

AI Ukraine 2018

Introduction to capsule networks

Andrii Babii, Ph.D. - apratster@gmail.com

Before we start

Hinton's articles:

- Dynamic Routing Between Capsules - Sabour, S., Frosst, N. and Hinton, G.E. (2017)
- Matrix capsules with EM routing - Hinton, G. E., Sabour, S. and Frosst, N. (2018)
- Optimizing Neural Networks that Generate Images - Tijmen Tieleman's disseration
- Transforming Auto-encoders - Hinton, G. E., Krizhevsky, A. and Wang, S. D. (2011)
- A parallel computation that assigns canonical object-based frames of reference. - Hinton, G.E. (1981)
- Shape representation in parallel systems - Hinton, G.E. (1981)

Glossary (by Sebastian Kwiatkowski)

<http://www.aisummary.com/blog/capsule-networks-glossary/>

Implementation

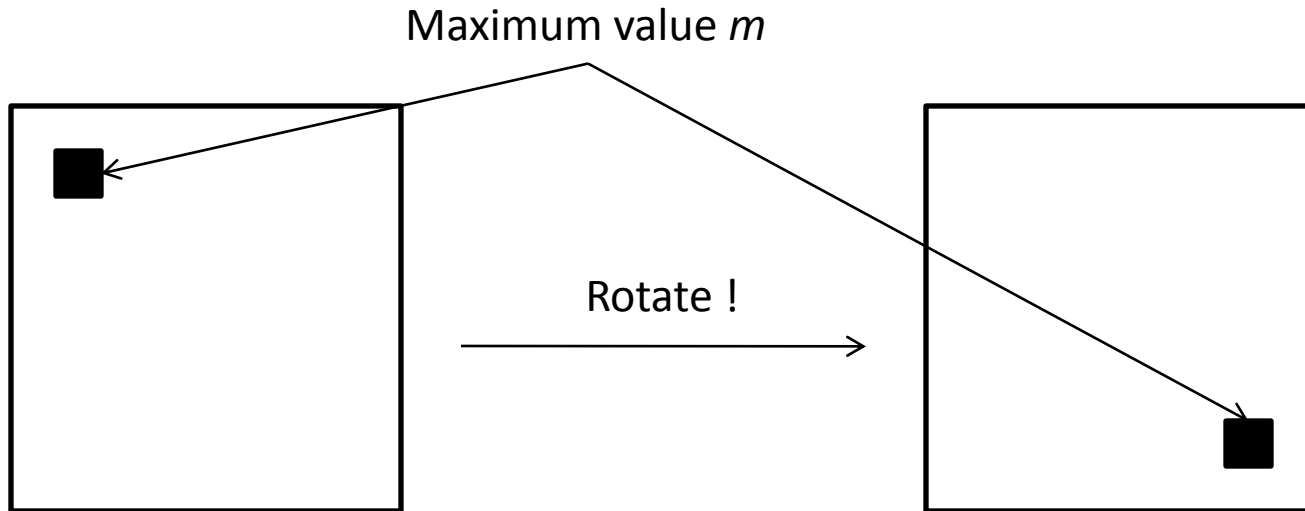
<https://github.com/Sarasra/models/tree/master/research/capsules>

For article “Dynamic Routing Between Capsules”

and more implementations links and info:

<https://github.com/sekwiatkowski/awesome-capsule-networks>

Equivariance and invariance

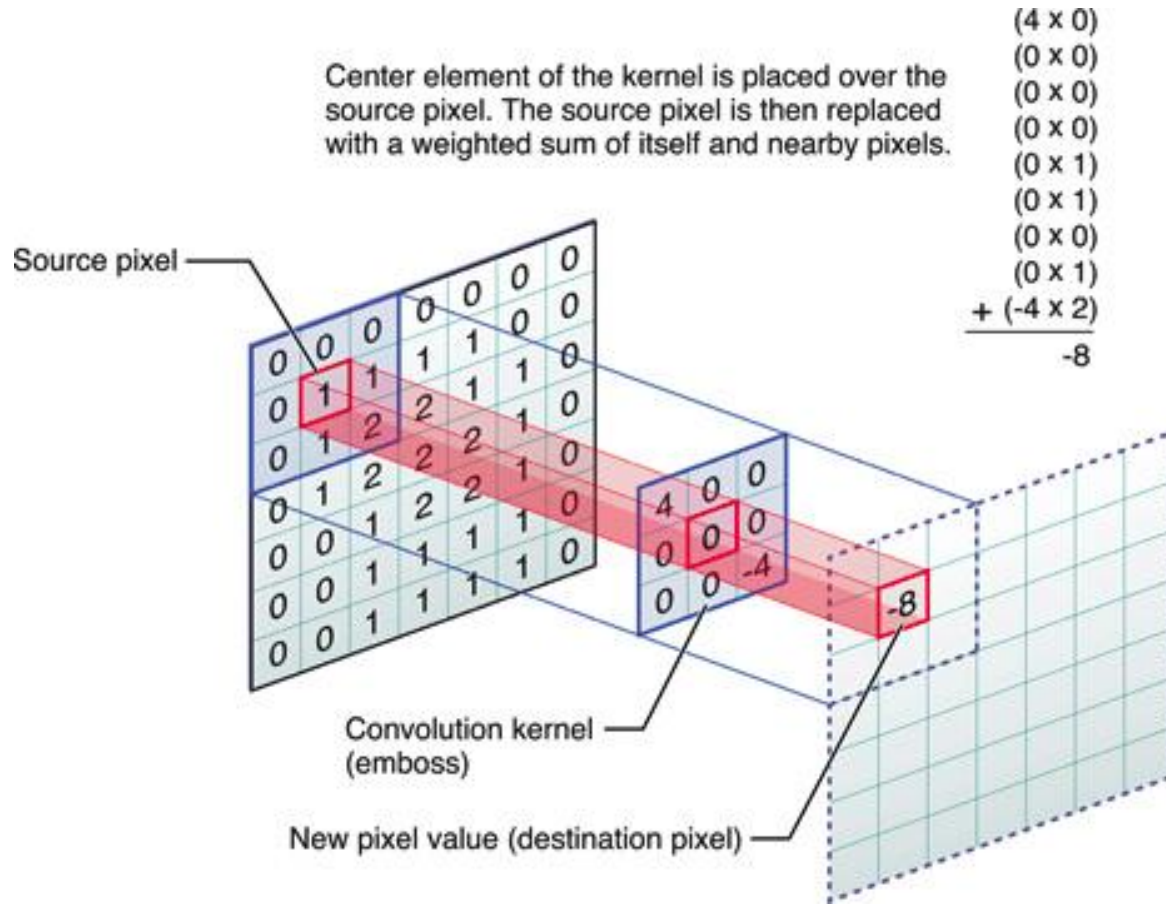


Maximum value m will be ***invariant*** to rotation, it is same.

Coordinates of the point with maximum value m , will vary "equally" with the distortion.
Coordinates of maximum will be ***equivariant***

Convolutional network

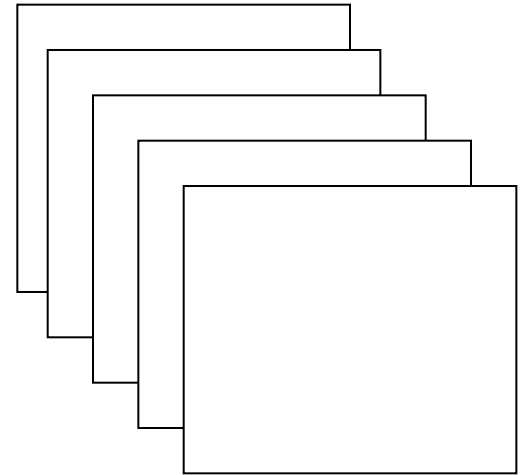
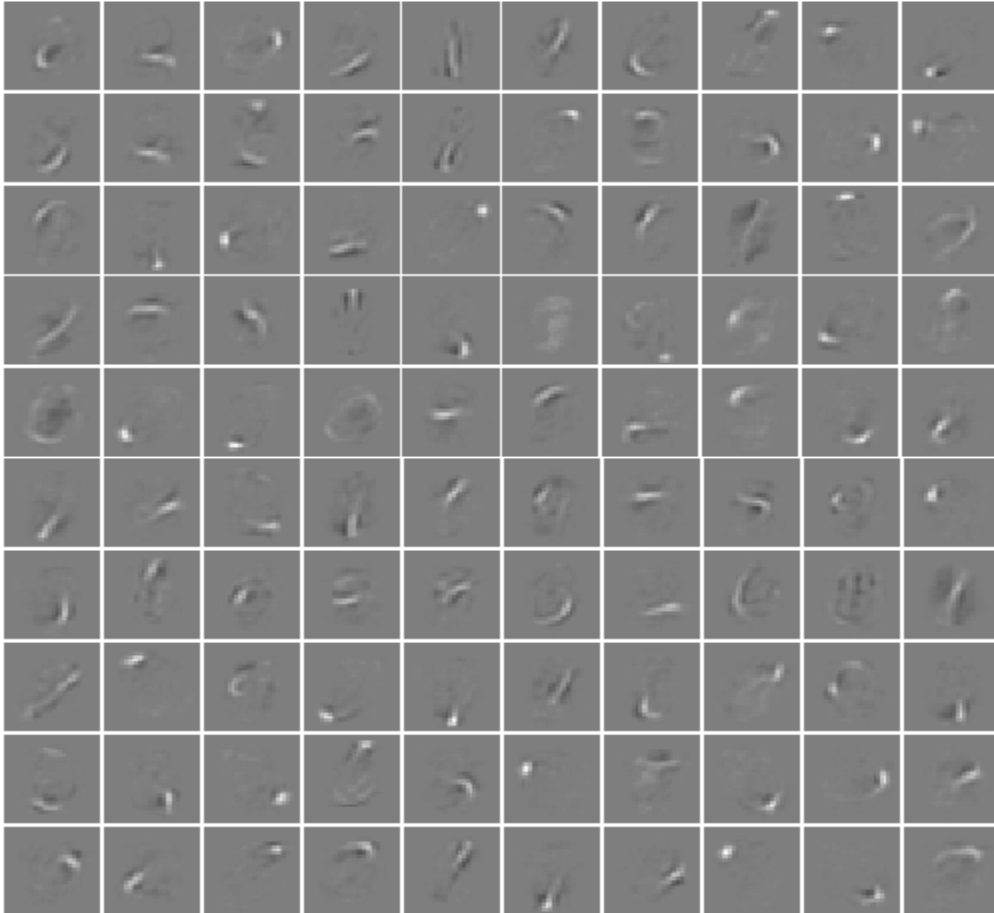
Convolution



<https://hcordeirodotcom.files.wordpress.com/2012/02/atividade-2-kernel-convolution.jpg>

Convolutional network

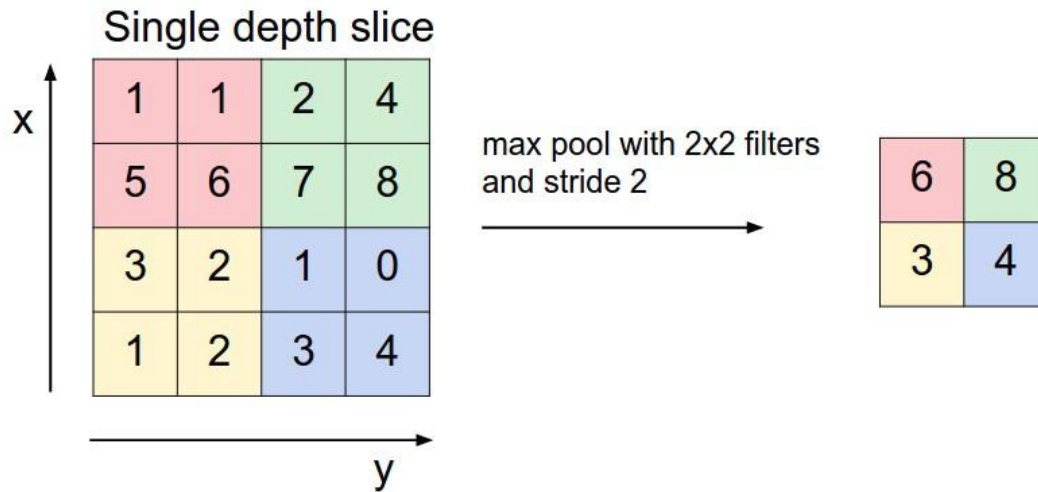
Feature map



<http://www.cs.toronto.edu/~ranzato/research/projects.html>

Convolutional network

Pooling layer



<http://cs231n.github.io/convolutional-networks/>

What wrong with convolutional networks?

