



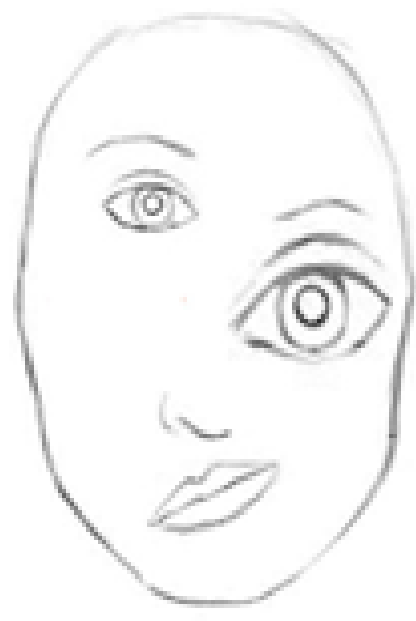
Hierarchical Neural Network

Suresh Sugumar
ssugumar@stanford.edu

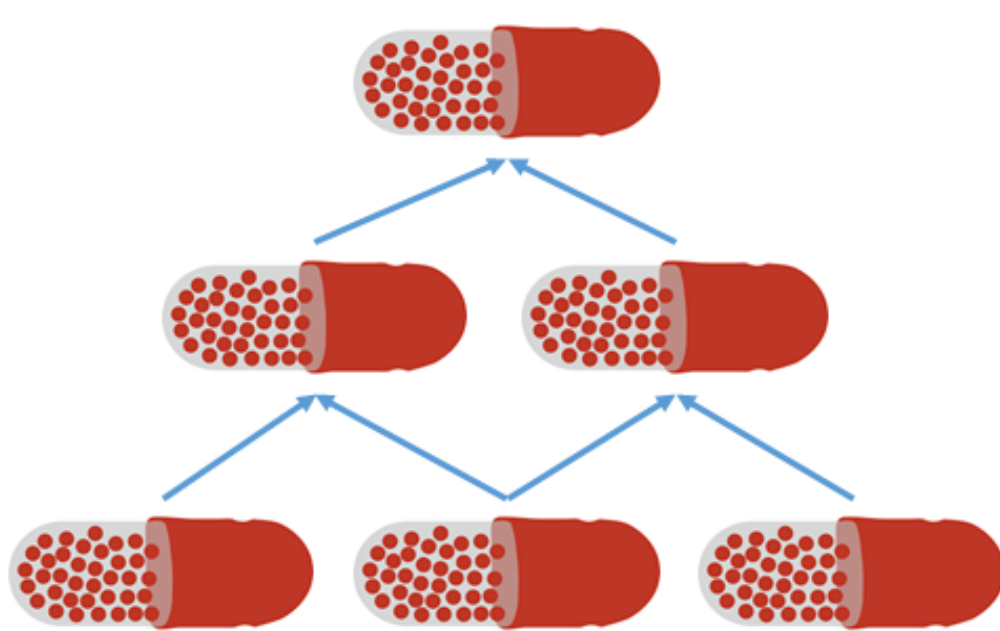
CS230
Deep Learning
Spring 2019

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



Hierarchical Neural Network
based on Capsules



Data & Features

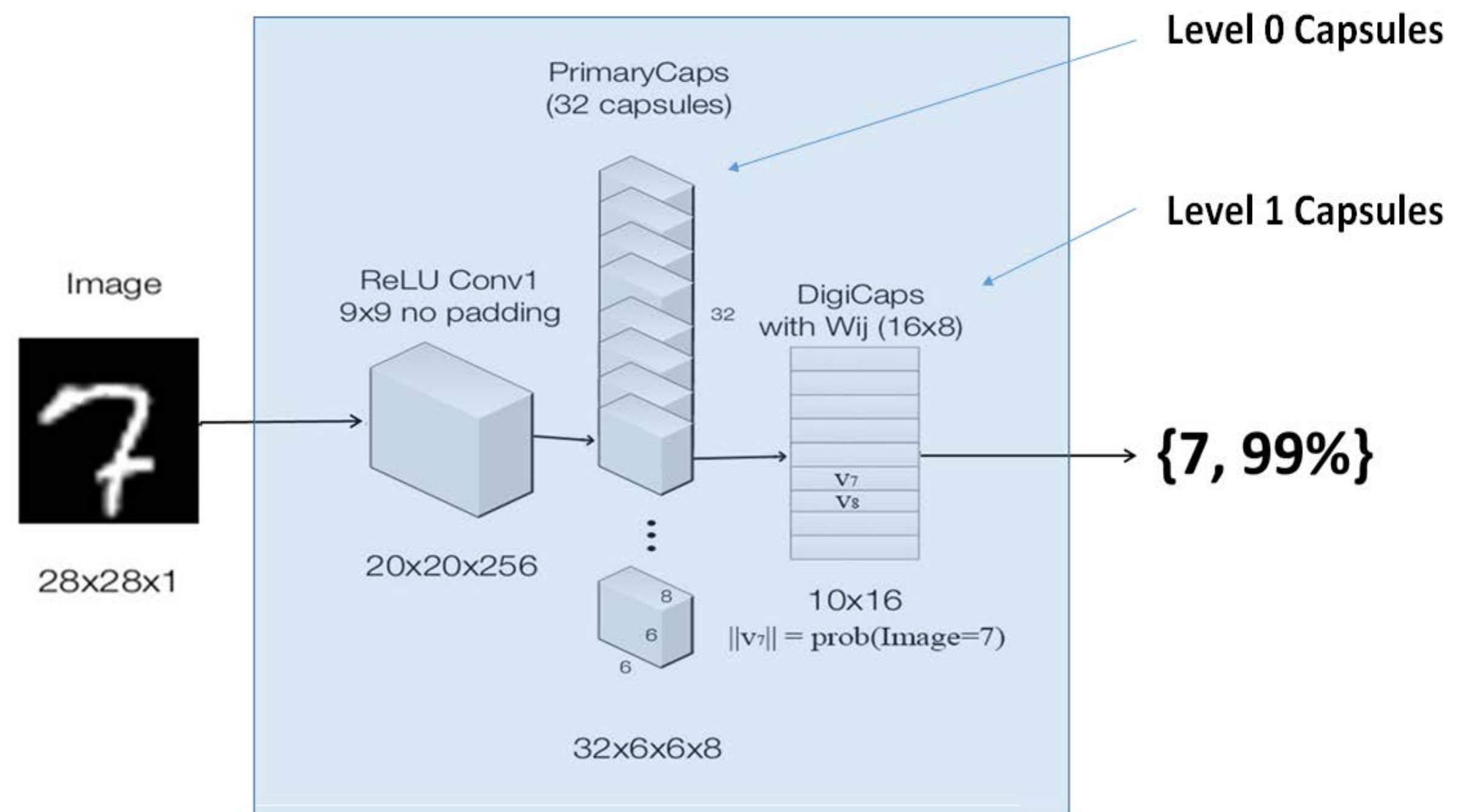
- Training is strictly performed on MNIST standard train dataset, and no augmented dataset has been trained
- Apart from the standard MNIST dataset, we generated additional 500+test images to test the translational invariance quality of both CNN and Capsule based HNN
- Rotation operation examples



- Overlap operation examples



Architecture & Model



Loss Function

$$L_c = T_c \max(0, m^+ - ||\mathbf{v}_c||)^2 + \lambda (1 - T_c) \max(0, ||\mathbf{v}_c|| - m^-)^2$$

Loss term for one DigitCap

calculated for correct DigitCap

calculated for incorrect DigitCaps

1 when correct DigitCap, 0 when incorrect

zero loss when correct prediction with probability greater than 0.9, non-zero otherwise

