CVPR LONG BEACH CALIFORNIA June 16-20, 2019

Center for Research in Computer Vision
UCF
UNIVERSITY OF CENTRAL FLORIDA

Capsule Networks: A Survey

Yogesh Rawat

CVPR Tutorial

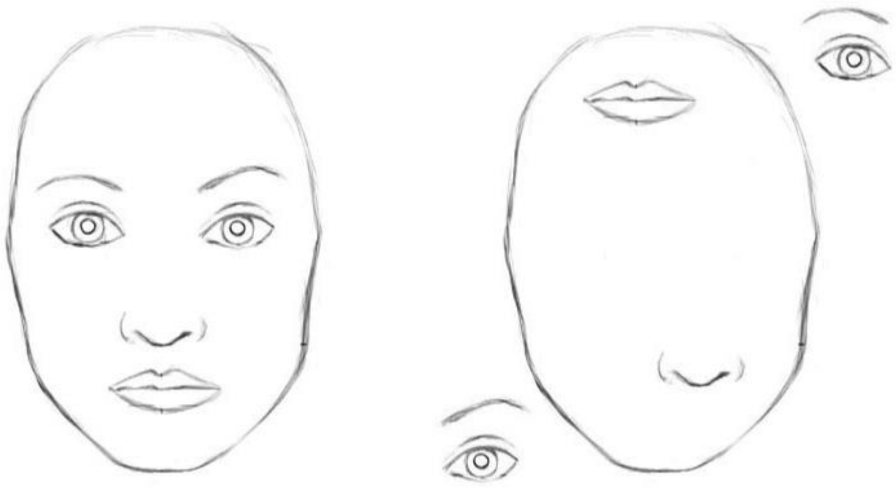
Sunday, June 16, 2019

Outline

- Introduction
- CapsNet
- EM routing
- Application[3D point cloud]

Introduction

- Motivation[CNN drawbacks]
  - *Internal data representation of a convolutional neural network does not take into account important spatial hierarchies between simple and complex objects.*



CNN:
neuron [scalar]
activation

CapsNet:
capsule [vector/matrix]
activation and state

Introduction

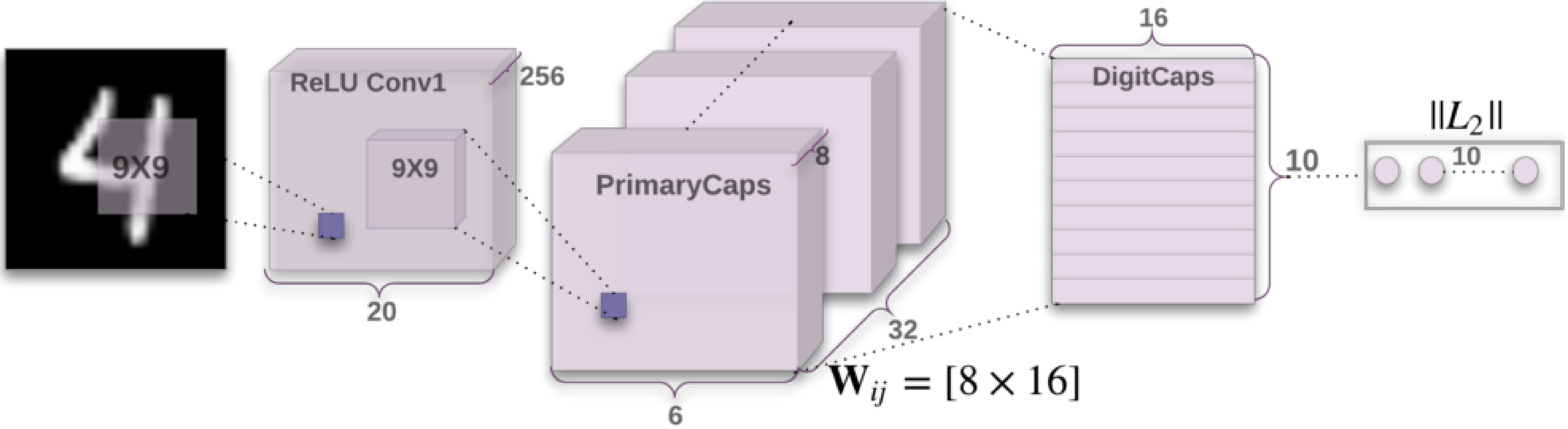
- Motivation
  - *Hinton: "The pooling operation used in convolutional neural networks is a big mistake and the fact that it works so well is a disaster."*

CNN:
- max pooling
- average pooling

CapsNet:
- dynamic routing(routing by agreement)
- EM routing

Introduction

- Motivation
  - Inverse rendering



CapsNet

- What is capsule?
  Capsules encapsulate all important information about the state of the feature they are detecting in **vector** form.

