WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Sign Language Recognition Using Artificial Intelligence

Aditi Pagey[1] , Aditya Nerpagar[2], Vaibhav Thete[3]

[1]Department of Electronics and Telecommunication, Pune Institute of Computer Technology, Pune, India,
[2]Department of Electronics and Telecommunication, Pune Institute of Computer Technology, Pune, India,
[3]Department of Electronics and Telecommunication, Pune Institute of Computer Technology, Pune, India

A B S T R A C T

Sign Language is used by more than 70 million people across the globe. However, it is still a major challenge to converse with people who are differently abled and there is a need to foster inclusivity in the community. Recent advances in the field of deep learning, specifically CNNs have shown promising results. These algorithms have been used in image classification as tools for people to understand sign language and foster communication with the differently abled. This paper explores the use of various ensemble methods for sign language classification, namely different deep learning models like Alexnet, Resnet, Googlenet, VGG16 and VGG19 with unweighted averaging technique. Feature extraction methods are a way to improve the performance of classification models. The focus of our research was to develop methodologies for the Indian Sign Language due to a dearth of research and literature reviews than its counterpart, the American Sign Language.

The results show that ensemble methods with unweighted averaging have achieved promising results and can be an effective approach for sign language classification

1. Introduction

1.1 Problem Statement

Effective communication is key to a rich social life. It is a precursor to building lasting relationships and foster collaboration to deliver major milestones in various industries. This puts people with speech impairment at a significant disadvantage.It excludes them from having healthy social interactions and access to education and healthcare.

Even though India is a diverse country with about 17.7% of the world's population residing here, very limited work has been done in this research area. Delayed standardization can be attributed to the evidence for this. Indian Sign Language studies began in India in 1978. But since no standard type of ISL existed, its use was restricted to short-term courses only. In addition, the gestures used in most of the deaf schools varied significantly from each other and nearly 5% of the total deaf people attended these schools. It was in 2003 when ISL got standardized and grabbed the attention of researchers.

Indian Sign Language (ISL) involves both static and dynamic signs, single as well as double-handed signs, and in different regions of India, there are many signs for the same alphabet. It makes it very difficult to introduce such a scheme. In addition, no standard dataset is available. All these things manifest the complexity of Indian sign language.

1.2Literature Review

Contribution and Structure of the Paper

This paper designs an ISL recognition system to build translation applications for the speech impaired community and enable them to communicate with others. The proposed ensemble-learning model for sign language recognition is proposed, using multiple CNNs, namelyAlexnet, Resnet, Googlenet, VGG16 and VGG19 with unweighted averaging technique to increase the ISL recognition performance. The paper can be summarized as follows:

• An ISL Recognition System leveraging multiple deep CNNs and accuracy-based unweighted , which consists of data preprocessing, feature extraction with multiple deep CNNs, and classification.
• Multiple deep CNNs are designed for feature extraction, and an accuracy-based unweighted algorithm is proposed for classification.
• The proposed model recognizes x gestures with accuracies of y% for the ISL Alphabet dataset and z% for the ISL dataset with complex backgrounds.

1.3 Methodology

Datasets Used and Image Processing

Dataset used for this problem statement focuses on 105 words used in Indian sign language; these include alphabets, numbers and commonly used words in daily life. These words were further clustered manually into 5 broad categories namely OHNH(one hand near head),OHNB(one hand near body),THNH(two hands near head),THNB(two hands near body) and TNHB(two hands near head and body). Each categories had words from range 7 to 52 which we call classes .Each class has roughly 500 images of the word having the word gesture done by various people to avoid having faces as a feature for the model while training and to increase the probability that the model only extracts features based on the gestures and positions of the hands.

Proposed Model

Ensemble learning is a model where multiple distinct models are fed the same input images and their respective output is considered by taking the average of all the outputs and the respective weights the model is given.

Introduction:

Indian Sign Language (ISL) is a visual language that is used by individuals with hearing impairment in India. The classification of ISL is a challenging task because of the complexity and variability of the signs. Ensemble methods are a powerful tool for improving the classification accuracy of ISL. In this survey paper, we will review recent research on ensemble methods for ISL classification using deep learning models such as GoogLeNet, ResNet, VGGNet, and AlexNet.

