American Sign Language Alphabet Recognition

Hao Wang
Tianchang Li
Jing Wen

What is American Sign Language?

American Sign Language is expressed by movements of the hands and face. It is primary language of many North Americans who are deaf and hard of hearing, and is used by many hearing people as well.

Why ASL important?

- Help deaf people communicate with others and receive messages.
- limitations.

Project Motivation

- Relieve hands of ASL translators.
- Assist online communication for people who use ASL.
- Encourage the interaction between deaf people and the public

Dataset

- ASL Alphabet from Kaggle
  - 3000 images for each alphabetic letter
  - 90% training and 10% valid
  - Another 1 image for each class for testing

- https://www.kaggle.com/grassknoted/asl-alphabet/data

A

D

H

Del

Nothing

Space

Project Procedure

- Random crop images into 64x64
- Transfer learning with ResNet and Inception
- Optimize structures and parameters

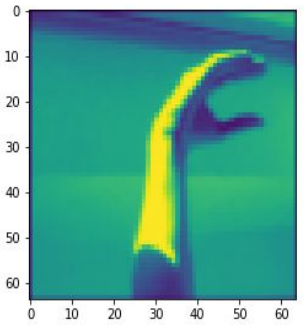
Models - ResNet

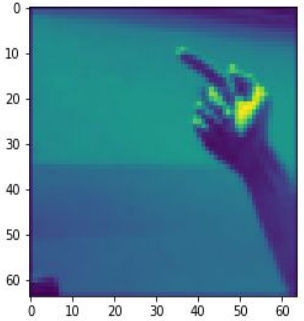
- 200 x 200 images, normalized
- Pre-trained ResNet with 29 classes
- Batch size = 256
- Accuracy > 96% after 7 epoches
- TOO SLOW

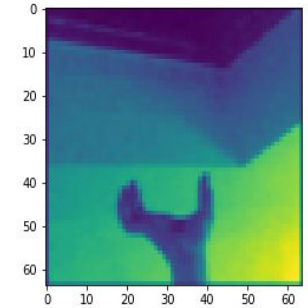
```
Epoch: 001/050 | Batch 000/312 | Cost: 3.6187
Epoch: 001/050 | Batch 120/312 | Cost: 1.2979
Epoch: 001/050 | Batch 240/312 | Cost: 0.8577
Epoch: 001/050 Train Acc.: 88.02% | Validation Acc.: 88.07%
Time elapsed: 147.42 min
Epoch: 002/050 | Batch 000/312 | Cost: 0.6861
Epoch: 002/050 | Batch 120/312 | Cost: 0.5177
Epoch: 002/050 | Batch 240/312 | Cost: 0.4941
Epoch: 002/050 Train Acc.: 92.47% | Validation Acc.: 92.24%
Time elapsed: 278.91 min
Epoch: 003/050 | Batch 000/312 | Cost: 0.3904
Epoch: 003/050 | Batch 120/312 | Cost: 0.3340
Epoch: 003/050 | Batch 240/312 | Cost: 0.2975
Epoch: 003/050 Train Acc.: 93.89% | Validation Acc.: 93.76%
Time elapsed: 406.97 min
Epoch: 004/050 | Batch 000/312 | Cost: 0.3646
Epoch: 004/050 | Batch 120/312 | Cost: 0.3013
Epoch: 004/050 | Batch 240/312 | Cost: 0.1960
Epoch: 004/050 Train Acc.: 95.14% | Validation Acc.: 95.01%
Time elapsed: 531.39 min
Epoch: 005/050 | Batch 000/312 | Cost: 0.3301
Epoch: 005/050 | Batch 120/312 | Cost: 0.2246
Epoch: 005/050 | Batch 240/312 | Cost: 0.2225
Epoch: 005/050 Train Acc.: 95.76% | Validation Acc.: 95.64%
Time elapsed: 659.01 min
Epoch: 006/050 | Batch 000/312 | Cost: 0.1563
Epoch: 006/050 | Batch 120/312 | Cost: 0.2355
Epoch: 006/050 | Batch 240/312 | Cost: 0.1807
Epoch: 006/050 Train Acc.: 96.21% | Validation Acc.: 95.96%
Time elapsed: 783.25 min
Epoch: 007/050 | Batch 000/312 | Cost: 0.1902
Epoch: 007/050 | Batch 120/312 | Cost: 0.1789
Epoch: 007/050 | Batch 240/312 | Cost: 0.1659
Epoch: 007/050 Train Acc.: 96.40% | Validation Acc.: 96.06%
Time elapsed: 907.61 min
Epoch: 008/050 | Batch 000/312 | Cost: 0.1289
Epoch: 008/050 | Batch 120/312 | Cost: 0.1582
Epoch: 008/050 | Batch 240/312 | Cost: 0.1731
Epoch: 008/050 Train Acc.: 96.89% | Validation Acc.: 96.43%
```

