



American Sign Language Recognition

AI Project Group 16:
Gianno Gomez
Mason Paradeza
Marianno Reynoso
Jingye Xu
Yuntong Zhang

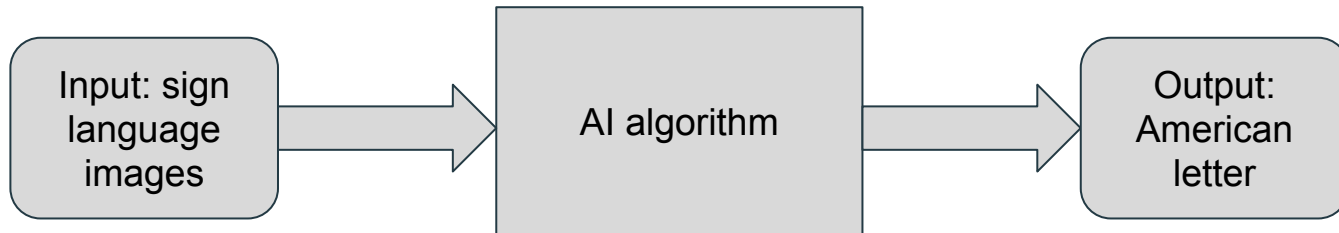


Problem introduction
Existing methods
Our method
Datasets
Evaluation results
Task allocation

Problem

American Sign Language: major communication method by hearing impaired people

Challenge: find an **efficient & accurate algorithm** to recognize sign language images



Existing work methods

Convolutional Neural Network (CNN):

- small vanilla CNN

- VGG-16

- GooLeNet

Support Vector Machine

Deep belief network

Our method

Transformer network: a novel *encoder-decoder* network architecture that use *self-attention* and *multi-head attention* mechanism that enable parallel processing for natural language processing (NLP) and produce wonderful results

Vision Transformer (ViT): applying transformer network for vision task - split an image to patches and pass them to the transformer network as a word in NLP

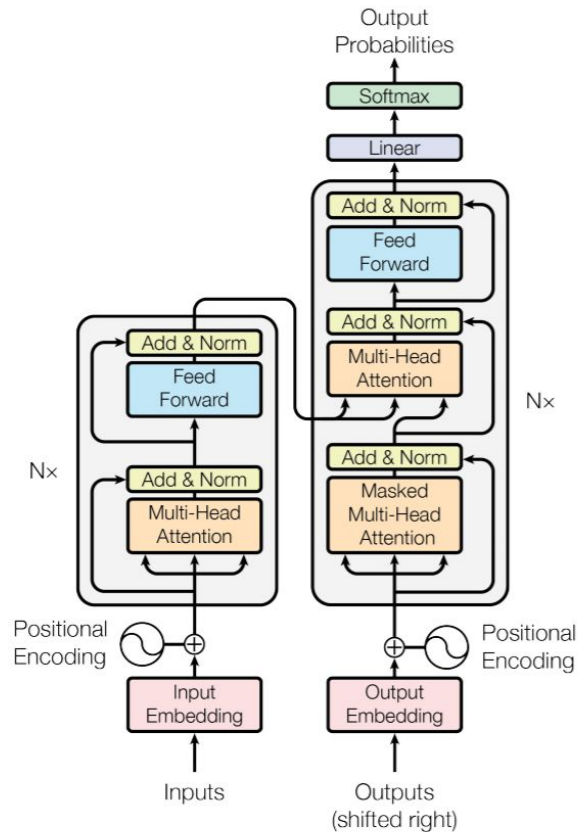
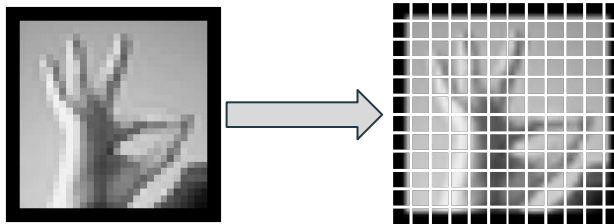


Figure 1: The Transformer - model architecture.

