Capsule Networks

Kai Lichtenberg

sentin

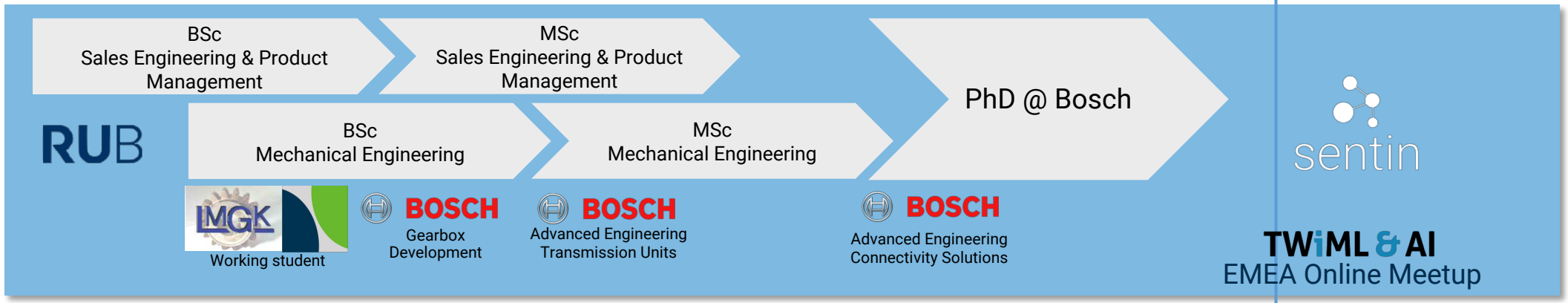
leverage your knowledge with data

19.12.18

Kai Lichtenberg, 33

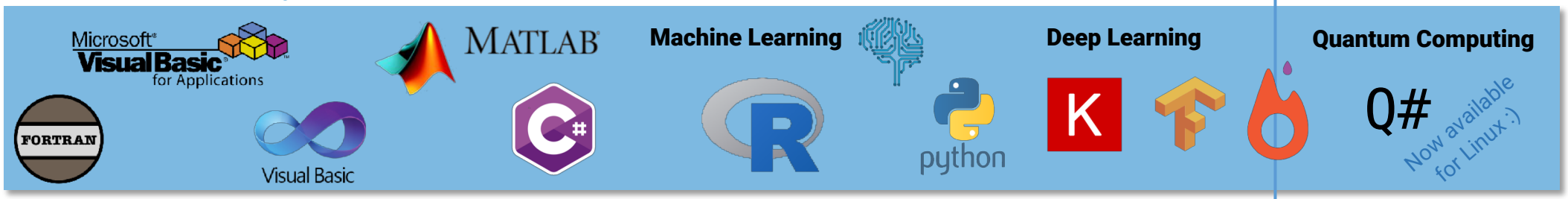
TODAY

Classic Stuff



Sales Engineer | Mechanical Engineer | Data Scientist | **Founder**, Data Scientist, Accountant, Office Manager, Janitor, …

