Global Journal of Computer and Engineering Technology (GJCET)

Vol. 2, No. 2, July, 2022 ISSN: 2767-1933

MACHINE AND DEEP LEARNING TECHNIQUES WITH HUMAN GESTURE RECOGNITION AND CLASSIFICATION FOR HUMAN-COMPUTER INTERACTION: A REVIEW

Ashutosh Mohite, AKS University, India (amohite306@gmail.com)
Akhilesh A. Waoo, AKS UniversityIndia (akhileshwaoo@gmail.com)
Akhilesh Kumar Shrivas, Guru GhasidasVishwavidyalaya, India (proffshrivas@gmail.com)

ABSTRACT

The primary goal of computers is to incorporate Human-Computer Interaction (HCI) into such a system that makes interacting with computers as instinctual as interacting with people. Incorporating gestures in to the communication methods is an important area of study in HCI. Human-Computer Interaction (HCI) relies heavily on gesture recognition and classification process. Deep learning is now widely used for a variety of image processing tasks after being recognized as an effective feature tool for resolving nonlinear problems. Deep learning has been introduced to recognize and classify images and has demonstrated good performance as a result of those successful applications. It is the domain of Machine and Deep Learning that is expanding Deep Neural Network (DNN). Among several DNN architectures, Convolutional Neural Networks (CNN) constitute the most widely used image analysis and categorization techniques. This survey study offers a thorough analysis of human gesture-based deep learning-based image identification and classification, and compares several approaches.

Keywords: Human-Computer Interaction, Deep Learning, Convolutional Neural Networks, Machine Learning, Deep Neural Networks, Computer Vision.

1. INTRODUCTION

Sign language, also known as gesture, is a unique form of communication that is frequently understudied. Because of the diverse range of potential uses, recognition of hand gestures is among a most significant human-computer research activities in computational intelligence (such as human movement recognition, face prediction, and gesture recognition). A variety of computer vision products, including human-computer interface, sign language recognition, palm action analysis, driver hand behavior monitoring, and virtual reality, require robust hand gesture recognition and identification in cluttered environments (Alnaim, N. et al., 2019). Noncontact gesture recognition is becoming increasingly important in interaction between humans and computers (HCI) applications as a result of computer technology and artificial intelligence's rapid advancement.

Deep learning algorithms are influenced by patterns of knowledge processing and communication found throughout biological nerve systems, such as neural networks with multiple hidden layers. They expertise at computer vision (CV) and can discover the features of the instructional unit quickly and precisely under complex situations (Barros, P. et al., 2017). The generalization theorem or probabilistic inference is commonly used to interpret deep neural networks (Ploue, G. et al., 2016). Deep learning has advanced rapidly over the last few years. In the areas of language processing, visual classification, and other areas, CNN has shown surprisingly effective results. Static and dynamic expressions are the two categories into which previous research has separated the study of hand gestures. While static gestures just need one image to deliver a meaningful message, dynamic gestures require a series of frames to carry out a single gesture. This research aims to address the problem of recognizing and distinguishing static hand gestures, in which the bare hand acquires different postures to express different meanings (Chung, H. et al., 2019). However, this challenge is exceedingly complicated, and retrieving the hand form is challenging due to the enormous variety of hand combinations and fluctuations in the image sensor's perspective. Additionally, detecting static hand movements is beneficial in a range of applications, such as sign language recognition with deaf and speech-impaired persons (Chen, et al., 2016), hand gesture instructions and operator hand tracking decrease interference (Das, et al., 2015), an alternative communication method for human-machine interaction (Dardas, et al.

2011), interaction with in-air publishing (Deng, et al., 2014), Communication between hands and objects in both virtual and augmented reality settings (Huang, et al., 2016), and several further uses. We have attempted to address the broad discussion of the various learning approaches used for gesture categorization and picture identification in this work as well as a review of other significant publications in the field.

