Capsule Networks

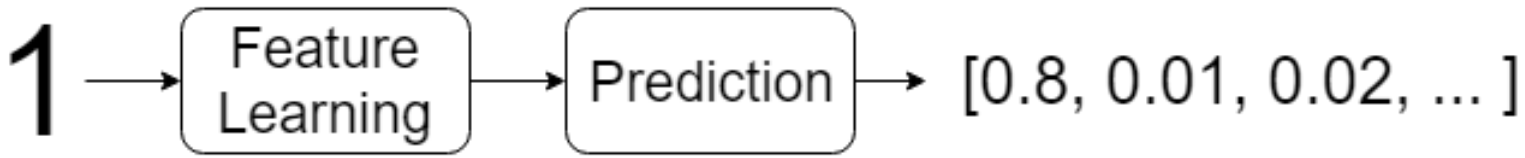
Suthee Chaidaroon

Santa Clara University
California, USA

April 11, 2018

Introduction

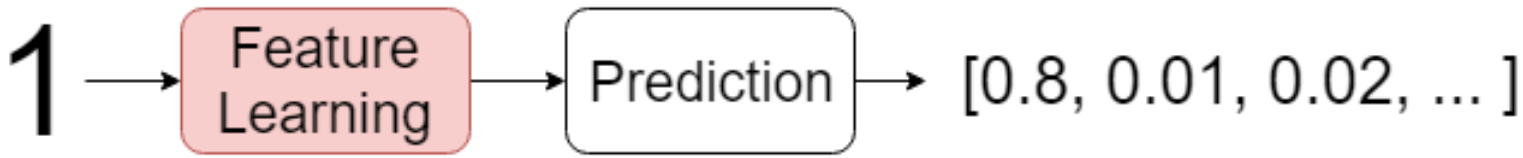
Digit Classification Task



Santa Clara University

Introduction

Digit Classification Task



Santa Clara University

Feature Learning

- Use Feed-forward neural networks.

Feature Learning

- Use Feed-forward neural networks.
- Use Convolutional neural networks.

Feature Learning

- Use Feed-forward neural networks.
- Use Convolutional neural networks.
- Use Capsule networks.

Convolutional Neural Network

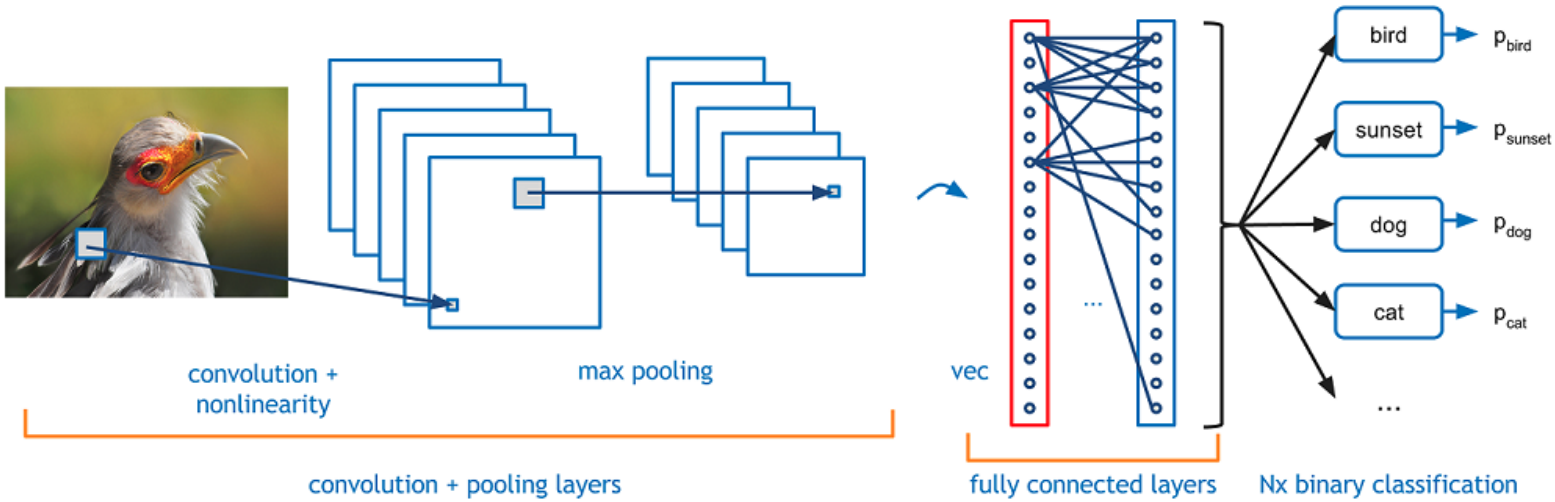
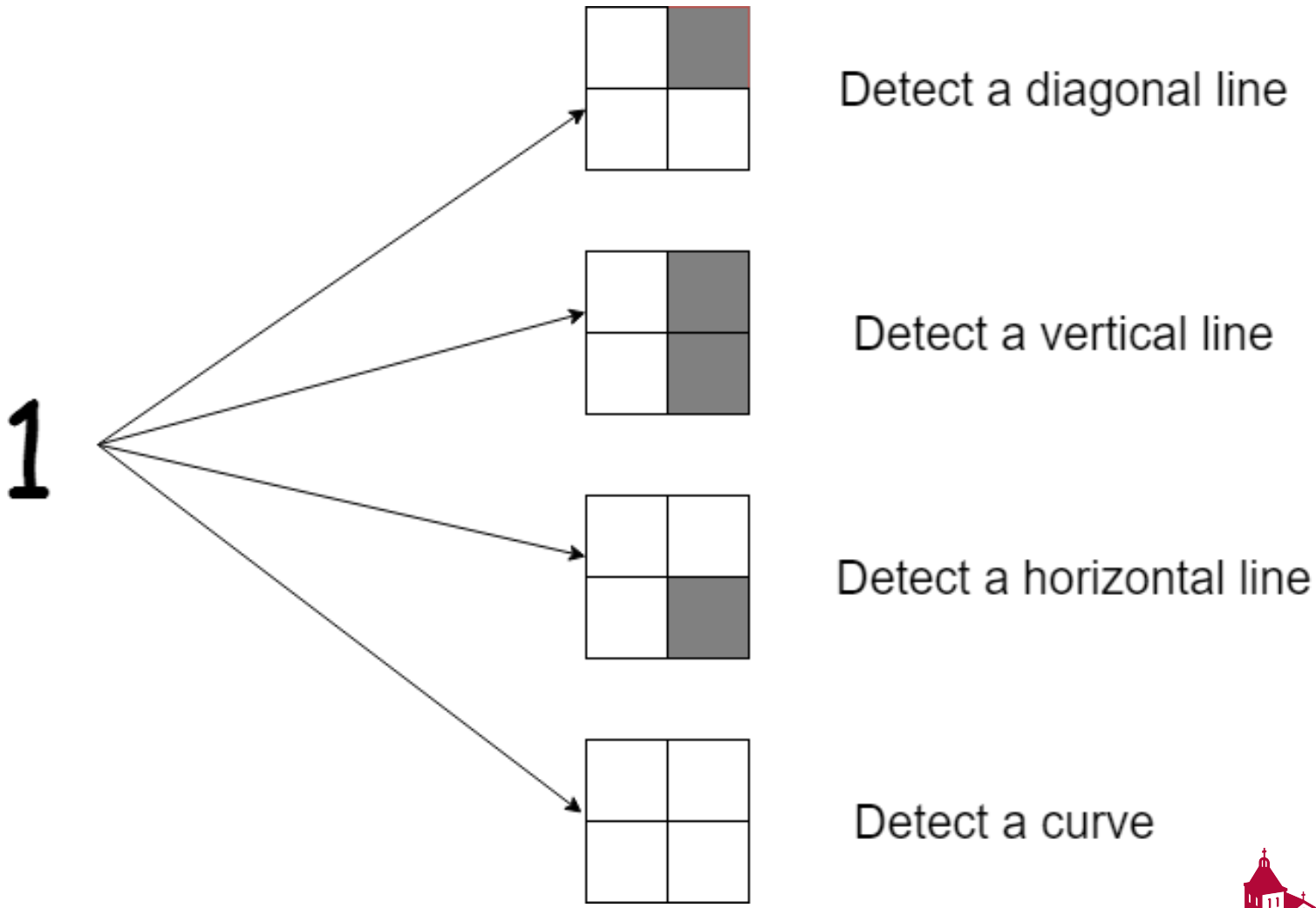