Language Journey



Kai Lichtenberg @kai_lichtenberg

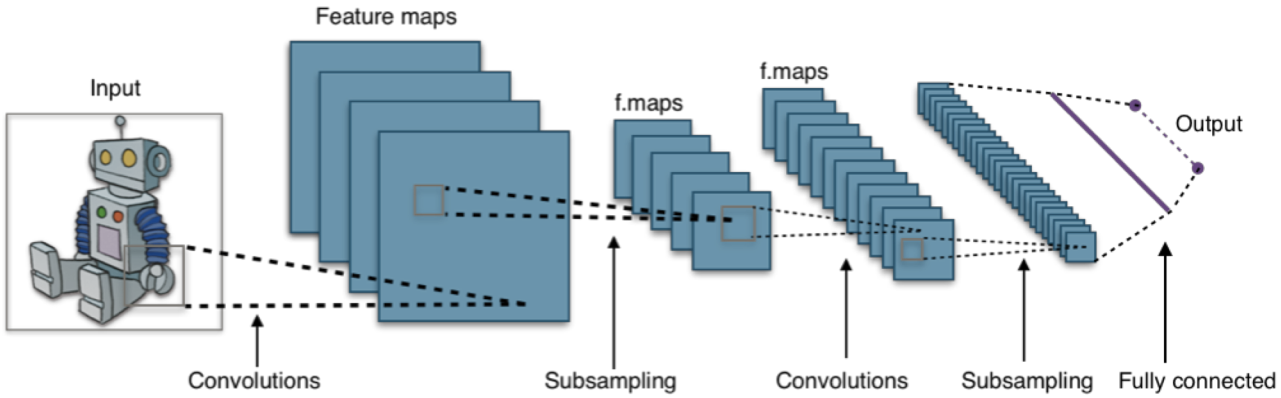
Agenda

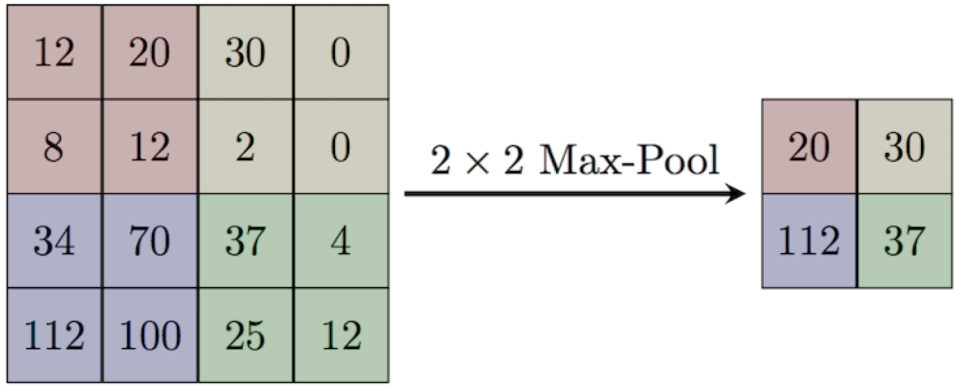
- What's wrong with Convolutional Neural Networks
- What is a capsule?
- CapsNet and Dynamic Routing (& Visualization)

Prerequisites: Some knowledge of how Convolutional Neural Networks are working

What is wrong with ConvNets?

Subsampling (max pooling) is an important concept in CNN's



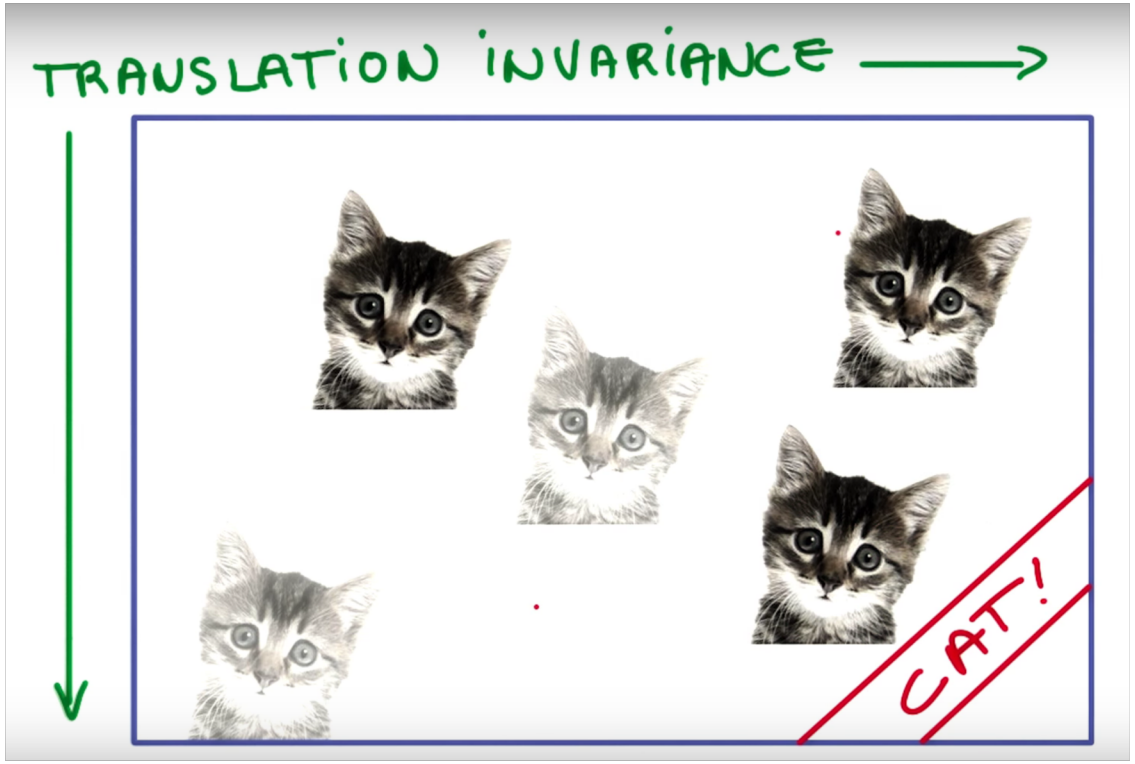


[Source:Wikipedia]

- Periodically inserted between successive convolutional layers
- Get translational invariance by throwing away spatial information
- Reduce the number of parameters used in the model
- There are various forms of pooling, but max pooling is the most popular

What is wrong with ConvNets?

