Recognition of American Sign Language Using Modified Deep Residual CNN with Modified Canny Edge Segmentation

N. Noor Alleema (nooralleema.n@gmail.com)
SRM Institute of Science and Technology, Ramapuram Chennai
S. Babeetha
SRM Institute of Science and Technology, Ramapuram Chennai
P. Santhosh Kumar
SRM Institute of Science and Technology, Ramapuram Chennai
Saravanan chandrasekaran
S Pandiaraj
SRM Institute of Science and Technology, Ramapuram Chennai
A. Ranjith Kumar
Lovely Professional University
K. Rajkumar
Jain University (Deemed-to-be University

Research Article

Keywords: Sign language recognition, Modified convolutional neural network, Modified canny edge detection, CNN classifier, Machine learning.

Posted Date: April 7th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1521209/v1

Abstract

American Sign Language (ASL) recognition system aims to recognise hand gestures' meaningful motions, and it is a crucial solution to communicate between the deaf community and hearing people. However, existing sign language recognising algorithms still have some drawbacks, such as the difficulty of recognising hand movements low recognition accuracy for most of the sign language recognition. To address this problem, a Modified Convolutional Neural Network (MCNN) deep residual 101 classifier-based American Sign Language identification system is developed. There are three main parts present in the method. The first part is preprocessing the images to remove the noise, enhance the contrast of the picture, adjust the contrast level and smoothen the picture using various filters. The second part is the segmentation, and it's used to partition the image using a modified canny edge detection method by removing all weak edges present in the image. Finally, classification will be done using the Modified CNN deep residual 101 classifiers. By classifying the image, the American Sign Language is accurately identified. This process is conducted through images. The outcome shows that the suggested approach has a 0.95 per cent accuracy and a 0.05 per cent False Positive Rate. Other CNN such as resNet 50 and resNet 18 reached 0.90% and 0.80% accuracy, respectively, which is lower than our proposed method. In addition, the resNet 101 classifier effectively recognises the difficult hand gestures through the image data and obtain high recognition accuracy for 36 signs is 0 to 9 numbers and a to z alphabets from American Sign Language.

1. Introduction

ASL is a natural language used by deaf community peoples throughout the United States and most of the Anglophone Canadians. American Sign Language is a well-organised visual language that includes facial expressions and hand movements. In the early nineteenth century, an American school for the deaf in West Hartford created American Sign Language (Snoddon, K. 2020). ASL is closely related to the French sign language. The sign language of the deaf people in France and French-speaking citizens in Switzerland is French sign language (Woodward, J. 1996). The majority of the educational interpreters and deaf community teachers who serve hearing-impaired students in the k-12 group are adults within university programs, and females who are learning American sign language and these learners are called second modality-second language learners (M2L2) (Villwock, A et al. 2021). American Sign Language has five expressive parameters: palm orientation, hand shape, non-manual makers, location, and movement. Children learn words by incorporating the input they receive into things and events around the environment (Lieberman, A.M et al. 2021). Deaf children who are learning American Sign Language will receive linguistic input, notes and contents for its moderators through visual cues (Schniedewind, E., et al. 2020). Additionally, the continuous movements of a particular area of the body give informative cues regarding signs (Hosain, AA, et al. 2020).

Deaf persons and their friends and relatives are the only ones who learn sign languages. Because most hearing individuals are inexperienced with sign language, deaf persons are isolated in society. To address this difficulty, scientists are working on a sign language system that can transform signs from video or

picture to speech or words automatically (Mahdikhanlou, K. and Ebrahimnezhad, H, 2020). Individuals who do not understand sign language have a hard time communicating with deaf people. To address this issue, researchers created a new language known as automatic sign language (Rastgoo, R et al.2020). In the United States, American Sign Language (ASL), pidgin signed English (PSE) and signing precise English (SEE) sign languages are used. For improved comprehension, PSE is a building block established by persons who use American Sign Language and people who use manually coded English. SEE is a manual communication method that aims to accurately portray the English language and grammar (Kraus, J.C. and Hague, A.K. 2020).

The most difficult task in ASL is to identify the sign correctly. This can be accomplished by teaching computers to recognise certain indications. The precision is determined by the approaches employed for prediction and categorisation, both of which are accomplished through machine learning. For recognising American Sign Language, a Convolutional Neural Network (CNN) and Support Vector Machine (SVM) are proposed. Machine learning is a branch of AI that allows computers to learn from data without having to be explicitly programmed (Rautaray, S.S. and Agrawal, A. 2015). The categorisation is the process of categorising observations into distinct groups based on previous data used to train the unit, then predicting the category connection of a new opinion in machine learning. To achieve classification, Neural Networks, Support Vector Machines, Decision Trees, Naive Bayes and other techniques are employed. Support Vector Machines are a sort of supervised learning model that uses classification and regression techniques to analyse data. Kernel Approach is a low-dimensional input-to-high-dimensional feature-space mapping technique that can be used in conjunction with SVM to do non-linear classification (Chen, Q. et al. 2007).

A Convolutional Neural Network (CNN) is a collection of one or more layers of deep feed-forwarding networks. CNN is very effective at dealing with data that is both visual and two-dimensional (Jia, Y. et al.2014). Vision-based and sensor systems are the two approaches used to recognise Sign language. To capture hand gestures, sensor-based techniques use a range of sensors. To track the hand emotions (IMUs), Flex sensors and Inertial Measurement Units are to be used (Lee, B.G. and Lee, S.M. 2018). The finger is fitted with a flex sensor that measures the bending angle and converts it to a digital output. The flex sensor, on the other hand, is unable to determine the orientation of the hand and fingers. To solve the problem, they used an IMU sensor (Wang, J. and Zhang, T. 2014). The IMU uses a template matching method to classify hand motions. IMU has a 91 per cent accuracy rate in recognising training gestures. For recognising sign language, a vision-based technique provides another strategy. Activities are recorded using a video camera or a visual sensor, which are then analysed using motion algorithms. For 48 ASL signals, the Dynamic Bayesian Network obtains the highest level of accuracy (DBN) (Mummadi, C.K. et al. 2018).