Figure: An example of a convolutional neural network[1]

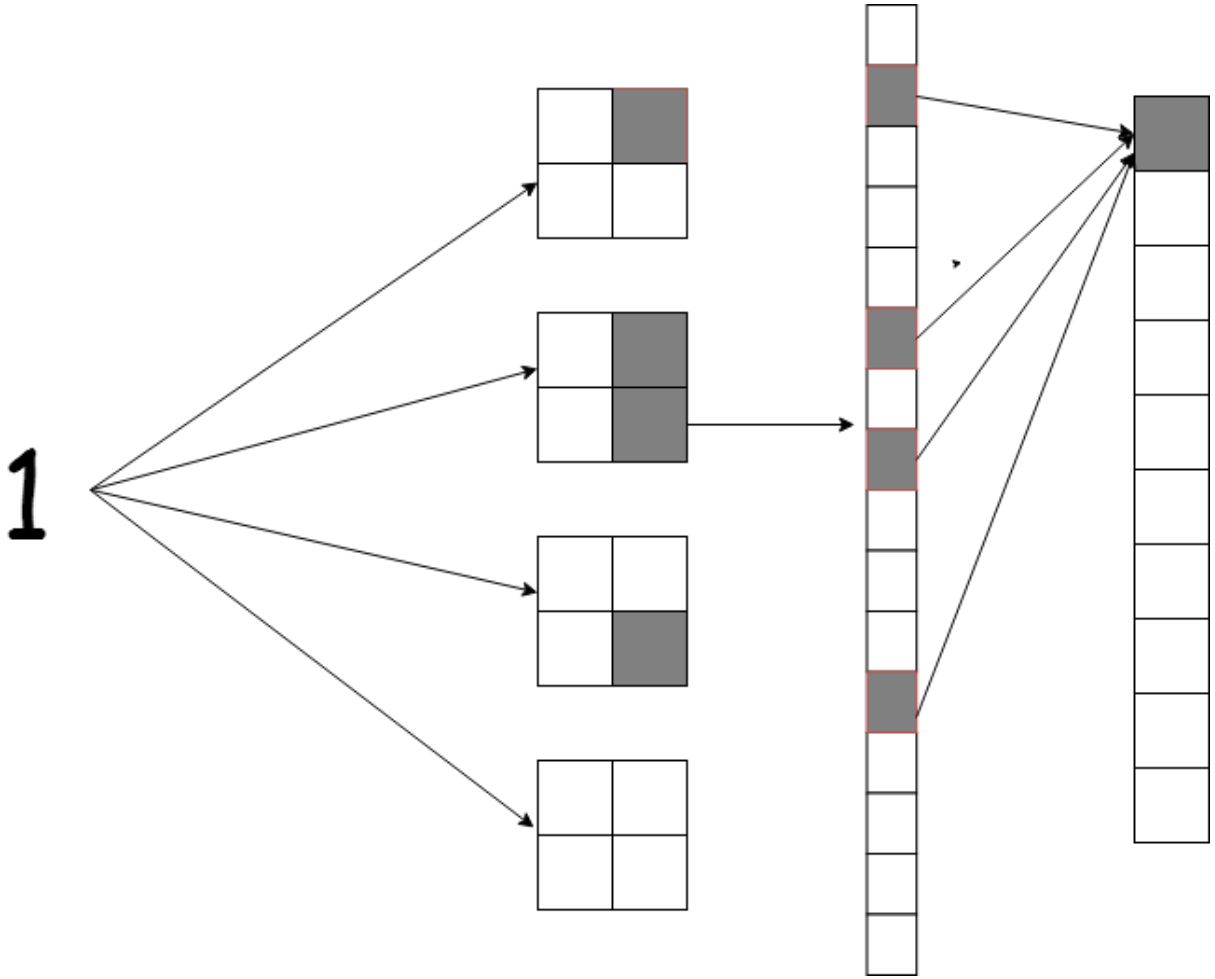
[1] Image credit: `https://qtmlresearch.com`

