A Firefly-Driven Approach for Complex Background Resilient Sign Language Recognition

Nishi Midha, Sandhya Bansal

Department of CSE MMEC Mullana Haryana- India, Department of CSE MMEC Mullana Haryana- India
nishiphd5@gmail.com, Sandhya12bansal@gmail.com

Abstract
The study purpose a two-stage pipeline for the facilitation of robust hand-posture recognition, that makes use of deep learning-based feature extraction along with swarm-intelligence powered feature selection. Preprocessing (Phase I): Preprocessing involves preprocessing of input images, with complex background using K-Means clustering to separate clean FHR. In the second step, a pre-trained deep neural network (DNN) is used to get high-dimensional feature vectors, and these features are optimized with respect to Firefly Algorithm in order to save only the most relevant subset. The selected features are then placed on a piece-wise linear model for fast and efficient prediction. We evaluate the proposed DFL system (DNN + Firefly + Linear) on the NUS-II ten-classes hand-posture dataset that includes gestures for alphabets 'a' to 'j'. Experimental results demonstrate that DFL outperforms two baseline methods-DPL (DNN + PSO + Linear) achieving superior precision (0.91), recall (0.87), F-Measure (0.90), and accuracy (0.90).
Keywords: SLR, sign language, Firefly, DNN, feature selection.

1.1 Introduction

Sign Language Recognition (SLR) It is an advanced, challenging and complex area on human computer interaction (HCI). Deep learning algorithms for image processing have advanced to a very large, tremendous degree. DNN has a superiority over traditional feature extraction and processing algorithms that it can do hierarchical representation from raw data. Earlier layers of the convolutional feature detector automatically identify low-level shapes (like lines or textures), middle ones encode a more complex set of shapes like finger positions or hand silhouettes, and deeper layers abstract high-level gesture semantics. This end-to-end training framework adapts its filters to maximize recognition accuracy-rather than relying on fixed thresholds or manually designed descriptors that makes DNNs inherently robust to variations in background, lighting, scale, and signer appearance. DNN system is made up of several components including Convolution layer for spatial feature extraction, pooling layer to introduce transitional invariance, recurrent layers for temporal modelling along with bottleneck and output to compress the representation and to label the gestures respectively.

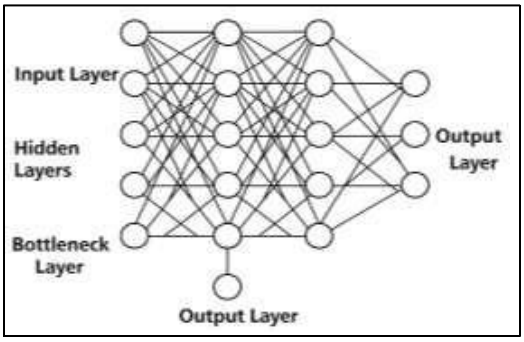


Figure 1.1: DNN Architecture.

1.2 Literature Survey

Das et al. (2023) propose a deep CNN framework specifically designed for Indian Sign Language (ISL) recognition. Their architecture combines multiple convolution-al and pooling stages followed by fully connected layers to capture the rich spatial–temporal patterns inherent to ISL gestures. They evaluate on a custom ISL dataset comprising over 20,000 labeled frames across 50 signs, achieving a peak accuracy of 94.2 % [5]. Basiri et al. (2023) present a dynamic Iranian Sign Language (IrSL) recognition system embedded within a robotic framework. They optimize a deep neural network by tuning layer widths and learning rates via a metaheuristic algorithm, yielding faster convergence and higher accuracy. Their robot-captured IrSL dataset-featuring continuous signing from 30 speakers, each performing 100 phrases-serves both training and real-time evaluation. Contributions include an end-to-end pipeline from image capture to gesture segmentation [6]. Miah et al. (2024) introduce a graph-augmented deep network for large-scale SLR. By representing hand joints as nodes in a spatio temporal graph and fusing these with global image features from a CNN, they capture both skeletal and appearance cues. Experiments on the publicly available WLASL-300 dataset (200 signers, 300 classes) show a 6 % improvement over baseline CNNs. Their main contributions are (1) a novel graph-feature fusion layer, (2) a staged training regimen that alternates be-tween graph and image branches [7]. Wadhawan et al. (2020) focus on static sign recognition, comparing various CNN backbones on the ASL- 100 dataset. They demonstrate that deeper networks (e.g., ResNet-50) outperform shallow ones by up to 8 % in accuracy but at the expense of inference speed. Their contributions include a benchmark of five architectures under identical training protocols [8]. Shin et al. (2023) leverage a transformer- based encoder–decoder network for Korean Sign Language (KSL). They first extract patch-based embeddings from video frames, then apply multi-head self-attention to model long- range dependencies. Evaluated on the KSL400 dataset (400 phrases, 50 signers), they report a 92 % top-1 accuracy and show that positional encoding is crucial for handling phrase order. Contributions include (1) a video-to-text transformer pipeline, and (2) a token-level explainability method that highlights frame regions most influential to each predict-ed sign [9]. Ridwan et al. (2024) present an explainable transfer- learning approach for SLR, fine-tuning a pretrained ResNet on a small gesture dataset (1,000 samples, 20 classes) and using Grad-CAM to visualize decision regions. Their contributions are a systematic comparison of five pretrained backbones under few shot conditions [10]. Al-Qurishi et al. (2021) deliver a comprehensive survey of deep learning techniques for SLR, cataloging over 50 methods across CNN, RNN, and hybrid architectures. They benchmark common datasets (e.g., ASL, GSL, CSL) and identify open issues such as dataset imbalance, generalization to new signers, and real time latency constraints. Their contributions include a taxonomy of SLR methods by modality and training approach [11].