Translational and other forms of invariance

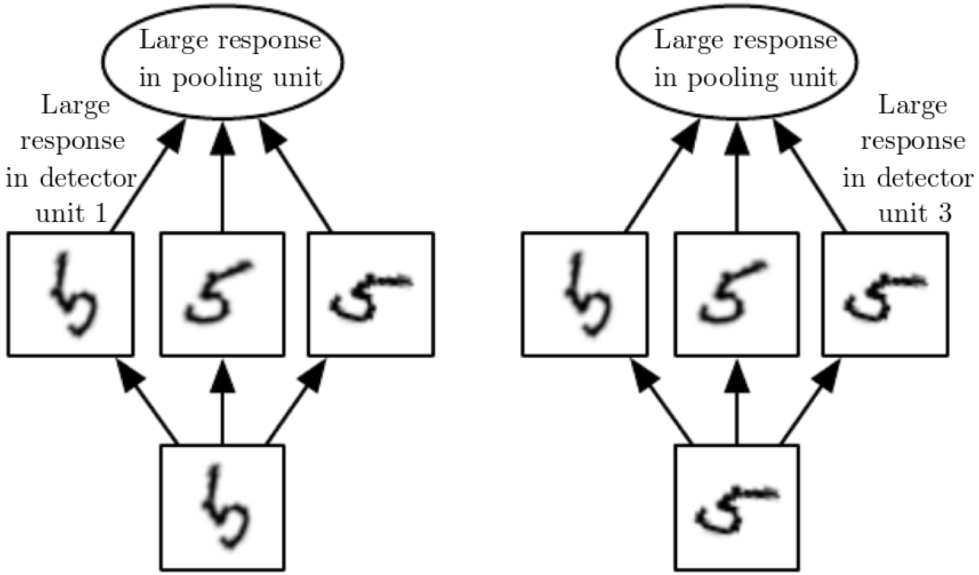


- It does not matter that much where the cat is located on the image
- At the end we know that there is a cat, but not where it was located on the original image
- No invariance for rotation, scale, viewpoint...

[Source:https://www.cc.gatech.edu/~san37/post/dlhc-cnn/]

What is wrong with ConvNets?

But wait! Why does it work anyway?



[Source: Deep Learning Textbook, Goodfellow et al.]

A crude way of routing information through the network

- Training on big data sets and augmentation
- Learning several filters for different versions of the digit
- Pool out the higher activation
- The result is the same regardless of which kind of 5 the network sees
- A “brute-force” method to create invariance for all kinds of transformations

What is wrong with ConvNets?

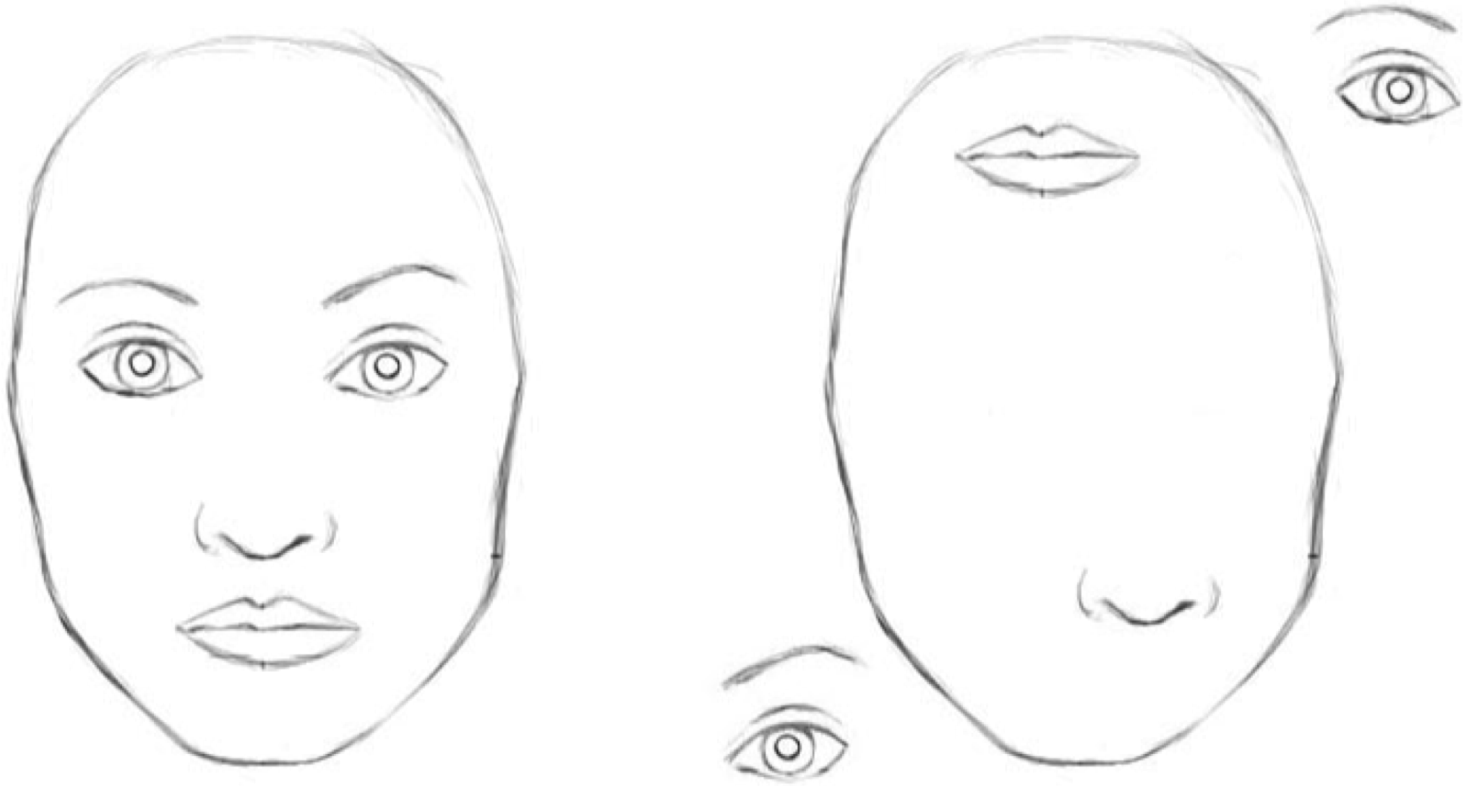
Awesome! So what's the problem? – DATA!!