Convolutional Layer



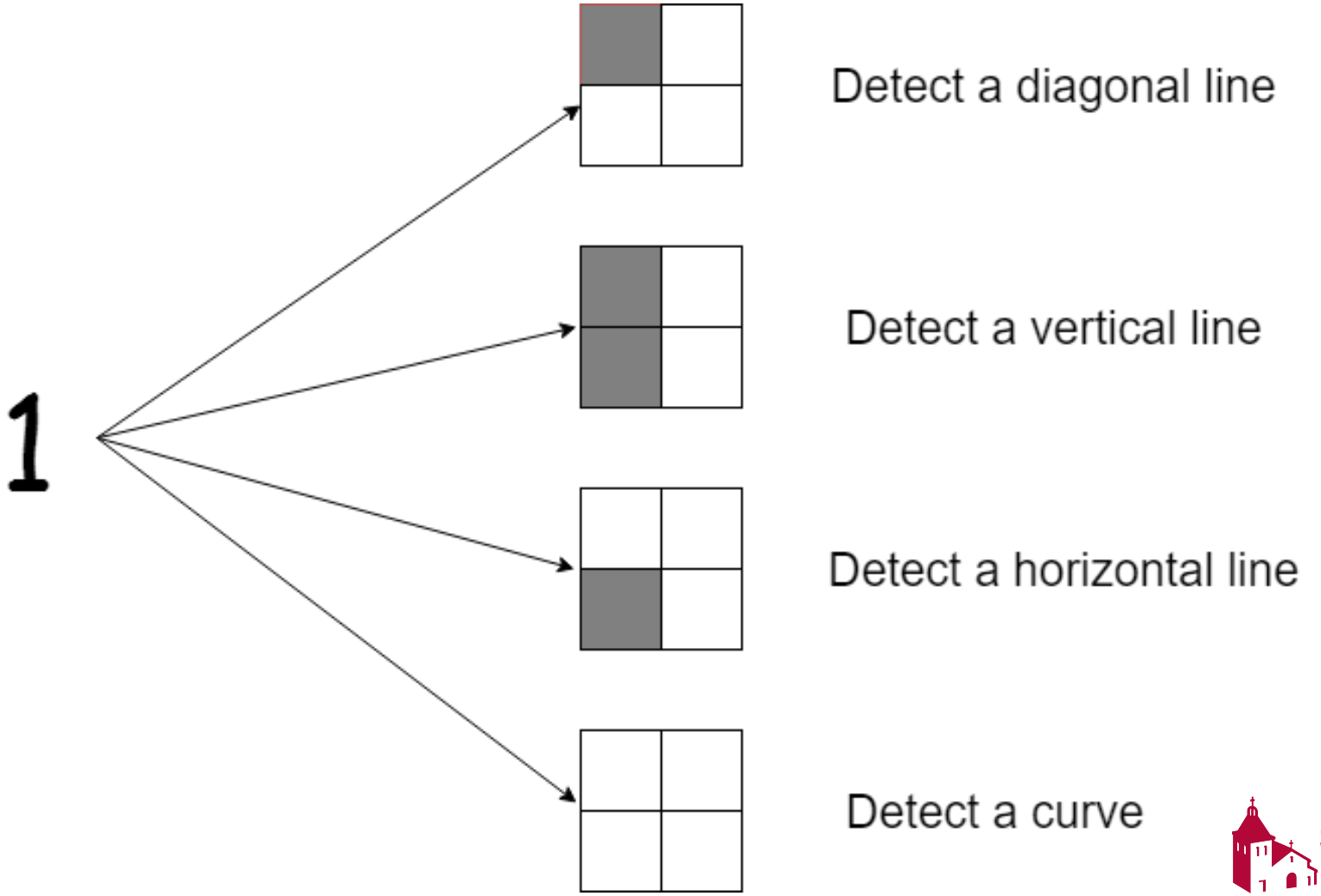
Santa Clara University

Convolutional Layer



Max Pooling Layer

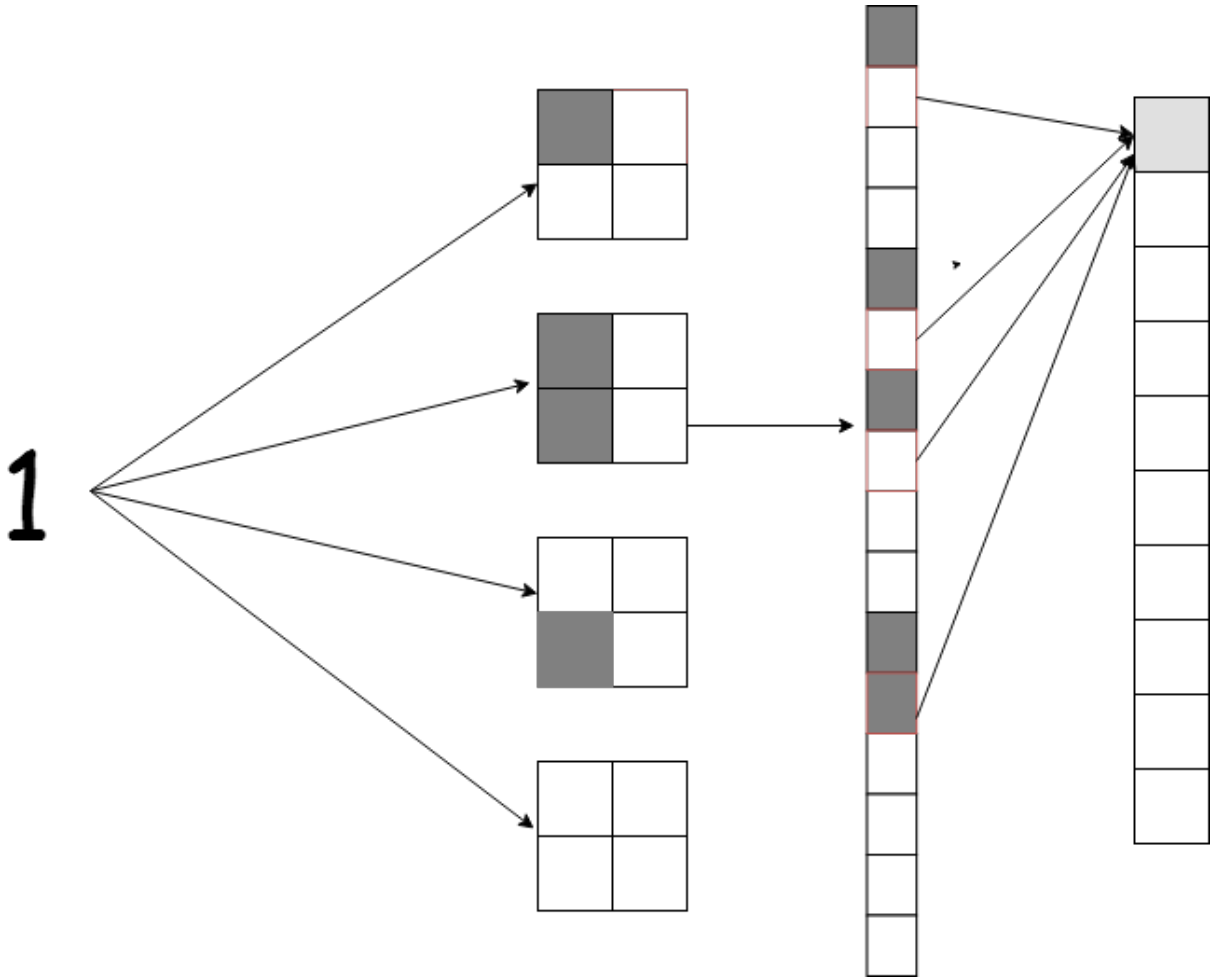
What if digit 1 is shifted to the left?



