Sign Language Recognition

Anastasios Sarafidis

mtn2028

The Project

- Classification of images containing the American Sign Alphabet
  - Deep Learning model to predict a letter from an image
  - CNN

Data

- ASL Alphabet
  - Public data gathered from Kaggle
  - Contains 3000 images for every letter
    - Plus 9000 for “Space”, “Delete” & “Nothing”
  - 87000 images in total

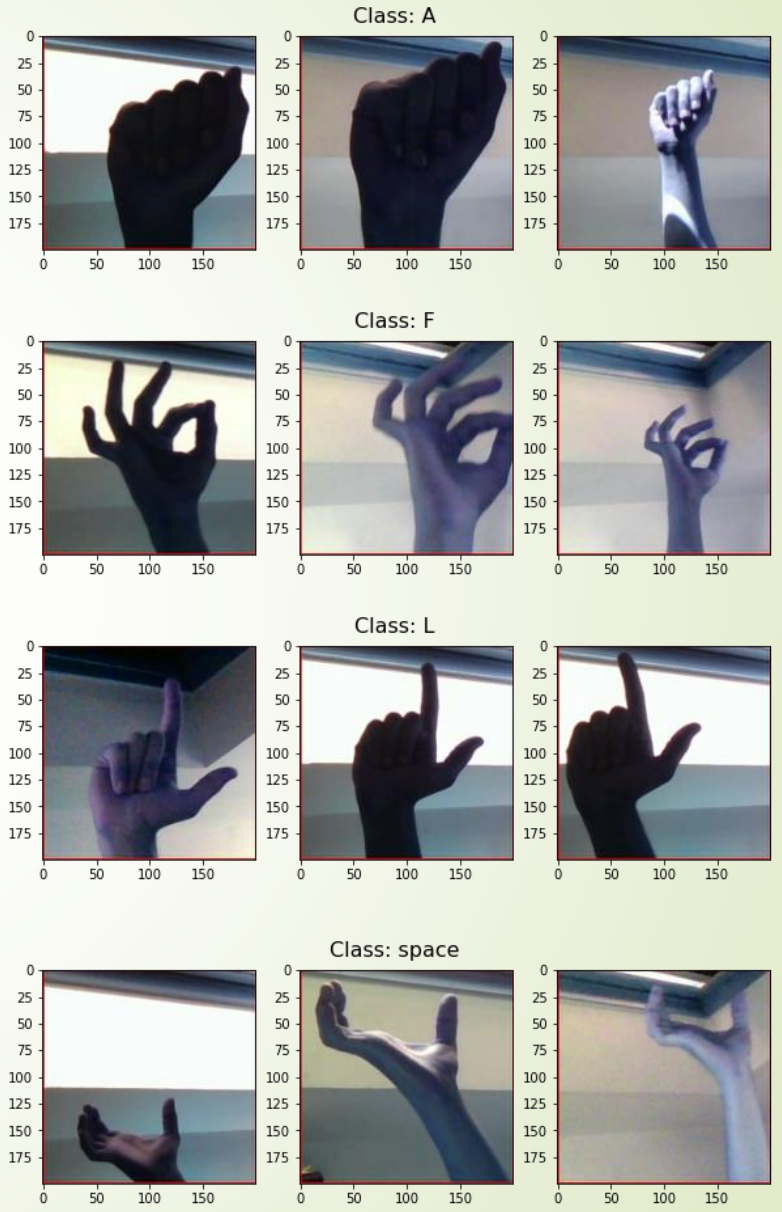


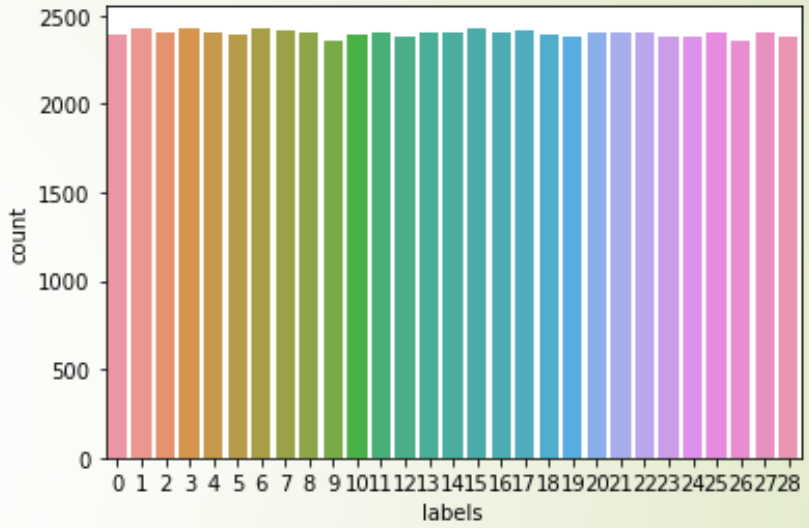
Image Preprocessing

I. Image size: 200x200 → 64x64

II. One-hot encoding on categorical labels

III. Split dataset to train/test sets

- Train set → 80% of initial dataset
- Test set → 20% of initial dataset



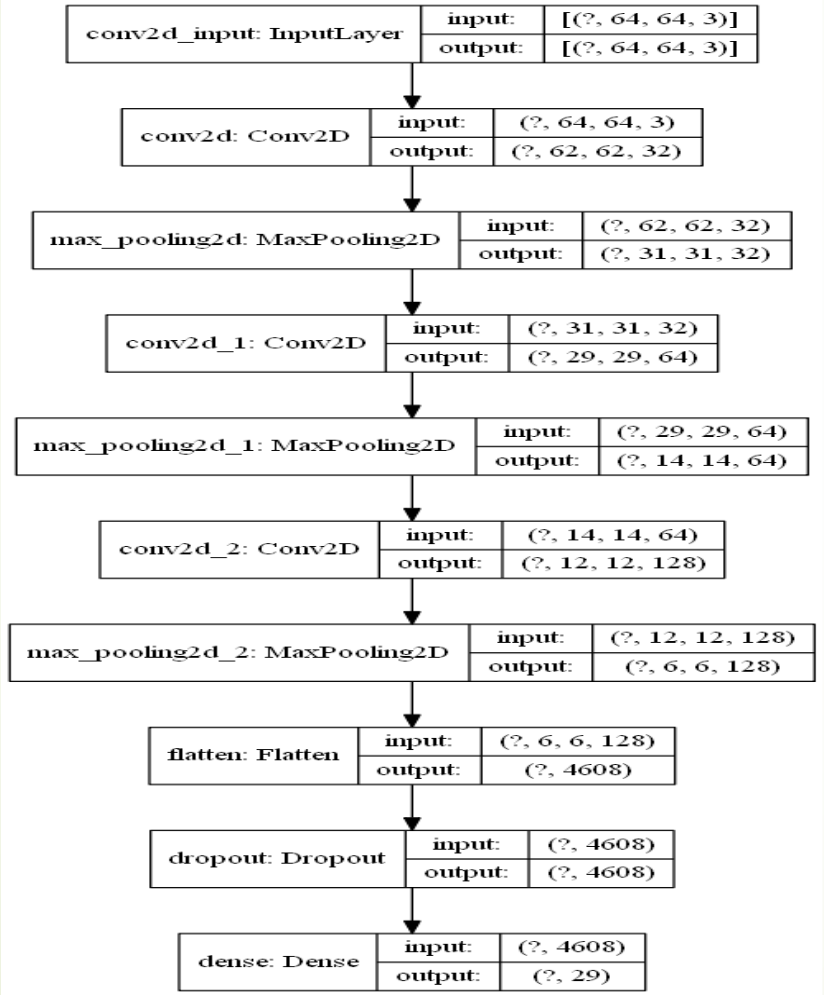
* Observation: Our dataset is balanced!

CNN Model

- Sequential CNN
  - Three 2D layers
  - Optimizer = Adam
  - Epochs = 20
  - Batch size = 64
  - Early Stop

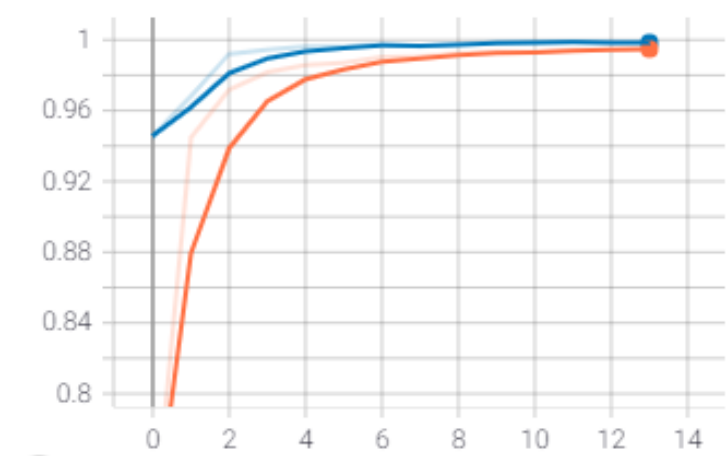
```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 62, 62, 32)        896
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 31, 31, 32)        0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 29, 29, 64)        18496
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 14, 14, 64)        0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 12, 12, 128)       73856
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 6, 6, 128)         0
_________________________________________________________________
flatten (Flatten)            (None, 4608)              0
_________________________________________________________________
dropout (Dropout)            (None, 4608)              0
_________________________________________________________________
dense (Dense)                (None, 29)                133661
=================================================================
Total params: 226,909
Trainable params: 226,909
Non-trainable params: 0
_________________________________________________________________
```

CNN Model

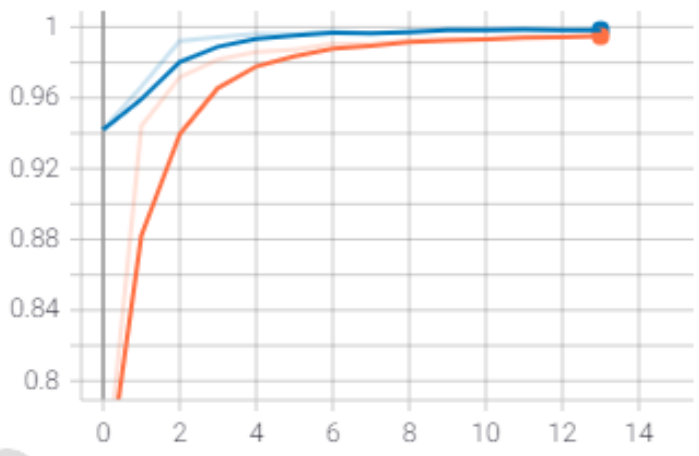

conv2d_input: InputLayer
input: [(?, 64, 64, 3)]
output: [(?, 64, 64, 3)]
conv2d: Conv2D
input: (?, 64, 64, 3)
output: (?, 62, 62, 32)
max_pooling2d: MaxPooling2D
input: (?, 62, 62, 32)
output: (?, 31, 31, 32)
conv2d_1: Conv2D
input: (?, 31, 31, 32)
output: (?, 29, 29, 64)
max_pooling2d_1: MaxPooling2D
input: (?, 29, 29, 64)
output: (?, 14, 14, 64)
conv2d_2: Conv2D
input: (?, 14, 14, 64)
output: (?, 12, 12, 128)
max_pooling2d_2: MaxPooling2D
input: (?, 12, 12, 128)
output: (?, 6, 6, 128)
flatten: Flatten
input: (?, 6, 6, 128)
output: (?, 4608)
dropout: Dropout
input: (?, 4608)
output: (?, 4608)
dense: Dense
input: (?, 4608)
output: (?, 29)

Results

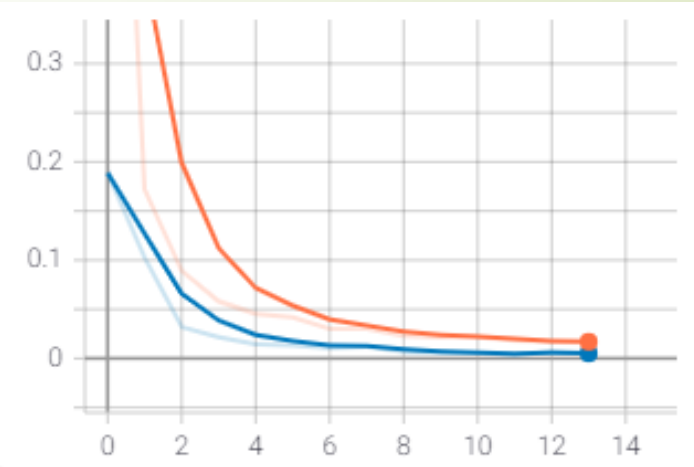
Accuracy



F1 Score



Loss



- ✓ Train: orange color
- ✓ Validation: blue color

Conclusion

- Metrics over 99%
- Loss less than 0.1%

Model almost accurate in recognizing American Sign Alphabet

THANK YOU FOR YOUR TIME!