2. RELATED WORKS

Using a number of methods, picture categorization and recognition may be implemented using machine learning or deep learning. Several research papers used various combinations of pre - processing and feature extraction techniques, like (Chang-Yi Kao, et al., 2011) presented a classifier-based Hidden Markov Model (HMM) technique for the way of hand motion is used to recognize hand gestures. They created eight different hand gestures either with one or even both hands. The face is initially located in their architecture, and then the palm of the hand is located from the skin region utilizing maximum circle plate mapping, and orientation is used as an important characteristic during feature extraction. The HMM recognition model achieved 96 % in this test.

Xie, B. et al. (2018) suggested an RGB-D static recognition of sign language approach based on fine-tuning Inception V3, that may also eliminate the processes of gesture categorization and feature extraction that were previously used in earlier systems. In prior approaches, the processes of gesture categorization and feature extraction were removed by Inception V3. In contrast to standard CNN algorithms, the authors have applied a two-stage training technique to fine-tune the system. The feature add layer of depth and RGB pictures is inserted into the CNN architecture using this manner. The proposed approach has the greatest accuracy 91.35%.

A CNN approach for identifying static gestures was proposed by Han et al. (2016). To begin, the researcher obtained 12,000 photos of 10 different movements. During the picture preprocessing step, the Gaussian skin model and background removal are employed to provide CNN training and test data. The experiment created a basic six-layer CNN with a classification rate of 93.8% (convolution layer, average pooling layer, downsampling layer, Dropout layer, complete connection layer, and last Softmax layer). The 10 categories of gesture data in this article, in contrast hand, are easy to recognize based on hand form and have a basic background. As a result, in order to achieve greater accuracy, the author may only utilize a basic CNN to categorize the RGB pictures.

Krizhevsky, A et al. (2012) trained a massive, deep convolutional neural network to classify the ImageNet LSVRC-2010 competition's 1.2 million high-resolution picture data into 1000 unique classes. On test data, they achieved top-1 and top-5 error rates of 37.5 % and 17.0 %, respectively, which is much better than the prior state-of-the-art. The neural network is made consisting of five convolutional layers, of which some have been preceded by max-pooling layers, then three fully-connected layers, the final of which is a 1000-way softmax. To accelerate training, we employed non-saturating nodes and a very efficient Provide evidence formulation of the CNN model. They used a recently developed regularization method called "dropout" to reduce fitting problem in the fully-connected layers, which proved to be very effective.

Pigou, L. et al. (2012) demonstrate an identification model based on the Microsoft Kinect, CNNs, and GPU momentum. Rather of developing intricate handmade features, CNNs may automate feature development. We can accurately detect 20 Italian gestures. The prediction model can generalize on individuals and situations that did not exist during training, with a cross-validation accuracy of 91.7 %. This model achieves a mean Weighted Score of 0.789 in the ChaLearn 2014 Focusing at Users gesture detection competition. The testing set's accuracy is 95.68%, with a rate of false positives of 4.13% due to noise motions. Because the validation set excludes the users and backdrops seen in the training set, the test result is larger. A team of researchers have shown that convolutional neural networks can distinguish distinct signals of a sign language successfully even when individuals and their surroundings are not included in the training set. Because the validation set lacks the users and backdrops seen in the trainings, the test result is higher.

Chen, Y., et al. (2016) discussed the features are extracted from Hyper Spectral Images (HSI) using the deep learning concept of Convolutional Neural Network (CNN). It employs a various pooling layer in CNN for feature extraction (nonlinear, invariant) from the HIS, which is useful for perfect image classification and target detection. It also identifies the broader issues with the HSI image features. To extract deep features from HSIs, the proposed technique employs several convolutional and pooling layers. These are useful for image classification and target identification.

3. DEEP LEARNING

Deep learning (DL) is a well-known research technique in fields like computer vision, audio synthesis, and natural language processing. It is a field that relies on self-improvement and learning via the study of computer algorithms. In contrast to machine learning, deep learning frequently uses artificial neural networks that are designed to imitate how humans think and learn. Deep learning is useful for speech recognition, language translation, and image categorization. Without the help of a person, it can solve any pattern recognition issue. This approach works best in situations when a lot of data has to be evaluated and human-like knowledge is needed.

