SIGN LANGUAGE DETECTION USING CONVOLUTIONAL NEURAL NETWORKS

ALIZIYAN, RIAAN SADIQ, AYAAN SADIQ, RAZEEN

OBJECTIVE

The goals of our project are to:

- Develop accurate sign language detection models using deep learning techniques.
- Enable real-time recognition of sign language gestures for improved communication accessibility.

DATASET

We have created a comprehensive dataset for our sign language project, consisting of 50x50 grayscale images for each alphabet letter. This dataset includes 1000 images for each of the 26 alphabetic signs, meticulously organized and annotated. By using grayscale images, we streamline processing and analysis, providing a robust foundation for research, education, accessibility, and technology development in the realm of sign language.



PREPROCESSING

The sign language project begins with a meticulously organized dataset consisting of 50x50 grayscale images representing each letter of the alphabet in American Sign Language (ASL). This dataset is pivotal for training a robust model capable of accurately recognizing and interpreting sign gestures. Prior to training, the images undergo essential preprocessing steps, including conversion to grayscale to simplify processing and normalization to standardize pixel values. Augmentation techniques such as rotation, shifting, zooming, and horizontal flipping are applied to the training dataset to enhance variability and improve the model's ability to generalize to unseen data. The model architecture is designed with sequential convolutional layers followed by batch normalization, maximizing efficiency in feature extraction and learning. Dropout layers are strategically inserted to mitigate overfitting during training.

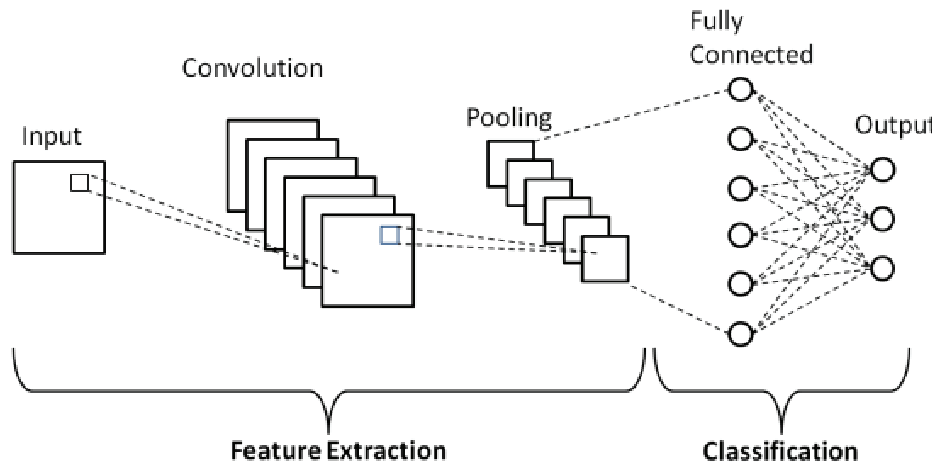
The final layers include dense connections culminating in a softmax output layer for multi-class classification of the ASL alphabet and a blank category. Training is executed over multiple epochs with early stopping mechanisms to ensure optimal performance and generalization capability.

```
# Convolutional layers with batch normalization
model.add(Conv2D(128, kernel_size=(3,3), activation='relu', input_shape=(50, 50, 1)))
model.add(BatchNormalization())
model.add(MaxPooling2D(pool_size=(2,2)))
model.add(Dropout(0.3))
```

```
train_generator = train_datagen.flow_from_directory(
    '/content/ASL-Dataset/splitdataset50x50/train',
    target_size=(50, 50),
    batch_size=batch_size,
    class_mode='categorical',
    color_mode='grayscale'
```

MODEL ARCHITECTURE

CNNs are ideal for the ASL gesture classification project because they excel at learning spatial hierarchies in images. By using convolutional layers, CNNs automatically extract features like edges and textures, crucial for recognizing ASL gestures without relying on manually crafted features. This approach enhances accuracy by capturing meaningful patterns directly from the input data, making the model effective and robust for real-world applications.



The CNN architecture for the ASL gesture classification project employs four Conv2D layers with increasing filter sizes, each followed by BatchNormalization and MaxPooling2D for feature extraction and dimensionality reduction. Dropout layers are used to mitigate overfitting. After flattening the features, three Dense layers progressively reduce in size, complemented by BatchNormalization and dropout for regularization. The final layer uses softmax activation for multi-class classification. This architecture aims to balance complexity and generalization, achieving effective classification of ASL gestures with high accuracy.

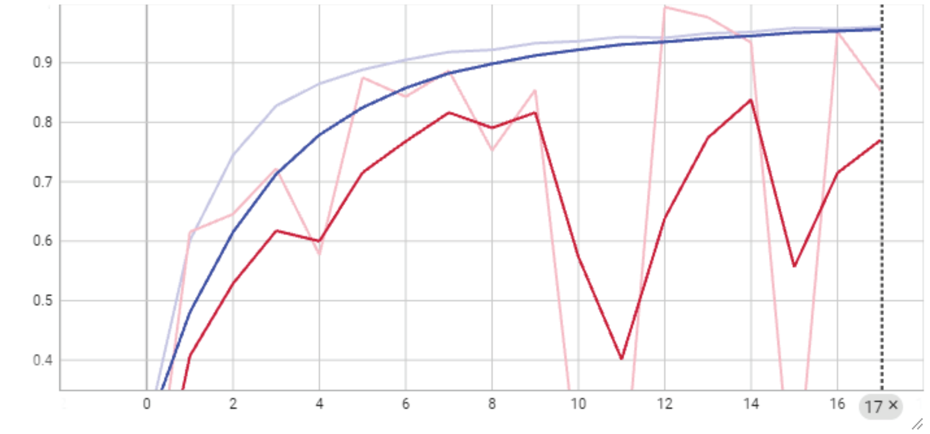
RESULTS

The model achieved a high test accuracy of approximately 96.05% in classifying sign language gestures. This indicates a robust performance in recognizing the various gestures. Below is a summary of the training process:

Training Accuracy: Consistently improved, reaching 96.05% by epoch 17.

Validation Accuracy: Showed variability, peaking at 85.34% before some fluctuations.

Test Accuracy: 99.31%, demonstrating the model's strong generalization ability.



Run ↑	Smoothed	Value	Step	Relative
train	0.9562	0.9605	17	10.24 min
validation	0.7702	0.8534	17	10.24 min

```
91/91 [==============================] - 2s 19ms/step - loss: 0.0267 - accuracy: 0.9931
Test accuracy: 0.9931318759918213
```

FUTURE WORK

Multi-Modal Approaches: Combine vision and NLP techniques for interactivecommunication.

Dataset Expansion: Include more diverse sign language gestures and variations.

Natural Language Processing (NLP):

- Transform the gesture detector into a sentence creator.
- Enable real-time conversations for more natural and interactive communication.

REFERENCES

1. Sharma, Shikhar, and Krishan Kumar. "ASL-3DCNN: American sign language recognition technique using 3-D convolutional neural networks." Multimedia Tools and Applications 80.17 (2021): 26319-26331.

2. O'shea, Keiron, and Ryan Nash. "An introduction to convolutional neural networks." arXiv preprint arXiv:1511.08458 (2015).