Image Classification with Convolutional Neural Networks

Interpreting Hand-Shape of ASL Interactions

Data Analytics Graduate Capstone

B. Kozura

Introduction

Brittany Kozura

- MSDA Student @WGU
- Software Engineer
- AI/ML Research Enthusiast

Fun fact: Hiked the Incan Trail to Machu Picchu

Analysis Problem

Context:

As machine learning research grows, there are some companies who are looking to harness Natural Language Processing for translational and interpreting work, also known as "Machine Translation".

Sign Languages present brand new problems as they do not utilize a spoken or written component.

This project attempts to achieve the first step in the process of Machine Translation for ASL – to classify hand shape in an 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% accuracy using Convolutional Neural Networks.

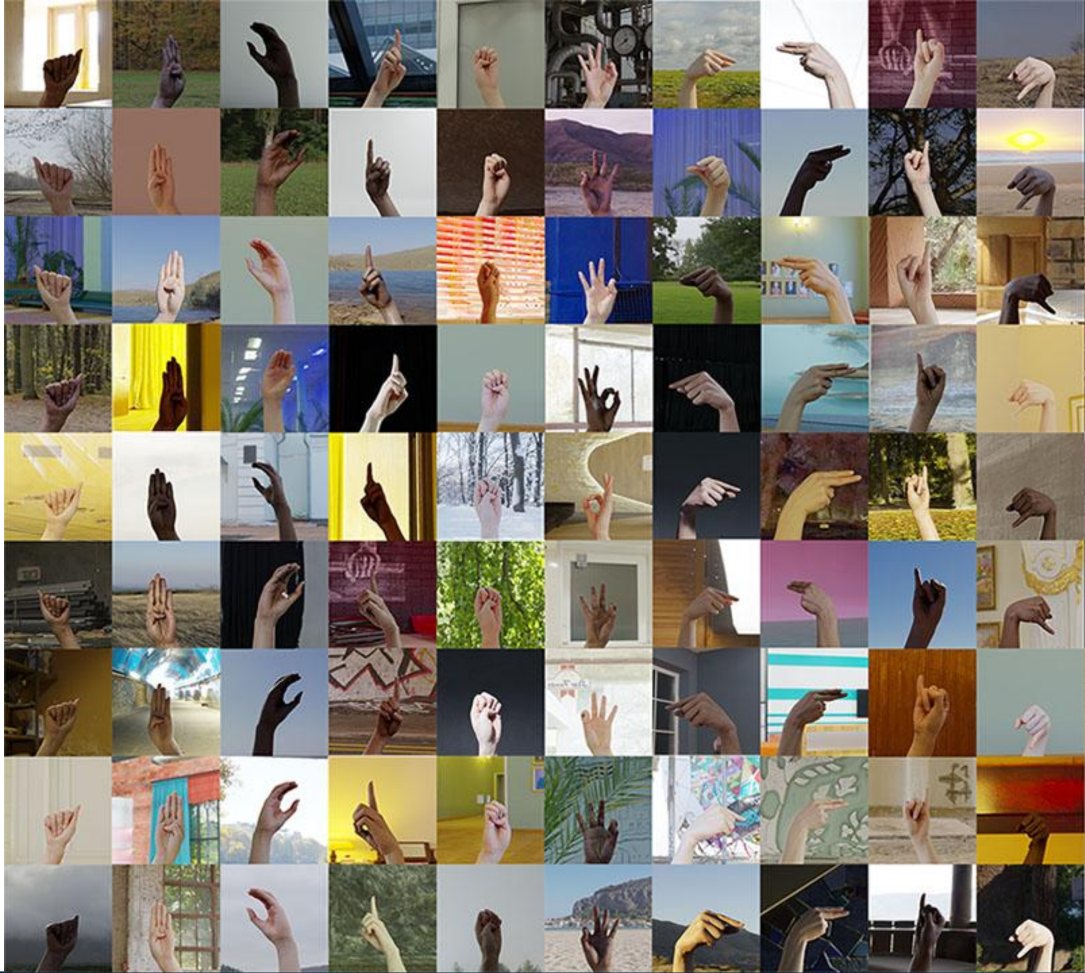
Analysis Process

1. Data collection
2. Parse image data
3. Image preprocessing
4. Build model
5. Train model
6. Test model
7. Evaluate performance
8. Plan for what's next