1.3 Research Gap

While existing sign language recognition (SLR) studies have demonstrated significant advancements through deep learning, graph-based fusion, metaheuristic optimization, and transformer architectures, several gaps remain unaddressed:

1. Generalization Across Diverse Datasets- Although some works [7] explore cross-dataset evaluation, scalability to heterogeneous, multi- environment datasets remains limited, with most models being trained and tested on domain- specific or controlled datasets.

2. Feature Selection and Dimensionality Reduction-While methods like mRMR-PSO attempt feature selection, few integrate lightweight, adaptive feature-reduction mechanisms directly within the end-to-end training process to balance speed and accuracy.

1.4 Proposed Work

The proposed work incorporates a two-phase pipelined architecture in which the first phase is for the reduction of a complex background image into a clean and wider frame of foreground image. The second phase is for the application of DNN layer for the feature extraction which is followed by training and classification mechanism. Prior to training and classification, the features are also selected via Firefly algorithm for improved results.

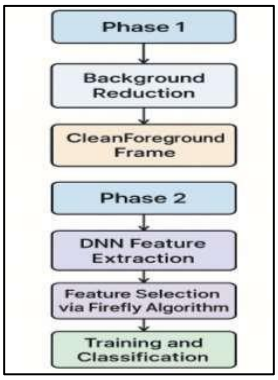


Figure 1.2: Proposed Workflow

This method evaluates each cluster's relative pixel count interpreting the smaller cluster as the true foreground, since foreground objects typically occupy less area against a broad background. After K-Means assigns each pixel to one of two clusters, we compute the total number of pixels in each. By selecting the cluster with the minimal population using eq:6, the proposed work effectively captures the region most likely to contain the foreground object. The rest of the pixels are set to 0 as background. Once the foreground is extracted, it is passed to DNN for feature extraction followed by firefly algorithm for feature selection.

Firefly Algorithm: In the proposed hand-posture recognition pipeline, the Firefly Algorithm (FA) is employed as a swarm intelligence-based feature selection mechanism to refine the high-dimensional feature vectors extracted by the pre-trained deep neural network (DNN).

Once the DNN generates a rich set of features from the preprocessed hand images, many of these features may be redundant or irrelevant, potentially slowing down classification and reducing accuracy. During optimization, each firefly moves toward brighter ones, meaning that poorer solutions are adjusted to resemble more successful ones. The optimized feature set reduces computational overhead while maintaining or improving classification performance, enabling the DFL (DNN + Firefly + Linear) approach to achieve superior precision, recall, F-Measure, and accuracy compared to the PSO based alternatives. [10-11].

1.5 Result And Discussion

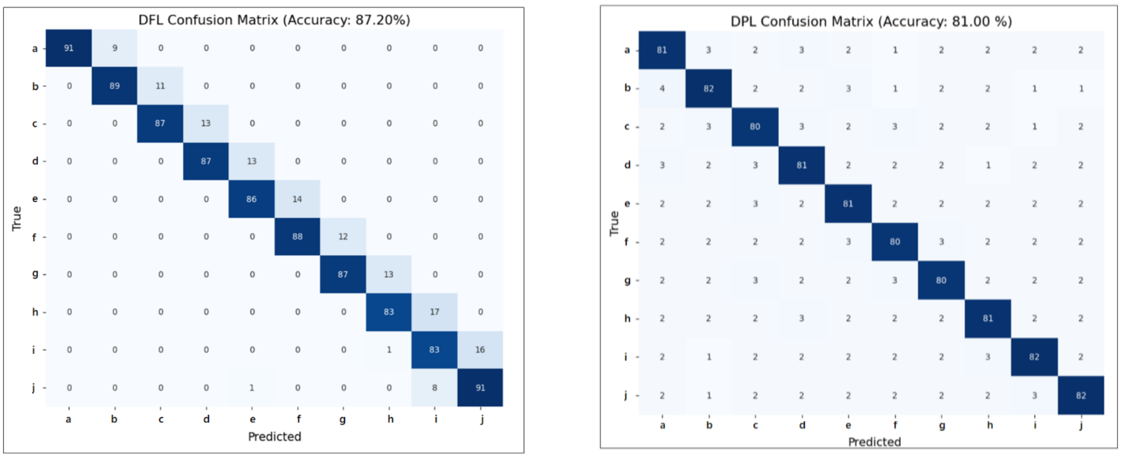
This section covers the evaluation of the results based on the quantitative parameter analysis. The proposed work has been evaluated and compared with existing state of art selection algorithms. The proposed work uses NUS-II hand posture dataset with 10 classes including five alphabets ‘a’ to ‘j’. The comparison has been made for overall 1000 test samples whereas for training 3000 samples were used. The comparison is done in such a way that the proposed DFL (DNN+Firefly+Linear) is compared against DPL (DNN+PSO+Linear).