  Vector: length & orientation

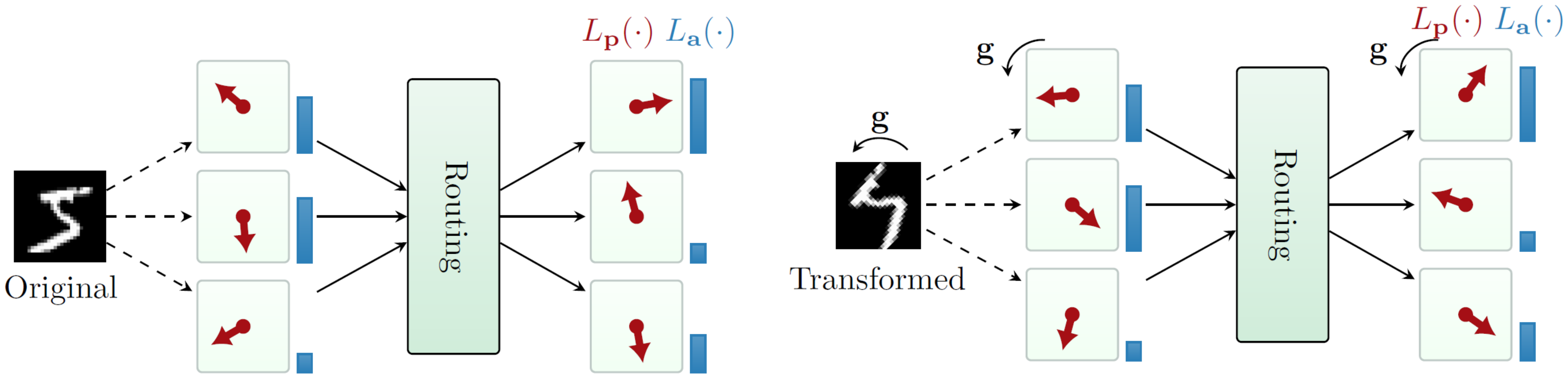
  想要达到的效果:

  when detected feature moves around the image or its state somehow changes, the probability still stays the same (length of vector does not change), but its orientation changes.

  [invariance & variance]

Generic routing by agreement

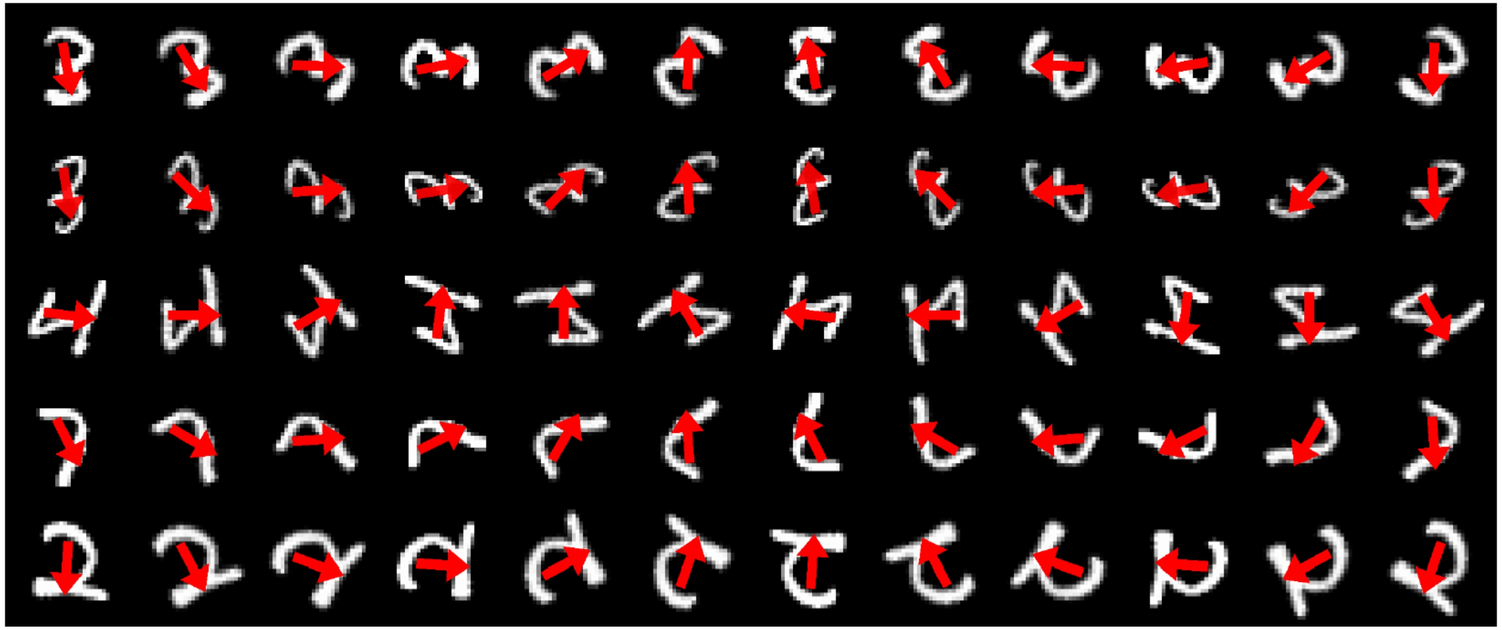
- Equivalance and invariance [5]



[5] Lenssen, Jan Eric, Matthias Fey, and Pascal Libuschewski. "Group equivariant capsule networks." Advances in Neural Information Processing Systems. 2018.

