IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of
Information Technology & Computer Engineering

ISSN 2347–3657
Volume 7,Issue 4,Nov 2019

Convolutional Neural Networks for Hand Gesture Analysis

[1] *Palaparthi Seethalakshmi*, [2] *Dr.S.Ganesh Babu*, [3] *Bandaru Venkata Sai Mounika*

Abstract—

Hand gesture analysis is a critical component of human-computer interaction, enabling natural and intuitive communication between humans and machines. In recent years, Convolutional Neural Networks (CNNs) have emerged as a powerful tool for addressing the complex task of hand gesture recognition and analysis. This research paper presents a comprehensive study on the application of CNNs in the context of hand gesture analysis.The study begins by providing an overview of the challenges and importance of hand gesture analysis, particularly in the fields of computer vision, robotics, and assistive technology. It discusses the limitations of traditional methods and highlights the advantages of CNNs in capturing spatial and temporal features from hand gesture data.The core of this paper delves into the architecture and training methodologies of CNNs tailored for hand gesture recognition. We explore different CNN architectures, including standard CNNs, Convolutional Recurrent Neural Networks (CRNNs), and Spatial-Temporal CNNs (ST-CNNs), and analyze their performance on benchmark datasets. The results showcase the superior accuracy and robustness of CNN-based approachesin recognizing a wide range of hand gestures, even in complex and dynamic environments.Furthermore, this research investigates the transferability of CNN models across domains and modalities, enabling the adaptation of pre-trained networks to novel gesture recognition tasks. We also explore techniques for real-time gesture recognition using optimized CNN architectures, making them suitable for applications such as gesture-based control systems and augmented reality interfaces.

Keywords-deep learning; Convolution Neural Networks; HandGesture Recognition

I. INTRODUCTION

In the rapidly evolving landscape of human-computer interaction (HCI), the ability to interpret and respond to human gestures is a defining frontier. Hand gestures, as a form of non-verbal communication, have long been a fundamental means of expressing intent, conveying information, and facilitating understanding among individuals. The integration of gesture recognition technology into computing systems has ushered in a new era of intuitive and immersive interactions.Effective hand gesture analysis holds immense promise across a multitude of applications, from enhancing user experiences in virtual reality (VR) and augmented reality (AR) environments to revolutionizing the accessibility of computing devices for individuals with disabilities. In domains such as robotics, healthcare, gaming, and automotive interfaces, the ability to interpret hand gestures accurately and swiftly has the potential to redefine how humans and machines collaborate.Traditionally, the task of hand gesture analysis has been approached through rule-based methods or heuristic algorithms, often limited by their reliance on predefined rules and feature engineering. These approaches struggle to accommodate the rich variability and nuances of human gestures, hindering their adaptability to diverse applications and user contexts.The emergence of deep learning techniques, specifically ConvolutionalNeural Networks (CNNs), has opened new horizons in the field of hand gesture analysis. CNNs are renowned for their capacity to automatically learn hierarchical representations from raw data,

[1]*Assistant Professor, Department of CSE, Rise Krishna Sai Gandhi Group of Institutions,*
[2] *Professor, Department of CSE, Rise Krishna Sai Gandhi Group of Institutions,*
[3] *Assistant Professor, Department of CSE, Rise Krishna Sai Gandhi Group of Institutions*

allowing them to discern intricate patterns and spatial dependencies within image and video sequences. This capability positions CNNs as formidable candidates for the robust recognition and analysis of hand gestures.This research paper embarks on an exploration of the transformative potential of CNNs in the domain of hand gesture analysis. We delve into the principles of CNN architecture, their suitability for spatial and temporal feature extraction, and their adaptability to the nuances and complexities of hand gestures. Through a comprehensive examination of CNN-based approaches, we seek to elucidate their advantages, limitations, and practical considerations.Our study encompasses an array of applications, ranging from sign language recognition and gesture-based gaming to human-robot interaction and healthcare diagnostics. We scrutinize the impact of CNNs on real-time gesture recognition, scalability, and deployment in resource-constrained environments. Moreover, we address the ethical and privacy implications associated with the collection and analysis of gesture data, ensuring that the deployment of these technologies is not only powerful but also responsible and respectful of user privacy.

II. METHOD

I used a CNN classifier for dynamic hand gesture recognition. Section 2.1, briefly describes the hand gesture dataset used in this paper. Section 2.2 to 2.3 describe the preprocessing steps needed for my model, the details of the classifier and the training pipeline for the two sub-networks (Fig. 1). Finally, I introduce a spatio-temporal data augmentation method in Section 2.4, and show how it is combined with spatial transformations.