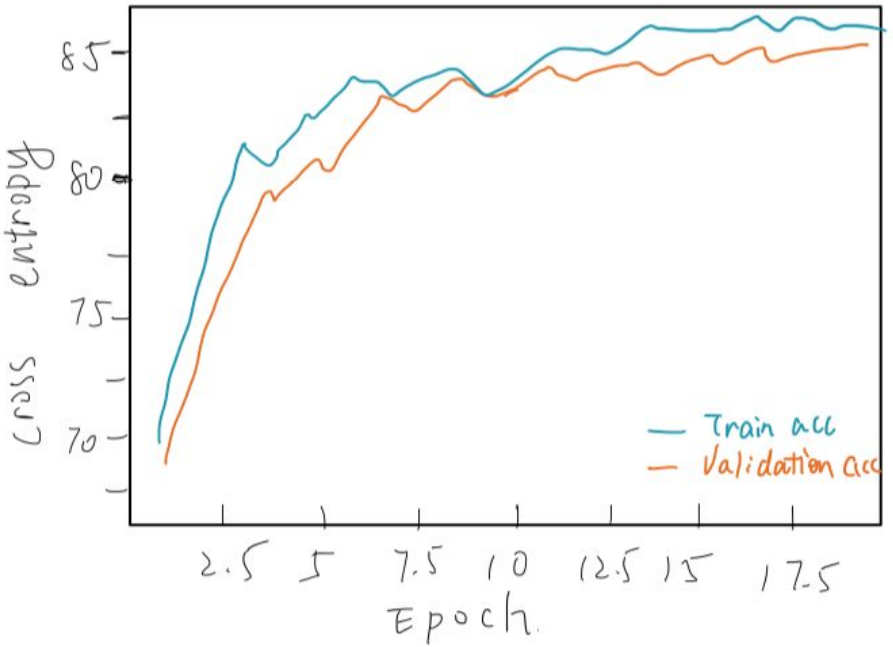
Models - ResNet

- Random crop into 64 x 64, normalized
- Pre-trained ResNet with 29 classes
- Batch size = 256
- Nice pattern in results
- Accuracy stuck around 85%

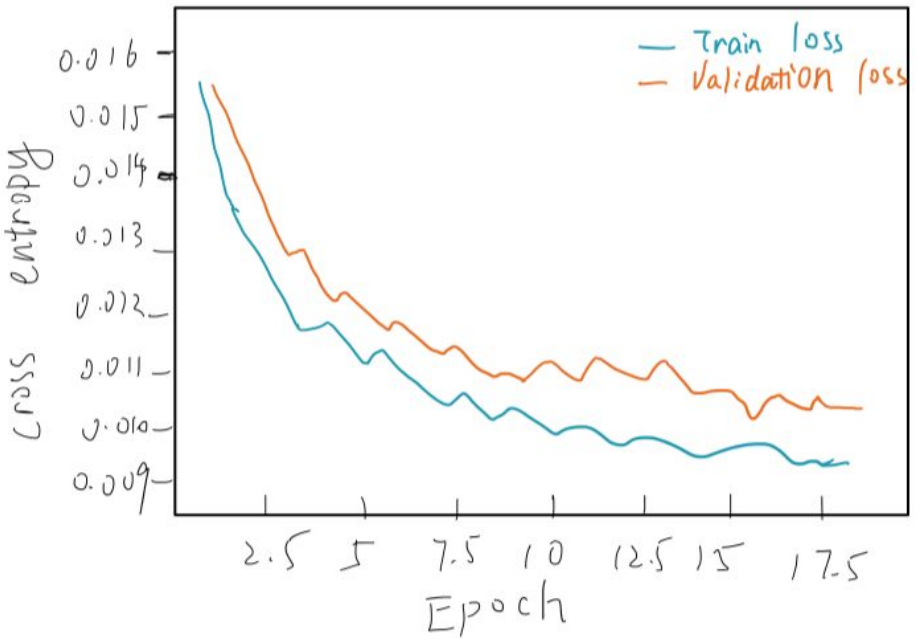








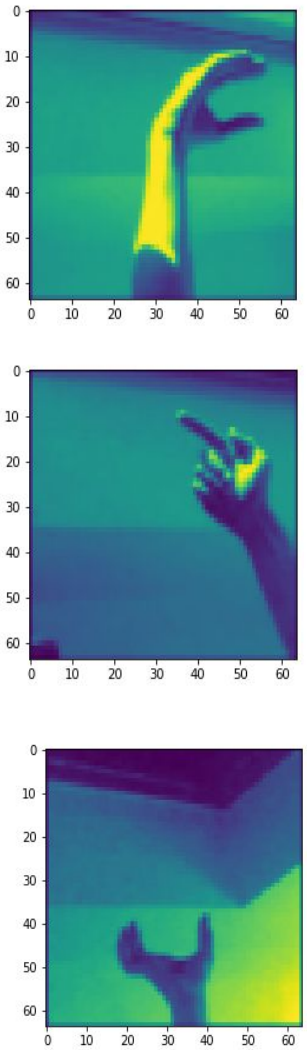
Lr = 1e-05 (pre-trained)
Batch size = 256
Epoch = 18