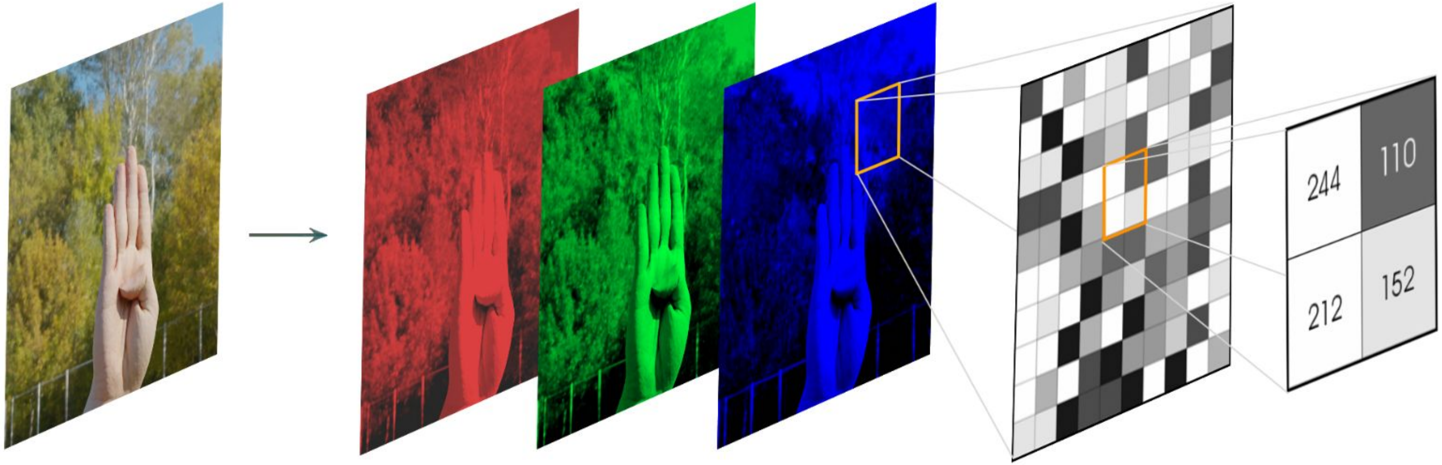
1. Image Data Source

"Synthetic ASL Alphabet" provided by Lexset

- 27,000 images
- A-Z + "Blank"
- various backgrounds, skintones, & lighting conditions