~13 Mio. | 1 K Classes + augmentation

[Source: http://karpathy.github.io/2014/09/02/what-i-learned-from-competing-against-a-convnet-on-imagenet/]

What is wrong with ConvNets?

Awesome! So what's the problem? Is it a face or not?



[Source: Max Pechyonkin – https://medium.com/ai%C2%B3-theory-practice-business/understanding-hintons-capsule-networks-part-i-intuition-b4b559d1159b]

What is wrong with ConvNets?

Max Pooling and the consequences

- Pooling was introduced in the 80's in a predecessor to CNN's (neo-cognitron) as a solution for invariance in handwritten digit recognition
- It works perfectly fine for this use case! Digits are 2D objects and pooling + multiple filters solves the problem!
- **The fact that it also works for complex 3D scenes and at imagenet scale was totally unexpected and Hinton describes that as complete disaster**
- He says that CNN's are eventually doomed because of this properties and research is not focusing on finding a new way

The main ideas behind capsules

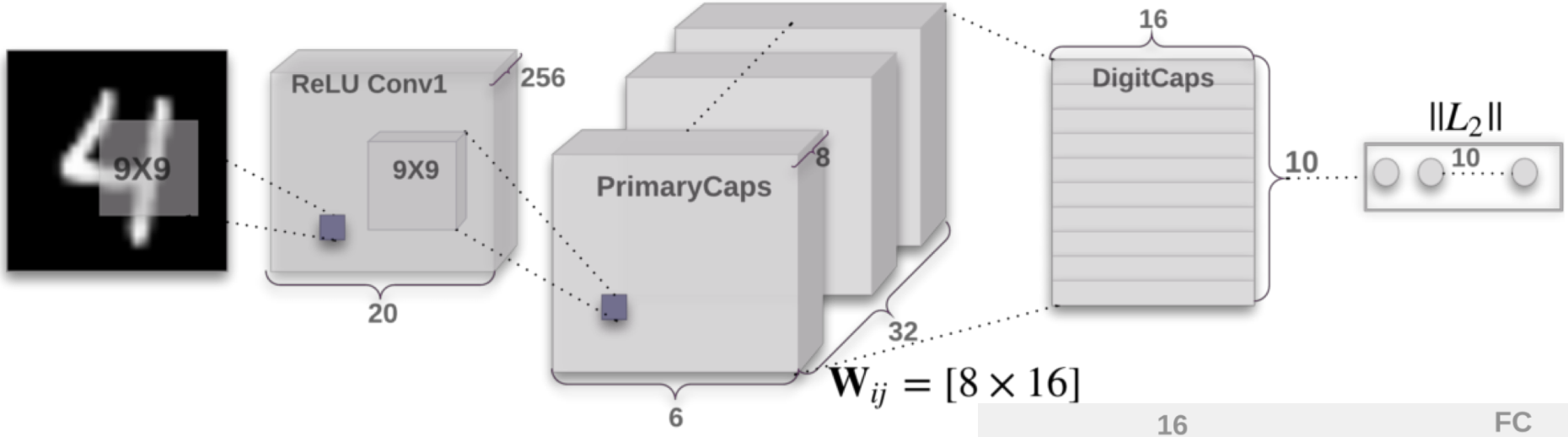
It's all about higher dimensional representation

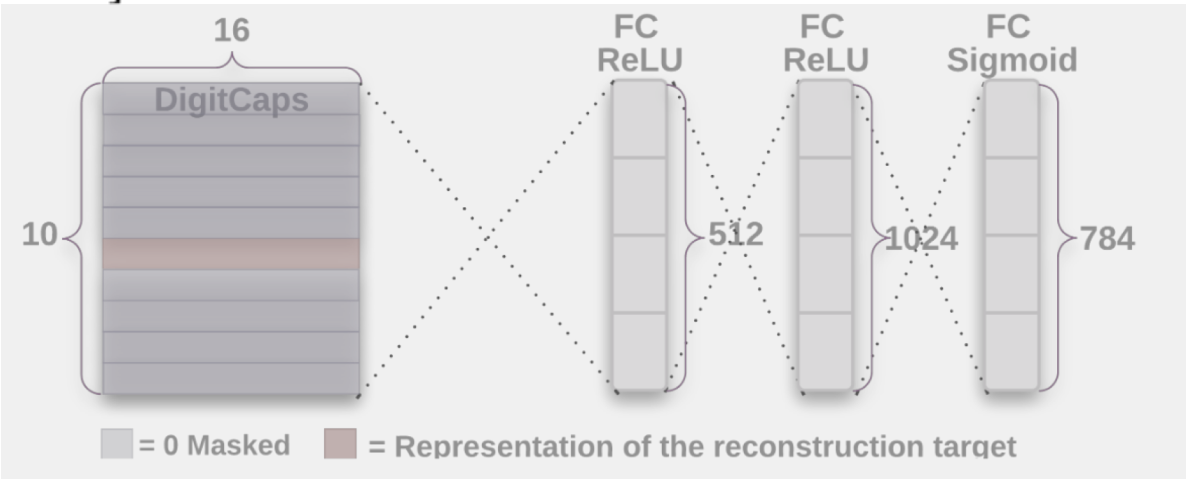
		Capsule	Neuron
	Input	vector : u_i	scalar : x_i
Opera-tions	Linear Transformation	$\hat{u}_{j\|i} = W_{ij}u_i + B_j$	$a_{j\|i} = w_{ij}x_i + b_j$
	(Weighting) & Summation	$s_j = \sum_i c_{ij}\hat{u}_{j\|i}$	$z_j = \sum_i a_{j\|i}$
	Non-Linearity	$v_j = \frac{\|s_j\|^2 s_j}{1 + \|s_j\|^2 \|s_j\|}$	$h_{w,b}(x) = f(z_j)$
	Output	vector : v_j	Scalar: h