Equivariance and Invariance

- Variation in pose vector



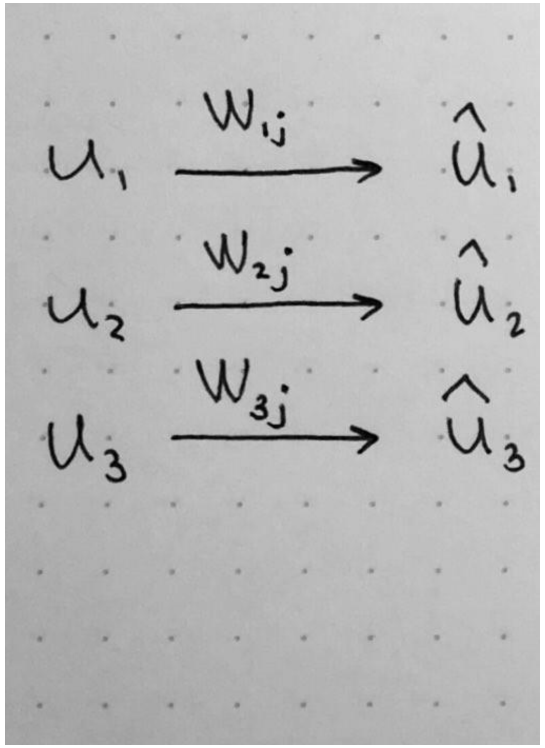
[5] Lenssen, Jan Eric, Matthias Fey, and Pascal Libuschewski. "Group equivariant capsule networks.“ Advances in Neural Information Processing Systems. 2018.

CapsNet

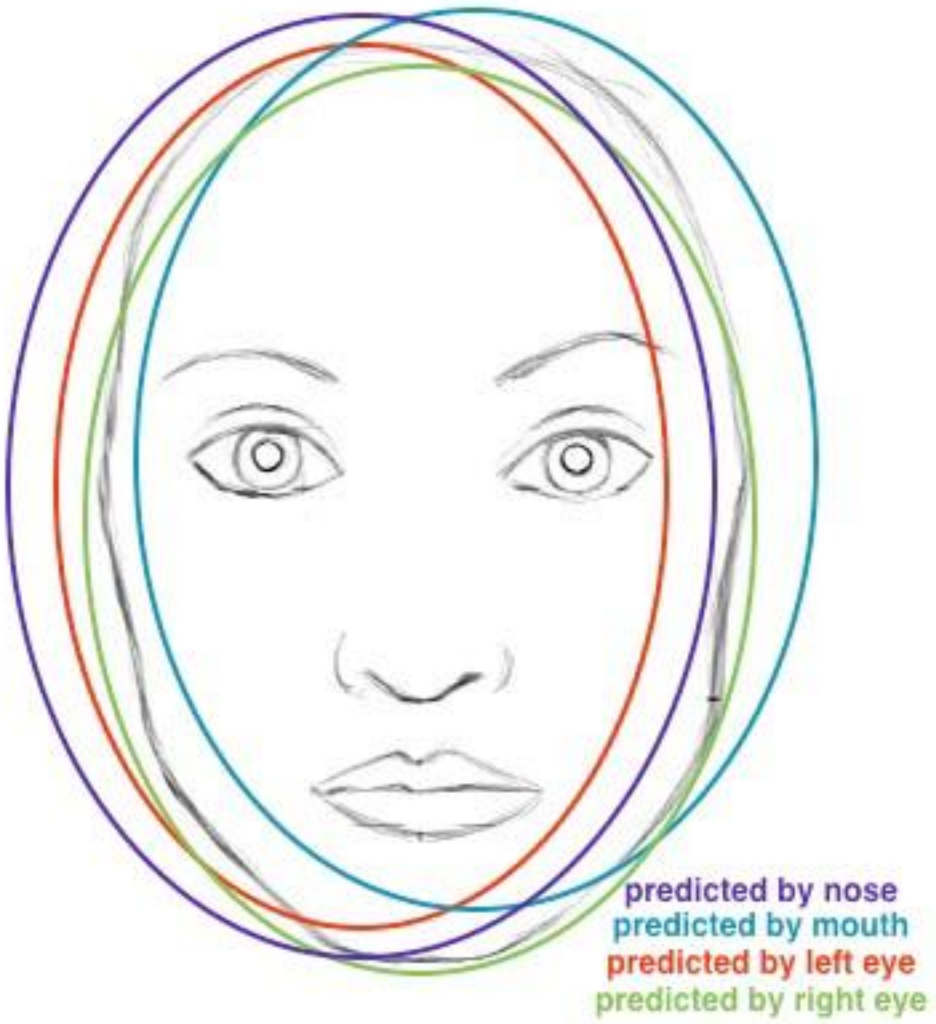
- How capsule works
  1. matrix multiplication of input vectors
  2. scalar weighting of input vectors
  3. sum of weighted input vectors
  4. vector-to-vector nonlinearity