2. Parse Image Data

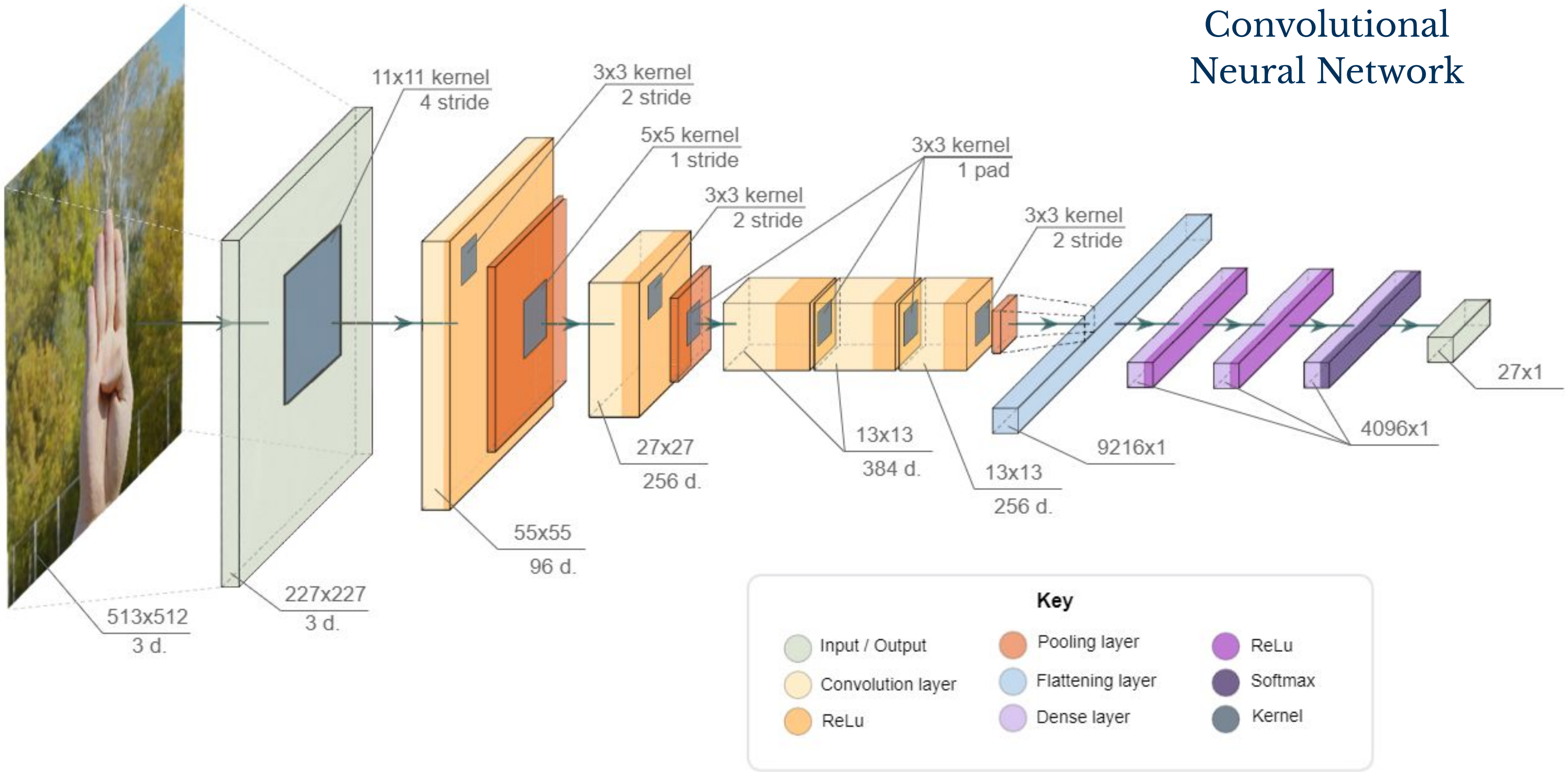


3. Image Preprocessing

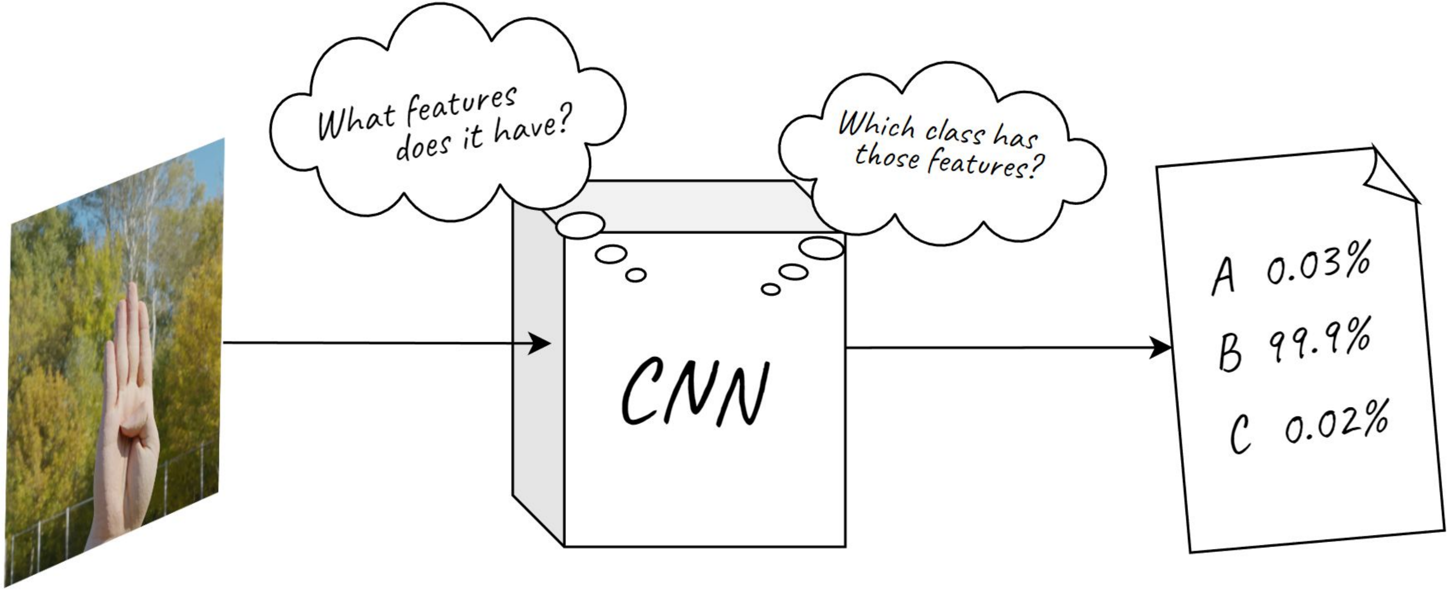
- Resizing
- Rotations
- Flipping
- Zooming
- Panning
- Shearing

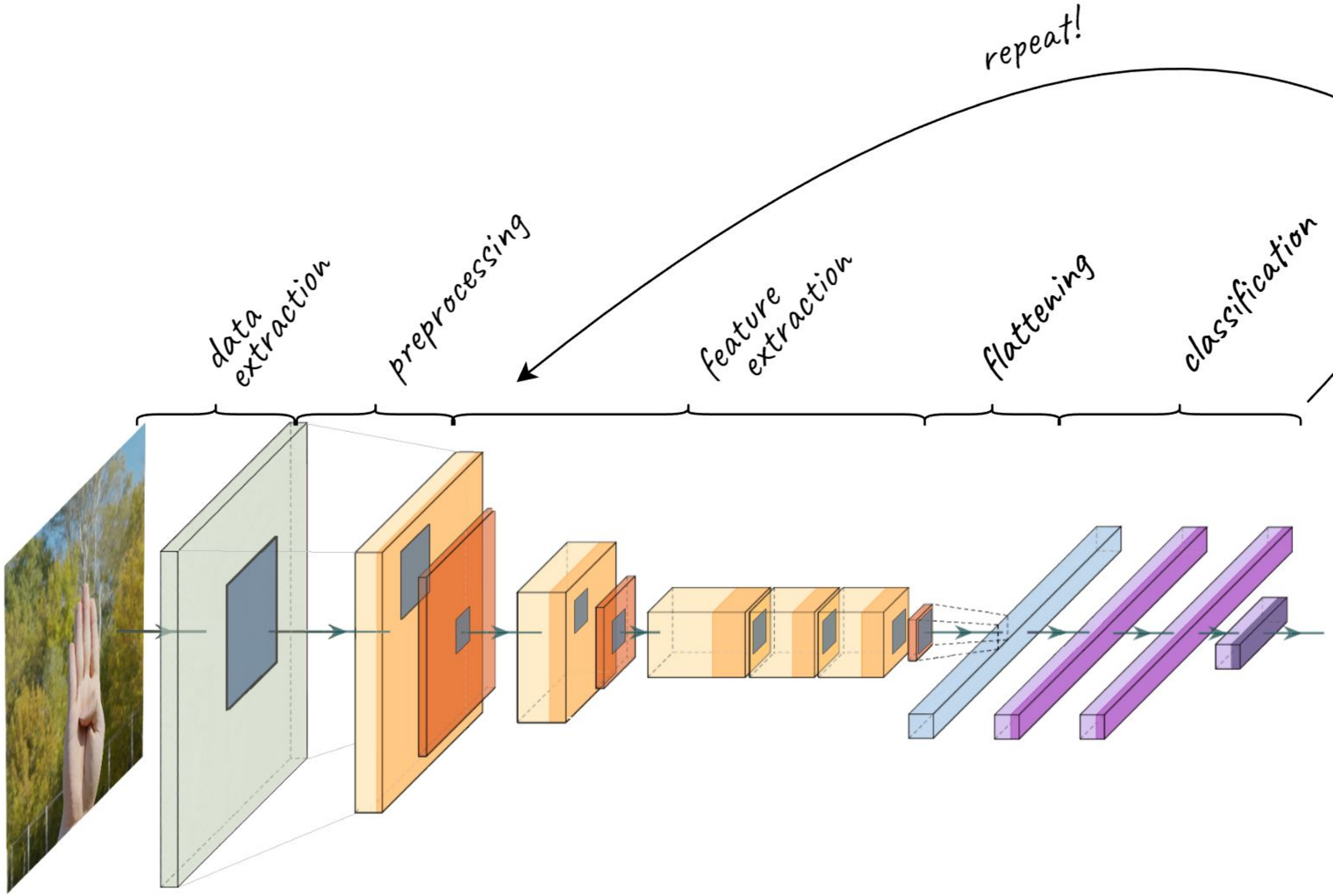
4. Build the Model

Convolutional Neural Network



How does a CNN work?




repeat!
data extraction
preprocessing
feature extraction
flattening
classification