Microsoft and leap motion have developed a revolutionary way for detecting and tracking hand and body motions with the introduction of Kinect and the leap motion controller mechanism. Kinect detects and tracks hands by recognising the human skeleton, whereas LMC's built-in cameras and infrared sensors are only used to detect and track hands (Bakken, J.P et al. 2020). However, similar to how every country

has its unique language, sign language is not universally accepted (Cheok, M.J. et al.2017). So modified CNN is proposed to recognise the sign language.

The remaining part of the paper contains: Portion 2 illustrates certain research papers that are related to recognising American Sign Language. Portion 3 contains the proposed methodology of the paper. Portion 4 contain the result and discussion, and Portion 5 contain the conclusion part of the paper.

2. Literature Survey

Sevgi Z. Gurbuz et al. (Gurbuz, S.Z. et al. 2020) had suggested Recognising American Sign Language with an RF sensor (ASL). The paper's major purpose was to use RF sensors in HCI applications for Deaf community members. A multi-frequency RF sensor network was used to provide non-invasive, non-contact ASL signing data that were not affected by lighting conditions. A machine learning (ML) approach is used to investigate linguistic aspects of RF ASL data. With a 72.5 per cent accuracy, data from an RF sensor network was used to classify 20 native ASL signs. With a 72.5 per cent accuracy, data from an RF sensor network was used to classify 20 native ASL signs.

Ankita Wadhawan et al (Wadhawan, A. and Kumar, P. 2020) had proposed a sign language recognition system for static signs using a deep learning algorithm. In the context of sign language recognition, estimation of the approach based on deep learning-based Convolutional Neural Network (CNN) provides good modelling of static signals. For this strategy, 35000 sign photos of 100 static signs were obtained from various operators. This method's accuracy was tested on roughly 50 CNN models, with the maximum accuracy for coloured and grayscale images being 99.72 and 99.90, respectively. Human sign recognition is a multidisciplinary topic that has yet to be solved.

C.K.M. Lee et al. (Lee, C.K. et al. 2021) had suggested utilising a recurrent neural network detection and training method for American Sign Language. The aim of the method was an ASL learning prototype. In the ASL alphabets, there are both static and dynamic signs. The classification approach employs a Long-Short Term Memory Recurrent Neural Network with the k-Nearest Neighbour approach and is based on the handling of input sequences. 100 samples were gathered for each letter, and the technique was trained with 2600 samples. The recognition rate for 26 ASL alphabets is 99.44 per cent accurate on average. The controller's sample frequency, on the other hand, is inconsistent. It needs to be post-processed to reduce the influence on real-time recognition systems.

Dongxu Li et al. (Li, D et al. 2020) had presented Deep Sign Language Recognition from Video by using a New Large-Scale Dataset and techniques. A new large-scale World Level American Sign Language (WLASL) video collection with over 2000 phrases has been added to the method. In this new big-scale data set, we used multiple deep learning algorithms to recognise word-level signs and evaluate their performance in large scale scenarios. The technique had a top 10 accuracy of 62.63 per cent on a total of 2000 words.

Maria Parelli et al. (Parelli M et al. 2020) used 3D hand pose estimation and deep learning to recognise sign language from RGB videos. The objective was to infer 3D coordinates for the hand joint from RGB data using a sophisticated architecture that was previously suggested in the literature for the problem of 3D human posture prediction. The use of RGB-D sensors and depth sensors to analyse 3D hand positions has received more attention. However, indicators are not caught in the great majority of SL corpora and in real-world circumstances because depth sensors and depth information are lacking.

Nikolas Adaloglou et al. (Adaloglou, N.M et al. 2021) had presented a sign language identification system based on a deep learning algorithm. To recognise the sign language, they used a computer vision-based technique. The method's goal was to provide information on how to recognise sign language by mapping unsegmented video streams to glosses. However, glosses generally display a significantly shorter time duration than actions.

Razieh Rastgoo et al. (Rastgoo, R et al.2020) had developed a deep cascaded model is utilised to recognise video-based isolated hand sign language recognition. Single Shot Detector (SSD), Convolutional Neural Network (CNN), and Long Short Term Memory (LSTM) deep learning approaches were utilised to develop a systematic cascaded model for sign language recognition that compensates for spatiotemporal hand-based input. In comparison to other models, the complexity of this one had the lowest parameters.

Khadijeh Mahdikhanlou and Hossein Ebrahimnezhad (Mahdikhanlou, K. and Ebrahimnezhad, H. 2020) had suggested that to recognise multimodal 3D American Sign Language, the static alphabet and integers, hand joints, and form coding are used. The method's goal is to recognise a Sign Language system employing cutting-edge multimedia that includes a webcam and a Leap Motion Controller (LMC). The approach computes two sets of characteristics, the first of which was the LMC sensor's calculation of angle at hand joints. The second set of attributes is derived from a webcam-provided hand shape contour. In any case, descriptors established for RGB photos perform poorly when it comes to portraying depth images due to their poor texture, high rate noise, and low resolution.

Razieh Rastgoo et al. (Rastgoo, R. et al. 2021) had presented Deep networks, and the SVD method can recognise real-time isolated hand sign language. They presented an innovative and rebellious route to apply SVD for integration of rated 3D keystrokes in order to gain more distinguishing traits. Each sign in the model had a recognition accuracy of more than 99 per cent, although some of the RKS-PERSIAN SIGN datasets were misclassified due to some examples with higher inter-class similarities that the model was unable to distinguish.

Maher Jebali et al. (Jebali, M., et al. 2021) had suggested Multimodal sensor fusion was employed for video-based continuous sign language recognition. They introduced a new method for recognising accurate word boundaries in an unbroken sign language. When compared to other challenges mentioned in the literature, the approach was effective in extracting independent signals from video. The method's accuracy was 95.18 per cent for one-hand gestures and 93.87 per cent for two-hand motions. Techniques

based on special devices did not arouse on a large scale due to the challenges created by the motion limit.

