Handwritten Letter Recognition

Jervis Muindi
jmuindi@cs.stanford.edu

OVERVIEW

Motivation: Although a lot communication is digital and electronic, there is still a considerable amount that occurs over traditional handwritten mediums

We want to develop an intelligent efficient system that recognizes handwritten letters. We implemented a variety of models and found that a simple CNN was best for this task.

DATA

EMNIST is an recent dataset that consists of handwritten letters. We use the EMNIST letters variant that has 120K+ handwritten english letters.

Example letters:

Data features: For the Support Vector machine, we scaled and normalized the input features. For the neural network we used the raw image inputs directly.

Data split: ~100K Train; ~20K Dev; ~20K Test

MODEL / RESULTS

LOGISTIC REGRESSION

- For baseline, we used Logistic Regression with softmax, a simple and fast to train model.

	Accuracy (%)	Training Time (s)
Baseline Logistic Regression	64.5	19

Table 1. Results for logistic regression

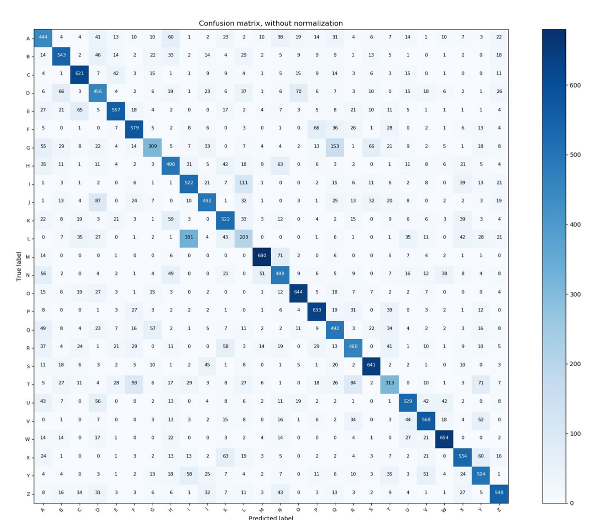

Figure 1. The confusion matrix for logistic regression

- Logistic regressions with softmax performs relatively well for a baseline
- One of the prevalent errors is model confusing 'G' with 'Q'

SUPPORT VECTOR MACHINE

- In Support Vector Machine with Kernel, some kernels may be better suited for the task of handwritten letter recognition. We implemented a SVM with four different kernels (Linear, RBF, Polynomial, Sigmoid) to identify the best one for this task empirically.

SVM Kernel	Accuracy (%)	Training Time (s)
Linear Kernel	41.6	521
RBF Kernel	70.4	619
Polynomial Kernel	73.0	556
Sigmoid Kernel	37.9	620

Parameters: polynomial degree of 3, Kernel coefficient gamma for {RBF, polynomial, sigmoid} set to 1/numfeatures

Table 2. Results for SVMs

- Best kernel is polynomial kernel with 73% accuracy
- However, Sigmoid SVM kernel performed worse than baseline model
- We suspect this is due to lack of parameter tuning: For the sigmoid kernel, if the chosen parameters are not well tuned, the algorithm can perform poorly[1]

NEURAL NETWORKS

- We looked at two types of neural networks: 1) Feed forward Neural Network 2) Convolutional Neural Network
- Feed forward network had 5 hidden layers with {512,256,128,64,32} hidden units.
- CNN network had 2 convolutional layers with max pool as shown below.