Geoffrey Hinton talk "What is wrong with convolutional neural nets ?"

<https://www.youtube.com/watch?v=rTawFwUvnLE>

“Feature extraction levels are interleaved with subsampling layers that pool the outputs of nearby feature detectors of the same type”

Slide:

- It is a bad fit to the psychology of shape perception: It does not explain why we assign intrinsic coordinate frames to objects and why they have such huge effects
- It solves the wrong problem: We want equivariance, not invariance. Disentangling rather than discarding.
- It fails to use the underlying linear structure: It does not make use of the natural linear manifold that perfectly handles the largest source of variance in images
- Pooling is a poor way to do dynamic routing: We need to route each part of the input to the neurons that know how to deal with it. Finding the best routing is equivalent to parsing the image

How it can be improved ?

Capsules + Dynamic Routing

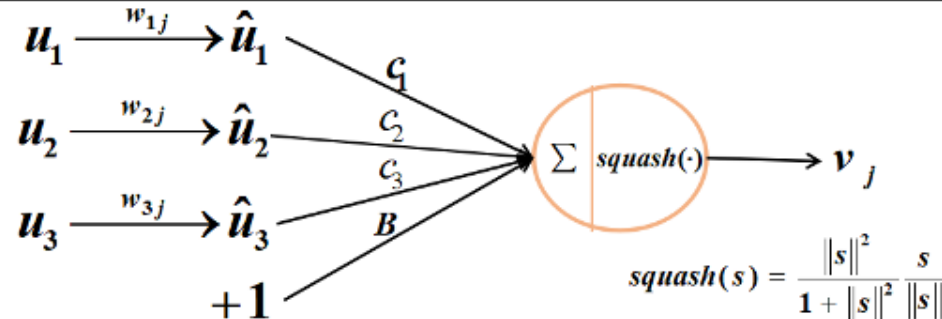
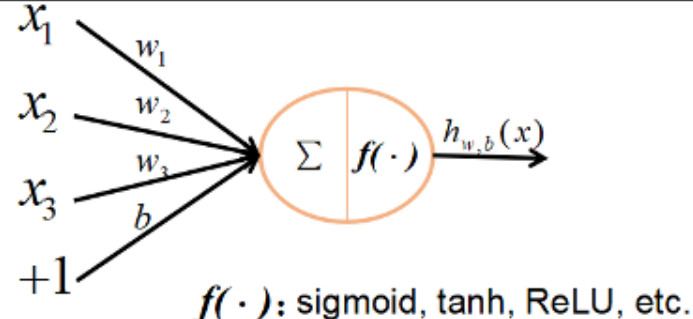
What we want to store:

Is this feature detected? (Probability)

Attributes of feature: pose (position, size, orientation), deformation, velocity, albedo, hue, texture, etc.

From **scalar values** to **vectors**!

Capsule

| | | capsule | VS. | traditional neuron |
|--|------------------------------|--|--|-------------------------------------|
| Input from low-level neurons/capsules | | vector(u_i) | | scalar(x_i) |
| Operations | Linear/Affine Transformation | $\hat{u}_{ji} = W_{ij} u_i + B_j$ (Eq. 2) | | $a_{ji} = w_{ij} x_i + b_j$ |
| | Weighting | $s_j = \sum_i c_{ij} \hat{u}_{ji}$ (Eq. 2) | | $z_j = \sum_{i=1}^3 1 \cdot a_{ji}$ |
| | Summation | | | |
| | Non-linearity activation | $v_j = \text{squash}(s_j)$ (Eq. 1) | | $h_{w,b}(x) = f(z_j)$ |
| output | | vector(v_j) | | scalar(h) |
|  <p style="text-align: center;"> $\text{squash}(s) = \frac{\ s\ ^2}{1 + \ s\ ^2} \frac{s}{\ s\ }$ </p> | | |  <p style="text-align: center;"> $f(\cdot)$: sigmoid, tanh, ReLU, etc. </p> | |

