2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI)

Hand Gesture Recognition with Convolution Neural Networks

Felix Zhan
USAOT
felixzhan@usaot.org

Abstract— **Hand gestures are the most common forms of communication and have great importance in our world. They can help in building safe and comfortable user interfaces for a multitude of applications. Various computer vision algorithms have employed color and depth camera for hand gesture recognition, but robust classification of gestures from different subjects is still challenging. I propose an algorithm for real-time hand gesture recognition using convolutional neural networks (CNNs). The proposed CNN achieves an average accuracy of 98.76% on the dataset comprising of 9 hand gestures and 500 images for each gesture.**

Keywords-deep learning; Convolution Neural Networks; Hand Gesture Recognition

I. INTRODUCTION

In recent years, robotics and artificial intelligence have been leveraged to increase the autonomy of people living with disabilities. In this context, the main objective is to improve the quality of life by enabling users to perform a wider range of day-to-day tasks more efficiently. In particular, hand gesture recognition has been recognized as a valuable technology for several application fields, especially for Sign Language Recognition (SLR). Sign languages comprise of complex hand movements, and even miniscule hand changes can have a variety of possible meanings. In response to this, in the last decade, many vision-based dynamic hand gesture recognition algorithms were introduced [1,2]. To recognize gestures, different features such as hand-crafted spatio-temporal descriptors [3] and articulated models [4], were used, along with gesture classifiers, hidden Markov models [5], conditional random fields [6] and support vector machines (SVM) [7] have been widely used. However, classification of gestures is unpredictable under varying lighting conditions, and from different subjects is still a challenging problem [8,9,10].

An intuitive approach for creating interfaces is to look at the muscle activity of the user. This activity can be recorded by the device using a camera. These recorded images can then be analyzed using deep learning algorithms to determine the sign.

Recently, classification with deep convolutional neural networks has been successful in various recognition challenges [11,12,13,14]. Multi-column deep CNNs that employ multiple parallel networks have been shown to improve recognition rates of single networks by 30-80% for various image classification tasks [15]. Similarly, for large scale video classification, Karpathy et al. [16] observed the best results on combining CNNs trained with two separate streams of the original and spatially cropped video frames.

Several authors have emphasized the importance of using many diverse training examples for CNNs [12, 17, 18]. They have proposed data augmentation strategies to prevent CNNs from overfitting when training with datasets containing limited diversity. Krizhevsky et al. [19] employed translation, horizontal flipping and RGB jittering of the training and testing images for classifying them into 1000 categories. Simonyan and Zisserman [18] employed similar spatial augmentation on each video frame to train CNNs for video-based human activity recognition. However, these data augmentation methods were limited to spatial variations. To add variations to video sequences containing dynamic motion, Pigou et al. [17] temporally translated the video frames in addition to applying spatial transformations. Other research motivated my ideas includes [20-65].

In this paper, I introduce a hand gesture recognition system that extracts hand components in the image and learns and predicts using 2D convolutional neural networks. To reduce potential over- fitting and improve generalization of the gesture classifier, I propose an effective spatio-temporal data augmentation method to deform the input volumes of hand gestures. The augmentation method also incorporates existing spatial augmentation techniques [12].

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.

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.

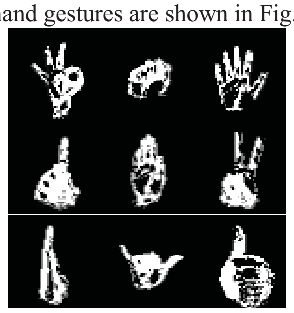

Figure 1

978-1-7281-1337-1/19/$31.00 ©2019 IEEE
DOI 10.1109/IRI.2019.00054

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 one of 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.

B.1 *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.

C.1 *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.

