Motivation

- CNN Pooling layer results in poor translational invariance
- CNN gets easily fooled on adversarial images, with the image with eyes deformed and dislocated, CNN still output a human face
- Goal is to build a new NN that is as good as CNN, and have better translational invariance qualities
- Capsule proposed by Hinton et., al. is a good candidate to build a hierarchical network for this purpose, and we shall compare the performance against CNN

Data & Features

- Training is strictly performed on MNIST standard train dataset, and no augmented dataset has been trained
- Apart from the standard MNIST dataset, we generated additional 500+test images to test the translational invariance quality of both CNN and Capsule based HNN
- Rotation operation examples

![Rotation Examples]

- Overlap operation examples

![Overlap Examples]

Architecture & Model

Loss Function

- Implemented capsule network architecture with dynamic routing in tensorflow/ keras and trained with MNIST

Results

Discussion & Future

- Capsule network performed equally well against CNN on standard MNIST database without any augmentation
- With augmented test data (rotation and overlaps), clearly capsule network outperformed the CNN predictions
- Capsule based hierarchical neural networks can perform image recognition tests better than CNN with a small training set
- Training a capsule based network takes significantly larger time than with a CNN – this needs to be investigated
- Test the capsule network on complex image processing application such as face recognition
- Investigate if capsule concept can be applied to other forms of tasks such as voice recognition, text sentiment analysis, etc.

References: