EXECUTIVE SUMMARY

Image Classification with Convolutional Neural Networks

Interpreting Hand-Shape of ASL Interactions

Data Analytics Graduate Capstone

Brittany Kozura
Western Governors University
D214

Problem Statement and Hypothesis

As machine learning research grows, there are some companies who are looking to harness Natural Language Processing for translational and interpreting work. AI development in this area of research, known as Machine Translation, has "led to very significant improvements in translation quality" in recent years (Stanford NLP Group, 2022). A new challenge in particular is presented when attempting NLP for sign languages, as they neither have a verbal nor written component. Therefore, new approaches to NLP must be considered when it comes to interpreting sign languages. This project acts as a predecessor to NLP for American Sign Language (ASL), where the first step in the process is to classify hand shape from an individual video frame or image. Our goal is to create a model that classifies hand shapes as a proof-of-concept for such a system.

Research Question

"To what extent can hand shape be accurately classified from images?"

Hypothesis

"Hand-shape in images can be classified with 90% or greater accuracy using Convolutional Neural Networks."

Summary of the Data Analysis Process

Data Collection

The dataset used for this analysis "Synthetic ASL Alphabet" was created by a software development company Lexset (Lexset, 2022) and was listed as Creative Commons Attribution-NonCommercial as a way to promote Lexset's synthetic data generator platform “Seahaven” which was used to create the data set. The data was obtained from open data repository Kaggle and was split into 'train' and 'test' data sets–each with images divided into folders by classification. It consists of 27,000 images across 27 categories (26 alphabet + "Blank").

Data Extraction and Preparation

RGB data is extracted from the images using Kera's internal `flow_images_from_directory` function, which allows us to include our preprocessing steps in the data extraction phase. During this process all data augmentation, resizing, and manipulations occur. Additionally, we split the dataset into a 72-19-10 train-validation-test split.

Model Design, Build, & Training

The data analysis technique that will be used to classify the images is a Convolutional Neural Network using Keras/Tensorflow. In image classification analyses, CNNs are industry standards since they perform "phenomenally well on computer vision tasks" (Rizvi, 2022). Instead of the analyst identifying the key features of the classes of images,

CNNs analyze images into three-dimensional matrices representing color components per pixel and "'learns' how to extract these features, and ultimately infer what object they constitute" (Google Developers, 2022). Because the network learns from the raw data rather than be interpreted by the analyst, it can identify features that a human analyst might not be able to identify.

Our CNN model is based on the original AlexNet architecture (Krizhevsky, 2017) with some additional features/modifications. The following diagram shows the full layer and kernel structure of the model.

