol. 24, no. 6, Jun. 2022, doi: 10.3390/e24060777.

[135] S. Z. Mirjalili, S. Mirjalili, S. Saremi, H. Faris, and I. Aljarah, "Grasshopper optimization algorithm for multi-objective optimization problems," *Applied Intelligence*, vol. 48, no. 4, pp. 805–820, Apr. 2018, doi: 10.1007/s10489-017-1019-8.

[136] S. Saremi, S. Mirjalili, and A. Lewis, "Grasshopper Optimisation Algorithm: Theory and application," *Advances in Engineering Software*, vol. 105, pp. 30–47, Mar. 2017, doi: 10.1016/j.advengsoft.2017.01.004.

[137] R. Guha, "Benchmarking gradient based Optimizers' sensitivity to Learning Rate," 2023, doi: 10.20944/preprints202301.0118.v1.

[138] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun ACM*, vol. 60, no. 6, pp. 84–90, Jun. 2017, doi: 10.1145/3065386.

[139] Q. Feng, X. Xu, and Z. Wang, "Deep learning-based small object detection: A survey," *Mathematical Biosciences and Engineering*, vol. 20, no. 4, pp. 6551–6590, 2023, doi: 10.3934/mbe.2023282.

[140] Z.-Q. Zhao, P. Zheng, S. Xu, and X. Wu, "Object Detection with Deep Learning: A Review," Jul. 2018, [Online]. Available: http://arxiv.org/abs/1807.05511

[141] G. Chandrarathne, K. Thanikasalam, and A. Pinidiyaarachchi, "A Comprehensive Study on Deep Image Classification with Small Datasets," in *Lecture Notes in Electrical Engineering*, 2020, vol. 619, pp. 93–106. doi: 10.1007/978-981-15-1289-6_9.

[142] N. Wojke, A. Bewley, and D. Paulus, "Simple Online and Realtime Tracking with a Deep Association Metric," Mar. 2017, [Online]. Available: http://arxiv.org/abs/1703.07402

[143] M. A. Khan *et al.*, "Gastrointestinal diseases segmentation and classification based on duo-deep architectures," *Pattern Recognit Lett*, vol. 131, pp. 193–204, Mar. 2020, doi: 10.1016/j.patrec.2019.12.024.

[144] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *IEEE Trans Pattern Anal Mach Intell*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017, doi: 10.1109/TPAMI.2016.2644615.

[145] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition." [Online]. Available: http://image-net.org/challenges/LSVRC/2015/

[146] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition." [Online]. Available: http://image-net.org/challenges/LSVRC/2015/

[147] J. Xu, Y. Pan, X. Pan, S. Hoi, Z. Yi, and Z. Xu, "RegNet: Self-Regulated Network for Image Classification," Jan. 2021, [Online]. Available: http://arxiv.org/abs/2101.00590

[148] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object Detection via Region-based Fully Convolutional Networks." [Online]. Available: https://github.com/daijifeng001/r-fcn.

[149] H. Sahbi, "Skeleton-based Hand-Gesture Recognition with Lightweight Graph Convolutional Networks," pp. 1–10, 2021.

[150] J. Li *et al.*, “Multi-human parsing in the wild,” *ArXiv*, pp. 1–19, 2017.

[151] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” Mar. 2017, [Online]. Available: http://arxiv.org/abs/1703.06870

[152] R. Girshick, “Fast R-CNN.” [Online]. Available: https://github.com/rbgirshick/

[153] C. Miron, A. Păsărică, D. Arotăritei, H. Costin, R. G. Bozomitu, and C. Rotariu, “Hand gesture detection using a stereo camera system and simulation of movement,” in *2017 10th International Symposium on Advanced Topics in Electrical Engineering, ATEE 2017*, Apr. 2017, pp. 297–300. doi: 10.1109/ATEE.2017.7905134.

[154] Y.-W. Shang and Y.-H. Qiu, “A Note on the Extended Rosenbrock Function.” [Online]. Available: http://direct.mit.edu/evco/article-pdf/14/1/119/1493646/evco.2006.14.1.119.pdf

[155] F. Kang, J. Li, and Z. Ma, “Rosenbrock artificial bee colony algorithm for accurate global optimization of numerical functions.”

[156] S. Alashhab, A. J. Gallego, and M. Á. Lozano, “Efficient Gesture Recognition for the Assistance of Visually Impaired People using Multi-Head Neural Networks,” 2022.

[157] F. Zhang *et al.*, “MediaPipe Hands: On-device Real-time Hand Tracking,” 2020.

[158] P. Molchanov, S. Gupta, K. Kim, and J. Kautz, “Hand gesture recognition with 3D convolutional neural networks,” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, vol. 2015-Octob, pp. 1–7, 2015, doi: 10.1109/CVPRW.2015.7301342.

[159] N. Nishida and H. Nakayama, “Multimodal gesture recognition using multi-stream recurrent neural network,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9431, pp. 682–694, 2016, doi: 10.1007/978-3-319-29451-3_54.

[160] M. Zollhöfer, “Commodity RGB-D Sensors: Data Acquisition,” *Advances in Computer Vision and Pattern Recognition*, pp. 3–13, 2019, doi: 10.1007/978-3-030-28603-3_1.