Santa Clara University

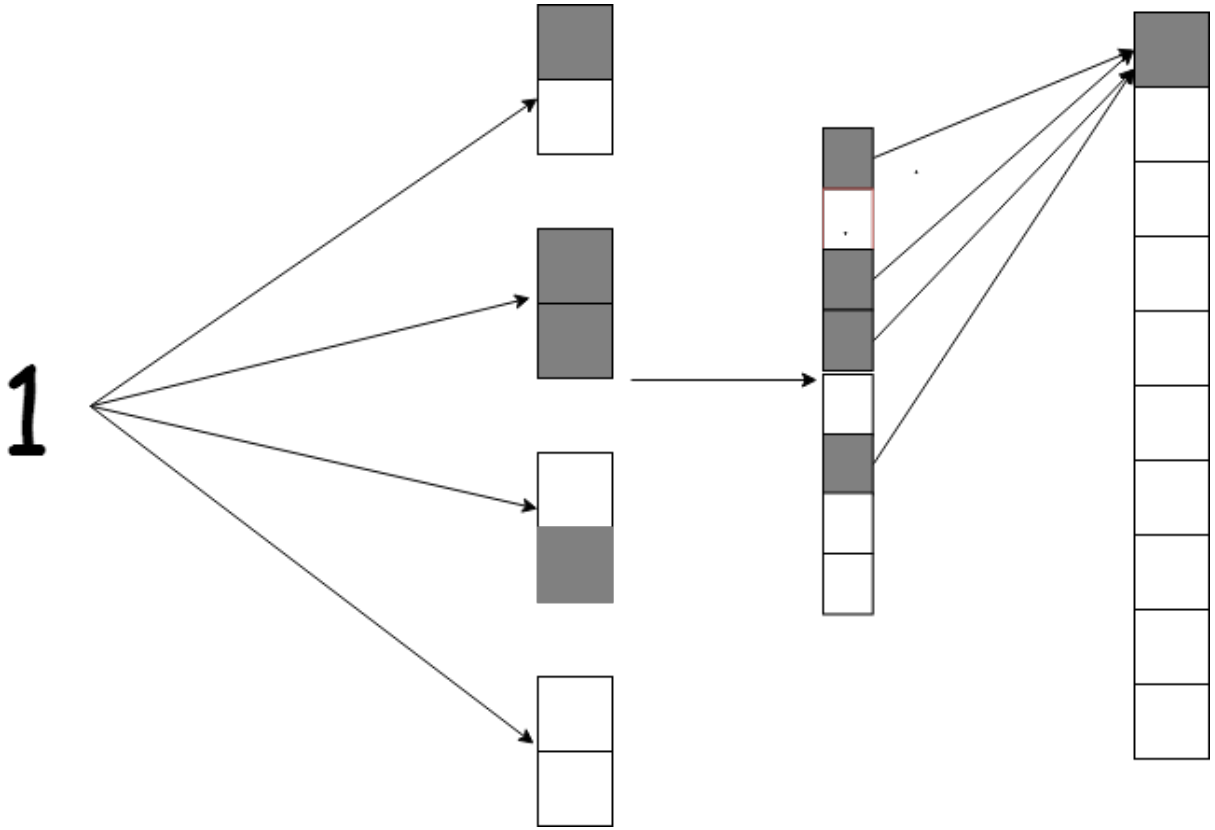
Max Pooling Layer

What if digit 1 is shifted to the left?

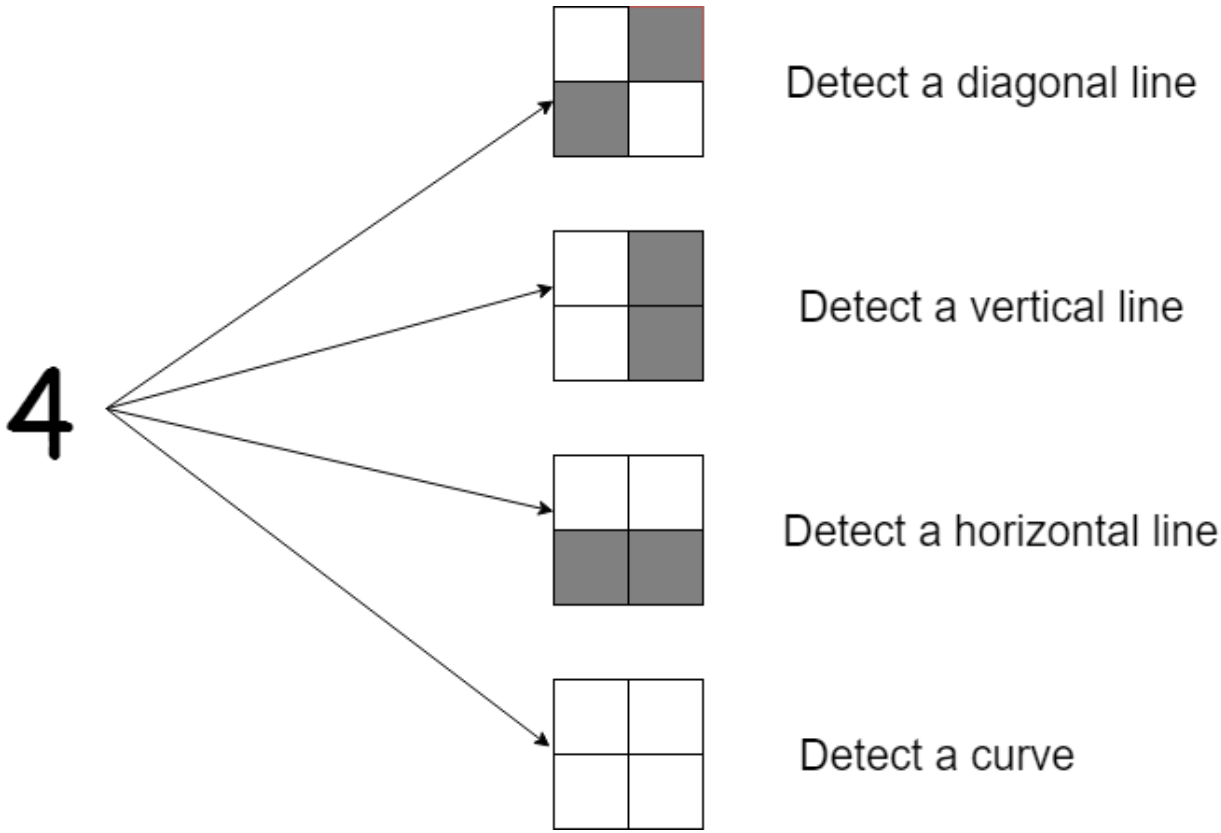


Max Pooling Layer

Sub-sample the feature map.