3. Proposed Methodology

The approach was designed to detect and recognise American Sign Language using machine learning algorithms. In order to solve this issue, a machine learning method and a Convolutional Neural Network were applied. Picture capture, image preprocessing, image segmentation, and image classification are the four steps of the suggested sign language recognition system. The architecture of the proposed system is illustrated in figure 1. The initial phase is picture acquisition, which involves collecting various sign images with a camera. Gaussian filter, contrast enhancement, equalisation of the histogram, and averaging filters are used to preprocess the gathered sign images. Then the preprocessed images are segmented using Modified Canny Edge Detector (MCED). Then the segmented images are classified, for classification machine learning algorithm is used, and the images are trained using a modified resNet 101 CNN classifier. The final CNN settings are fine-tuned till the outcome is accurate enough.

Image preprocessing is the method it contains different morphological operations. In this phase, four methods can be used to preprocess the image that are Gaussian filter, contrast, enhancement, equalisation of the histogram, averaging filter.

(i)Gaussian filter

The first method in preprocessing is noise removal. In order to remove noise from the image Gaussian filter is used, and it is also used to correct the blurriness of the image and used for smoothing the image (Kaur, S et al. 2021).

$$pG(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2^2}}$$

1

σ Denoted as a standard deviation, the amount of smoothing is determined by the Gaussian standard deviation, and μ is a mean value.

(ii)Contrast enhancement

The next method in preprocessing is contrast enhancement, and it is used to make the image features stand out more clearly. Image quality should be improved, and enhancing contrast level of the images and brightness of the image must be preserved (Luque-Chang A., 2021).

$$I_n(x) = I_n(x) + a(\bar{I1} - \bar{I2})(I - I_n(x))I_1(x)$$

2

X is the pixel location In represent the pixel intensity value. $I1, I2$ represent the average pixel intensity.

(iii)Equalisation of histogram

Histogram equalisation is the next approach to preprocessing. It is utilised to enhance the contrast of a picture by altering the intensity distribution of the histogram (Reddy, K.S and Jaya, T. 2021).

$$D_i = \left[\sum_{j=0}^{i} N_j \right] * \frac{maximumintensitylevel}{noofpixels}$$

3

N represents the number of pixels i, j represents intensity levels.

(iv)Averaging Filter

The next preprocessing approach is average filtering, which smooths pictures by lowering the difference in intensity between neighbouring pixels. The average filter analyses the image pixel by pixel, substituting the average value of surrounding pixels for each value. The averaging filter is derived using Eq. 4, (Shedbalkar J et al. 2021)

$$y[n] = 1 / N \sum_{i=0}^{N-1} x[n - i]$$

4

N represents the length of the average x(n) present input, x(n-i) represents previous input and y(n) represents present output.

3.2 Image Segmentation

The image segmentation is utilised to split the image into different portions according to its features and characteristics. In this method, Modified Canny Edge Detector is used for image segmentation. MCDE is used to detect the edges and hence defines the boundaries of an object. In MCDE, double thresholding and edge tracking methods will be used then the final segmentation image will be produced.

(i)Modified Canny Edge Detection (MCDE)

Step 1

Use a gradient in the x and y dimensions to find the edge direction. The direction is found using 3 by 3 convolution kernels in CDE, although CDE does not provide an exact direction, so we used MCDE. Unnecessary features can be ignored by using a larger Sobel operator without affecting the detection of actual edges. The middle pixel technique calculates the gradient intensity of non-edge pixels using a pair

of 55 convolution kernels. The first pixel of interest (POI) will be drawn in the centre of the adjacent pixels. The distance between the centre pixel of interest and the relative distance between the centre pixels of interest is then calculated.

Figure (2a)

The Taylor expansion of bivariate functions, which is utilised in nonparametric regression, was applied in this study to estimate the regression function and local linear kernel estimation, which uses Eq. 5 to evaluate the gradient (G x and G y). (Qin, X, 2021).

$$y_{i\approx\hat{y1}} + \Delta x * Gx + \Delta y * Gy$$

5

Where x and y represent pixel distances, Yi is a measure of how interested a pixel is in one of its neighbours.

The pixel of interest is (0,0), and a 25 by 3 matrix was created to define the x and y directions, with the first column entries being 1's and the next two columns occupied with relative pixel locations. It was calculated using Eq. 6 (Qin, X, 2021).

W (I, j) = exp (-d (I, j)/2k), k (6)

Use the weighted residual sum of squares to get the gradient intensity, which is defined in Eq. 7 [32]

Q=(Y-Xβ)^1 $w(Y - X\beta)$ (7)

Equation 8 achieves the second and third parts by decreasing Q by subtracting the horizontal and vertical gradients (Gx, Gy). (Qin, X, 2021)

$$\hat{\beta} = \left(X' WX\right) X' WY = AY$$

8

Closest neighbouring pixel to the Pixel of Interest. The intercept is represented by the first row of a 25 by 3 matrix. The gradient along the x-axis is shown in the second row, while the gradient along the y axis is shown in the third row.

After that, the edge pixel must be located. The convolutional kernel fully maps every pixel in CDE. This will have an impact on the pixel's precision.

Figure (2b)

To address this problem in MCDE edge pixels, different operators are used. In order to generate an appropriate convolution kernel using Eq. 9. An alternative form of surrounding pixels should be utilised to

determine the gradient of the pixel at (0,0). Then, without making any alterations, repeat the procedure for each and every point in the original image. (Qin, X, 2021)

$$SSQ = \sum_{i=1}^{n} (yi - \hat{y}1)^2 \quad (9)$$

Step 2

Non-maxima suppression creates a narrow line for the edge by tracking along the edge path and suppressing any pixel value that is not designated an edge.

The magnitude of a single-pixel q with a 45-degree edge will be compared to the magnitudes of r and p, where r and pare interpolated values based on the two nearest pixels. For example, Eq. 10 can be used to calculate r. (Qin, X, 2021)

$$R = \alpha b + (1 - \alpha) a \quad (10)$$

where, a and b represent pixel's original magnitudes closest to r.

$$\alpha = \left|G_x\right| / \left|G_y\right|$$

Step 3

The dual thresholding approach is often used to prevent streaks and to filter out the weak gradient values generated by noise fluctuations.

Step 4

Edge tracking is a technique for identifying strong and weak edges in an image and subsequently removing any weak edges.

Step 5

The final processed image will be displayed once we set the removed weak edges in the previous phase to zero.