4. DEEP LEARNING BASED RECOGNITION

Artificial intelligence, because of the learning role concept, provides a good and efficient method that is employed in a broad range of current applications. (Kaur, et al., 2016) describes deep learning makes precise predictions by using several layers to understand the data. (Kao,et al., 2011) recommended seven well-known hand movements that a mobile camera might record, yielding 24,698 picture frames. For hand classification, feature extraction and an adjusted deep convolutional neural network (ADCNN) were used. The experiment evaluates the training data at 100% as well as the test dataset at 99% over the course of 15,598 seconds. Other devices that were put out tracked the hand using a camera. After that, the backdrop was eliminated using morphology and the skin colour (Y-Cb-Cr) approach. ROI was also monitored using kernel correlation filters (KCF). The final image is fed into a deep convolutional neural network (CNN). The CNN model was used to compare the performance of two modified Alex Net and VGG Net models. The accuracy rate for training and testing data in (Krizhevsky, A. et al., 2012) was 99.90% and 95.61%, respectively. A novel deep convolutional neural network-based process for classifying hand gestures in which the resized image is directly consumed into the network after passing through the segmentation and detection stages. The system in (Le et al., 2016) runs in real time and gets results with a simple background of 97.1% and a complex background of 85.3%. The Kinect sensor's image was used to segment colour pictures, which were then combined using convolution neural networks. The neural network's threshold and weights were then modified using an error back propagation approach. The SVM classification method was incorporated into the network to enhance the outcomes in (Li, et al., 2019). Another research trained the CNN to detect seven hand motions with an average identification rate of 95.96% by first filtering away non-skin colours in a picture using the Gaussian Mixture Model (GMM) (Lin, et al., 2014). The proposed system employs a long-term recurrent convolutional network-based activity classification model that is fed several frames from the recorded video sequence. The exemplary frames are extracted using a de-convolutional neural network based on semantic segmentation. In (Lu, et al., 2017), they train the de-convolutional network using tiled picture patterns and tiled parameters. (Mohammed et al., 2019) present a dual convolutional neural network (DC-CNN) in which the original image is preprocessed to detect the hand's edge before being fed into the network. Each of the two-channel CNNs has its own weight and softmax classifier for categorizing the output data. The suggested system's identification accuracy rate is 98.02%. Last but not least, (Murphy, et al., 2012) suggests a new neural net for skeleton-based sign language identification built on the foundation of SPD manifold learning. Figure 1 shows an illustration of a deep learning convolution neural network.

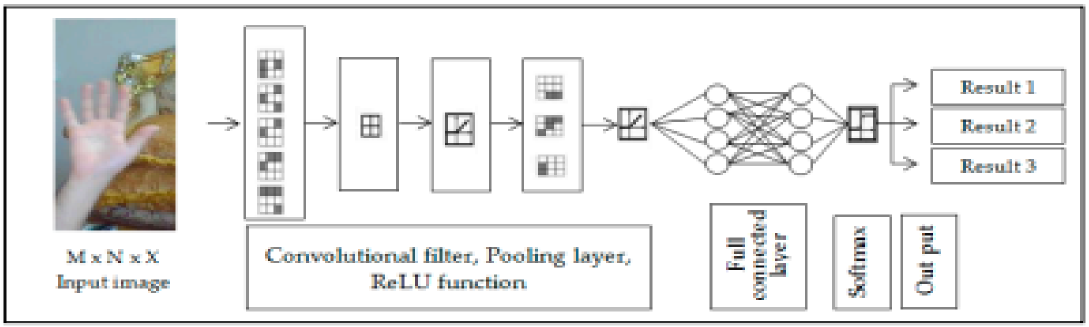

Figure 1: Example of a deep learn convolution neural network for image classification.