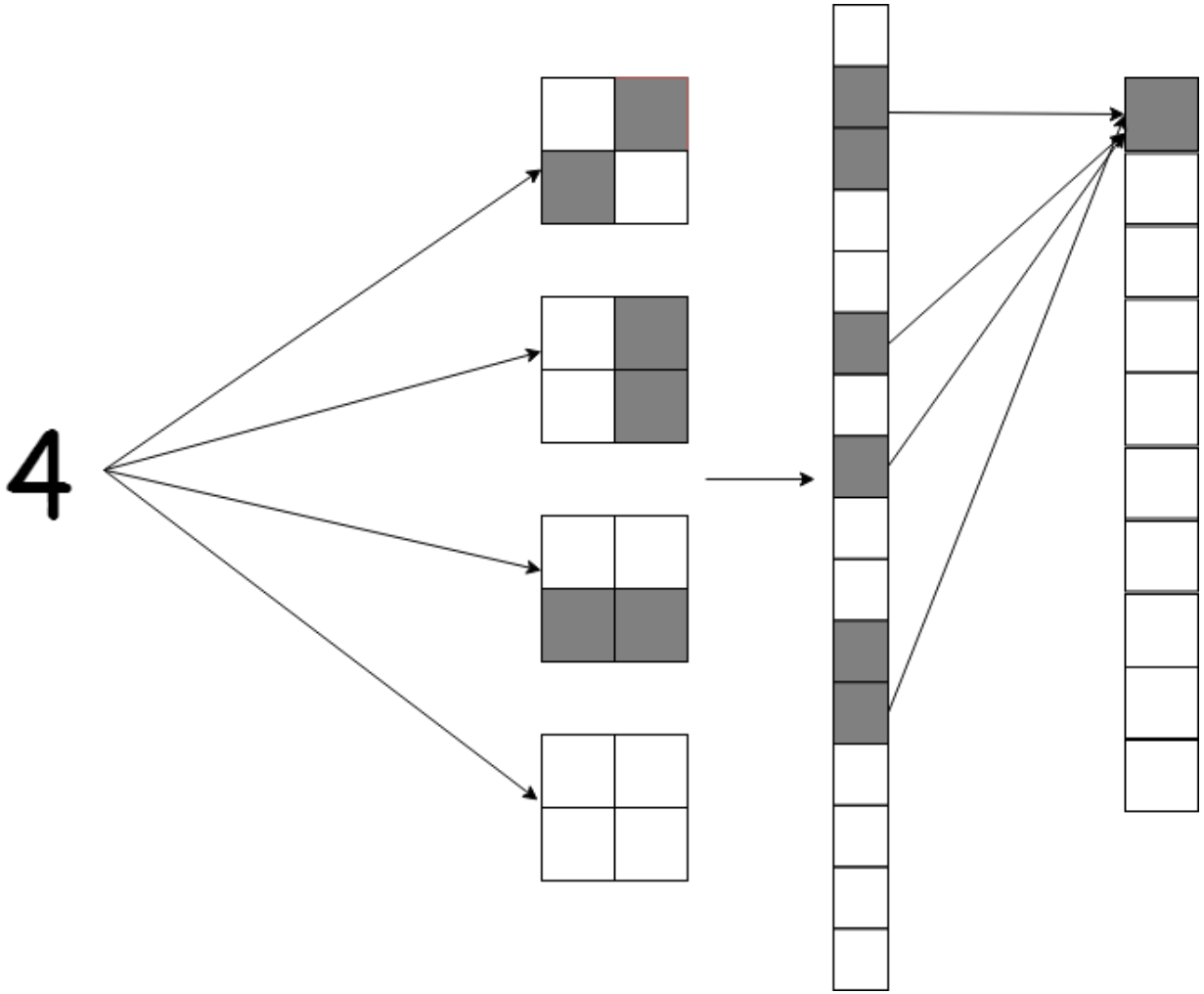


Try to recognize digit 4



Santa Clara University

Try to recognize digit 4



What is the limitation of CNNs?

Why does CNNs poorly distinguish digit 4 from digit 1?

Capsule Networks

- Model a part-whole relationship.
- Each capsule represents one unique feature.
- A group of low-level capsules represents "part" of the object.
- A high-level capsule represents an object.
- Model a hierarchical structure of visual features.

Santa Clara University

Try to recognize a face using CNNs

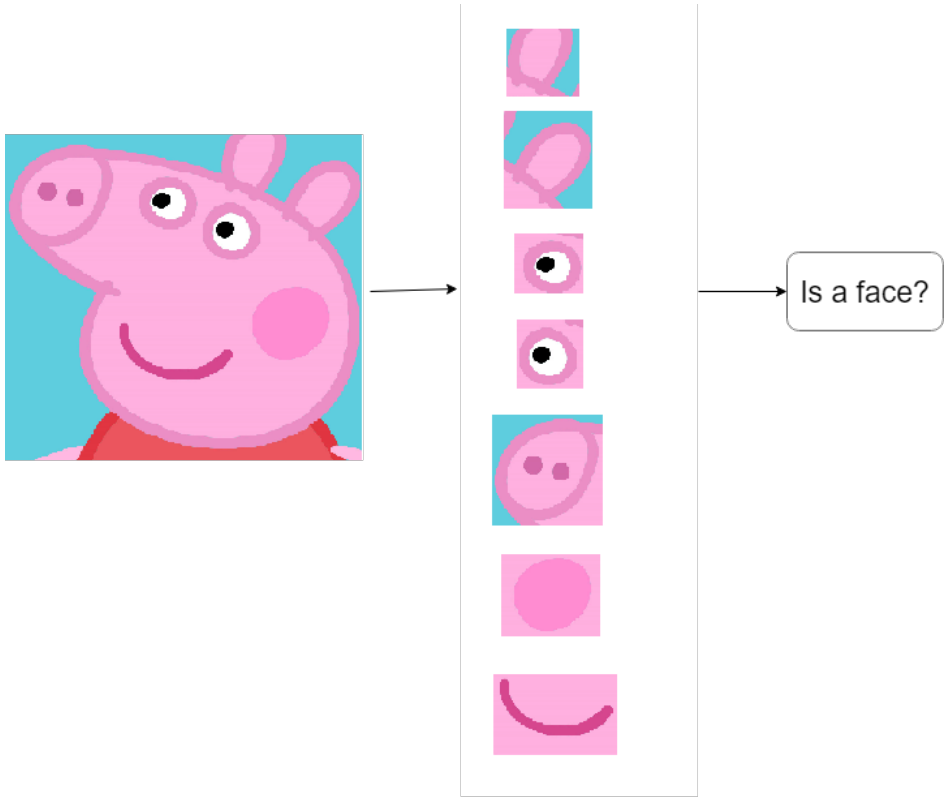


Figure: Detecting Peppa pig's face[2]