3.3 Classification

Image classification is the process of employing rules to locate and label groups of pixels or vectors inside an image. The image is classified using a modified deep residual CNN classifier. CNN represents a significant advancement in image identification. CNNs have an input layer, an output layer, and a hidden layer that includes convolutional, ReLU, and pooling layers, as well as fully connected layers.

(i)Modified deep residual CNN

In this proposed method, modified deep residual 101 CNN is used to recognise the 36 classes that are A to Z and 0 to 9. A Convolutional Neural Network with 101 deep layers, ResNet 101 is a Convolutional Neural Network. The ImageNet database can be used to load a pre-trained version of the network that has been trained on over a million images. We test the suggested method's ability to recognise various types of sign images.

Figure 3 explains the first layer of the Convolutional Neural Network is the convolution layer, and it performs convolution operation to the input. Then the sum result will be filtered using an activation function like ReLU and passes the output to the pooling layer. The procedure of the convolution layer will be calculated using Eq. 11(Liew, W.S. et al. 2021)

$$O_{i(x,y)} = f^{1}\left(\sum_{i=0}^{m-Im-I}\sum_{j=0} w(i,j).\, I(x+i, y+j) + bias\right)$$

11

The pooling layer is being used to minimise the number of variables in the network and the number of dimensions in the feature map. To extract features such as edges, points, and so on, max pooling is utilised. The neurons from the current layer to the next layer are then connected using a fully connected layer. The training dataset classifies an input image into various categories. The result of the neural network is classified using Eq. 12, (Liew, W.S. et al. 2021)

$$y = \sigma\left(w^{L}..\sigma\left(w^{2}\sigma\left(w^{1}x + b^{1}\right) + b^{2}\right)\cdots + b^{L}\right)$$

12

x is input and σ is an activation function, and w is denoted as a network parameter and b denoted as bias.

Original resNet 101 has 101 deep layers, and a network results with different structures and produces a different outcome. Several changes to resNet 101 decreased the size, computational cost, and the number of residual blocks, and the ReLU activation between each residual block was deleted(https://www.kaggle.com/ayuraj/american-sign-language-dataset). To keep the output size of each layer consistent, a max-pooling layer was added before each convolutional block. The upgraded resNet 101 contains fewer filters and lower complexity parameters than the original resNet 101. The computation time is affected by the amount of data and the size of the neural network. The modified resNet 101 is employed in this proposed solution due to the high processing complexity of deeper and more advanced neural networks. Modified resNet 101 reduces training time and interference while preserving as much of the original performance as possible. Finally, in this proposed method, 35 classes are classified using modified resNet 101. Then all 35 signs are identified using this proposed ASL system.

4. Result And Discussion

The goal of this approach is to recognise sign language in order for deaf individuals to understand it. The proposed modified residual 101 CNN system is evaluated in this section. The proposed CNN classifier for recognising the sign language has been simulated using MatLab 2020b with CPU: Intel Core i5, GPU: NVidia GeForce GTX 1650, RAM: 16GB. CNN examines the image entered to determine whether it has been recognised in the sign language. An optimisation is utilised to solve the CNNs optimal problem to increase its performance and accuracy. The image was initially taken and preprocessed to remove irrelevant things from the image. Then the image was segmented to partition image into several parts. Then the image was classified using a modified CNN resNet 101 classifier to detect whether the signs had been recognised or not.

Here we used these steps for removing unwanted things from the image and adding some filters to enhance and smoothening the images to improve accuracy and computational cost etc. The next phase is preprocessing stage, which involves removing noise from the image, enhancing the image and smoothening the image etc. In the proposed approach, there are four steps to preprocessing the image: Gaussian filter, contrast enhancement, equalisation of the histogram, averaging filter. We started by removing unwanted noise from the original image with a Gaussian filter. The image will then be enhanced with contrast enhancement to boost the image's quality and brightness. Following that, equalisation of the histogram will take place, and it is used to adjust the contrast level of the image. The averaging filter is the final stage, and it is used to smooth the image by lowering the difference in intensity between adjacent pixels.

The next phase is segmentation which is used to partition the images into several parts according to their features and properties. In this proposed approach, Modified Canny Edge Detector is utilised to classify the edges and boundaries of an image. Then double thresholding method is used to eliminate streaking and filter out all weak gradient values. Then Edge tracking method is used to remove all weak edges present in the image then all weak edges are removed from the image. The segmented image is then created to improve the method's effectiveness and accuracy.

Original images of a raw dataset of ASL images that is 0 to 9 and a to z are placed in Table 1.

In Table 2, we applied a Gaussian filter to the original images in order to remove noises from the images and correct the blurriness of the image.

In Table 3, we applied contrast enhancement to the filtered photos, which is used to make image features stand out more clearly, increase image quality, and brighten the image.

Table 5 shows how we used an averaging filter on the augmented image to smooth it out by lowering the intensity differences between neighbouring pixels.

In Table 6, the modified canny edge detection method is used for the segmentation process, and it is used to extract the edges by removing all weak edges from the preprocessed image.

A statistics accuracy relates to how close to its true value. This is significant because results inaccuracies might be caused by malfunctioning equipment, human mistakes, or insufficient data processing. The accuracy of proposed and existing approaches are sketched in Fig. 4(a). Proposed resNet 101 achieved 0.95% accuracy, and the existing method of resNet 50 and resNet 18 achieved 0.90% and 0.80% accuracy, respectively. This clearly shows how the proposed system is more advantages than an existing system. Figure 4(b) shows the suggested system's sensitivity. The suggested system resNet 101 reached 0.90% of accuracy, and the existing systems resNet 50 and resNet 18 reached 0.81% and 0.75%, respectively. The specificity of the suggested system is shown in Fig. 4(c), and in specificity, the suggested resNet 101 give 0.92% of accuracy, and the existing system resNet 50 and resNet 18 give 0.84% and 0.71% of accuracy, respectively. Figure 4 illustrates a comparison of planned and existing systems in terms of accuracy, sensitivity, and specificity.