- Instead of scalar values the outputs are vectors
- This idea is pretty old
- A proper architecture and a good way of information routing was not found for a long time
- Until last year: CapsNet and Dynamic Routing Between Capsules by Hinton et al.

The CapsNet Architecture for MNIST

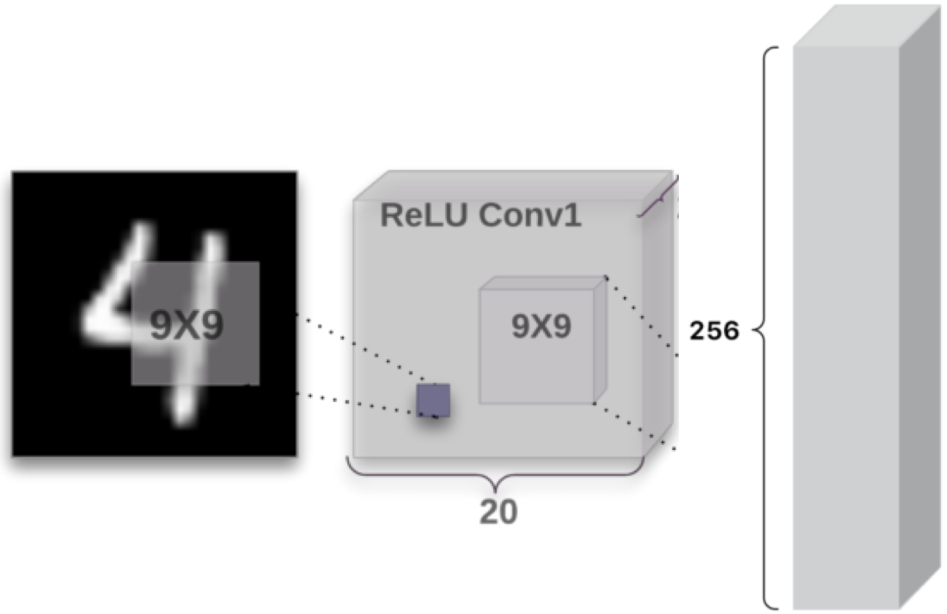
It's all about higher dimensional representation





The CapsNet Architecture for MNIST

We are starting with a convolutional layer!



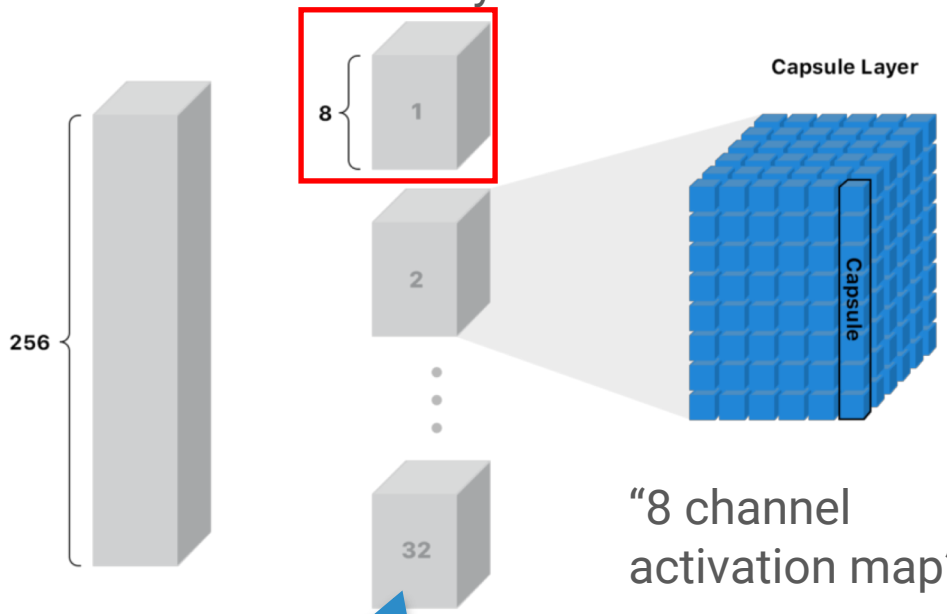
- A classic convolutional layer:
  - 256 filters (9 by 9 convolution) with stride 1
  - → 256 activation maps (20 by 20)
- That leverages the replication of learned filters through space

Demo!