Capsule = New Version Neuron!
vector in, vector out VS. scalar in, scalar out

Squash

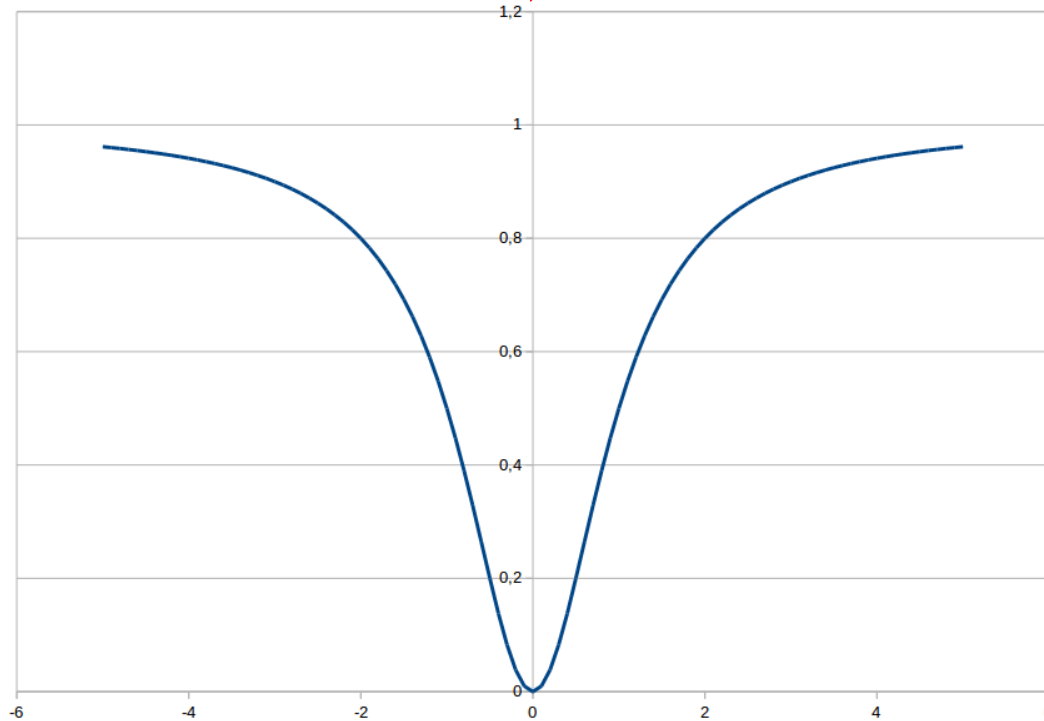
Squash function:

$$\mathbf{v}_j = \frac{\|\mathbf{s}_j\|^2}{1 + \|\mathbf{s}_j\|^2} \frac{\mathbf{s}_j}{\|\mathbf{s}_j\|}$$

Squash

Squash function:

$$\mathbf{v}_j = \frac{\|\mathbf{s}_j\|^2}{1 + \|\mathbf{s}_j\|^2} \frac{\mathbf{s}_j}{\|\mathbf{s}_j\|}$$



Dynamic Routing Between Capsules - Sabour, S., Frosst, N. and Hinton, G.E. (2017)