CapsNet

1. matrix multiplication of input vectors



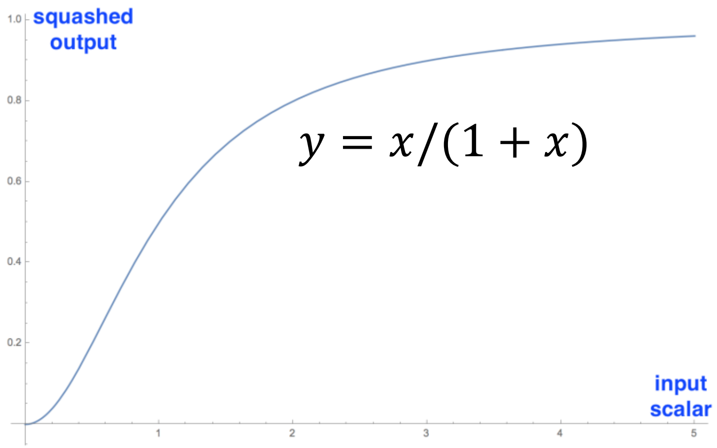
$\hat{u_i}$: predict vectors
u_i: output of last layer capsules



CapsNet

2. Scalar Weighting of Input Vectors
   Question: How to determine the weights/coefficients?
   Answer: Dynamic routing
3. Sum of Weighted Input Vectors
4. “Squash”: Novel Vector-to-Vector Nonlinearity

$$\mathbf{v}_j = \underbrace{\boxed{\frac{\|\mathbf{s}_j\|^2}{1+\|\mathbf{s}_j\|^2}}}_{\text{additional “squashing”}} \underbrace{\boxed{\frac{\mathbf{s}_j}{\|\mathbf{s}_j\|}}}_{\text{unit scaling}}$$



CapsNet

- Dynamic routing

  "*Lower level capsule will send its input to the higher level capsule that "agrees" with its input. This is the essence of the dynamic routing algorithm.*" – routing by agreement

Procedure 1 Routing algorithm.

procedure ROUTING($\hat{\boldsymbol{u}}_{j|i}$, r, l)
    for all capsule i in layer l and capsule j in layer $(l+1)$: $b_{ij} \leftarrow 0$.
    for r iterations **do**
        for all capsule i in layer l: $\mathbf{c}_i \leftarrow \mathtt{softmax}(\mathbf{b}_i)$ ▷ `softmax` computes Eq. 3
        for all capsule j in layer $(l+1)$: $\mathbf{s}_j \leftarrow \sum_i c_{ij} \hat{\mathbf{u}}_{j|i}$
        for all capsule j in layer $(l+1)$: $\mathbf{v}_j \leftarrow \mathtt{squash}(\mathbf{s}_j)$ ▷ `squash` computes Eq. 1
        for all capsule i in layer l and capsule j in layer $(l+1)$: $b_{ij} \leftarrow b_{ij} + \hat{\mathbf{u}}_{j|i}.\mathbf{v}_j$
    return $\mathbf{v}_j$

CapsNet

- Dynamic routing questions
  - What is agreement?
  - Softmax along which axis? [My experiment result]

    99.60 v.s. 99.37 on MNIST

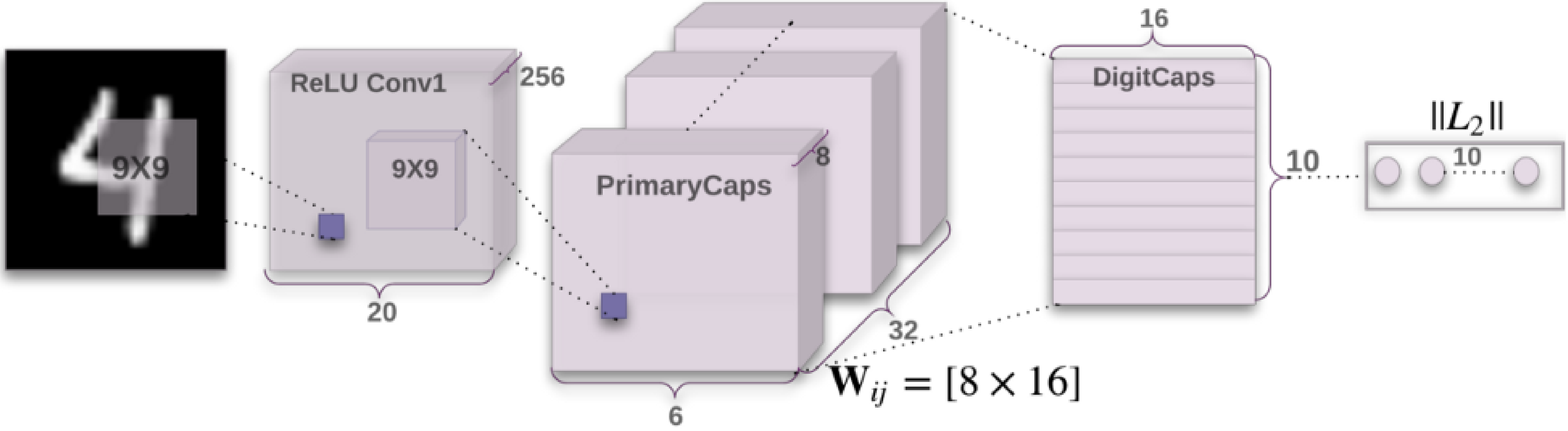
  - How many iterations to use?
    - two many iterations: overfit
    - suggestion: 3
  - Trainable parameters?
    W_{ij}, C_{ij}