In statistical analysis, the discrepancy between the calculated and real value is referred to as "error". The error of the proposed approach is analysed and sketched in Fig. 5(a). In the proposed system, 0.05% of error occurred, and in the existing system, 0.1% and 0.2% errors occurred in resNet 50 and resNet 18, respectively. The precision of the proposed and existing approach is sketched in Fig. 5(b), and the precision has a well defined harmonic mean. The precision of the proposed resNet 101 is 0.94%, and the precision of the existing approach resNet 50 and resNet 18 is 0.89% 0.79%, respectively. The F1 score of the proposed approach is analysed and sketched in Fig. 5(c). The F1 score of the proposed resNet 101 is 0.90% and F1 score of the existing approach resNet 50 and resNet 18 is 0.85% ,0.80% respectively. F1 score is a machine learning metric that was used to calculate the system's binary categories and quantify the data set's accuracy. The comparison of existing and proposed analysis of error, precision and F1 score is analysed and sketched in Fig. 5. When compared to the existing methodology, the new method has a high precision value.

In Fig. 6, the comparison of proposed and existing approaches of kappa, MCC and FRR is analysed and sketched. Matthews correlation coefficient (MCC) is used to assess the validity of two binary classifies. In the proposed approach, the MMC value of resNet 101 is 0.88%, and the MMC value of the existing approach is 0.70% and 0.67%, respectively. The MMC values are analysed and sketched in Fig. 6(a). Comparing the observed values of a training dataset to the predicted value is how the statistical analysis of kappa is stated. In the proposed approach, kappa is sketched in Fig. 6(b), and the value of kappa in resNet 101 is 0.89%, and the kappa value of existing system resNet 50 and resNet 18 is 0.83% and 0.75%, respectively. The FRR is analysed and sketched in Fig. 6(c). In the proposed approach, the value of FRR in resNet 101 is 0.05%. In existing approach, the value of FRR in resNet 50 and resNet 18 is 0.08% 0.12%, respectively. The proposed system has a high kappa value when compared to existing techniques

The experiment's outcome shows that the resNet 101 classifier can discover the optimal solution of the modified CNN classification system, which increases the feature expression, efficiency, greatest detection rate without changing the characteristics of the image. American Sign Language recognition is still a challenging problem even for recent and modern approaches. In this system, we proposed a method based on a modified CNN classifier using resNet 101. According to the outcomes, the suggested modified

CNN based resNet 101 classifier is best suited for recognising the sign language and should be used in real-time applications

5. Conclusion

Sign Language recognition is necessary for deaf and dumb people to communicate with others. This paper presents a method in order to recognise the American Sign Language using newly introduced Modified deep residual CNN, which can detect and track the hand gestures for 36 signs. Pre-processing and segmentation will be done in this method by using images, preprocessing will be performed using various filters, and segmentation will be performed using a modified canny edge detection algorithm. The upgraded resNet 101 contains fewer filters and lower complexity parameters than the original resNet 101. The computation time is affected by the amount of data and the size of the neural network. This suggested approach achieves low error rates when compared to other CNNs such as resNet 50 and resNet 18, and it will be used to recognise challenging hand movements through images, and it will speed up the training and interference while maintaining as much of the original performance as possible. Modified deep residual CNN has the following advantages: computational efficiency, quick processing, and cheap computational cost. The resNet 101 classifier is utilised to recognise 36 signs from American Sign Language in this proposed technique. The proposed system has a 0.95 per cent accuracy rate, which is higher than the present system. Recognition of words, sentences, and hand movements will be added in future work.

Declarations

Funding: There is no funding provided to prepare the manuscript.

Conflict of Interest: The process of writing and the content of the article does not give grounds for raising the issue of a conflict of interest.

Data availability statement: If all data, models, and code generated or used during the study appear in the submitted article and no data needs to be specifically requested.

Code availability: No code is available for this manuscript

References

1. Adaloglou, N.M, Chatzis, T., Papastratis, .I, Stergioulas, A., Papadopoulos, G.T., Zacharopoulou, V, Xydopoulos, G., Antzakas, K., Papazachariou, D. and none Daras, P. (2021). A Comprehensive Study on Deep Learning-based Methods for Sign Language Recognition. *IEEE Transactions on Multimedia.*
2. Bakken, J.P., Varidireddy. N. and Uskov. V.L. (2020) Smart Universities: Gesture Recognition Systems for College Students with Disabilities. *In Smart Education and e-Learning* 2020: 393–411, Springer, Singapore.

3. Chen, Q., Georganas, N.D., Petriu, E.M. (2007). Real-time vision based hand gesture recognition using haar-like features. In: 2007 IEEE instrumentation and measurement technology conference IMTC 2007, pp 1–6.
4. Cheok, M.J., Omar .Z, Jaward, M.H. (2017) A review of hand gestures and sign language recognition techniques. *Int. J. Mach. Learn. Cybern*. 8, 1–23. [CrossRef]
5. Gurbuz, S.Z., Gurbuz, A.C., Malaia, E.A, Griffin, D.J. Crawford CS, Rahman MM, Kurtoglu, E,. Aksu, R., Macks ,T. and Mdrafı, R., (2020) American Sign Language recognition using rf sensing. *IEEE Sensors Journal*, 21(3), 3763–3775.
6. Hosain, A..A, Santhalingam PS, Pathak P, Rangwala H and Kosecka J (2020) Fine hand: Learning hand shapes for American Sign Language recognition. ArXiv preprint arXiv: 2003–08753.
7. https://www.kaggle.com/ayuraj/american-sign-language-dataset
8. Jebali, M., Dakhli, A. and Jemni, M. (2021) Vision-based continuous sign language recognition using multimodal sensor fusion. *Evolving Systems*, pp.1–14.
9. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R. et al (2014) Caffe: convolutional architecture for fast feature embedding. In: *Proceedings of the 22nd ACM international conference on Multimedia*, pp 675–678.
10. Kaur ,S., Singla, J. and Singh, A. (2021, February) Review on Medical Image Denoising Techniques. *In 2021 International Conference on Innovative Practices in Technology and Management (ICIPTM)* (pp. 61–66). IEEE.
11. Kraus, J.C. and Hague, A.K. (2020) S.igning Exact English Transliteration: Effects of Accuracy and Lag Time on Message Intelligibility. *The Journal of Deaf Studies and Deaf Education*, 25(2), 199–211.
12. Lee, B.G. and Lee, S.M. (2018) Smart Wearable Hand Device for Sign Language Interpretation system with Sensors Fusion. *IEEE Sens. J.*, 18 ,1224–1232. [CrossRef].
13. Lee, C.K., Ng, K.K., Chen, C.H., Lau, H,C,, Chung, S.Y. and Tsoi, T, (2021) American Sign Language recognition and training method with recurrent neural network. *Expert Systems with Applications*, 167, 114403.
14. Li, D, Rodriguez, C., Yu, X. and Li, H .(2020) Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison. *In Proceedings of the IEEE/CVF winter conference on applications of computer vision, p*p. 1459–1469.
15. Lieberman, A.M., Fitch, A. and Borovsky, A. (2021) Flexible fast-mapping: Deaf children dynamically allocate visual attention to learn novel words in American Sign Language. *Developmental Science*, p.13166.
16. Liew, W.S., Tang, T.B., Lin, C.H. and Lu, C.K. (2021). Automatic colonic polyp detection using integration of modified deep residual convolutional neural network and ensemble learning approaches. *Computer Methods and Programs in Biomedicine*, 206,106114.
17. Luque-Chang, A., Cuevas, E., Pérez-Cisneros, M., Fausto, F., Valdivia-Gonzalez A and Sarkar R (2021) Moth swarm algorithm for image contrast enhancement. *Knowledge-Based Systems*, 212, 106607.