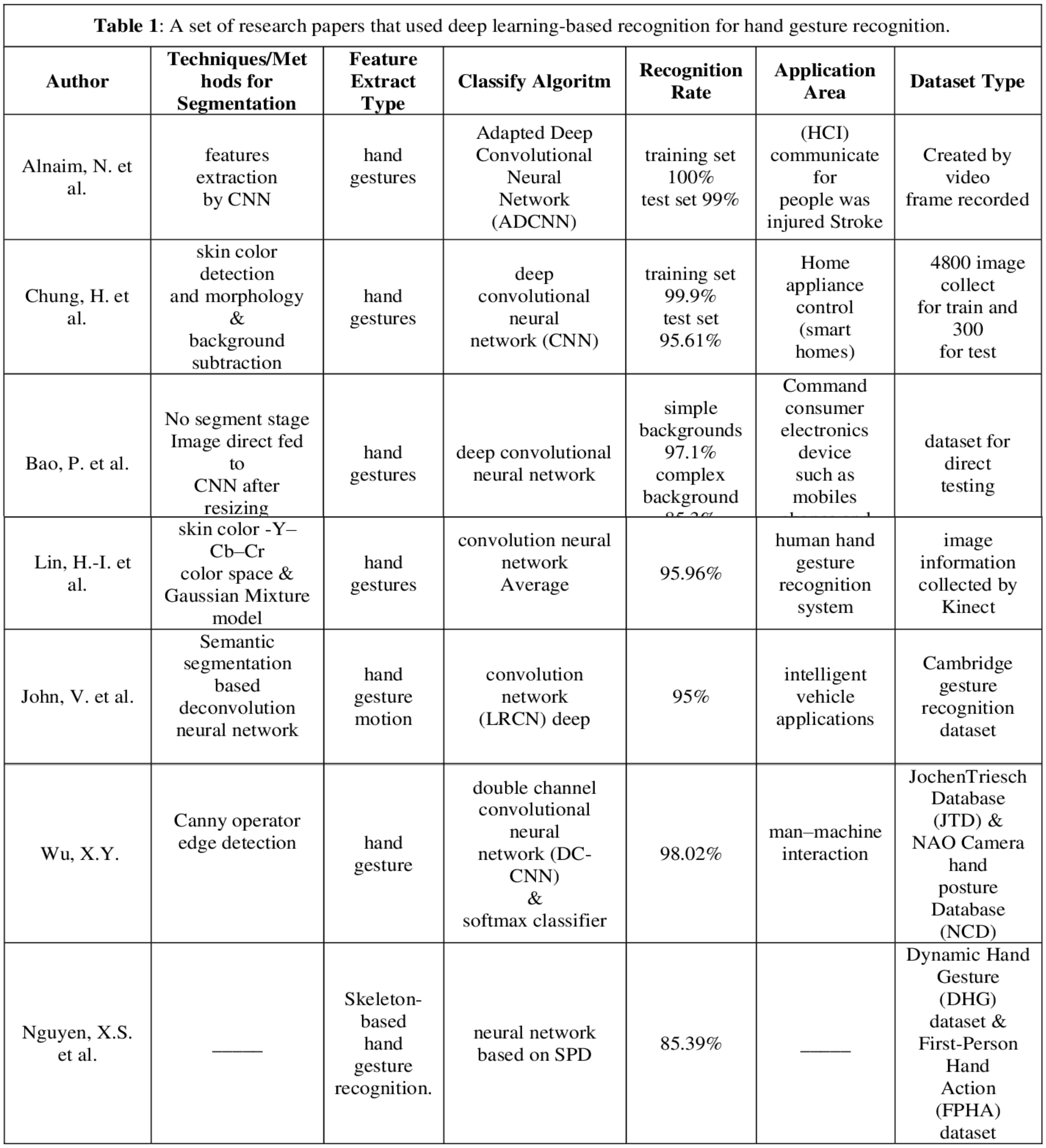

Table 1: A set of research papers that used deep learning-based recognition for hand gesture recognition.

Author	**Techniques/Methods for Segmentation**	**Feature Extract Type**	**Classify Algoritm**	**Recognition Rate**	**Application Area**	**Dataset Type**
Alnaim, N. et al.	features extraction by CNN	hand gestures	Adapted Deep Convolutional Neural Network (ADCNN)	training set 100% test set 99%	(HCI) communicate for people was injured Stroke	Created by video frame recorded
Chung, H. et al.	skin color detection and morphology & background subtraction	hand gestures	deep convolutional neural network (CNN)	training set 99.9% test set 95.61%	Home appliance control (smart homes)	4800 image collect for train and 300 for test
Bao, P. et al.	No segment stage Image direct fed to CNN after resizing	hand gestures	deep convolutional neural network	simple backgrounds 97.1% complex background 85.3%	Command consumer electronics device such as mobiles	dataset for direct testing
Lin, H.-I. et al.	skin color -Y–Cb–Cr color space & Gaussian Mixture model	hand gestures	convolution neural network Average	95.96%	human hand gesture recognition system	image information collected by Kinect
John, V. et al.	Semantic segmentation based deconvolution neural network	hand gesture motion	convolution network (LRCN) deep	95%	intelligent vehicle applications	Cambridge gesture recognition dataset
Wu, X.Y.	Canny operator edge detection	hand gesture	double channel convolutional neural network (DC-CNN) & softmax classifier	98.02%	man–machine interaction	JochenTriesch Database (JTD) & NAO Camera hand posture Database (NCD)
Nguyen, X.S. et al.	_____	Skeleton-based hand gesture recognition.	neural network based on SPD	85.39%	_____	Dynamic Hand Gesture (DHG) dataset & First-Person Hand Action (FPHA) dataset

5. MOTION AND SKELETON BASED RECOGNITION

Motion-based recognition, which extracts the object from a succession of visual frames, can be utilised for detection. In (Nakjai, P. et al., 2019), the backdrop and skin colour were distinguished using the CAMShift algorithm, and gesture recognition was projected to benefit from the naive Bayes classifier. The recommended approach must overcome difficulties such variable lighting, where shifts in light have an impact on the outcome of the skin part. The level of expression flexibility presents another difficulty since it directly influences the output result by altering rotation. If the hand is in the corner of the image and the dots which must cover the palm are not on the hand, the captured user motion may be unsuccessful. The skin-like colour in the study of (Nguyen, et al., 2019) is the key issue that still needs to be resolved since it has an adverse effect on the overall system and renders the results invalid.

In skeleton-based identification, model parameters are specified that can aid in the detection of complicated characteristics. The hand model establishes geometric qualities and restrictions and readily translates parameters and correlations in order to concentrate on geometric and statistical aspects, whilst the numerous skeletal data formats for classification may be employed for the hand model. In (Oudah, et al., 2020), a dynamic hand motion is provided utilizing a skeleton-based technique using a depth and structural dataset. A linear kernel and supervised learning (SVM) are then utilized to identify the data. The SVM algorithm outperformed the HMM in some specifications such as elapsed time and recognition accuracy rate. Another dynamic identification system that used Kinect sensor depth information for acquisition and categorization was proposed in (Oudah, et al. 2020). Table 2 lists studies that applied skeleton-based recognition and motion-based detection to hand gesture applications.

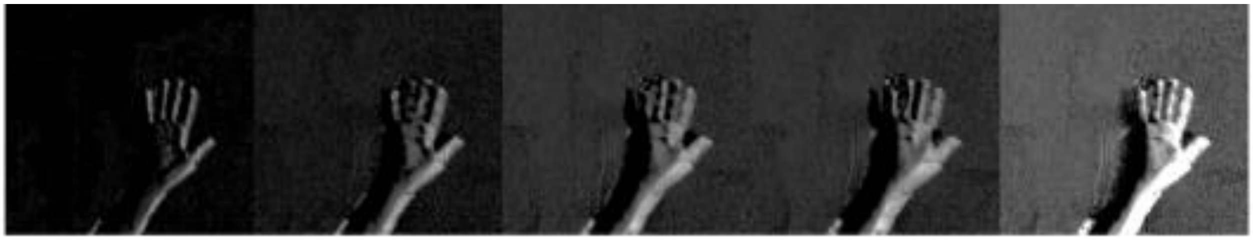

Figure 2: Example of hand motion identification where the moving item, such as the hand, is recovered from a stationary backdrop using frame difference calculation to get hand characteristics