CapsNet

- Network architecture
  - Encoder
  - Decoder[regularization]

CapsNet

- Part I. Encoder



CapsNet

- Part I. Encoder

CapsNet Loss Function

calculated for correct DigitCap

calculated for incorrect DigitCaps

loss term for one DigitCap

L2 norm

L2 norm

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

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

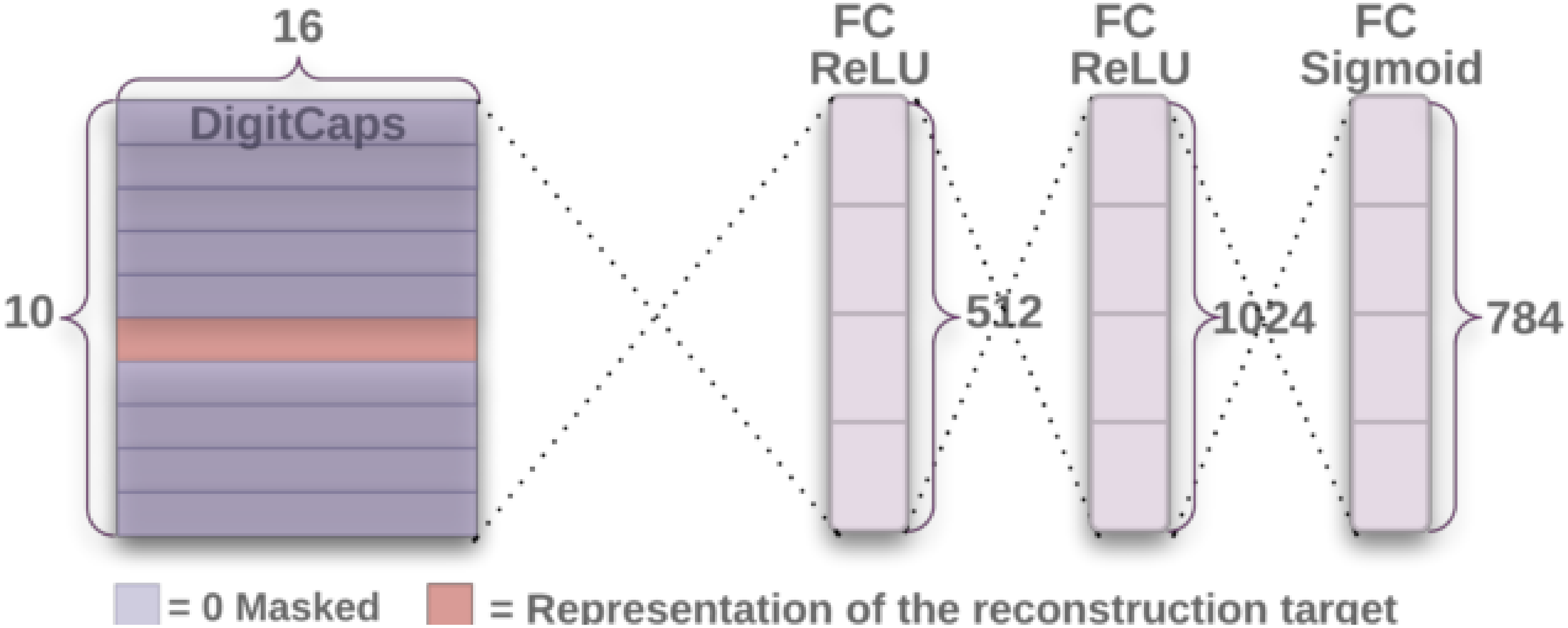
1 when incorrect DigitCap, 0 when correct

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

Note: correct DigitCap is one that matches training label, for each training example there will be 1 correct and 9 incorrect DigitCaps