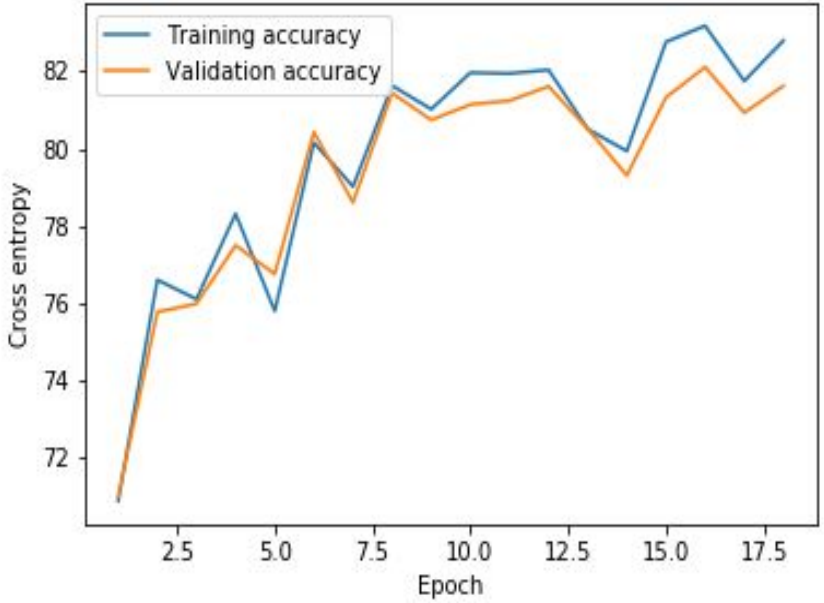


Accuracy and loss plots

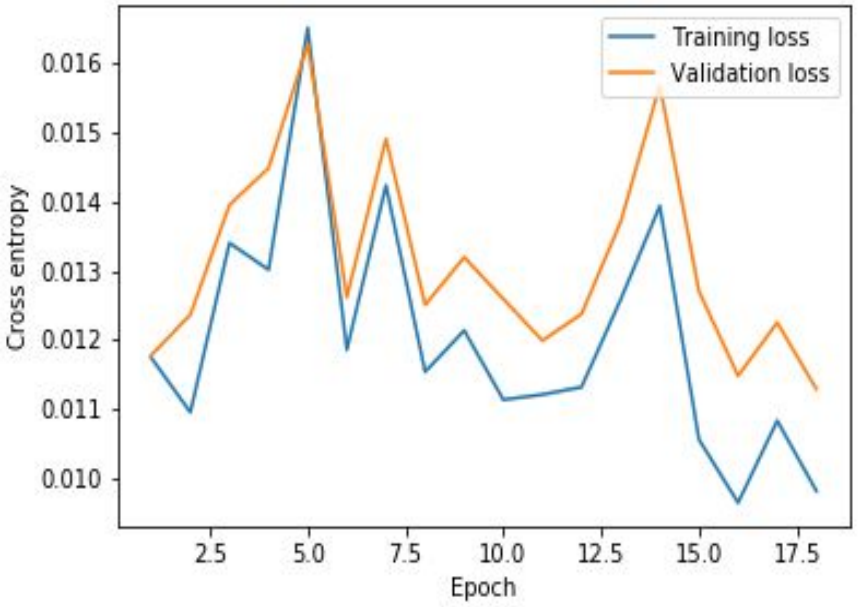
Models - ResNet

- Random crop into 64 x 64, normalized
- Batch size = 256
- Initial lr (1e-5 --> 0.1)
- Exponential lr decay (0.95)
- Loss fluctuates a lot





Lr = 0.1
Batch size = 256
Exp Decay = 0.95
Epoch = 18



Accuracy and loss plots

Model Evaluation

1. In general not a hard task for ResNet. Challenge is to maintain accuracy on small images
2. Achieved 85% so far
3. Try to break the local min by fine tuning the lr and batch size
4. Extend the training epochs to further improve the performance

Improvement & Conclusion

Change the background of pictures to avoid overfitting.

Do some common transformations to collected images like rotation and light adjustment to generalize the algorithm.

Include hierarchy feature in the model, like using Hyper Net to connect feature maps together.

Improvement & Conclusion

As we can see from the test accuracy, the accuracy is pretty high (about 85% overall).

But the running time is also unexpectedly long (it takes about 6 hours to train model). To solve this time-consuming problem, two measures can be taken:

1. Looking for more efficient models like Inception v-3 and v-4. Inception v-3 should achieve the same performance with less parameters, thus saving training time (This is exactly the model we are building now).
2. Because of the high prediction accuracy, pixels of images can be properly reduced.