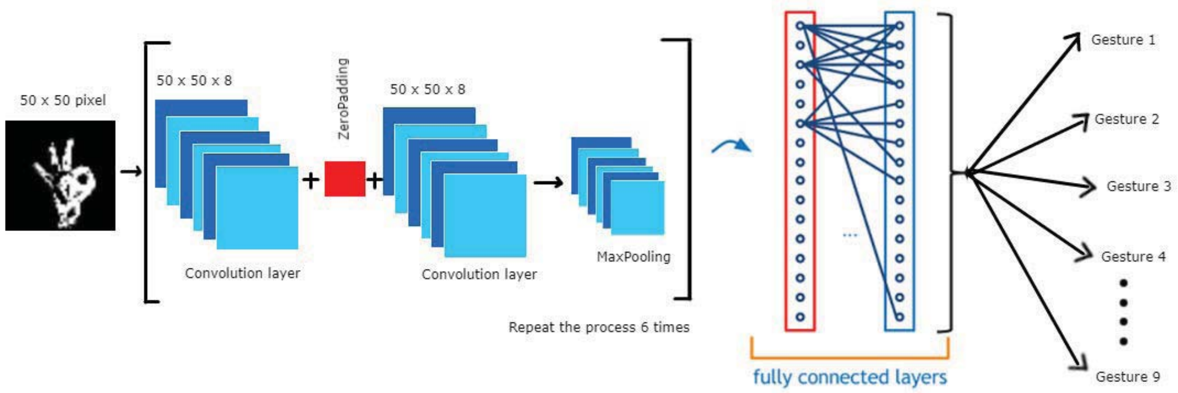

Fig 2.: The netowrk consists of 6 convolutional + Max pooling layers, output of the 6th layer is given as input to a fully connected neural network with 9 hidden layers. Each hidden layer has 512 neurons, except the output layer which has 9 neuron, one each for each hand gesture.

D. *SPATIO-TEMPORAL DATA AUGMENTATION*

The dataset has 4500 gestures for training, which are not enough to prevent overfitting. To avoid overfitting, I performed spatio-temporal data augmentation. I have performed horizontal mirroring of the images to generate a new set of data as shown in Fig 3.

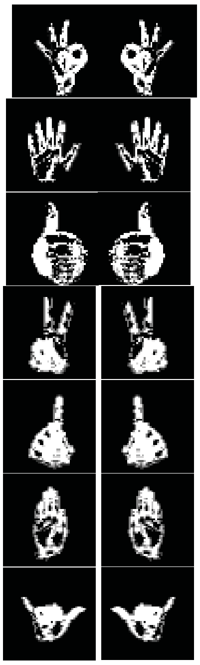

Fig 3. Spatio-Temporal data augmentation

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. Sign language 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.