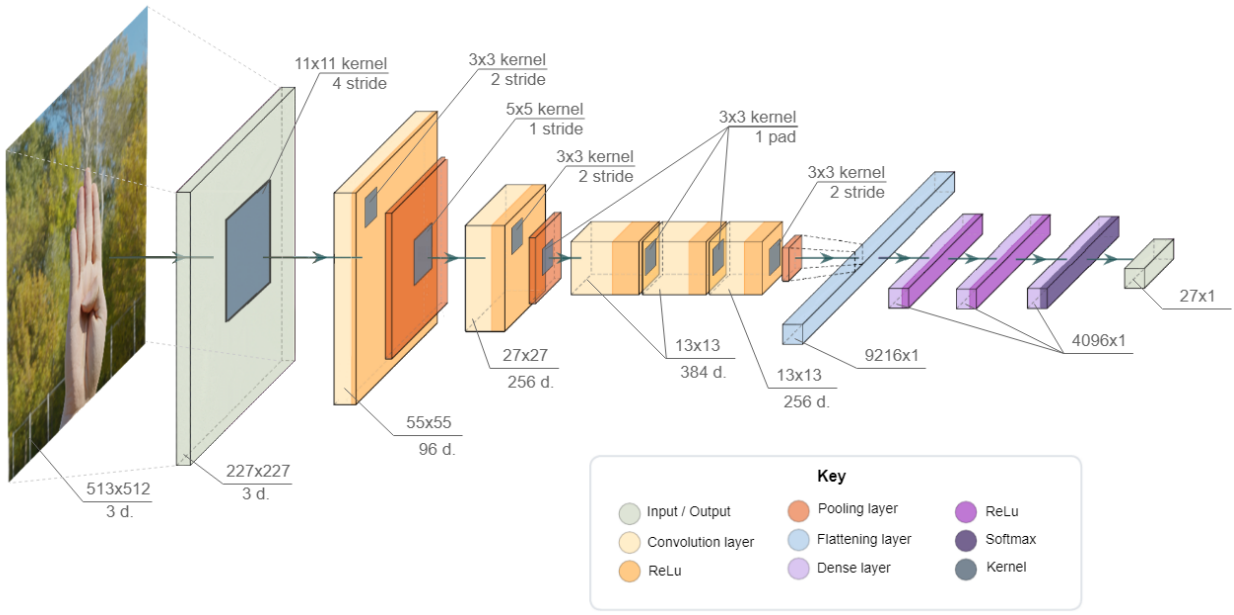

Diagram Representing the structure of the CNN

Our model training resulted in an accuracy of over 99% for training data and over 94% for validation data.

Outline of Findings

Model Evaluation

We evaluated the system based on accuracy and Kappa, both of which came out at over 93%, so our hypothesis was correct.

Test Set Evaluation Metrics	
Precision	93.96%
Recall	93.78%
F1	93.78%
Accuracy	93.78%
Kappa	93.54%

Limitations of Analysis

Limitations of Synthetic Data

- Synthetic Data tends to result in more 'ideal' data rather than real-world data.

  Solution: Augment data with more diverse pre-processing and real-world images.

Limitations of CNNs (Rizvi, 2020)

- CNNs require a large amount of data.
- CNNs do not encode the position and orientation of objects in images.
- CNNs do not incorporate time-series data.

  Solution: Hybridize with other models in the next phase

Proposed Next Steps

Continued Training with more Real-World Data

- Expand Training & Testing Data: Enhance with "Real-World" data and augment more synthetic data that is "less-than-ideal" to get more diversity.

Hybridize with Other Models

- Position Tracking : Add an additional network (potentially an RNN) to locate the position of the hands in the images.
- Analysis over Time: Implement a time-series component to track the location and shape of the hand over time.

Benefits

Market Opportunities

Sign language interpreter industry is worth over a billion dollars, but is "dominated by small, local providers, with several large, national players."(Hickey & Leske, 2021). The market is so saturated with small firms, that an ASL interpreting firm with as little as $5 million in revenue "would be considered a big player in the industry" (Hickey & Leske, 2021). This means there is ample opportunity to provide real-time translation services

Expanding Accessibility

Greater access for deaf and hard-of-hearing from underprivileged communities. The average ASL interpreter costs $50 - $150/ hour (Hickey & Leske, 2021). Having a technical solution to ASL interpretation means that accessibility becomes more affordable.

Sources

Chollet François, Kalinowski, T., & Allaire, J. J. (2022). Deep learning with R. Manning.

Google Developers. (2022, July 18). ML Practicum: Image Classification. Google Developers. Retrieved August 2, 2022, from https://developers.google.com/machine-learning/ practica/image-classification/convolutional-neural-networks

Hickey, S., & Leske, H. (2021, April 22). *ASL interpreting: What you need to know about the ASL services market*. Nimdzi. Retrieved September 21, 2022, from https://www.nimdzi.com/asl-interpreting/

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional Neural Networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386

Lexset. (May 2022). Synthetic ASL Alphabet, 1.0. Retrieved Sept 10, 2022 from https://www.kaggle.com/datasets/lexset/synthetic-asl-alphabet.

Rizvi, M. S. Z. (2020, October 19). CNN image classification: Image Classification using CNN. Analytics Vidhya. Retrieved August 23, 2022, from https://www.analyticsvidhya.com/blog/2020/02/learn-image-classification-cnn-convolutional-neural-networks-3-datasets/

The Stanford NLP Group. (2022). Machine Translation. The Stanford Natural Language Processing Group. Retrieved September 19, 2022, from https://nlp.stanford.edu/projects/mt.shtml