[161] Y. Meraihi, A. B. Gabis, S. Mirjalili, and A. Ramdane-Cherif, "Grasshopper optimization algorithm: Theory, variants, and applications," *IEEE Access*, vol. 9, pp. 50001–50024, 2021, doi: 10.1109/ACCESS.2021.3067597.

[162] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN : Towards Real-Time Object Detection with Region Proposal Networks," vol. 39, no. 6, pp. 1137–1149, 2017.

[163] M. R-cnn, P. Doll, and R. Girshick, "Mask R-CNN".

[164] V. J. Hodge, R. Hawkins, and R. Alexander, "Deep reinforcement learning for drone navigation using sensor data," *Neural Comput Appl*, vol. 0123456789, 2020, doi: 10.1007/s00521-020-05097-x.

[165] C. C. Tsai, C. C. Kuo, and Y. L. Chen, "3D Hand Gesture Recognition for Drone Control in Unity," *IEEE International Conference on Automation Science and Engineering*, vol. 2020-Augus, pp. 985–988, 2020, doi: 10.1109/CASE48305.2020.9216807.

[166] W. Koch, R. Mancuso, R. West, and A. Bestavros, "Reinforcement learning for UAV attitude control," *ACM Transactions on Cyber-Physical Systems*, vol. 3, no. 2, 2019, doi: 10.1145/3301273.

[167] N. Mohamed, M. B. Mustafa, and N. Jomhari, "A Review of the Hand Gesture Recognition System: Current Progress and Future Directions," *IEEE Access*, vol. 9. Institute of Electrical and Electronics Engineers Inc., pp. 157422–157436, 2021. doi: 10.1109/ACCESS.2021.3129650.

[168] A. Mujahid *et al.*, "Real-time hand gesture recognition based on deep learning YOLOv3 model," *Applied Sciences (Switzerland)*, vol. 11, no. 9, May 2021, doi: 10.3390/app11094164.

[169] J. J. Bird, "Statistical and Spatio-temporal Hand Gesture Features for Sign Language Recognition using the Leap Motion Sensor," pp. 1–13, 2022.

[170] L. Ge, Y. Cai, J. Weng, and J. Yuan, "Hand PointNet: 3D Hand Pose Estimation Using Point Sets," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 8417–8426, 2018, doi: 10.1109/CVPR.2018.00878.

[171] N. Plath, M. Toussaint, and S. Nakajima, “Multi-class image segmentation using conditional random fields and global classication,” *ACM International Conference Proceeding Series*, vol. 382, no. May, 2009, doi: 10.1145/1553374.1553479.

[172] J. Malik *et al.*, “DeepHPS: End-to-end estimation of 3D hand pose and shape by learning from synthetic depth,” *Proceedings - 2018 International Conference on 3D Vision, 3DV 2018*, pp. 110–119, 2018, doi: 10.1109/3DV.2018.00023.

[173] N. Dhanachandra, K. Manglem, and Y. J. Chanu, “Image Segmentation Using K-means Clustering Algorithm and Subtractive Clustering Algorithm,” *Procedia Comput Sci*, vol. 54, pp. 764–771, 2015, doi: 10.1016/j.procs.2015.06.090.

APPENDIX A

LIST OF PUBLICATIONS

Conference:

1. Khan, Fawad Salam, Mohd Norzali Haji Mohd, Raja Masood Larik, Muhammad Danial Khan, Muhammad Inam Abbasi, and Susama Bagchi. "A Smart Flight Controller based on Reinforcement Learning for Unmanned Aerial Vehicle (UAV)." In *2021 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, pp. 203-208. IEEE, 2021.

Journals:

1. Khan, Fawad Salam, Mohd Norzali Haji Mohd, Saiful Azrin BM Zulkifli, Ghulam E Mustafa, Suhail Kazi Abro and Dur Muhammad Soomro "Deep Reinforcement Learning Based Unmanned Aerial Vehicle (UAV) Control Using 3D Hand Gestures" *CMC-Computers, Material & Continua* 72(2022): 5741-5759 **(IF 3.77)**
2. Khan, Fawad Salam, Mohd Norzali Haji Mohd, Dur Muhammad Soomro, Susama Bagchi, and M. Danial Khan. "3D Hand Gestures Segmentation and Optimized Classification Using Deep Learning." *IEEE Access* 9(2021):131614-131624 **(IF3.36)**

APPENDIX B

VITA

FAWAD SALAM KHAN received a Bachelor's degree (B.S.) in Computer Engineering from the Sir Syed University of Engineering and Technology, Karachi Pakistan in 2003 and a Masters of Engineering (ME) in Computer Systems from the NED University of Engineering and Technology, Karachi, Pakistan in 2011. He also completed a second master's in Computer Science from International Islamic University, Islamabad Pakistan in 2020. He is currently pursuing a Ph.D. in the Faculty of Electrical and Electronic Engineering (FKEE) at UTHM. His research interests in image classification, optimization techniques, reinforcement learning, pattern recognition, computer vision, machine learning, and deep learning. He received various gold and silver medals for different AI projects in Malaysia. He is an IEEE graduate student member and professional member of PEC Pakistan. He has also done various IT training in Singapore and Germany. He has more than 18 years of experience in industry and academia. Currently affiliated with Pakistan's top AI Company CONVSYS (Pvt) Ltd. as the Director and CEO heading different AI, Machine, and Deep Learning Projects in Silicon Valley (USA), Malaysia, and the UK. He is also teaching various computer engineering and computer sciences subjects in different universities in Pakistan.