[25] Pravin Chopade, Justin Zhan, Marwan Bikdash, Node attributes and edge structure for large-scale big data network analytics and community detection, 2015 IEEE International Symposium on Technologies for Homeland Security, pp. 1-8, 2015.
[26] Pravin Chopade, Justin Zhan, A framework for community detection in large networks using game-theoretic modeling, IEEE Transactions on Big Data, Vol. 3, Issue 3, pp. 276-288, 2017.
[27] Pravin Chopade, Justin Zhan, Structural and functional analytics for community detection in large-scale complex networks, Journal of Big Data, Vol. 2, Issue 1, , 2015
[28] Matin Pirouz, Justin Zhan, Shahab Tayeb, An optimized approach for community detection and ranking, Journal of Big Data, Vol. 3, Issue 1, pp. 22, 2016.
[29] Matin Pirouz, Justin Zhan, Optimized relativity search: node reduction in personalized page rank estimation for large graphs, Journal of Big Data, Vol. 3, Issue 1, 2016.
[30] Shahab Tayeb, Matin Pirouz, Brittany Cozzens, Richard Huang, Maxwell Jay, Kyle Khembunjong, Sahan Paliskara, Felix Zhan, Mark Zhang, Justin Zhan, Shahram Latifi, Toward data quality analytics in signature verification using a convolutional neural network, 2017 IEEE International Conference on Big Data, pp. 2644-2651, 2017.
[31] Haysam Selim, Justin Zhan, Towards shortest path identification on large networks, Journal of Big Data, Vol. 3, Issue 1, pp. 10, 2016
[32] Xian-Ming Xu, Justin Zhan, Hai-tao Zhu, Using social networks to organize researcher community, International Conference on Intelligence and Security Informatics, pp. 421-427, 2008.
[33] Felix Zhan, Gabriella Laines, Sarah Deniz, Sahan Paliskara, Irvin Ochoa, Idania Guerra, Shahab Tayeb, Carter Chiu, Matin Pirouz, Elliott Ploutz, Justin Zhan, Laxmi Gewali, Paul Oh, Prediction of online social networks users' behaviors with a game theoretic approach, pp. 1-2, 2018 15th IEEE Annual Consumer Communications & Networking Conference.
[34] Felix Zhan, Brandon Waters, Maria Mijangos, LeAnn Chung, Raghav Bhagat, Tanvi Bhagat, Matin Pirouz, Carter Chiu, Shahab Tayeb, Elliott Ploutz, Justin Zhan, Laxmi Gewali, An efficient alternative to personalized page rank for friend recommendations, pp. 1-2, 2018 15th IEEE Annual Consumer Communications & Networking Conference.
[35] Justin Zhan, Xing Fang, A novel trust computing system for social networks, IEEE Third International Conference on Social Computing, pp. 1284-1289, 2011.
[36] Justin Zhan, Secure collaborative social networks, IEEE Transactions on Systems, Man, and Cybernetics, Part C, Vol. 40, Issue 6, pp. 682-689, 2010.
[37] Justin Zhan, Xing Fang, Social computing: the state of the art, International Journal of Social Computing and Cyber-Physical Systems, Vol. 1, Issue 1, pp. 1-12, 2011.
[38] Justin Zhan, Xing Fang, A computational trust framework for social computing (a position paper for panel discussion on social computing foundations), IEEE Second International Conference on Social Computing, pp. 264-269, 2010.
[39] Felix Zhan, Gabriella Laines, Sarah Deniz, Sahan Paliskara, Irvin Ochoa, Idania Guerra, Shahab Tayeb, University of Nevada Las Vegas AEOP/STEM/REAP/RET Programs Technical Report, Vol. 2, pp. 26-30, 2017.
[40] Felix Zhan, Brandon Waters, Maria Mijangos, Raghav Bhagat, Tanvi Bhagat, A Low Cost, High Speed Alternative to Personalized Page Rank for Friend Recommendations, http://aeop.asecamps.com/wp-content/uploads/2017/07/TeamC.pdf.
[41] Felix Zhan, How to Optimize Social Network Influence, 2019 IEEE International Conference on Artificial Intelligence and Knowledge Engineering, Cagliari, Italy, June 3-5, 2019.
[42] Justin Zhan, Gary Blosser, Chris Yang, Lisa Singh, Privacy-preserving collaborative social networks, International Conference on Intelligence and Security Informatics, pp. 114-125, 2008.
[43] Justin Zhan, Xing Fang, Trust maximization in social networks, International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction, pp. 205-211, 2011.