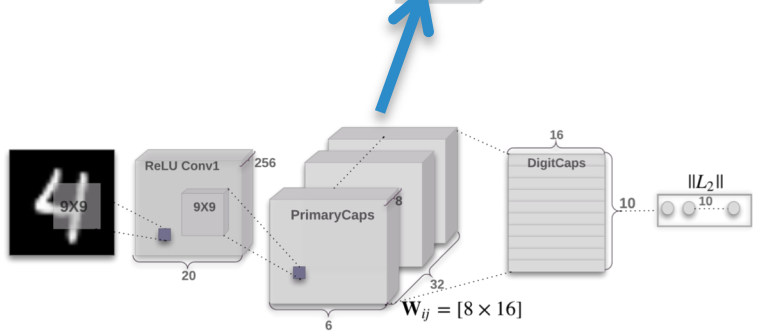
The CapsNet Architecture for MNIST

The PrimaryCaps Layer: A convolutional capsule layer

8 dimensional 9 by 9 conv filter



“8 channel activation map”

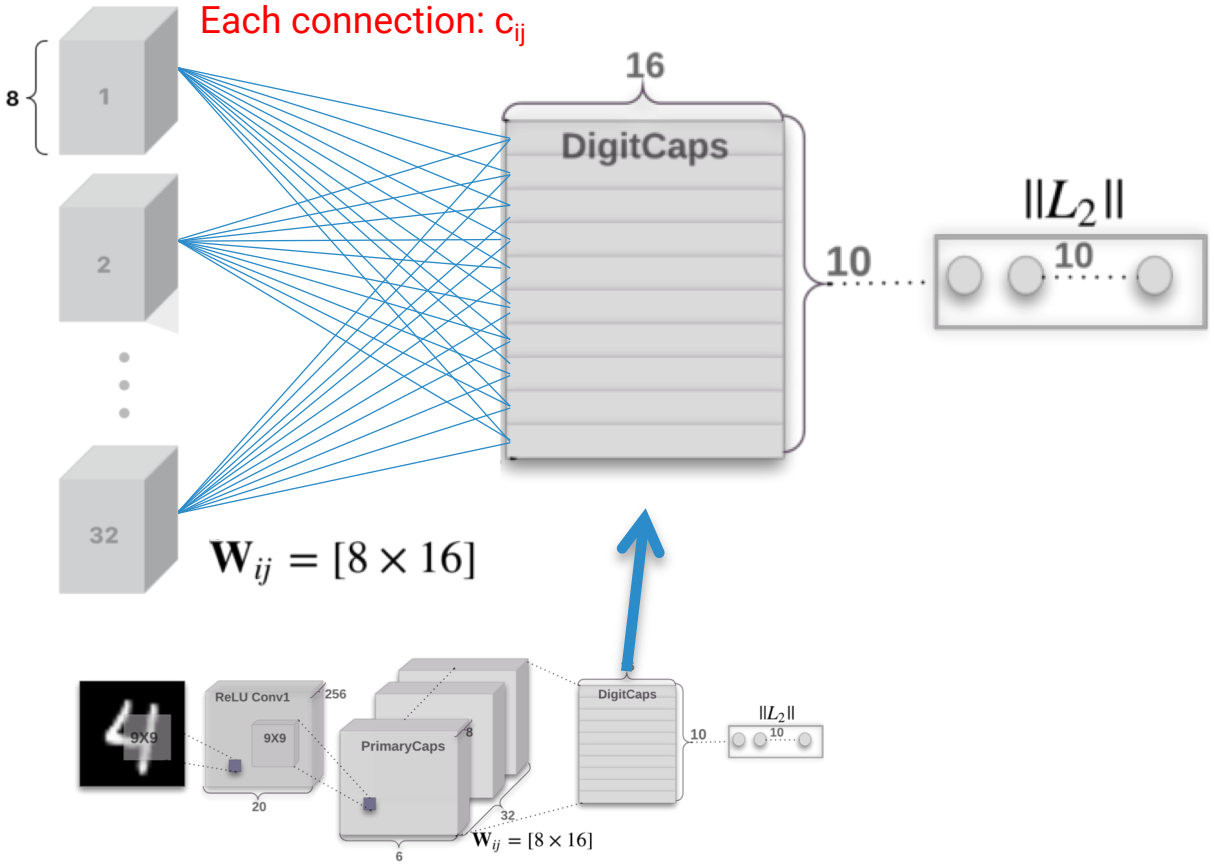


- Similar to a classic convolutional layer, only one dimension higher:
  - 32 8D filters (9 by 9 convolution) with stride 2
  - → 32 8D activation maps with 36 capsules (vectors) each
- Very similar to a “sliced” classic convolutional layer

[Source: Hinton et al. – Dynamic Routing between Capules and Nick Bourdakos - https://medium.freecodecamp.org/understanding-capsule-networks-ais-alluring-new-architecture-bdb228173ddc]