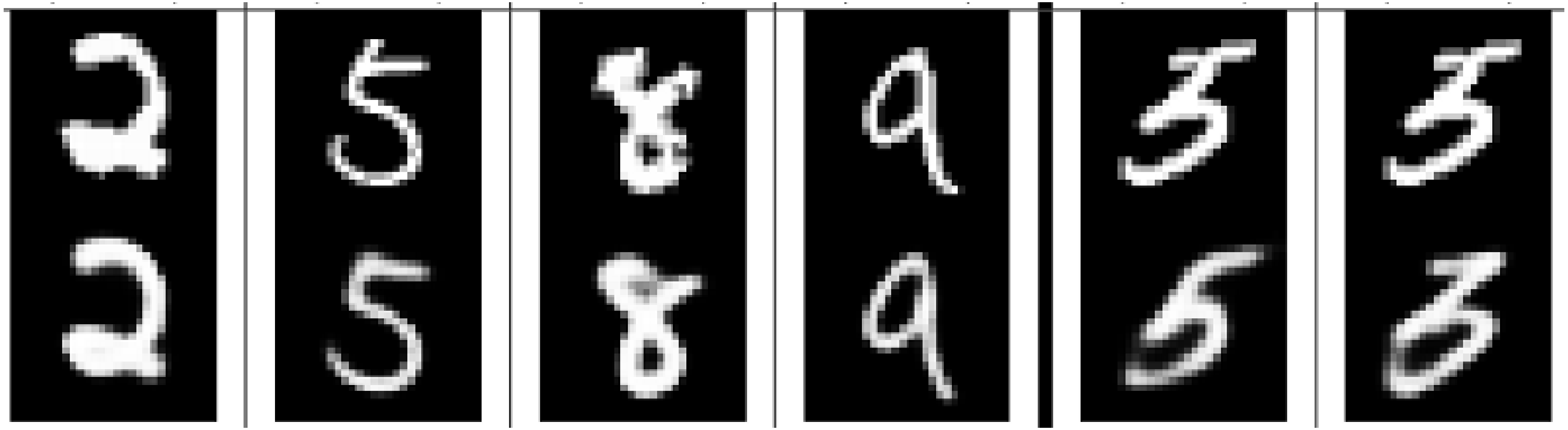
CapsNet

- Part II. Decoder



CapsNet

- Part II. Decoder

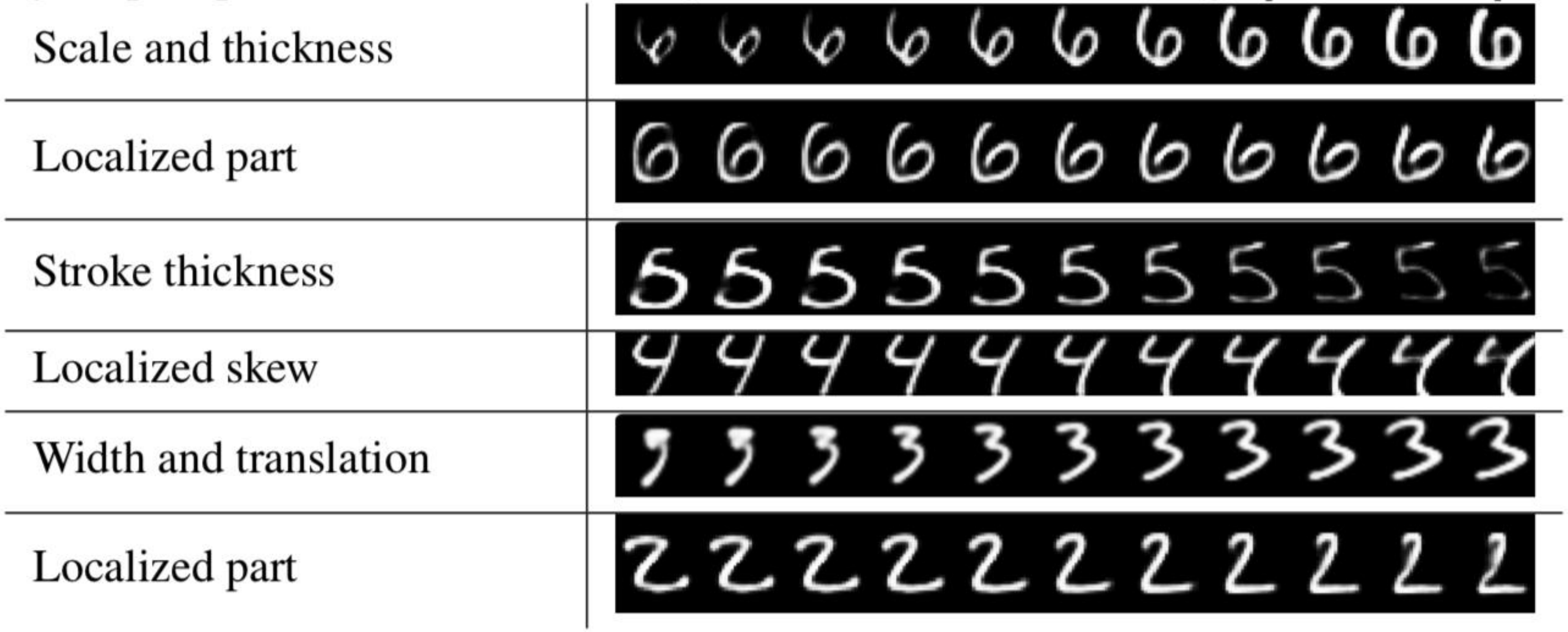


Question: why a decoder is necessary?
Loss: MSE

CapsNet Experiments

- What the individual dimensions of a capsule represent

Figure 4: Dimension perturbations. Each row shows the reconstruction when one of the 16 dimensions in the DigitCaps representation is tweaked by intervals of 0.05 in the range $[-0.25, 0.25]$.