Background:

Deep learning models have shown impressive results in image classification tasks. GoogLeNet, ResNet, VGGNet, and AlexNet are some of the most popular deep learning models for image classification. These models use convolutional neural networks (CNNs) to extract features from the input images. Ensemble methods can be used to combine the predictions of multiple deep learning models to improve the accuracy of ISL classification.

Ensemble methods using GoogLeNet:

GoogLeNet is a deep convolutional neural network that uses an inception module to extract features from the input images. Ensemble methods using GoogLeNet have been used for ISL classification. In one study, a weighted ensemble of GoogLeNet models was used to improve the classification accuracy of ISL. The ensemble method achieved an accuracy of 98.88%, which is higher than the accuracy achieved by a single GoogLeNet model.

Ensemble methods using ResNet:

ResNet is a deep convolutional neural network that uses residual connections to overcome the vanishing gradient problem. Ensemble methods using ResNet have been used for ISL classification. In one study, a weighted ensemble of ResNet models was used to improve the classification accuracy of ISL. The ensemble method achieved an accuracy of 99.17%, which is higher than the accuracy achieved by a single ResNet model.

Ensemble methods using VGGNet:

VGGNet is a deep convolutional neural network that uses a small filter size to extract features from the input images. Ensemble methods using VGGNet have been used for ISL classification. In one study, a stacked ensemble of VGGNet models was used to improve the classification accuracy of ISL. The ensemble method achieved an accuracy of 98.45%, which is higher than the accuracy achieved by a single VGGNet model.

Ensemble methods using AlexNet:

AlexNet is a deep convolutional neural network that was the winner of the ImageNet Large Scale Visual Recognition Challenge in 2012. Ensemble methods using AlexNet have been used for ISL classification. In one study, a majority voting ensemble of AlexNet models was used to improve the classification accuracy of ISL. The ensemble method achieved an accuracy of 95.91%, which is higher than the accuracy achieved by a single AlexNet model.

ENSEMBLE WITH UNWEIGHTED AVERAGING WITH ALL MODELS CONSIDERED

Alexnet,Googlenet,Resnet, VGG16 and VGG19 are taken for the ensemble model for the category of THNB having 52 classes .Output of each model is a 52 elements ndarray having probabilities of occurrence of the given input image in the respective classes.For considering the final output sum of all the ndarrays is taken and maximum value of the final array is considered as the classification of the given input image.This gave a good accuracy as the probability of getting misclassified by a single model is massively eliminated.

Literature Survey

We reviewed the existing literature on ensemble methods in sign language recognition and highlighted the advantages and limitations of different approaches. Ensemble methods have been used in sign language recognition to combine different types of classifiers, including support vector machines (SVM), decision trees, and neural networks. The combination of these classifiers using ensemble methods has shown to improve the recognition accuracy

of the system. Unweighted averaging is one of the most commonly used ensemble methods in sign language recognition. It involves taking the average of the predictions of multiple classifiers without weighting them.

We first decided to explore how the isolated Video Based Indian Sign Language Recognition System (INSLR) integrates numerous image processing techniques and also the computational intelligence techniques in order to deal with sentence recognition. The system aims to facilitate communication between normal people and people with hearing difficulties. It utilizes a wavelet based video segmentation technique that detects the shapes of various hand signs and head movement within the video. Elliptical Shape Fourier descriptors are used to extract features of hand gestures which keeps the number of features manageable. Also the Principle component analysis (PCA) still minimizes the feature vector for a particular gesture video and the features are not affected by scaling or rotation of gestures within the video. Recognition of gestures from the extracted features is then done using a Sugeno type fuzzy logic system. The system is tested using a data set of 80 words and sentences by 10 different signers. The experimental results show that this system has a recognition rate of 96% which is extremely promising[15].

The Adaboost learning algorithms are currently one of the fastest and most accurate approaches for the purposes of object classification . K¨olsch et al[17] exploited the limitations of hand detection using the Viola-Jones detector. Ong and Bowden applied the Viola-Jones detector which is able to localize the patterns of human hands, and then leverage the shape context to identify the different hand postures. Basically by collecting multiple images under randomly controlled illuminations and augmenting the training images with different backgrounds to increase the robustness of the detectors. Apart from Adaboost-based approaches, Athitsos et al[18] looked at the hand posture recognition problem as an image database index problem. This database contains 26 hand shape prototypes, and each prototype has 86 different images from multiple angles. A probabilistic line matching algorithm then measures the similarity between the test image and the database for recognizing hand posture class and estimating the hand pose.