The CapsNet Architecture for MNIST

The DigitCaps Layer: No convolution, capsules only



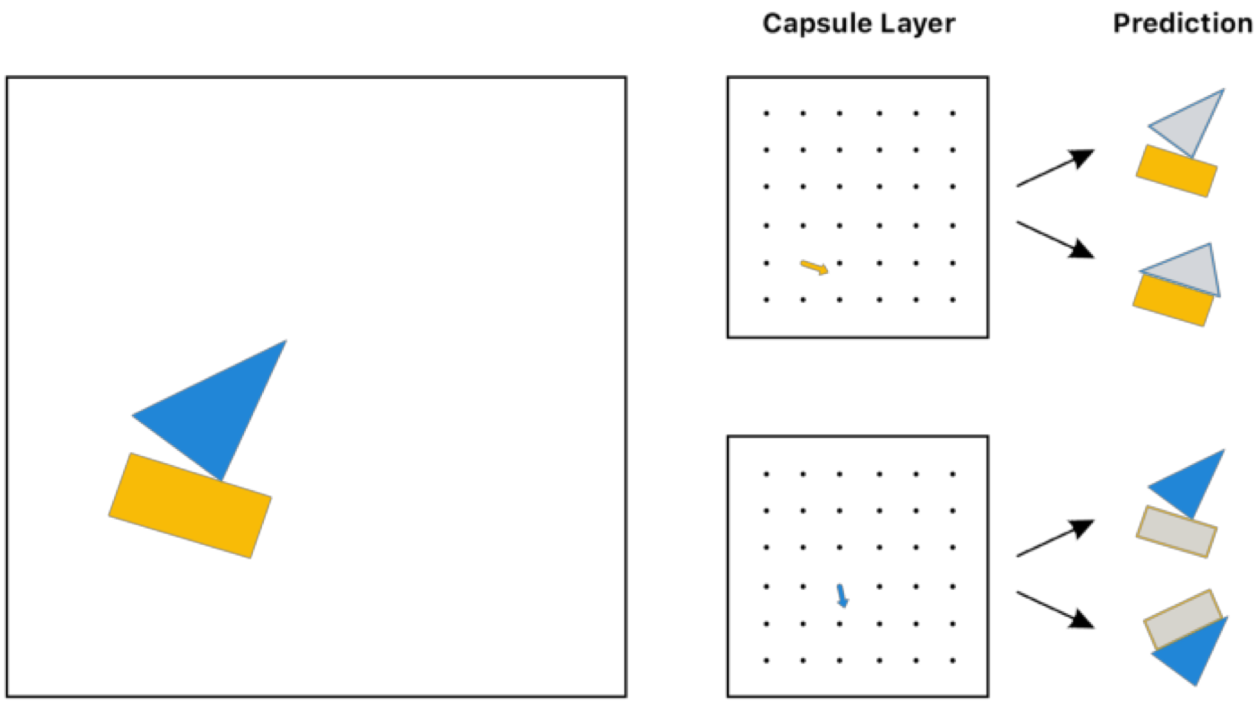
[Source: Hinton et al. – Dynamic Routing between Capules and Nick Bourdakos - https://medium.freecodecamp.org/understanding-capsule-networks-ais-alluring-new-architecture-bdb228173ddc]

- Fully connected!
- No scalar weight, but an 8x16 weight matrix for each connection [8]*[8x16]=[16]
- Each capsule with each DigitCaps capsule. 11,520 calculations (1152 for each Digit)!!
- We need some routing! Each connection has an additional scalar routing weight c_{ij}

Demo!