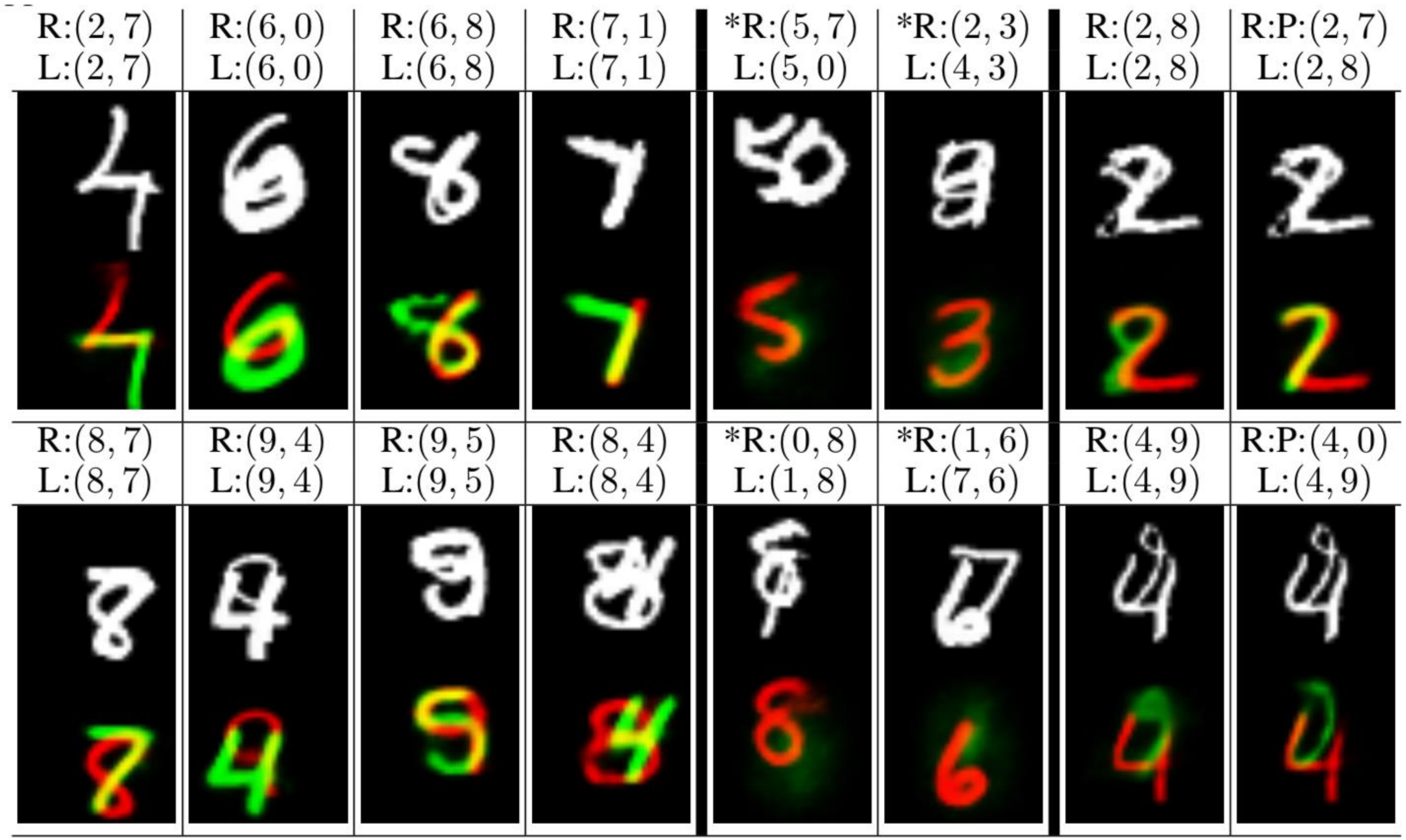


CapsNet Experiments

- Robustness to Affine Transformations
  - SOTA CNN:
    99.22 % -> 66 %
  - CapsNet:
    99.23% -> 79%

CapsNet Experiments

- Segmenting highly overlapping digits[MultiMNIST]