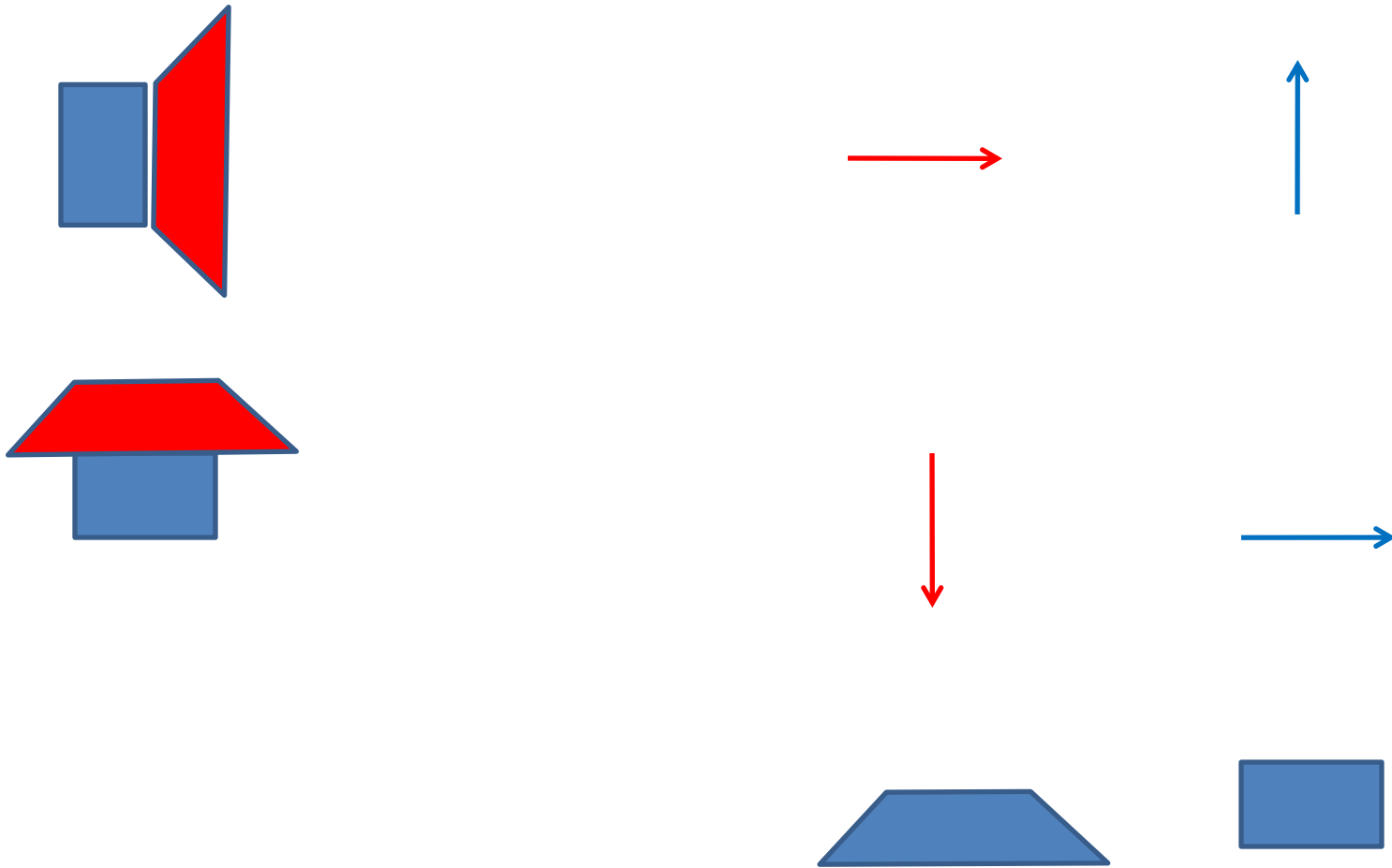
Routing

Procedure 1 Routing algorithm.