Dynamic Routing By Agreement

The DigitCaps Layer: No convolution, capsules only

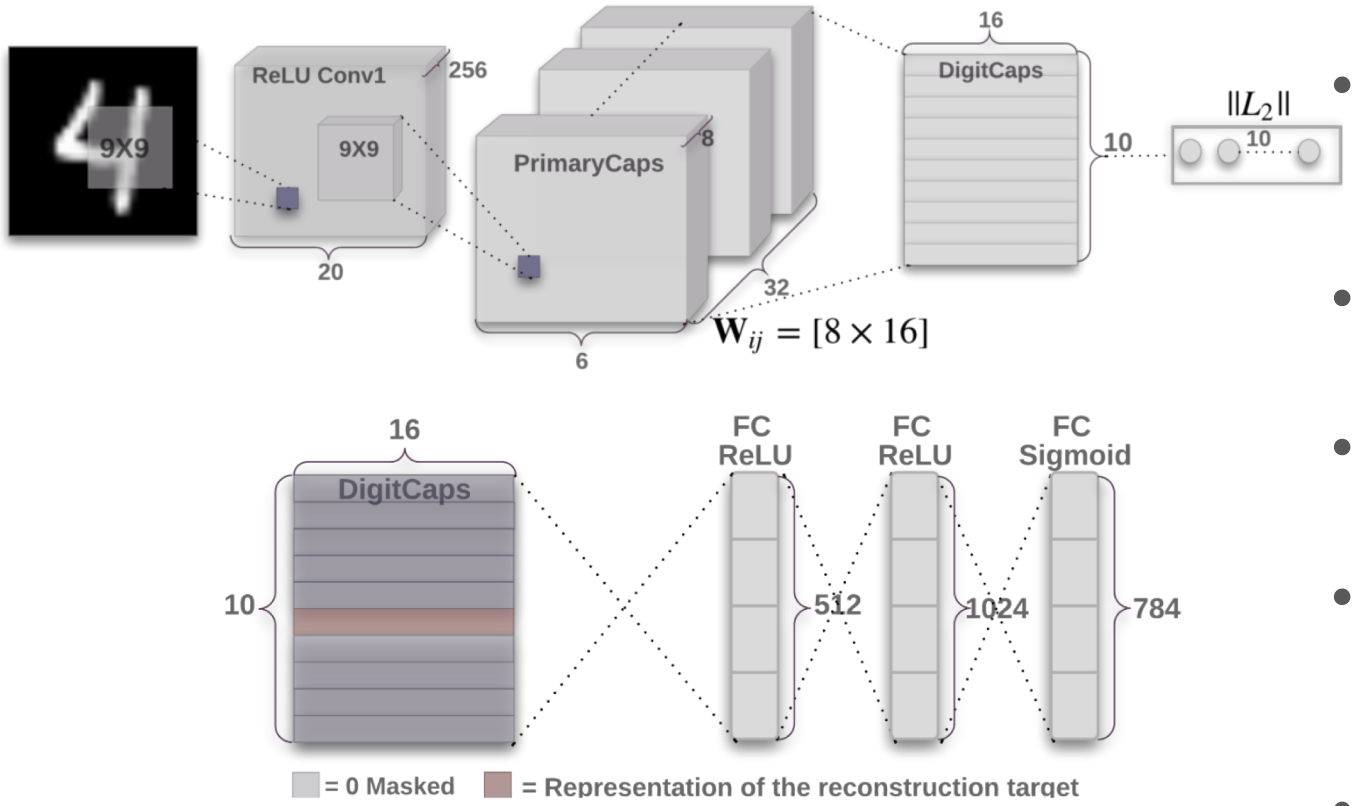


[Source: Aurelien Geron – https://www.youtube.com/watch?v=pPN8d0E3900]

- A ship and a house which can be composed by a rectangle and a triangle (btw: a CNN would fail!)
- One capsule learns to detect triangles the other learns to detect rectangles
- Both predict for the class house and the class boat
- They agree on the orientation for the boat and update the weight c_{ij} accordingly
- This process is done in a loop (~7 times) for each picture!
- During initial training with backprop priors are calculated for c_{ij}

How to train

Backprop – What else?

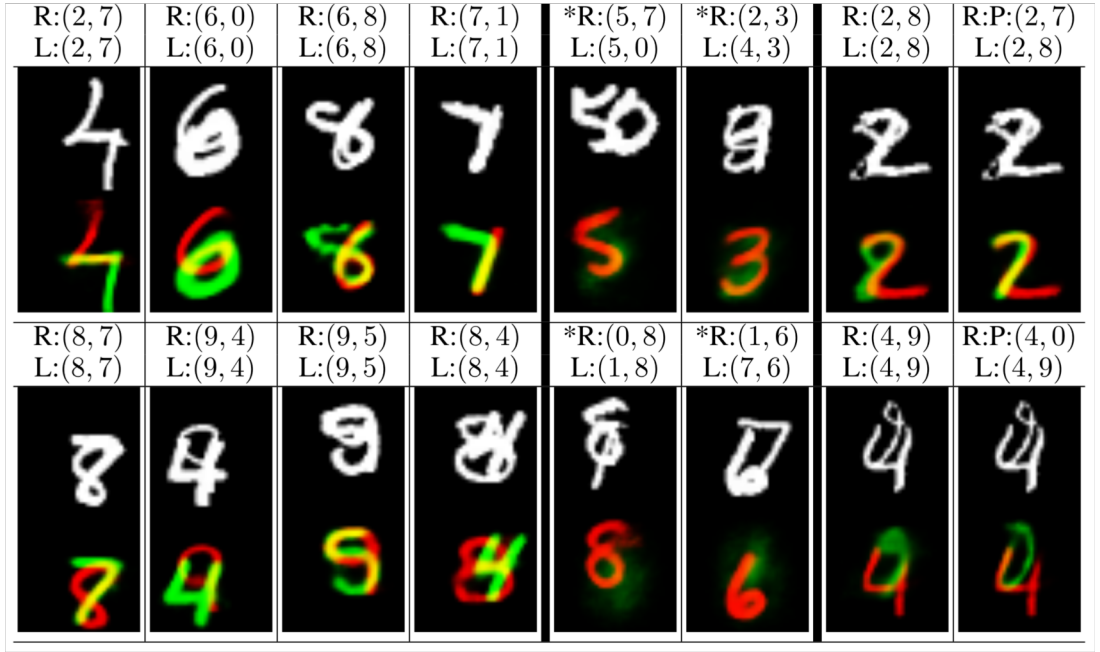


- Training is done with backpropagation and the priors for the weights c_{ij} are trained during this process
- Due to squashing the length of each DigitCaps vector is between 0 and 1 an corresponds to the probability
- Classic cross entropy or margin loss for multiple instance classification
- Additional: Reconstructor network that tries to reconstruct the original image
- This helps regularizing and make sure the capsules learn the instantiation parameters that are necessary to reconstruct the image
- Final loss: margin loss + 0.0005 * reconstruction loss

[Source: Hinton et al. – Dynamic Routing between Capules]

Wrap Up

- Requires less data and reaches high accuracy on MNIST
- Works very well for overlapping objects
- Very promising for image segmentation
- Fewer parameters than convolutional networks (though more computations needed)
- Works very well for overlapping objects
- Robust to changes in viewpoint, lighting, etc.
- The result is interpretable (See the visualization tool)
- Harder to attack



Capsule Networks are in a very early development stage and there's a lot of stuff that is not tested right now. It's maybe comparable to the beginning of convolutional networks. Who knows?