[44] Justin Zhan, Vivek Guidibande, Sai Phani Krishna Parsa, Identification of top-K influential communities in big networks, Journal of Big Data, Vol. 3, Issue 1, pp. 16, 2016.
[45] Justin Zhan, Sweta Gurung, Sai Phani Krishna Parsa, Identification of top-K nodes in large networks using Katz centrality, Journal of Big Data, Vol. 4, Issue 1, pp. 16, 2017.
[46] Justin Zhan, Xing Fang, Peter Killion, Trust optimization in task-oriented social networks, 2011 IEEE Symposium on Computational Intelligence in Cyber Security, pp. 137-143, 2011.
[47] Carter Chiu and Justin Zhan, Deep Learning for Link Prediction in Dynamic Networks Using Weak Estimators, IEEE Access, Volume 6, Issue 1, pp., 2018.
[48] Moinak Bhaduri and Justin Zhan, Using Empirical Recurrences Rates Ratio for Time Series Data Similarity, IEEE Access, Volume 6, Issue 1., pp.30855-30864, 2018.
[49] Jimmy Ming-Tai Wu, Justin, Zhan, and Sanket Chobe, Mining Association Rules for Low Frequency Itemsets, PLoS ONE 13(7): e0198066. , 2018.
[50] Payam Ezatpoor, Justin Zhan, Jimmy Ming-Tai Wu, and Carter Chiu, Finding Top-k Dominance on Incomplete Big Data Using MapReduce Framework, IEEE Access, Volume 6, Issue 1, pp. 7872-7887, 2018.
[51] Pravin Chopade and Justin Zhan, Towards A Framework for Community Detection in Large Networks using Game-Theoretic Modeling, IEEE Transactions on Big Data, Volume: 3, Issue: 3, pp.276-288, 2017.
Moinak Bhaduri, Justin Zhan, and Carter Chiu, A Weak Estimator For Dynamic Systems, IEEE Access, Volume 5, Issue 1, pp. 27354-27365, 2017.
[52] Matin Pirouz and Justin Zhan, Toward Efficient Hub-Less Real Time Personalized PageRank, IEEE Access, Volume 5, Issue 1, pp. 26364-26375, 2017.
[53] Moinak Bhaduri, Justin Zhan, Carter Chiu, and Felix Zhan, A Novel Online and Non-Parametric Approach for Drift Detection in Big Data, IEEE Access, Volume 5, Issue 1, pp. 15883-15892, 2017.
[54] Carter Chiu, Justin Zhan, and Felix Zhan, Uncovering Suspicious Activity from Partially Paired and Incomplete Multimodal Data, IEEE Access, Volume 5, Issue 1, pp. 13689 - 13698 ,2017.
[55] Jimmy Ming-Tai Wu, Justin Zhan, Jerry Lin, Ant Colony System Sanitization Approach to Hiding Sensitive Itemsets, IEEE Access, Vol. 5, No. 1, pp. 10024–10039, 2017.
[56] Justin Zhan and Binay Dahal, Using Deep Learning for Short Text Understanding, Journal of Big Data, 4: 34, pp. 1-15, 2017.
[57] Justin Zhan, Timothy Rafalski, Gennady Stashkevich, Edward Verenich, Vaccination Allocation in Large Dynamic Networks, Journal of Big Data, 4:2, 2017.
[58] Zhan, J., Oommen, J., and Crisostomo, J., Anomaly Detection in Dynamic Systems Using Weak Estimator, *ACM Transaction on Internet Technology*, Vol. 11, No. 1, pp. 53-69, 2011.
[59] Zhan, J., Hsieh C., Wang, I., Hsu, T., Liau, C., and Wang D., Privacy-Preserving Collaborative Recommender Systems, *IEEE Transaction on Systems, Man, and Cybernetics*, Part C, Volume 40, Issue 4, pp. 472-476, 2010.
[60] Wang, I., Shen, C., Zhan, J., Hsu, T., Liau, C. and Wang, D., Empirical Evaluations of Secure Scalar Product, *IEEE Transactions on Systems, Man, and Cybernetics*, Part C, Vol. 39, Issue 4, pp. 440-447, 2009.
[61] Andrea Hart, Brianna Smith, Sean Smith, Elijah Sales, Jacqueline Hernandez-Camargo, Yarlin Mayor Garcia, Felix Zhan, Lori Griswold, Brian Dunkelberger, Michael R. Schwob, Sharang Chaudhry, Justin Zhan, Laxmi Gewali, Paul Oh, Resolving Intravoxel White Matter Structures in the Human Brain Using Regularized Regression and Clustering, Journal of Big Data, 2019.
[62] Jimmy Ming-Thai Wu, Justin Zhan, and Jerry Chun-Wei Lin, An ACO-based Approach to Mine High-Utility Itemsets, Knowledge-Based Systems, Vol. 116, pp. 102-113, 2017.
[63] Justin Zhan, Vivek Gudibande, and Sai Phani Krishna Parsa, Idenfication of Top-K Influential Communities in Large Networks, *Journal of Big Data*, 3:16, 2016.
[64] Matin Pirouz, Justin Zhan, Node Reduction in Personalized Page Rank Estimation for Large Graphs, *Journal of Big Data*, 3-12, 2016.
[65] Haysam Selim, Justin Zhan, Towards Shortest Path Identification on Large Networks, *Journal of Big Data*, 3-10, 2016.