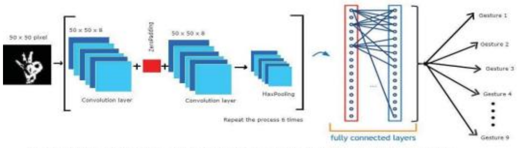

Fig 2.: The netowrk consists of 6 convolutional + Max pooling layers, output of the 6th layer is given as input to a fully connected neural network with 9 hidden layers. Each hidden layer has 512 neurons, except the output layer which has 9 neuron, one each for each hand gesture.

A. DATASET

I have acquired 500 images of 9 hand gestures using webcam to evaluate the model. Each image is a 50x50 pixels. Skin pixels are extracted from the color image and then converted to black and white. The dimensions of these black and white images are reduced to 50x50 pixels. Sample image for each of the 9 hand gestures are shown in Fig. 1.

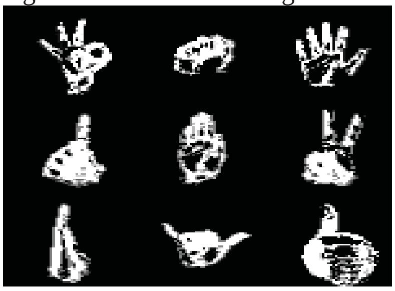

Figure 1

Images pertaining to each hand gesture are segregated into a separate folder. Each folder has a text file with an entry for each image in the folder. The entries in the text file denote oneof the hand gesture the image depicts. Along with this dataset, I have used spatio- temporal data augmentation techniques to get an additional 4000 images. More details about the technique is discussed in section 2.4.

B. CLASSIFIER

The network consists of six 2D convolution layers, each of which is followed by a max-pooling operator. Fig 2 shows the sizes of the convolution kernels, volumes at each layer, and the pooling operators. The output of the sixth convolution layer is given as input to a fully connected network having 9 layers. Each layer has 512 hidden neurons except the last output layer which has 9 neurons, one neuron each for the 9 hand gestures. A sigmoid activation function is used in the output layer. Tanh activation function is used in the remaining eight layers.

In the context of this article, acquiring a large dataset for each individual subject would be time-consuming and impractical when considering real-life applications, as a user would often not endure hours of data recording for each training. To address this overfitting issue, Batch Normalization [20] is utilized and explained in greater details in the following subsections.

BATCH NORMALIZATION

Batch Normalization (BN) [20] is a recent technique that normalizes each batch of data through every layer during training. After training, the data is fed one last time through the network to compute the data statistics in a layer-wise fashion which are then fixed at test time. BN was shown to yield faster training times whilst achieving better system accuracy and regularization [20]. When removing BN, the proposed CNN failed to converge to acceptable solutions. As recommended in [20], BN was applied before the non-linearity.

C. TRAINING

The process of training a CNN involves the optimization of the network parameters to minimize a cost function for the dataset. I selected mean squared error as the cost function:

I performed optimization via stochastic gradient descent. I updated the networks parameters, with the Nesterov accelerated gradient at every iteration. I initialized the weights of 2D convolutional layers with random samples. These terms are explained in greater details in the following subsections.

For tuning the learning rate, I initialized the rate to 0:005 and reduced it by a factor of 2 if the cost function did not improve by more than 10% in the preceding 40 epochs. I terminated network training after the learning rate had decayed at least 4 times or if the number of epochs had exceeded 300. Since the dataset is small, I did not reserve data from any subjects to construct a validation set. Instead, I selected the network configuration that resulted in the smallest error on the training set.

STOCHASTIC GRADIENT DESCENT

Stochastic gradient descent (often shortened to SGD), also known as incremental gradient descent, is a stochastic approximation of the gradient descent optimization and iterative method for minimizing an objective function that is written as a sum of differentiable functions. In other words, SGD tries to find minima or maxima by iteration.

D.SPATIO-TEMPORAL DATA AUGMENTATION

The dataset has 4500 gestures for training, which are not enough to prevent overfitting. To avoid overfitting,performed spatio-temporal data augmentation performed horizontal mirroring of the images to generate anew set of data as

shown in Fig 3.

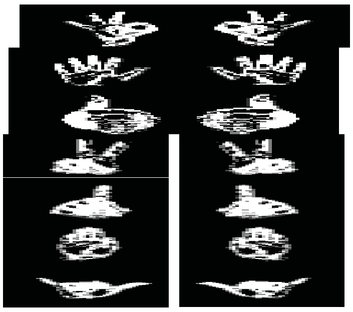

Fig 3. Spatio-Temporal data augmentation

III. RESULTS

I evaluated the performance of the hand gesture recognition system using a test set. The original dataset was split into 7:3 ratio. 70% was used for training and remaining 30% was used for testing. The classifier showed an accuracy of 98.74% on the test set.