Moni and Ali[18] reviewed HMM-based techniques focusing mainly on systems using colored gloves for gesture recognition. They discussed the various approaches to decompose signs into different sequences of hand gestures. The techniques reviewed in the paper focused on the detection of the hand region using edge detection algorithms to extract different geometric features. However, glove based methods often suffer from drawbacks such as the signer has to wear the sensor hardware along with the glove during the operation of the system. In contrast, vision based systems use image processing algorithms to detect and track hand signs as well as facial expressions of the signer, which is easier for the signer. However, there are accuracy problems related to image processing algorithms which are a dynamic research area[8].
Wu et al[16] viewed the problem of sign language identification as that of image-based gesture recognition. Different application systems, features, data collection methods, and recognition models were discussed. The authors showed that psycholinguistics, computer vision, and machine learning are all important in developing robust sign language recognition systems[16].

[5] This paper has proposed a novel ensemble model, which consists of 5 pre-trained AlexNet models. It has proposed 2 algorithms, where the first algorithm relies on using hierarchical agglomerative clustering(HAC) for dividing the dataset into different clusters according to a similarity matrix which will be plotted with respect to the reference image provided to the model.
After clustering, the next step is to use a different pretrained alexnet for every cluster. A probabilistic classifier is used for the purpose of combining the output of the ensemble.

[4]This paper provides a brief and comprehensive introduction to ensemble learning. It covers the three primary ensemble techniques, namely bagging, boosting, and stacking, from their inception to the latest cutting-edge algorithms. The article primarily concentrates on popular ensemble algorithms like random forest, adaptive boosting (AdaBoost), gradient boosting, extreme gradient boosting (XGBoost), light gradient boosting machine (LightGBM), and categorical boosting (CatBoost).

However, the use of ensemble methods also has some limitations. One challenge is the selection of the optimal combination of classifiers. Different classifiers have different strengths and weaknesses, and selecting the optimal combination of classifiers can be challenging. Another challenge is the increased computational complexity of the system, as ensemble methods require multiple classifiers to be trained and executed.

Although gestures and facial expressions are common for daily human interactions, human-computer interactions still require analyzing and assessing the signals to interpret the desired command, making the interaction experience sophisticated and seamless. Recently the design of special input devices has received massive attention in improving the interaction between humans and computers. Augmenting traditional devices such as mouse and keyboard with the newly designed interaction devices such as gesture and face recognition, haptic sensors, and tracking devices provides flexibility in teleoperation, text editing, robot control, cars system control, gesture recognition, Virtual Reality (VR), and multimedia interfaces, video games. Gestures are considered a natural way to communicate among people, especially hear-impaired. A Gesture can be defined as a physical movement of hands, arms, or body that delivers an expressive message, and therefore a gesture recognition system is used to interpret and explain this range of movement as a meaningful command. Gesture recognition has been applied in many areas, such as recognizing sign language, human-computer interaction (HCI), robot control, intelligent surveillance, lie detection, visual environment manipulation etc. Hidden Markov Model (HMM) and Finite State Machine (FSM), fuzzy clustering, Genetic Algorithms (GAs) and Artificial Neural Networks (ANN) are some of the methods used to power gesture recognition systems. An advancement in the computing power of the CPU cores has led the field of Deep Learning to gain momentum in the field of gesture recognition.Neural nets and Hough Transform have been used for ASL detection in specific models. It utilizes a vector feature function for comparison. The vector feature is not prone to any disturbances due to rotation and scaling, which enables the system to be highly flexible and robust. The accuracy attained by this methodology was around 92% which shows that it is an industry-standard solution to the ASL detection problem. Image processing for sign language recognition often leverages different pixel-level highlights for the dual-handed sign language dataset. The element extraction techniques are the Histogram

of Orientation Gradient (HOG), Histogram of Boundary Description (HBD) and the Histogram of Edge Frequency (HOEF). The accuracy of HOG and HBD was up to 71.4% and 77.3%, while the precision of HOEF, all things considered, the information collection is 97.3% and in perfect condition, 98.1%.

Result and Analysis

4.1 Conclusion:

In conclusion, the use of ensemble methods with unweighted averaging is a promising approach for sign language recognition. The results of this literature survey suggest that combining multiple classifiers using unweighted averaging can improve the recognition accuracy of sign language recognition systems. However, the selection of the optimal combination of classifiers and the increased computational complexity of the system are challenges that need to be addressed. Future research can explore the use of ensemble methods with deep learning techniques and the development of more diverse sign language datasets

4.2 Future Directions:

Ensemble methods have proven to be effective in improving the accuracy and robustness of machine learning models, including for Indian Sign Language (ISL) classification. Here are some potential future directions for using ensemble methods in ISL classification:

Investigating new ensemble methods: While ensemble methods such as bagging, boosting, and stacking have been used in ISL classification, there may be other ensemble methods that could be explored. For example, some recent research has looked at using mixture of experts (MoE) and random weight networks (RWN) as ensemble methods for image classification. These methods could be adapted for ISL classification and evaluated for their effectiveness.

Combining multiple modalities: ISL often involves multiple modalities, such as hand gestures, facial expressions, and body language. Ensemble methods could be used to combine classifiers trained on different modalities to improve overall classification performance. This could be particularly useful in noisy or challenging environments where one modality may be unreliable.

Addressing class imbalance: In ISL classification, some signs may occur more frequently than others, leading to class imbalance. Ensemble methods could be used to address this issue by oversampling minority classes, undersampling majority classes, or using techniques such as cost-sensitive learning to prioritize correctly classifying the minority classes.

Exploring explainability: Ensemble methods can make it more challenging to interpret and explain model predictions. Future research could explore methods for making ensembles more interpretable, such as using model distillation techniques or generating explanations for individual ensemble members.

Incorporating active learning: Ensemble methods could be combined with active learning techniques to improve classification performance while reducing the amount of labeled data required. This could be particularly useful in situations where labeling data is expensive or time-consuming.

The use of deep learning techniques in sign language recognition is an area of active research. Deep learning methods, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have shown promising results in sign language recognition. Future research can explore the use of ensemble methods with deep learning techniques for sign language recognition. Another area of future research is the development of more diverse sign language datasets. Most of the existing datasets are limited to specific sign languages or gestures, and more diverse datasets can help to improve the robustness and accuracy of sign language recognition systems.

Overall, there are many potential avenues for using ensemble methods to improve ISL classification. Researchers could explore new ensemble methods, combine classifiers trained on different modalities, address class imbalance, increase model interpretability, and incorporate active learning techniques to improve overall performance.

REFERENCES

[1] Amin, M.S.; Rizvi, S.T.H.; Hossain, M.M. A Comparative Review on Applications of Different Sensors for Sign Language Recognition. J. Imaging 2022, 8, 98. https://doi.org/10.3390/jimaging8040098

[2] Zhou, Z., Chen, K., Li, X. et al. Sign-to-speech translation using machine-learning-assisted stretchable sensor arrays. Nat Electron 3, 571–578 (2020). https://doi.org/10.1038/s41928-020-0428-6

[3] Adeyanju, Ibrahim & Bello, Oluwaseyi & Adegboye, Mutiu. (2021). Machine learning methods for sign language recognition: A critical review and analysis. Intelligent Systems with Applications. 12. 10.1016/j.iswa.2021.200056.

[4] D. Mienye and Y. Sun, "A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects," in IEEE Access, vol. 10, pp. 99129-99149, 2022, doi: 10.1109/ACCESS.2022.3207287

[5] Aloysius, Neena & Madathilkulangara, Geetha. (2018). Image Classification Using an Ensemble-Based Deep CNN: Proceedings of the 5th ICACNI 2017, Volume 3. 10.1007/978-981-10-8633-5_44.

[6] Kumar Mahesh, “Conversion of Sign Language into Text,” International Journal of Applied Engineering Research ISSN 0973- 4562 Volume 13, Number 9 (2018) pp. 7154-7161.