[2]Image source: `http://www.nickjr.com/peppa-pig/`

Try to recognize a face using CNNs

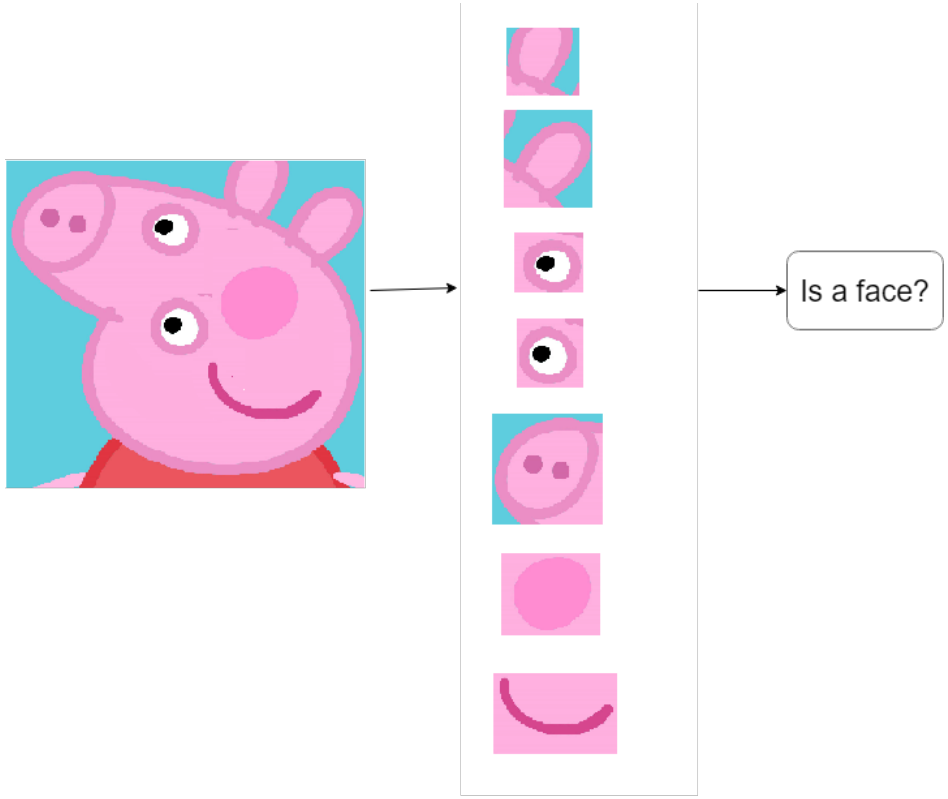
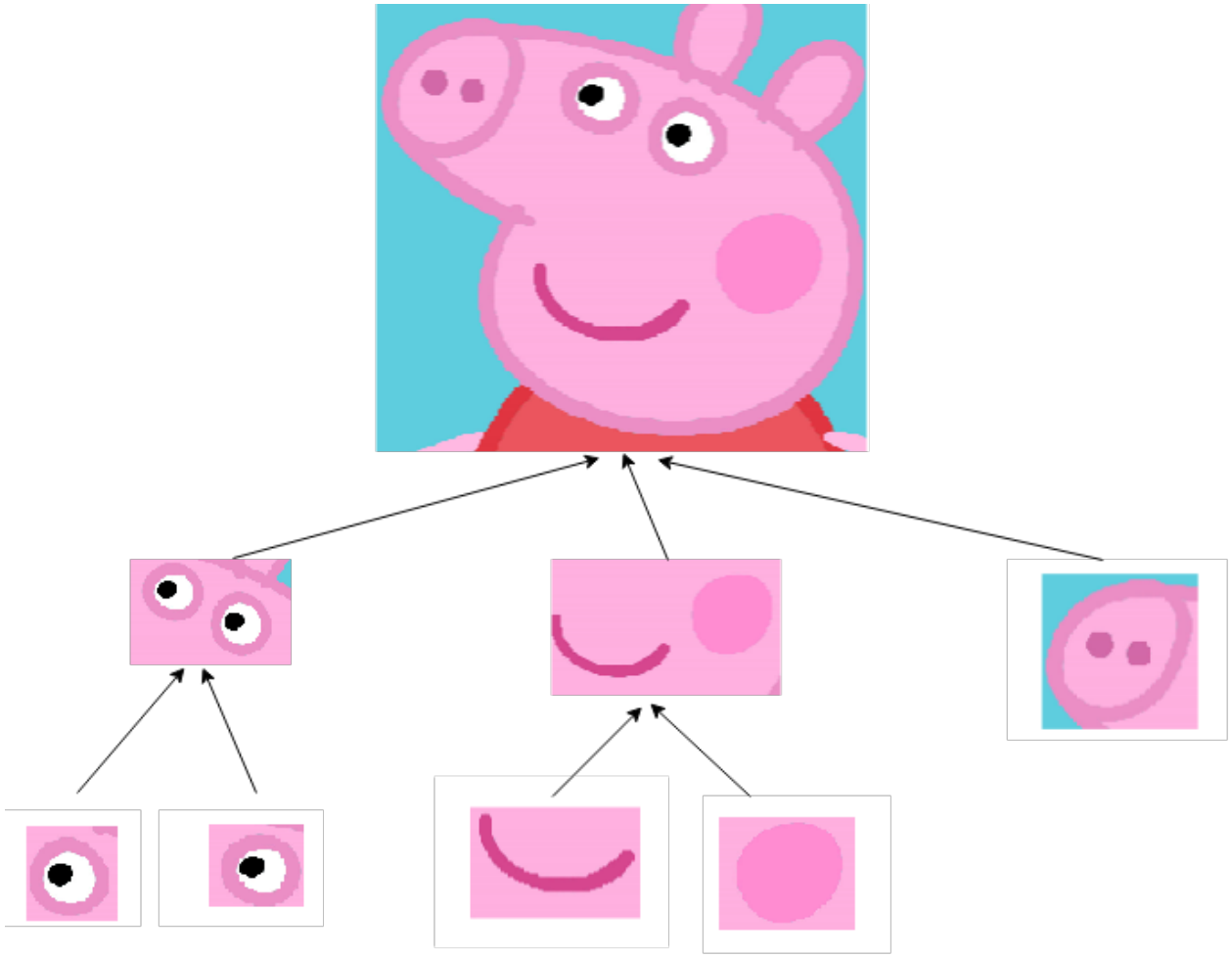
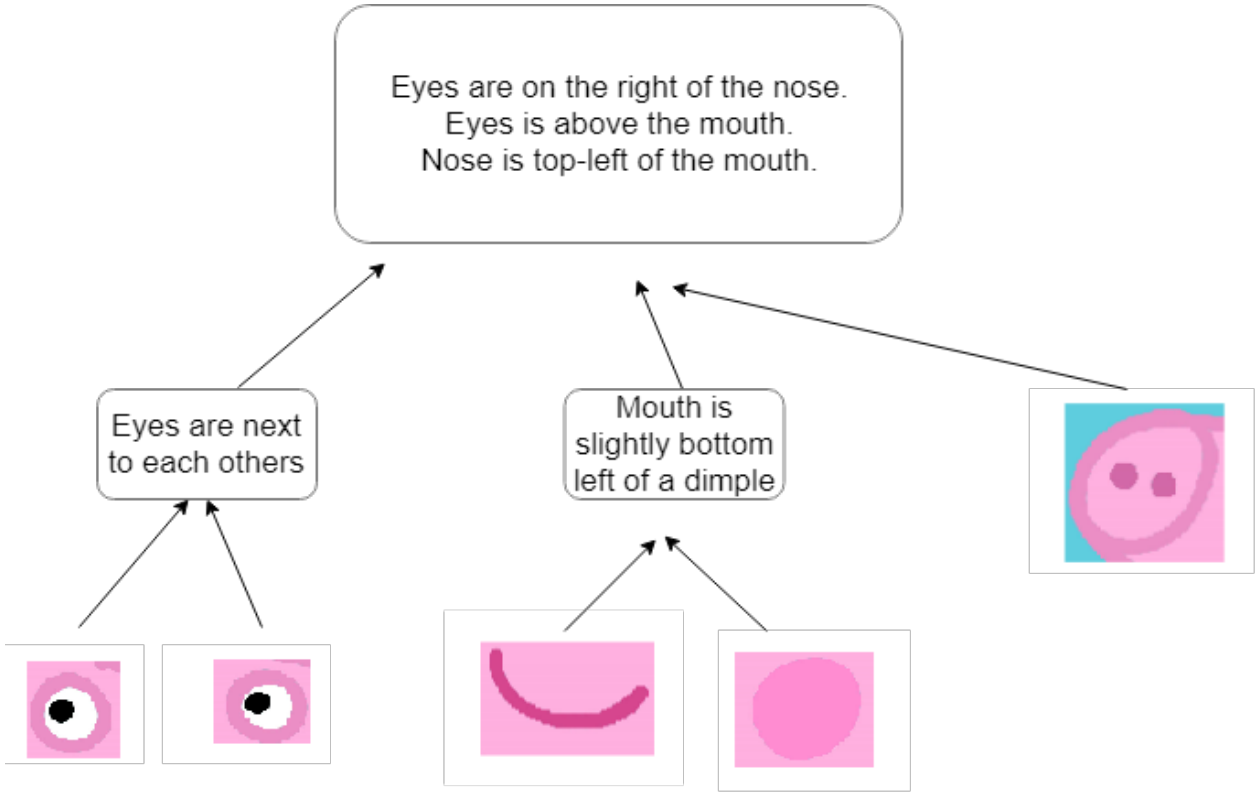