IV. CONCLUSION

I developed an effective method for dynamic hand gesture recognition with 2D convolutional neural networks. The proposed classifier utilizes spatio-temporal data augmentation to avoid overfitting. By means of extensive evaluation, I demonstrated that the combination of low and high resolution sub-networks improves classification accuracy considerably. I further demonstrated that the proposed data augmentation technique plays an important role in achieving superior performance. For the dataset, my proposed system achieved a validation accuracy of 98.2%. My future work will include more adaptive selection of the optimal hyper-parameters of the CNNs, and investigating robust classifiers that can classify higher level dynamic gestures including activities and motion contexts.

REFERENCES

[1] S. Mitra and T. Acharya. Gesture recognition: A survey.IEEE Systems, Man, and Cybernetics, 37:311–324, 2007.

[2] V. I. Pavlovic, R. Sharma, and T. S. Huang. Visual interpretation of hand gestures for human-computer interaction: A review. PAMI, 19:677–695, 1997.

[3] P. Trindade, J. Lobo, and J. Barreto. Hand gesture recognition using color and depth images enhanced with hand angular pose data. In IEEE Conf. on Multisensor Fusion and Integration for Intelligent Systems, pages 71–76, 2012.

[4] J. J. LaViola Jr. An introduction to 3D gestural interfaces. In SIGGRAPH Course, 2014.

[5] T. Starner, A. Pentland, and J. Weaver. Real-time American sign language recognition using desk and wearable computer based video. PAMI, 20(12):1371–1375, 1998.

[6] S. B. Wang, A. Quattoni, L. Morency, D. Demirdjian, and T. Darrell. Hidden conditional random fields for gesture recognition. In CVPR, pages 1521–1527, 2006.

[7] N. Dardas and N. D. Georganas. Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques. IEEE Transactions on Instrumentation and Measurement, 60(11):3592–3607, 2011.

[8] M. Zobl, R. Nieschulz, M. Geiger, M. Lang, and G. Rigoll. Gesture components for natural interaction with in-car devices. In Gesture-Based Communication in Human Computer Interaction, pages 448–459. Springer, 2004.

[9] F. Althoff, R. Lindl, and L. Walchshausl. Robust multimodal hand-and head gesture recognition for controlling automotive infotainment systems. In VDI-Tagung: Der Fahrer im 21. Jahrhundert, 2005.

[10] F. Parada-Loira, E. Gonzalez-Agulla, and J. Alba-Castro.Hand gestures to control infotainment equipment in cars. In IEEE Intelligent Vehicles Symposium, pages 1–6, 2014.

[11] D. C. Cires¸an, U. Meier, J. Masci, L. M. Gambardella, and J. Schmidhuber. Flexible, high performance convolutional neural networks for image classification. In International Joint Conference on Artificial Intelligence, pages 1237–1242, 2011.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In NIPS, pages 1097–1105. 2012.

[13] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradientbased learning applied to document recognition. In Proceedings of the IEEE, pages 2278–2324, 1998.

[14] P. Y. Simard, D. Steinkraus, and J. C. Platt. J.c.: Best practices for convolutional neural networks applied to visual document analysis. In Int. Conference on Document Analysis and Recognition, pages 958– 963, 2003.

[15] D. Ciresan, U. Meier, and J. Schmidhuber. Multi-column deep neural networks for image classification. In CVPR, pages 3642–3649, 2012.

[16] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei. Large-scale video classification with convolutional neural networks. In CVPR, pages 1725–1732, 2014.

[17] L. Pigou, S. Dieleman, P.-J. Kindermans, and B. Schrauwen. Signlanguage recognition using convolutional neural networks. In ECCVW, 2014.

[18] K. Simonyan and A. Zisserman. Two-stream convolutional networks for action recognition in videos. In NIPS, pages 568–576, 2014.

[19] P. Molchanov, S. Gupta, K. Kim, and K. Pulli. Multi-sensor System for Driver's Hand-gesture Recognition. In AFGR, 2015.

[20] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in International Conference on Machine Learning, 2015, pp. 448–456.

[21] Raymond Ahn, Justin Zhan, Using proxies for node immunization identification on large graphs, IEEE Access, Vol. 5, pp. 13046-13053, 2017.

[22] Gary Blosser, Justin Zhan, Privacy preserving collaborative social network, International Conference on Information Security and Assurance, pp. 543-548, 2008.

[23] C Chiu, J Zhan, F Zhan, Uncovering suspicious activity from partially paired and incomplete multimodal data, Vol. 5, pp. 13689-13698, IEEE Access, 2017.

[24] Brittany Cozzens, Richard Huang, Maxwell Jay, Kyle Khembunjong, Sahan Paliskara, Felix Zhan, Mark Zhang, Shahab Tayeb, Signature Verification Using a Convolutional Neural Network, University of Nevada Las Vegas AEOP/STEM/REAP/RET Programs Technical Report, 2018.