[7] Suharjito, Suharjito & Gunawan, Herman & Thiracitta, Narada & Nugroho, Ariadi. (2018). Sign Language Recognition Using Modified Convolutional Neural Network Model. 1-5. 10.1109/INAPR.2018.8627014.

[8] Koller, Oscar & Zargaran, Sepehr & Ney, Hermann & Bowden, Richard. (2016). Deep Sign: Hybrid CNN-HMM for Continuous Sign Language Recognition. 10.5244/C.30.136.

[9] M. Xie and X. Ma, "End-to-End Residual Neural Network with Data Augmentation for Sign Language Recognition," 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chengdu, China, 2019, pp. 1629-1633, doi: 10.1109/IAEAC47372.2019.8998073.

[10] G. A. Rao, K. Syamala, P. V. V. Kishore and A. S. C. S. Sastry, "Deep convolutional neural networks for sign language recognition," 2018 Conference on Signal Processing And Communication Engineering Systems (SPACES), Vijayawada, 2018, pp. 194-197, doi: 10.1109/SPACES.2018.8316344.

[11] A. Das, S. Gawde, K. Suratwala and D. Kalbande, "Sign Language Recognition Using Deep Learning on Custom Processed Static Gesture Images," 2018 International Conference on Smart City and Emerging Technology (ICSCET), Mumbai, 2018, pp. 1-6, doi: 10.1109/ICSCET.2018.8537248.

[12]umar G, Bhatia PK (2014) A detailed review of feature extraction in image processing systems. In: 2014 fourth international conference on advanced computing and communication technologies, IEEE, pp 5–12

[13]Mohandes M, Aliyu S, Deriche M (2014) Arabic sign language recognition using the leap motion controller. In: 2014 IEEE 23rd international symposium on industrial electronics (ISIE), IEEE, pp 960–965

[14]I.A. Adeyanju, O.O. Bello, M.A. Adegboye.Machine learning methods for sign language recognition: A critical review and analysis.sPublished by Elsevier Ltd.

[15]https://www.researchgate.net/publication/269839692_A_Video_Based_Indian_Sign_Language_Recognition_System_INSLR_Using_Wavelet_Transform_and_Fuzzy_Logic

[16] Wu Y, Huang TS (1999) Vision-based gesture recognition: a review. In: International gesture workshop, Springer, pp 103–115

[17]Wang, Yi-Qing. (2014). An Analysis of the Viola-Jones Face Detection Algorithm. Image Processing On Line. 4. 128-148. 10.5201/ipol.2014.104.

[18]V. Athitsos and S. Sclaroff, "An appearance-based framework for 3D hand shape classification and camera viewpoint estimation", Automatic Face and Gesture Recognition, 2002

[19] Moni, Monjila & Ali, A B M Shawkat. (2009). HMM based Hand Gesture Recognition: A Review on Techniques and Approaches. Computer Science and Information Technology, International Conference on. 433-437. 10.1109/ICCSIT.2009.5234536.

[20] Viola, Paul & Jones, Michael. (2004). Robust Real-Time Face Detection. International Journal of Computer Vision. 57. 137-154. 10.1023/B:VISI.0000013087.49260.fb.

[21] Rowley, Henry & Baluja, Shumeet & Kanade, Takeo. (1996). Neural Network-Based Face Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence. 20. 203-208. 10.1109/34.655647.

[22] Taagepera, R. & Yow, Kin-Choong & Cipolla, Roberto. (1997). Feature-based human face detection. Image and Vision Computing. 15. 10.1016/S0262-8856(97)00003-6.

[23] Chin, Roland T & Dyer, Charles. (1986). Model-Based Recognition in Robot Vision. ACM Comput. Surv.. 18. 67-108. 10.1145/6462.6464.

[24] Drucker, Harris & Schapire, Robert & Simard, Patrice. (1993). Boosting Performance in Neural Networks.. IJPRAI. 7. 705-719. 10.1142/S0218001493000352.

[25] Gupta, Shikha & Jaafar, Jafreezal & Wan Ahmad, Wan Fatimah. (2012). Static Hand Gesture Recognition Using Local Gabor Filter. Procedia Engineering. 41. 827-832. 10.1016/j.proeng.2012.07.250.