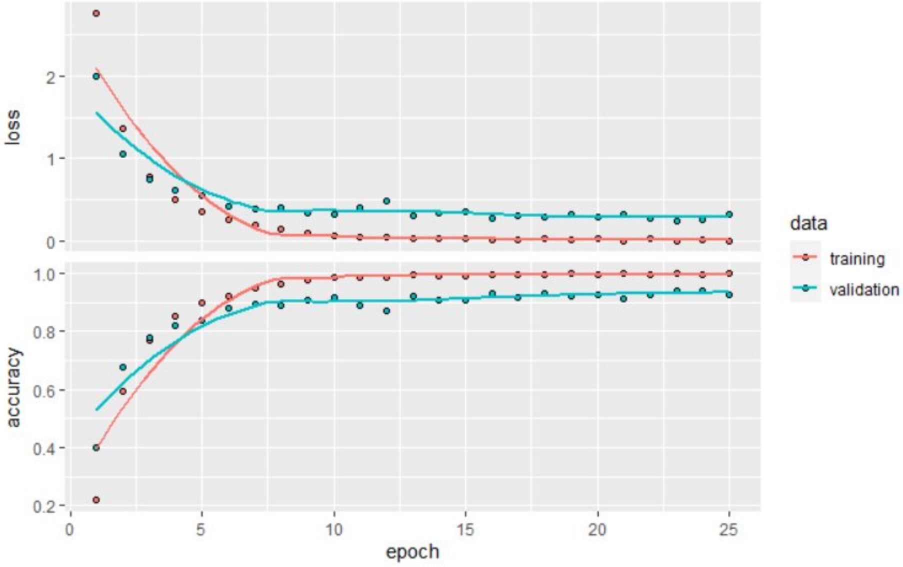
5. Train the Model

training set:

accuracy: 99.94%
loss: 0.0040

validation set:

accuracy: 94.07%
loss: 0.2587



6. Results

Test Data: Actual vs Predicted Class

Actual \ Predicted	A	B	Blank	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	91	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	7	0
B	0	99	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Blank	0	0	96	2	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
C	0	0	0	98	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0
D	0	0	1	0	94	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	3	0	0
E	0	0	0	0	0	96	0	0	0	0	0	0	0	2	1	0	0	0	0	0	1	0	0	0	0	0	0
F	0	0	0	0	0	0	99	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	1	0	0	98	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
H	0	0	0	1	0	0	0	0	96	0	0	0	0	0	0	0	0	2	0	0	1	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	0	97	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0
J	0	0	0	0	0	0	0	0	0	0	91	0	1	0	0	0	1	2	0	0	0	0	0	0	0	5	0
K	0	1	0	0	0	0	0	0	0	0	0	94	0	1	0	0	0	0	1	0	0	2	0	0	0	1	0
L	0	0	0	0	0	0	1	0	0	0	1	0	98	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M	0	0	0	0	0	2	0	0	0	0	0	1	0	91	1	0	0	1	0	0	3	0	1	0	0	0	0
N	1	0	0	0	2	1	0	0	0	0	0	0	0	4	89	0	0	0	0	0	1	1	0	0	1	0	0
O	0	0	0	1	0	1	1	0	0	1	0	0	0	0	0	94	0	0	0	1	0	0	0	0	1	0	0
P	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	94	3	0	0	0	0	0	0	0	1	0
Q	0	0	0	0	1	0	0	0	0	0	2	0	0	0	0	0	5	91	0	0	0	0	0	0	0	1	0
R	0	1	0	0	0	0	1	0	0	1	0	0	0	2	0	1	0	0	89	0	0	1	0	0	4	0	0
S	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	13	0	0	1	79	4	0	0	0	1	0	0
T	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	2	95	0	0	0	0	0	0
U	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	1	0	2	0	0	94	0	0	1	0	0
V	0	1	0	1	1	0	1	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	91	2	0	0	0
W	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	3	94	0	0	0
X	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	2	0	95	0	1
Y	2	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	96	0
Z	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	2	1	0	0	93

precision	0.939569
recall	0.937778
f1	0.937799
accuracy	0.937778
rand_accuracy	0.037037
kappa	0.935385

Limitations

Limitations of Synthetic Data

- Tends to result in more 'ideal' data rather than real-world data

Solution: augment image data with more diverse pre-processing and real-world images

Limitations of CNNs

- Require a lot of data
- Do not encode the position and orientation of objects in images
- Do not incorporate time-series data

Solution: Hybridize with RNN model in the next phase

Proposed Path Forward

Continue Training

Expand Training & Testing Data

- Enhance with "Real-World" data
- Augment more synthetic data that is "less-than-ideal" to get more diversity

Hybridize with RNN Models

Position Tracking

- Add ability 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 [1]
- Is "dominated by small, local providers, with several large, national players." [1]
- As little as $5 million in revenue "would be considered a big player in the industry" [1]
- Ample opportunity to provide real-time translation services

Expanding Accessibility

- Greater access for deaf and hard-of-hearing from underprivileged communities
- Average ASL interpreter costs $50 - $150/ hour [1]
- Having a technical solution to ASL interpretation => 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

[1] 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

Thanks for listening!

Related Links

[Analysis Report]

[Executive Summary]

[Presentation Recording]