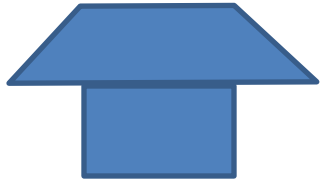
```
1: procedure ROUTING( $\hat{\mathbf{u}}_{j|i}, r, l$ )
2:   for all capsule  $i$  in layer  $l$  and capsule  $j$  in layer  $(l + 1)$ :  $b_{ij} \leftarrow 0$ .
3:   for  $r$  iterations do
4:     for all capsule  $i$  in layer  $l$ :  $\mathbf{c}_i \leftarrow \text{softmax}(\mathbf{b}_i)$  ▷ softmax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

Dynamic Routing Between Capsules - Sabour, S., Frosst, N. and Hinton, G.E. (2017)

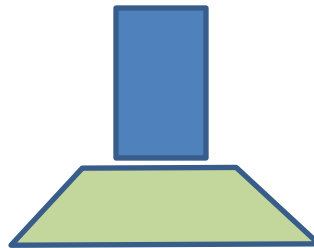
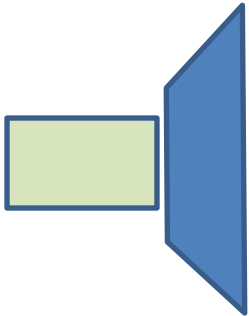
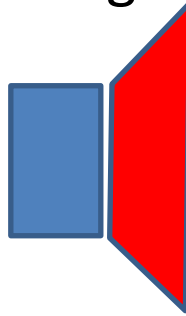
Routing



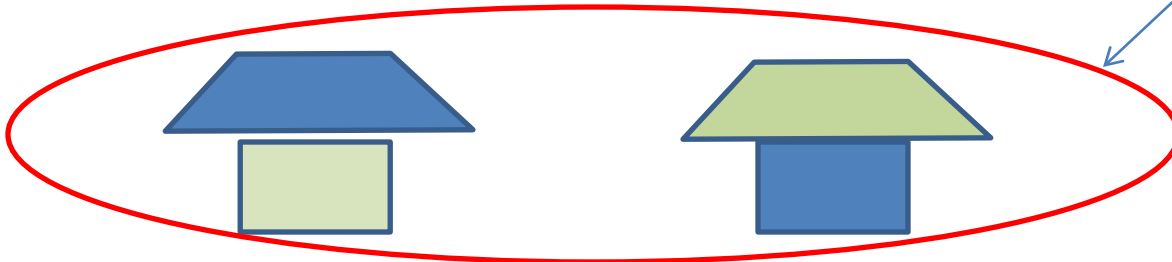
Routing



?

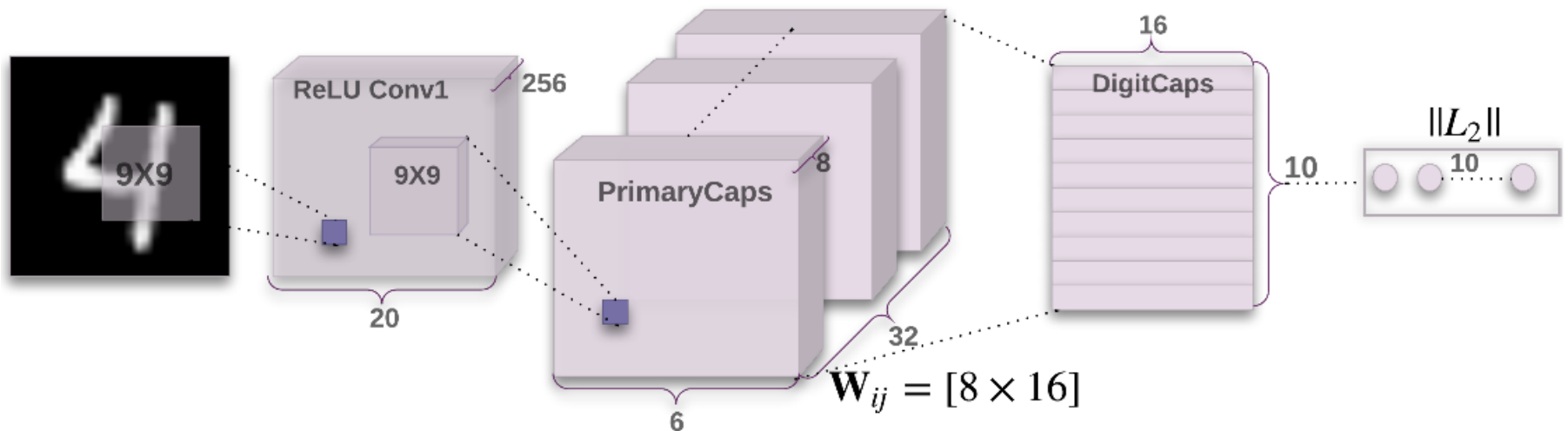


Agree from both capsules



How it can be solved ?

Architecture of simple Capsule Net for MNIST:



$$L_k = T_k \max(0, m^+ - \|\mathbf{v}_k\|)^2 + \lambda (1 - T_k) \max(0, \|\mathbf{v}_k\| - m^-)^2$$

Dynamic Routing Between Capsules - Sabour, S., Frosst, N. and Hinton, G.E. (2017)

Examples

Dynamic Routing Between Capsules - Sabour, S., Frosst, N. and Hinton, G.E. (2017)
(MNIST)

Fashion MNIST

<https://github.com/XifengGuo/CapsNet-Fashion-MNIST>

Capsule Deep Neural Network for Recognition of Historical Graffiti Handwriting
N Gordienko, Y Kochura, V Taran, G Peng

Food detection, face detection and lot more:

Advantages and disadvantages

- + Good preliminary results (MNIST).
 - + Requires less training data.
 - + Works good with overlapping objects.
 - + Can detect partially visible objects.
 - + Results are interpretable, components hierarchy can be mapped.
 - + Equivariance
-
- No known yet accuracy on large data sources CIFAR10? Accuracy seems low.
 - Really **slow** training time (so far).
 - Non linear squashing reflect the probability nature not so good as we want.
 - Cosine of an angle for measuring the agreement?

Questions, Discussion...