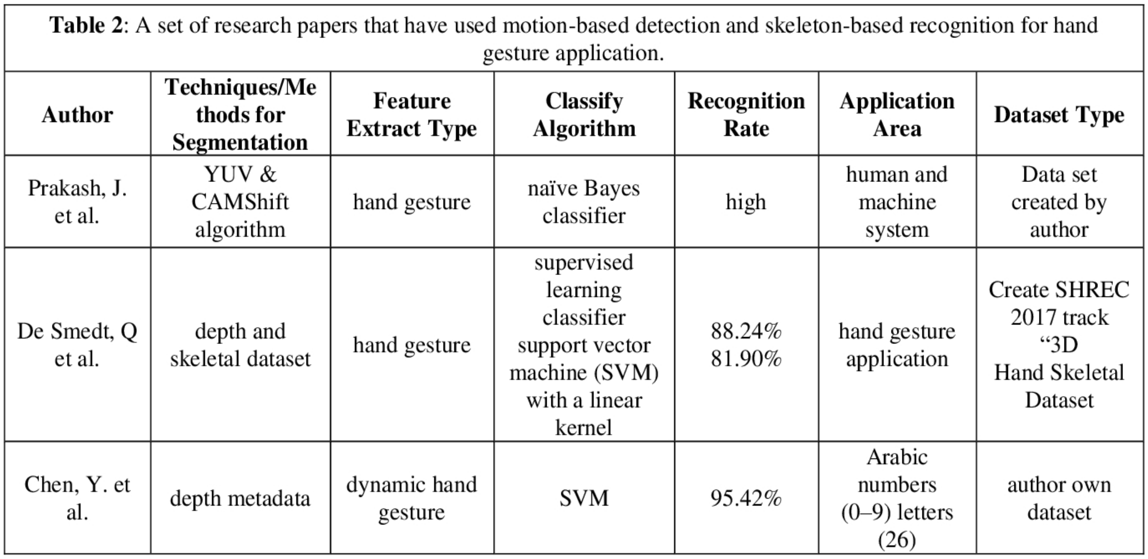

Table 2: A set of research papers that have used motion-based detection and skeleton-based recognition for hand gesture application.

Author	Techniques/Methods for Segmentation	Feature Extract Type	Classify Algorithm	Recognition Rate	Application Area	Dataset Type
Prakash, J. et al.	YUV & CAMShift algorithm	hand gesture	naïve Bayes classifier	high	human and machine system	Data set created by author
De Smedt, Q et al.	depth and skeletal dataset	hand gesture	supervised learning classifier support vector machine (SVM) with a linear kernel	88.24% 81.90%	hand gesture application	Create SHREC 2017 track “3D Hand Skeletal Dataset
Chen, Y. et al.	depth metadata	dynamic hand gesture	SVM	95.42%	Arabic numbers (0–9) letters (26)	author own dataset

6. RESEARCH GAPS AND DIFFICULTIES

The majority of research studies have focused mostly on sign language and computer applications, thus it is simple to recognize the knowledge gaps from the previous sections. From the other hand, many research publications focus on developing new algorithms or frameworks for recognizing sign language. The researcher's major challenge has been creating a solid framework that solves the most fundamental issues with the fewest restrictions and yields a precise and trustworthy solution. Deep learning and artificial intelligence methods are used in real time to match communication gestures with dataset movements that already include specific postures or gestures. Even while certain techniques can recognize a lot of motions, there might be major limitations in some circumstances, such as forgetting specific gestures due to differences in classification algorithm performance.

7. CONCLUSION

Deep learning methods and their applications in the field of gesture image processing are thoroughly discussed in this paper. It is concluded that deep learning methods that use convolutional neural networks are gaining recognition in all disciplines of hand gesture image analysis, including classification, detection, and segmentation. Each of the techniques discussed above has benefits and drawbacks, and they may perform well in some challenges while falling short in others.

REFERENCES

Alnaim, N.,Abbod, M.&Albar, A.(2019). Hand Gesture Recognition Using Convolutional Neural Network for People Who Have Experienced A Stroke. In Proceedings of the 2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), Ankara, Turkey, 11–13 October 2019, pp. 1–6.