Vaswani, Ashish, et al. "Attention is all you need." *Advances in neural information processing systems* 30 (2017).

Dosovitskiy, Alexey, et al. "An image is worth 16x16 words: Transformers for image recognition at scale." *arXiv preprint arXiv:2010.11929* (2020).

Datasets

[Sign language MNIST](#)

Contain grayscale images (28*28) of America Sign Language from letter A to Z (except for J and Z because of gesture motions).

Train: 27455

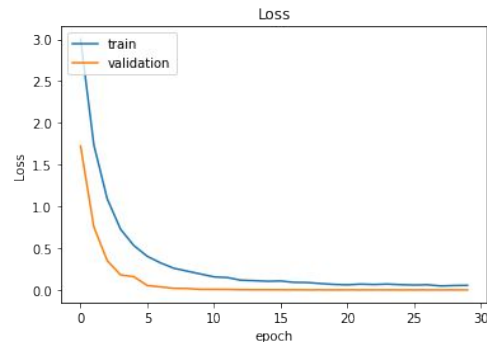
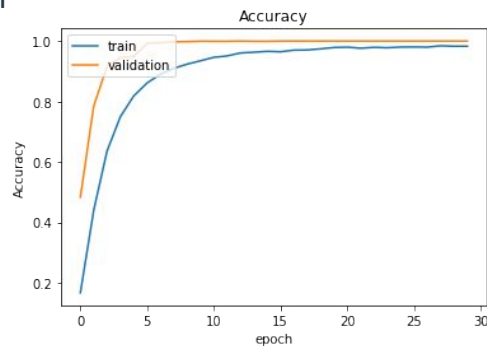
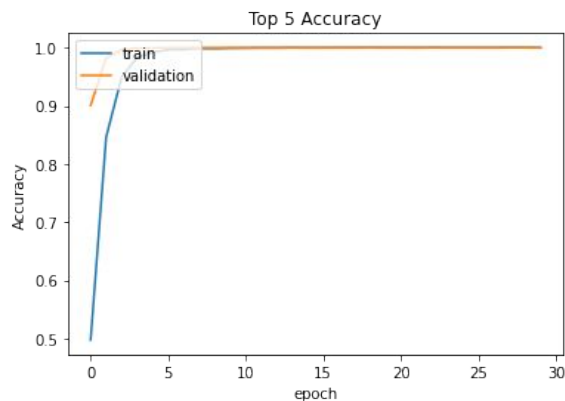
Test: 7172



Evaluation

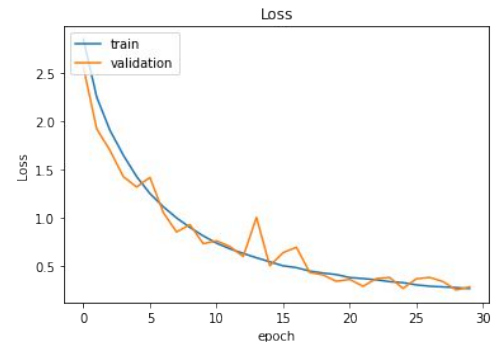
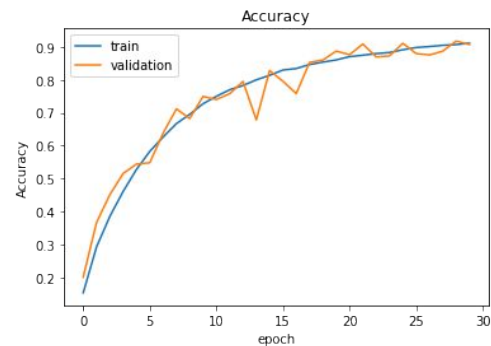
the proposed method - a transformer model

Accuracy on test data: 96.49%



an existing method - a CNN model

Accuracy 90.5%



Task allocation

We worked on the proposed method together. Jingye finish most of the CNN model code and Yuntong finish most of the transformer model code.

For the final deliverables:

1. **Gianno** is responsible for the web page showcase
2. **Mason** and **Marianno** are responsible for the poster
3. **Jingye** and **Yuntong** are responsible for the presentation video