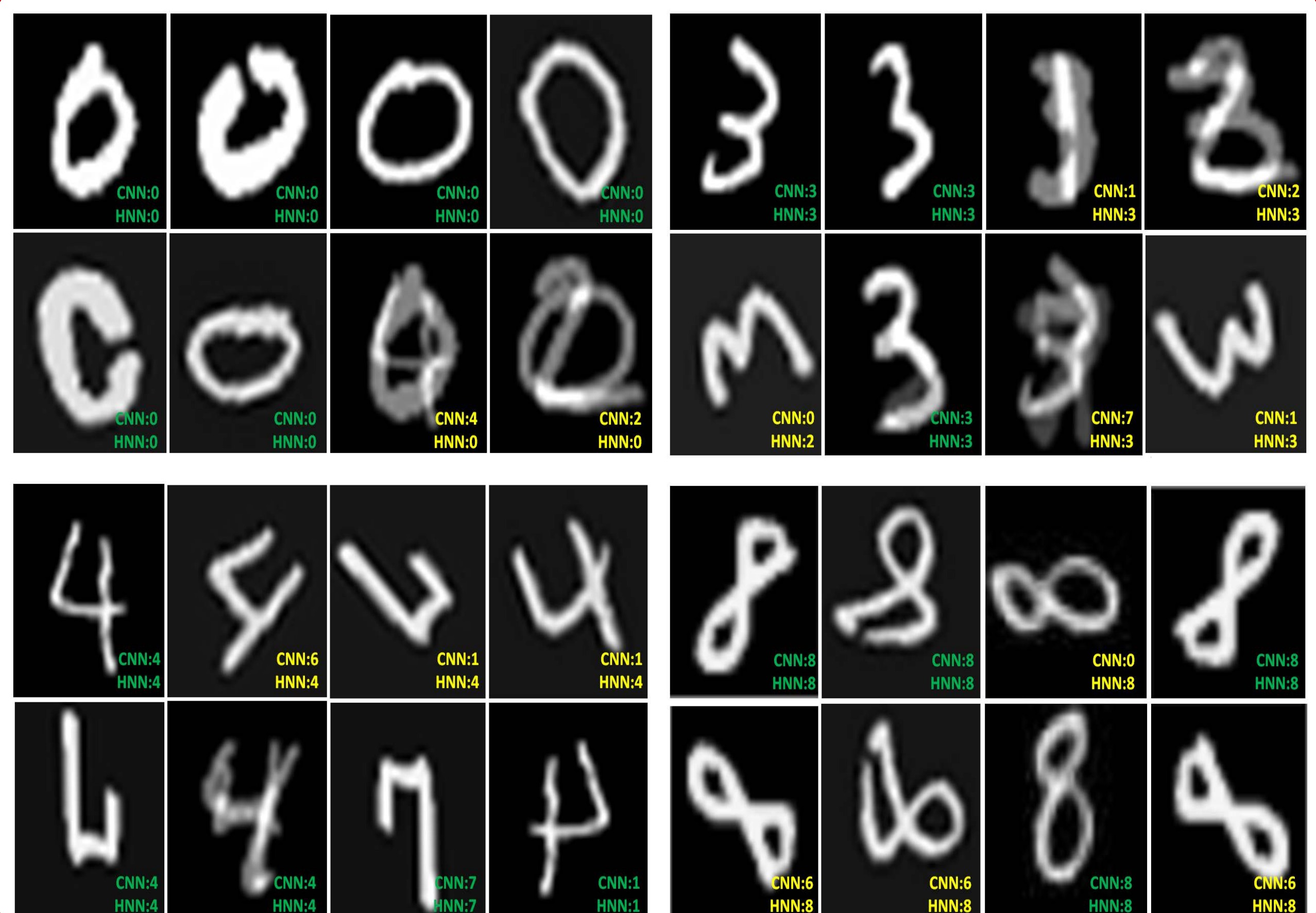
0.5 constant used for numerical stability

1 when incorrect DigitCap, 0 when correct

zero loss when incorrect prediction with probability less than 0.1, non-zero otherwise

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.

GitHub <https://github.com/sureshsugumar/HierNet>

YouTube <https://youtu.be/TL81f10X6vg>

References:

- Geoffrey E Hinton et., al. Dynamic Routing Between Capsules. [arXiv:1710.09829](https://arxiv.org/abs/1710.09829), 2017.