18. Mahdikhanlou, K. and Ebrahimnezhad, H (.2020) Multimodal 3D American Sign Language recognition for static alphabet and numbers using hand joints and shape coding. *Multimedia Tools and Applications*, 79, 22235–22259.
19. Mahdikhanlou, K. and Ebrahimnezhad, H. (2020) Multimodal 3D American Sign Language recognition for static alphabet and numbers using hand joints and shape coding. *Multimedia Tools and Applications*, 79, 22235–22259.
20. Mummadi, C.K., Leo, F.P.P. and Verma, K.D., Kasireddy, S.. Scholl, P.M, Kempfle J., Laerhoven (2018) KV Real-Time and Embedded Detection of Hand Gesture with an IMU-Based Glove. *Informatics*, 5, 28. [CrossRef]
21. Parelli M, Papadimitriou, K., Potamianos, G., Pavlakos, G. and Maragos, P (2020, August) Exploiting 3d hand pose estimation in deep learning-based sign language recognition from rgb videos. *In European Conference on Computer Vision*, Springer, Cham, pp. 249–263.
22. Qin, X. (2021) A modified canny edge detector based on weighted least squares. *Computational Statistics*, 36(1), 641–659.
23. Rastgoo, R., Kiani, K. and Escalera, S. (2020). Video-based isolated hand sign language recognition using a deep cascaded model. *Multimedia Tools and Applications*, 79, 22965–22987.
24. Rastgoo, R., Kiani, K. and Escalera, S. (2021) Real-time isolated hand sign language recognition using deep networks and SVD. *Journal of Ambient Intelligence and Humanized Computing*, pp.1–21.
25. Rastgoo,. R, Kiani, K. and Escalera, S. (2020) Video-based isolated hand sign language recognition using a deep cascaded model. *Multimedia Tools and Applications*, 79, 22965–22987.
26. Rautaray, S.S. and Agrawal, A. (2015) Vision based hand gesture recognition for human computer interaction: a survey. *Artif Intell Rev* 43(1),1–54.
27. Reddy, K.S and Jaya, T. (2021) De-noising and enhancement of MRI medical images using Gaussian filter and histogram equalisation. *Materials Today: Proceedings*.
28. Schniedewind, E., Lindsay, R. and Snow, S. (2020) Ask and ye shall not receive: Interpreter-related access barriers reported by Deaf users of American Sign Language. *Disability and Health Journal*, 13(4),100932.,
29. Shedbalkar J, Prabhushetty, K. and Inchalc, A. (2021). A Comparative Analysis of Filters for Noise Reduction and Smoothening of Brain MRI Images. *In 2021 6th International for Convergence in Technology (I2CT)* (pp. 1–6). IEEE.
30. Snoddon, K. (2020). Sign language planning and policy in Ontario teacher education. *Language Policy*, pp.1–22.
31. Villwock, A., Wilkinson, E., Piñar, P. and Morford, J.P. (2021) Language development in deaf bilinguals: Deaf middle school students co-activate written English and American Sign Language during lexical processing. *Cognition*, 211, 104642.
32. Wadhawan, A. and Kumar, P. (2020) Deep learning-based sign language recognition system for static signs. *Neural Computing and Applications*, 32(12), 7957–7968.

33. Wang, J. and Zhang, T. (2014) An ARM-Based Embedded Gesture Recognition System Using a Data Glove. *In Proceedings of the 26th Chinese Control and Decision conference (CCDC), Changsha*, China, 31 May–2 June 2014, 1580–1584.
34. Woodward, J. (1996) Modern Standard Thai Sign Language, influence from ASL, and its relationship to original Thai sign varieties. *Sign Language Studies*, 92(1), 227–252.

Tables

Table1 to 6 are available in the Supplementary Files section.

Figures

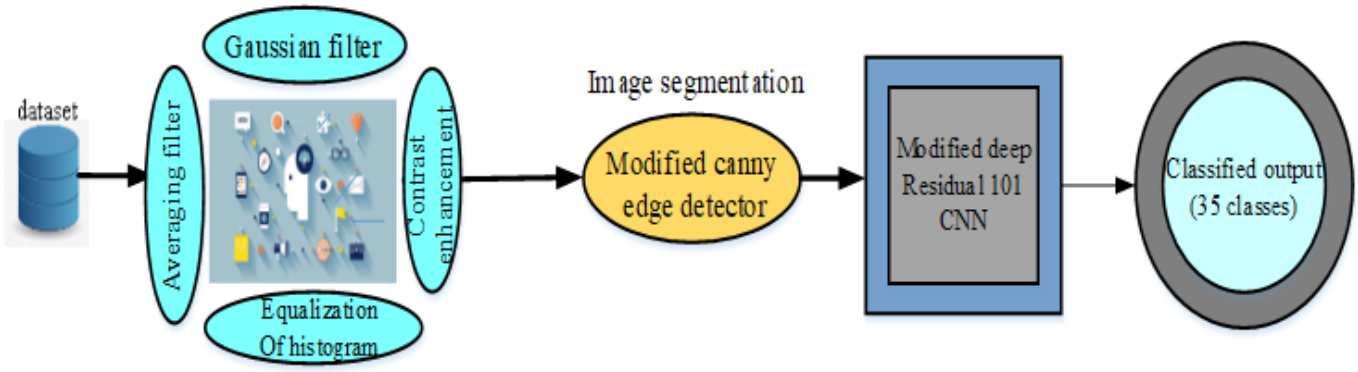


Figure 1

Architecture of the proposed system

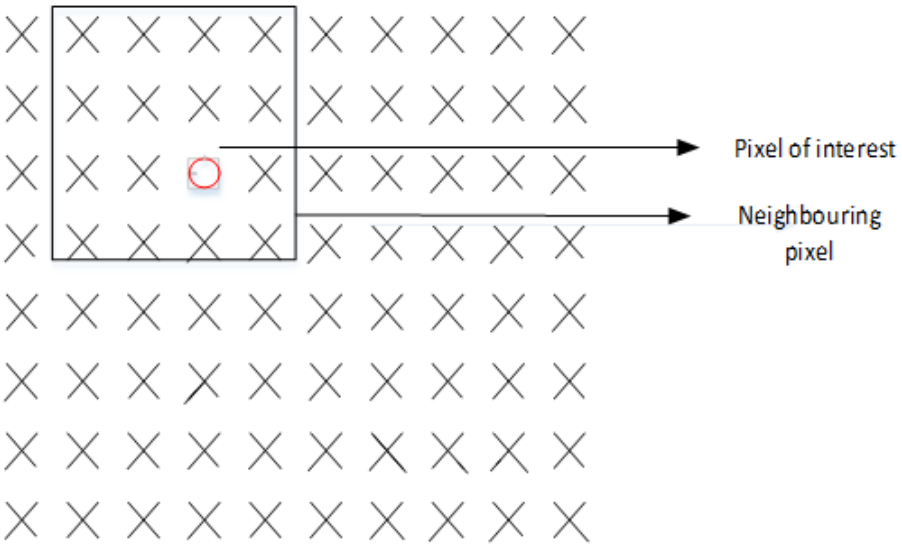


Fig. (2a)

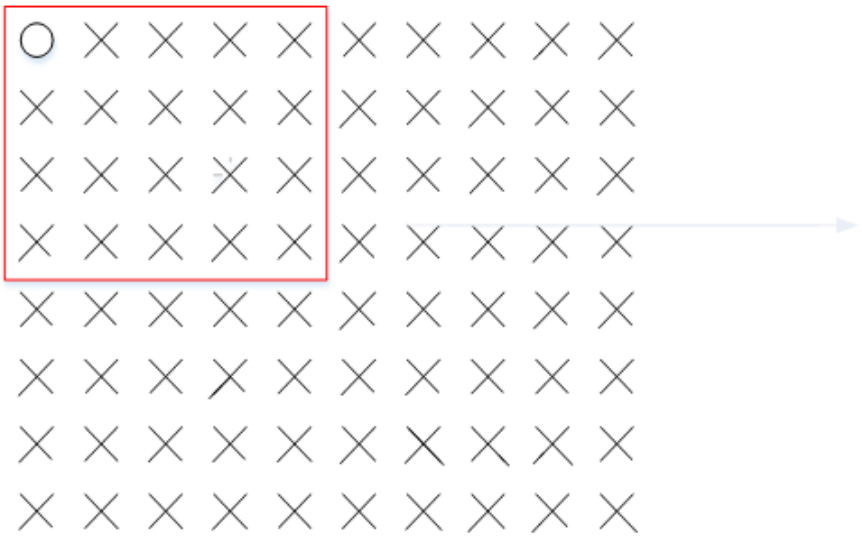


Fig. (2b)

Figure 2

Legend not included with this version

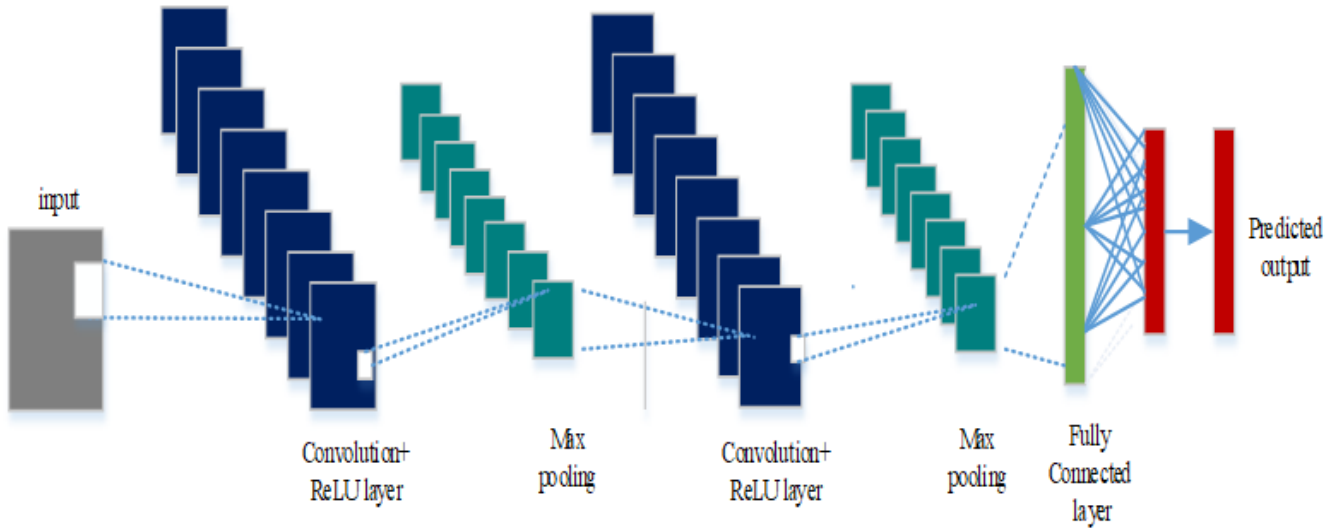
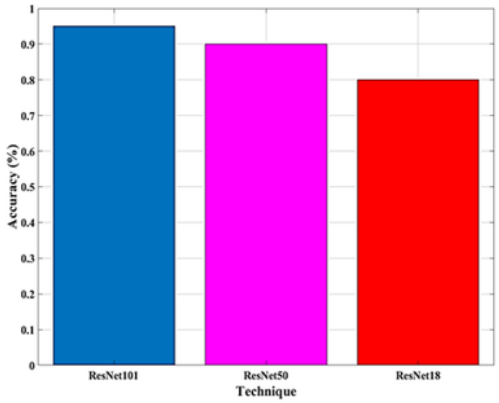
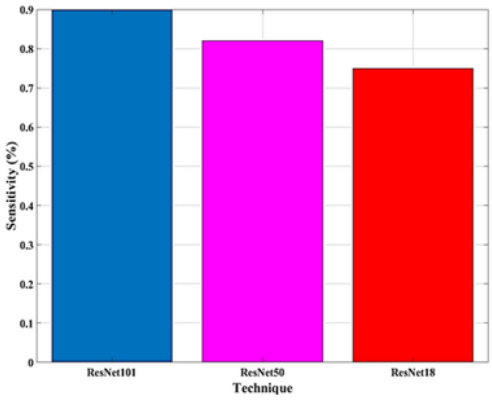