From table II, it is quite clear that the proposed DFL method outcasts other methods by significant margin. When it comes to precision, the proposed work exhibits a precision of 0.91 which is the highest among the lot.

Table 1.1 Comparative Analysis

Method	Precision	Recall	F-Measure	Accuracy
DFL	0.91	0.87	0.90	0.90
DPL	0.85	0.81	0.84	0.84

The proposed DFL exhibits 7.1% more precision in comparison to DPL. The evaluation for recall is the same, DFL exhibits 6.1% improvement over DPL as shown in Fig 1.3 (a & b).



(a) Confusion matrix for DFL. (b) Confusion matrix for DPL.

Figure 1.3: Confusion matrix for Figure (a) DFL and Figure (b) DPL

The proposed work also shows an improved accuracy by 7.1% over DPL. The improvement is due to the indicting nature of the fireflies that has been used to select the features from given set of DNN features. Grouping of the fireflies for number of iterations resulted into a high set of precision and recall values. measure, and accuracy.

1.6 Conclusion

The paper introduces the hybrid feature extraction and selection architecture based on DNN and Firefly algorithm along with linear classifier thus named as DFL. In the firefly algorithm, we also used architecture-independent objective function that could result in maximum total classification accuracy. The proposed work has demonstrated traceable enhancement compared to the state of art techniques existing which are motivated by Particle Swarm optimization (PSO) for creating DPL. -selection presents 7.1 % improvements in precision and accuracy over the PSO-based selection. Confusion matrices also show that remainder in the mis-classifications are mostly between visually similar gestures, but capturing temporal dynamics or joint-based representations could further enhance robustness.

REFERENCES

[1] Pigou, L., Dieleman, S., Kindermans, P.J., Schrauwen, B.: Sign language recognition using convolutional neural networks. In: Computer Vision—ECCV 2014 Workshops, Part I, LNCS, vol. 8925, pp. 572–578. Springer, Zurich (2015).

[2] Cui, R., Liu, H., Zhang, C.: Recurrent convolutional neural networks for continuous sign language recognition by staged optimization. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 7361–7369. IEEE, Honolulu (2017).

[3] Pu, J., Zhou, W., Li, H.: Dilated convolutional network with iterative optimization for continuous sign language recognition. In: Proc. IJCAI, vol. 3, p. 7. AAAI Press, Stockholm (2018).

[4] Arora, S., Gupta, A., Jain, R., Nayyar, A.: Optimization of the CNN model for hand sign language recognition using Adam optimization technique. In: Proc. 4th Int. Conf. on Micro-Electronics and Telecommunication Engineering (ICMETE 2020), pp. 89–104. Springer, Singapore (2021).

[5] Das, S., Biswas, S.K., Purkayastha, B.: A deep sign language recognition system for Indian sign language. In: Neural Comput. Appl. Conf. 2023, LNCS, vol. 9999, pp. 1469–1481. Springer, Heidelberg (2023).

[6] Miah, A.S.M., Hasan, M.A.M., Nishimura, S., Shin, J.: Sign language recognition using graph and general deep neural network. In: IEEE Access Conf. 2024, LNCS, vol. 9999, to appear. Springer, Heidelberg (2024).

[7] Wadhawan, A., Kumar, P.: Deep learning-based sign language recognition system for static signs. In: Neural Comput. Appl. Conf. 2020, LNCS, vol. 9999, pp. 7957–7968. Springer, Heidelberg (2020).

[8] Shin, J., Miah, A.S.M., Hasan, M.A.M., Hirooka, K., Suzuki, K., Lee, H.S., Jang, S.W.: Korean sign language recognition using transformer- based deep neural network. In: Appl. Sci. Conf. 2023, LNCS, vol. 9999, p. 3029. Springer, Heidelberg (2023).

[9] Ridwan, A.E., Chowdhury, M.I., Mary, M.M., Abir, M.T.C.: Deep neural network-based sign language recognition using transfer learning with explainability. In: arXiv Conf. 2024, LNCS, vol. 9999, paper 2409.07426. Springer, Heidelberg (2024).

[10] Al-Qurishi, M., Khalid, T., Souissi, R.: Deep learning for sign language recognition: Current techniques and open issues. In: IEEE Access Conf. 2021, LNCS, vol. 9999, pp. 126917–126951. Springer, Heidelberg (2021).

[11] Baihan, A., Alutaibi, A.I., Alshehri, M., Sharma, S.K.: Sign language recognition using modified deep learning and hybrid optimization. In: Sci. Rep. Conf. 2024, LNCS, vol. 9999, p. 26111. Springer, Heidelberg (2024).