Barros, P.,Parisi, G. I., Weber, C.&Wermter, S. (2017). Emotion-modulated attention improves expression recognition: a deep learning model, Neurocomputing, vol. 253, pp. 104–114.

Bao, P.,Maqueda, A.I., del-Blanco, C.R.&García, N. (2017). Tiny hand gesture recognition without localization via a deep convolutional network. IEEE Trans. Consum. Electron, 63, pp.251–257.

Bengio, Y.,Courville, A. Vincent, P. (2013). Representation Learning: A Review and New Perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence. 35 (8): 1798–1828.

Chung, H., Chung, Y.& Tsai, W.(2019). An efficient hand gesture recognition system based on deep CNN. In Proceedings of the 2019 IEEE International Conference on Industrial Technology (ICIT), Melbourne, Australia, 13–15 February 2019, pp. 853–858.

Chen, J., Ou, Q., Chi, Z.& Fu, H. (2016). Smile detection in the wild with deep convolutional neural networks, Machine Vision and Applications, vol. 28, no. 1-2, pp. 173–183.

Chen, Y.,Luo, B., Chen, Y.-L., Liang, G.& Wu, X. (2015). A real-time dynamic hand gesture recognition system using kinect sensor. In Proceedings of the 2015 IEEE International Conference on Robotics and Biomimetics (ROBIO), Zhuhai, China, 6–9 December 2015, pp. 2026–2030.

Cybenko (1989). Approximations by superpositions of sigmoidal functions" (PDF). Mathematics of Control, Signals, and Systems. 2 (4), pp. 303–314.

Das, N.,Ohn-Bar, E.& Trivedi, M.M. (2015). On Performance Evaluation of Driver Hand Detection Algorithms: Challenges, Dataset, and Metrics. In Proceedings of the IEEE Conference on Intelligent Transportation Systems, (ITSC), Las Palmas, Spain, 15–18, pp. 2953–2958.

Dardas, N.H.&Georganas, N.D. (2011). Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques. IEEE Trans. Instrum. Meas, 60, 3592–3607.

De Smedt, Q.,Wannous, H.,Vandeborre, J.-P.,Guerry, J.,Saux, B.L.&Filliat, D. (2017). 3D hand gesture recognition using a depth and skeletal dataset: SHREC'17 track. In Proceedings of the Workshop on 3D Object Retrieval, Lyon, France, 23–24 April 2017, pp. 33–38.

Deng, L.& Yu, D. (2014). Deep Learning: Methods and Applications" (PDF). Foundations and Trends in Signal Processing. 7 (3–4), pp.1–199.

Huang, Y., Liu, X., Zhang, X.& Jin, L. (2016) A Pointing Gesture Based Egocentric Interaction System: Dataset, Approach and Application. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Las Vegas, NV, USA, pp. 370–377.

Han, M., Chen, J., Li, L.& Chang, Y. (2016). Visual hand gesture recognition with convolution neural network. IEEE/acis Int. Conf. on Software Engineering, Artificial Intelligence, NETWORKING and Parallel/distributed Computing, Shanghai, China, pp. 287–291.

Hornik, Kurt (1991). Approximation Capabilities of Multilayer Feedforward Networks". Neural Networks. Vol.4 (2), pp. 251–257.

Hornik, Kurt (1991). Approximation Capabilities of Multilayer Feedforward Networks". Neural Networks. Vol.4 (2), pp. 251–257.

Hassoun, Mohamad H. (1995). Fundamentals of Artificial Neural Networks. MIT Press. p. 48. ISBN 978-0-262-08239-6.

John, V.,Boyali, A.,Mita, S.,Imanishi, M.&Sanma, N. (2016). Deep learning-based fast hand gesture recognition using representative frames. In Proceedings of the 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Gold Coast, Australia, pp. 1–8.

Kang, B., Tan, K.H., Jiang, N., Tai, H.S.,Tre_er, D.& Nguyen, T. (2017). Hand segmentation for hand-object interaction from depth map. In Proceedings of the 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP 2017), Montreal, QC, Canada, 14–16 November 2017, pp. 259–263.