Figure 3

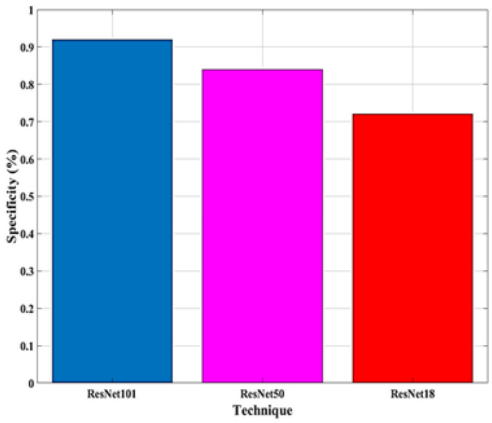
Generic Architecture of Convolution Neural Network



(a)



(b)

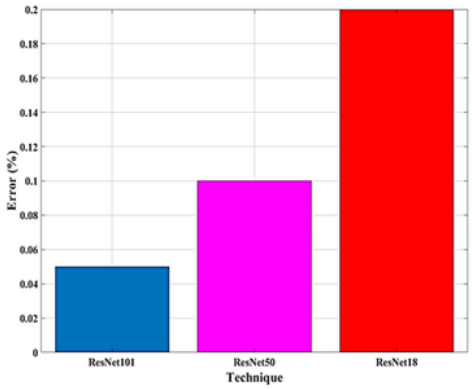


(c)

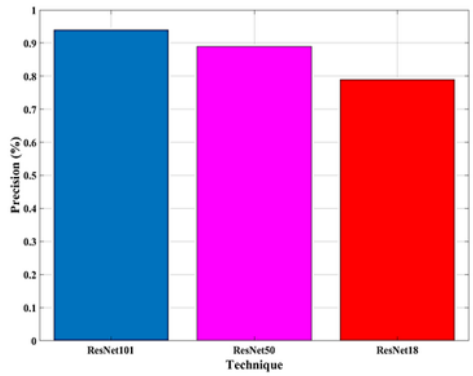
Figure 4

Comparison of proposed and existing system (a) accuracy

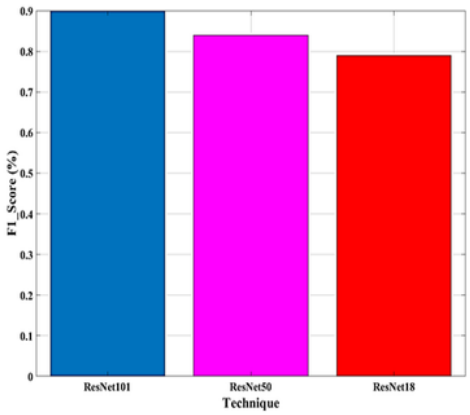
(b) sensitivity (c) specificity



(a)



(b)

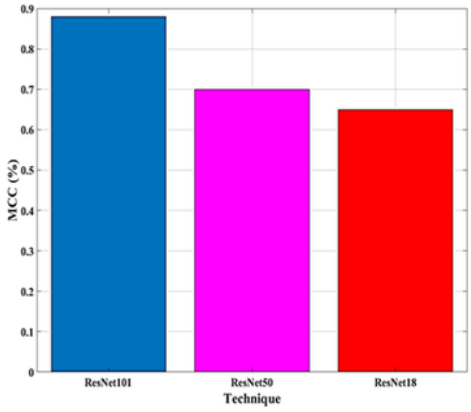


(c)

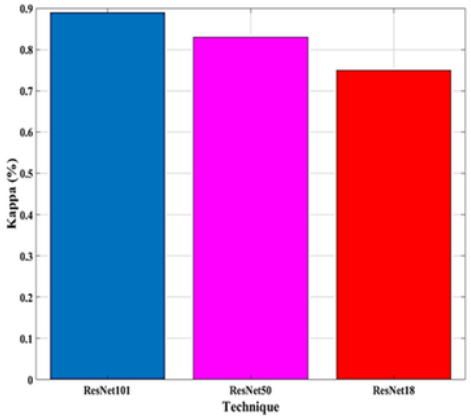
Figure 5

Comparison of proposed and existing approaches (a) error

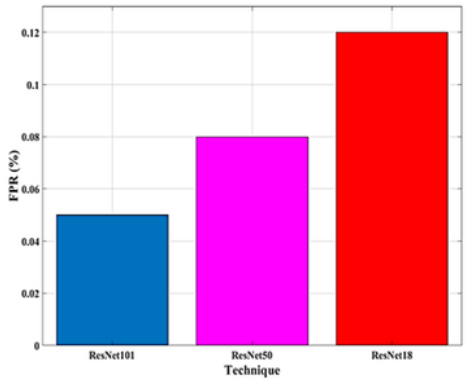
(b) precision (c) F1 score



(a)



(b)



(c)

Figure 6

comparison of proposed and existing approaches (a) MCC (b) Kappa (c) FRR

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table1to6.docx