CapsNet Experiments

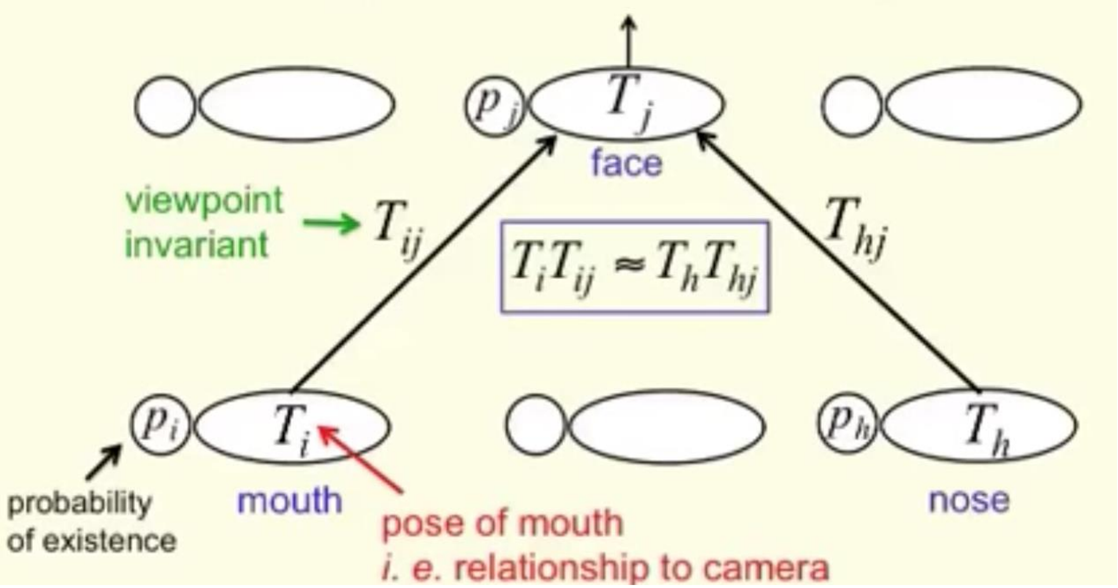
- Overall

Table 1: CapsNet classification test accuracy. The MNIST average and standard deviation results are reported from 3 trials.

Method	Routing	Reconstruction	MNIST (%)	MultiMNIST (%)
Baseline	-	-	0.39	8.1
CapsNet	1	no	$0.34_{\pm 0.032}$	-
CapsNet	1	yes	$0.29_{\pm 0.011}$	7.5
CapsNet	3	no	$0.35_{\pm 0.036}$	-
CapsNet	3	yes	$\mathbf{0.25}_{\pm 0.005}$	**5.2**

Dynamic vs EM routing

- Both are iterative
- Dynamic
  - Squash function
  - Cosine similarity
- EM
  - Existence probability
  - Distribution



[3] Sabour, Sara, Nicholas Frosst, and Geoffrey E. Hinton. "Dynamic routing between capsules." *Advances in neural information processing systems*. 2017.
[4] Hinton, Geoffrey E., Sara Sabour, and Nicholas Frosst. "Matrix capsules with EM routing." ICLR (2018).
Image source: https://jhui.github.io

References

1. Hinton, Geoffrey E., Alex Krizhevsky, and Sida D. Wang. "Transforming auto-encoders." In International Conference on Artificial Neural Networks, 2011.
2. Kulkarni, Tejas D., William F. Whitney, Pushmeet Kohli, and Josh Tenenbaum. "Deep convolutional inverse graphics network." In Advances in neural information processing systems, 2015..
3. Sabour, Sara, Nicholas Frosst, and Geoffrey E. Hinton. "Dynamic routing between capsules." *Advances in neural information processing systems*. 2017.
4. Hinton, Geoffrey E., Sara Sabour, and Nicholas Frosst. "Matrix capsules with EM routing." ICLR (2018).
5. Lenssen, Jan Eric, Matthias Fey, and Pascal Libuschewski. "Group equivariant capsule networks." Advances in Neural Information Processing Systems. 2018.
6. Neural Network Encapsulation http://openaccess.thecvf.com/content_ECCV_2018/papers/Hongyang_Li_Neural_Network_Encapsulation_ECCV_2018_paper.pdf ECCV 2018
7. Zhang, Suofei, Quan Zhou, and Xiaofu Wu. "Fast dynamic routing based on weighted kernel density estimation." International Symposium on Artificial Intelligence and Robotics. Springer, Cham, 2018.
8. Duarte, Kevin, Yogesh Rawat, and Mubarak Shah. "Videocapsulenet: A simplified network for action detection." Advances in Neural Information Processing Systems. 2018.
9. Zhao, Wei, et al. "Investigating capsule networks with dynamic routing for text classification." *arXiv preprint arXiv:1804.00538*(2018).
10. Xinyi, Zhang, and Lihui Chen. "Capsule Graph Neural Network." ICLR (2018).
11. 3D Point-Capsule Networks https://arxiv.org/abs/1812.10775 CVPR 2019
12. Multi-modal capsules
13. DeepCaps: Going Deeper with Capsule Networks https://arxiv.org/abs/1904.09546 CVPR 2019
14. CapProNet: Deep Feature Learning via Orthogonal Projections onto Capsule Subspaces https://papers.nips.cc/paper/7823-cappronet-deep-feature-learning-via-orthogonal-projections-onto-capsule-subspaces.pdf NIPS 2018
15. LaLonde, Rodney, and Ulas Bagci. "Capsules for Object Segmentation."Conference on Medical Imaging with Deep Learning (MIDL 2018), (2018)

Applications

16. Afshar, Parnian, Arash Mohammadi, and Konstantinos N. Plataniotis. "Brain tumor type classification via capsule networks." 2018 25th IEEE International Conference on Image Processing (ICIP). IEEE, 2018.
17. Iesmantas, Tomas, and Robertas Alzbutas. "Convolutional capsule network for classification of breast cancer histology images." *International Conference Image Analysis and Recognition*. Springer, Cham, 2018.
18. Zhang, Xinsong, et al. "Multi-labeled Relation Extraction with Attentive Capsule Network." *AAAI 2018*.
19. Zhang, Ningyu, et al. "Attention-Based Capsule Networks with Dynamic Routing for Relation Extraction." EMNLP 2018.
20. Frosst, Nicholas, Sara Sabour, and Geoffrey Hinton. "DARCCC: Detecting Adversaries by Reconstruction from Class Conditional Capsules." arXiv preprint arXiv:1811.06969(2018).