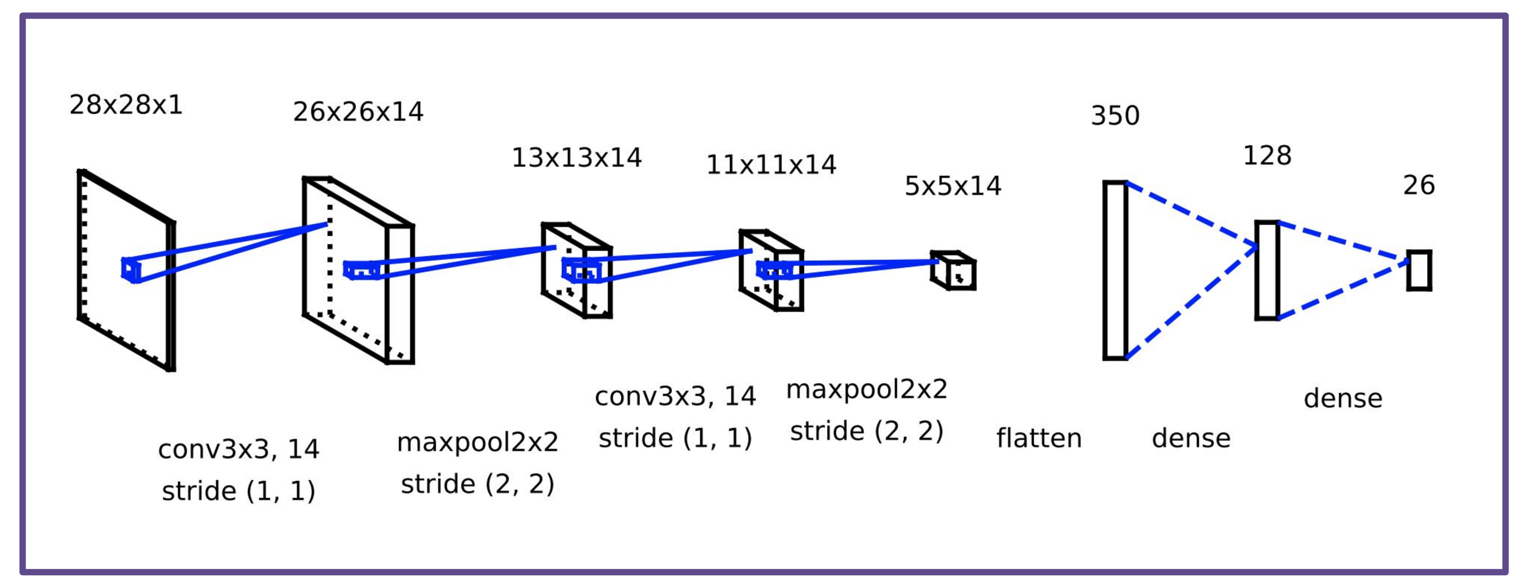

Figure 2. CNN model architecture

Neural Network Model	Accuracy (%)	Training Time (s)
Feed forward Network	90.3	1671
Convolutional Neural Network	91.1	1294

Table 3. Results for Neural Networks

- CNN model performed best attaining 91% accuracy
- The most common remaining error that model is making is confusing 'I' with 'L'. Makes sense especially considering lowercase: 'i' vs 'l'

TRANSFER LEARNING

- Training a deep-learning model from scratch can be expensive. Transfer learning allows one to leverage a pre-trained model on one task to use in another (related) task.
- We explored using pre-trained ResNet model from the ImageNet competition for handwritten letter recognition