Figure: Detecting an incorrect Peppa pig's face

Learn Part-Whole Relationship

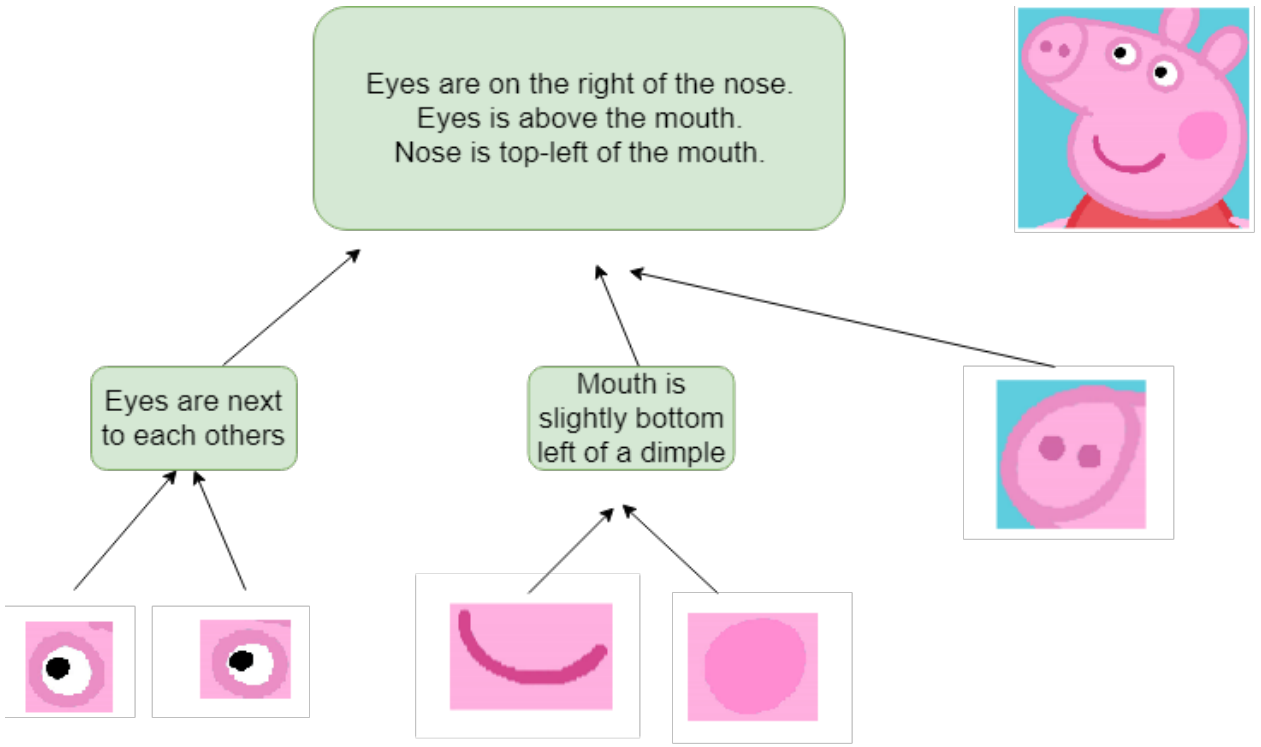


Learn Part-Whole Relationship

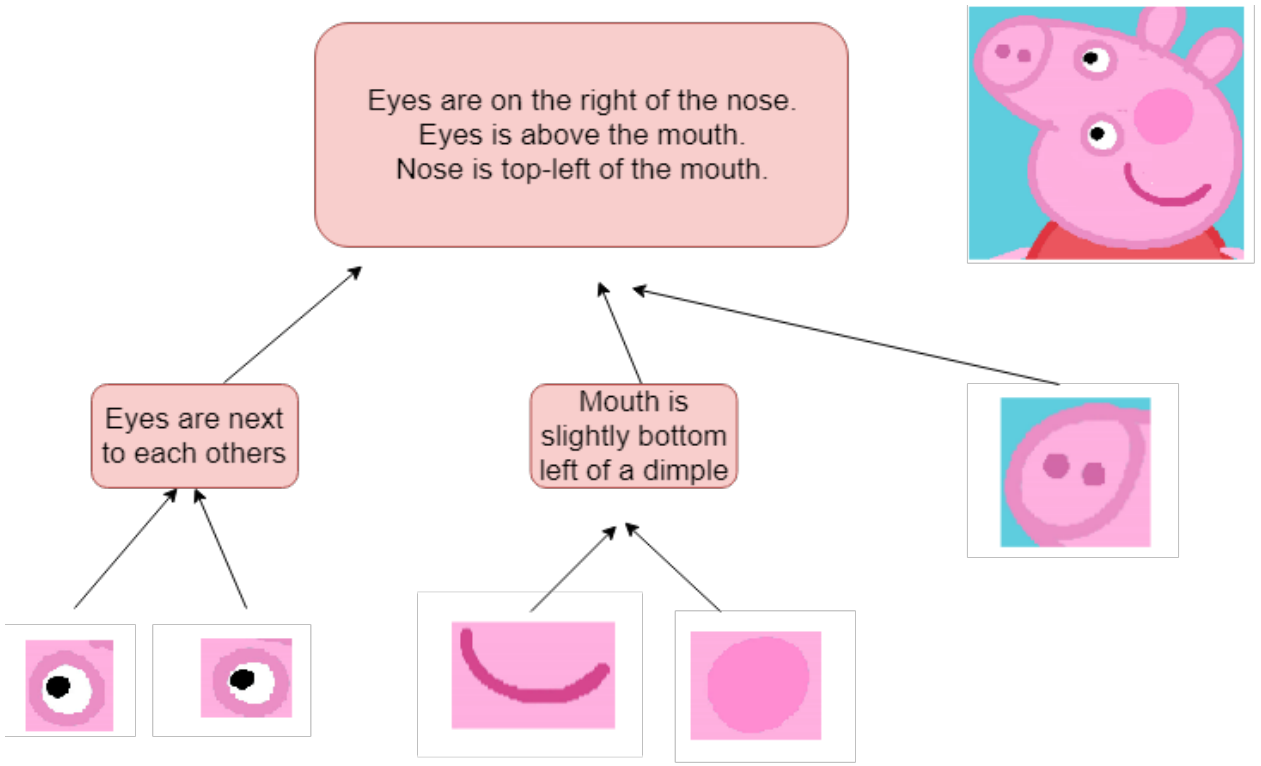


Santa Clara University

Learn Part-Whole Relationship



Learn Part-Whole Relationship



Santa Clara University

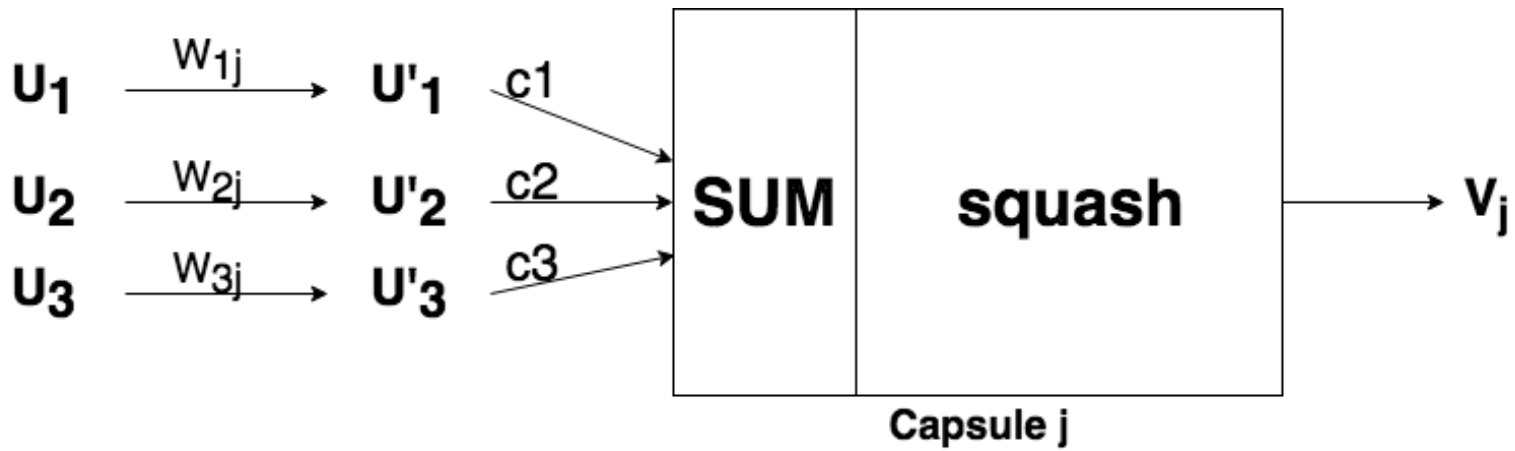
Capsule Approach

- Try to understand 3D space.