Kaur, H.& Rani, J. (2016). A review: Study of various techniques of Hand gesture recognition. In Proceedings of the 2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES), Delhi, India, 4–6 July 2016, pp. 1–5.

Kao, C.Y.&ShyurngFahn, C. (2011). A Human-Machine Interaction Technique: Hand Gesture Recognition Based on Hidden Markov Models with Trajectory of Hand Motion, Procedia Engineering vol. 15, pp. 3739 – 3743.

Krizhevsky, A.,Sutskever, I.& Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.

Le, T.H.N., Zhu, C.,Zheng, Y.,Luu, K.&Savvides, M. (2016). Robust hand detection in Vehicles. In Proceedings of the International Conference on Pattern Recognition, Cancun, Mexico, 4–8, pp. 573–578.

Li, G., Tang, H., Sun, Y., Kong, J., Jiang, G., Jiang, D., Tao, B.,Xu, S.& Liu, H.(2019). Hand gesture recognition based on convolution neural network. Cluster Compute. 22, pp.2719–2729.

Lin, H.-I., Hsu, M.-H.& Chen, W.K.(2014). Human hand gesture recognition using a convolution neural network. In Proceedings of the 2014 IEEE International Conference on Automation Science and Engineering (CASE), Taipei, Taiwan, 18–22 August 2014, pp. 1038–1043.

Lu, Z., Pu, H., Wang, F., Hu, Z., & Wang, L. (2017). The Expressive Power of Neural Networks: A View from the Width Archived 2019-02-13 at the Wayback Machine. Neural Information Processing Systems, pp.6231-6239.

Mohammed, A. A. Q.,Lv, J.& Islam, M. S. (2019). A deep learning-based end-to-end composite system for hand detection and gesture recognition, Sensors, Vol.19(23), pp.5282.

Murphy, Kevin P. (24 August 2012). Machine Learning: A Probabilistic Perspective. MIT Press. ISBN 978-0-262-01802-9.

Nakjai, P.&Katanyukul, T.(2019). Hand Sign Recognition for Thai Finger Spelling: An Application of Convolution Neural Network. J. Signal Process, pp.131–146.

Nguyen, X.S.,Brun, L.,Lézoray, O.,Bougleux, S. (2019). A neural network based on SPD manifold learning for skeleton-based hand gesture recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, pp. 12036–12045.

Oudah, M., Al-Naji, A.&Chahl, J. (2020). Hand gesture recognition based on computer vision: a review of techniques. journal of Imaging, vol.6(8), 73.

Prakash, J.&Gautam, U.K. (2019). Hand Gesture Recognition. Int. J. Recent Technol. Eng. 2019, 7, pp.54–59.

Ploue, G.&Cretu, A.M.(2016). Static and dynamic hand gesture recognition in depth data using dynamic time warping. IEEE Trans. Instrum. Meas, 65, 305–316.

Pigou, L.,Dieleman, S.,Kindermans, P. J.&Schrauwen, B. (2014). Sign language recognition using convolutional neural networks. In European conference on computer vision, Springer, Cham, pp. 572-578.

Prakash, J.&Gautam, U.K. (2019). Hand Gesture Recognition. Int. J. Recent Technol. Eng. 2019, 7, pp.54–59.

Schmidhuber, J. (2015). Deep Learning in Neural Networks: An Overview. Neural Networks. 61, pp.85–117.

Tao, W.,Leu, M.C.&Yin, Z. (2018). American Sign Language alphabet recognition using Convolutional Neural Networks with multiview augmentation and inference fusion. Eng. Appl. Artif. Intell. 76, 202–213.

Wu, X.Y. (2019). A hand gesture recognition algorithm based on DC-CNN. Multimed. Tools Appl., 1–13.

Xie, B., He, X.&Li, Y. (2018). RGB-D static gesture recognition based on convolutional neural network. The Journal of Engineering, 2018(16), pp. 1515-1520.

Zhang, C.,Tian, Y.,Guo, X.& Liu, J. (2018). DAAL: deep activation-based attribute learning for action recognition in depth videos, Computer Vision and Image Understanding, vol. 167, pp. 37–49.