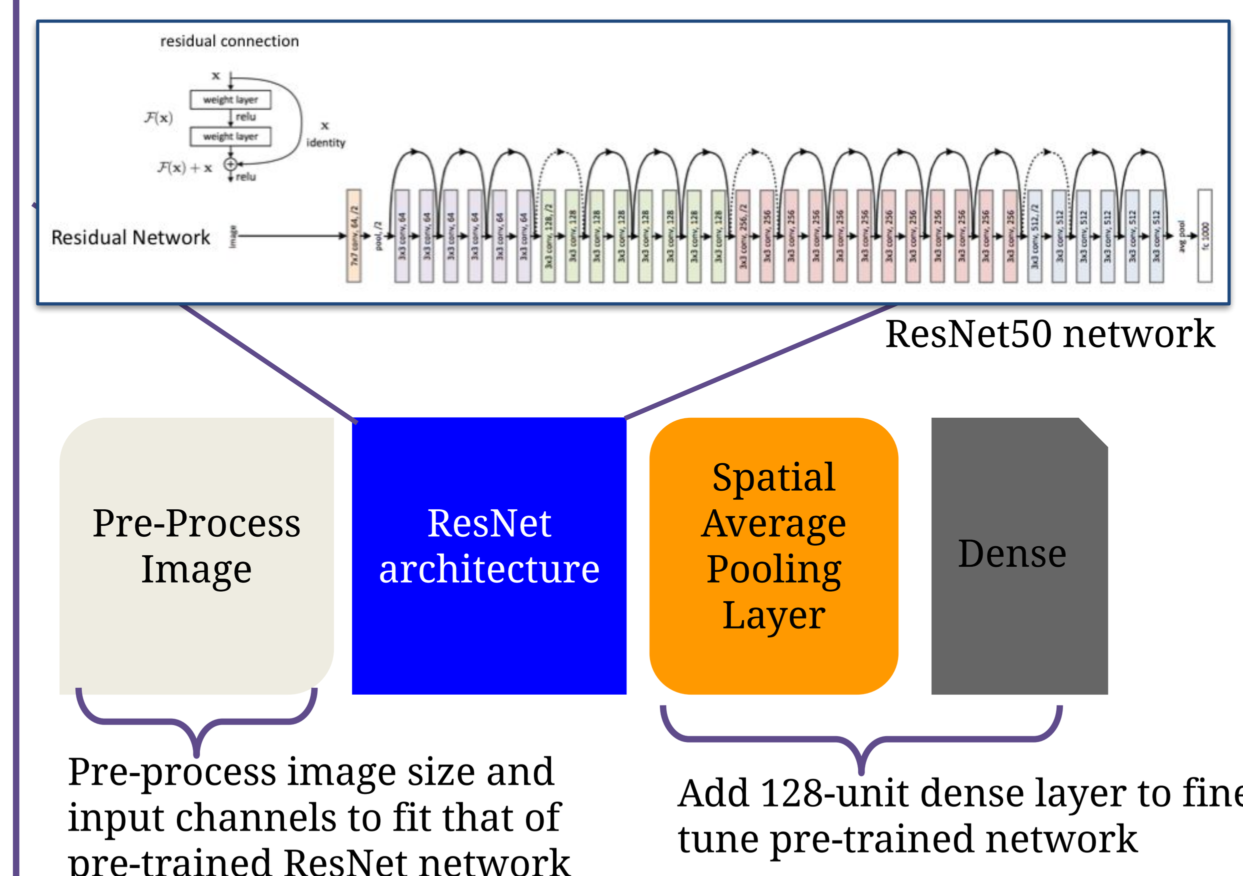

Figure 3. A diagram of how we carried out transfer learning

Model	Accuracy (%)	Training Time (s)
ResNet50	13.6	73823

Table 4. Results for transfer learning

DISCUSSION

Training Time Vs Accuracy tradeoff

- Judging purely on the basis of training time, Logistic regression was the fastest with a training time of less than a minute, followed by SVM.
- However, assuming one has more resources available to train more sophisticated models, we find that our CNN performs the best with an accuracy of more than 90%.

Transfer Learning

- Transfer learning model was expensive to train/tune due to the deep nature of the source models. Only able to train for 2 full epochs.
- Some recent research suggests that transfer learning with ImageNet models[2] may not always work.
- In our case, we find transfer learning from ResNet did not yield positive results in training time/ accuracy.

FUTURE WORK

- Work on improving model accuracy to improve performance even more.
- Explore using this model as a foundation in a system to recognize handwritten handwriting all together.
- Explore a different approach of transfer learning: using transfer learning as a fixed feature extractor for logistic regression

REFERENCES

[1] R. Amami, D. B. Ayed, and N. Ellouze. Practical selection of svm supervised parameters with different feature representations for vowel recognition. arXivpreprint arXiv:1507.06020, 2015.
[2] Simon Kornblith, Jonathan Shlens, and Quoc V. Le. Do better imagenet models transfer better? CoRR, abs/1805.08974, 2018. URL http://arxiv.org/abs/1805.08974