- Capsule encapsulates all important information about the state of the features they are detecting in a vector form.
- A vector length is a probability of detecting a feature.
- A vector direction is the state of the detected feature.

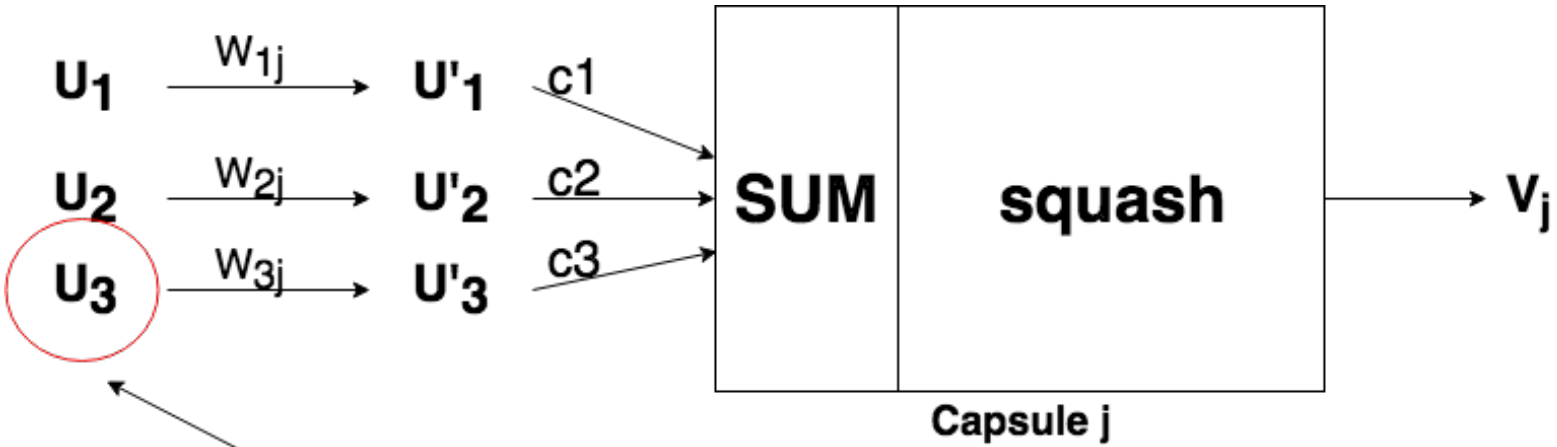
Capsule Approach

- Show a Capsule v.s. Traditional Neural.
- Draw a capsule layer.
- Draw a one-layer feedforward neural.

Capsule

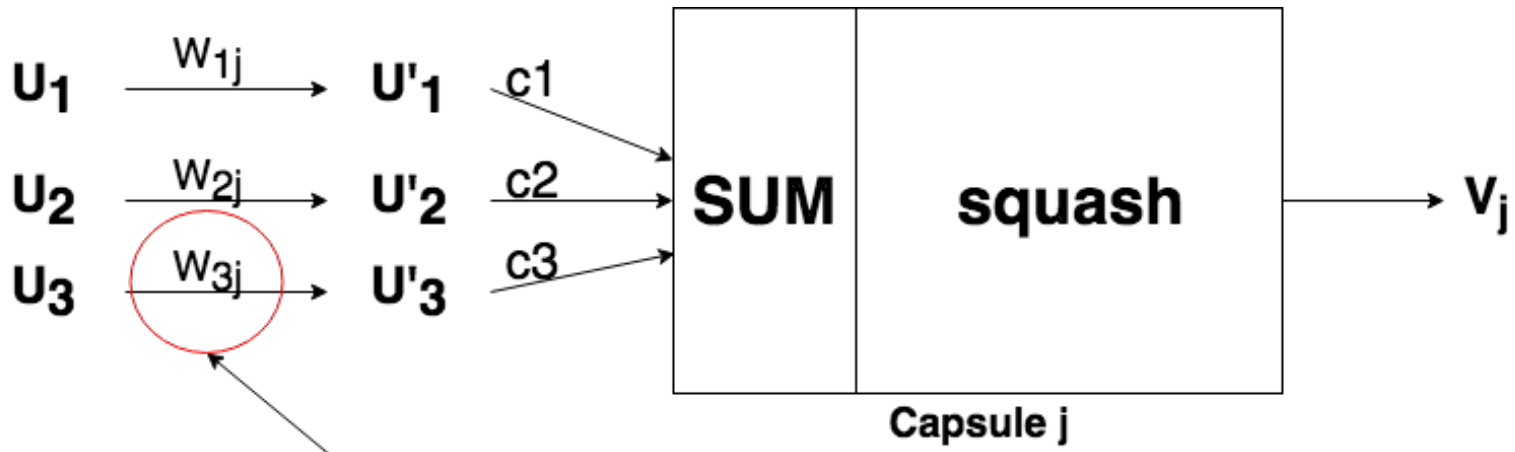


Capsule



a vector that encodes the state of the feature.

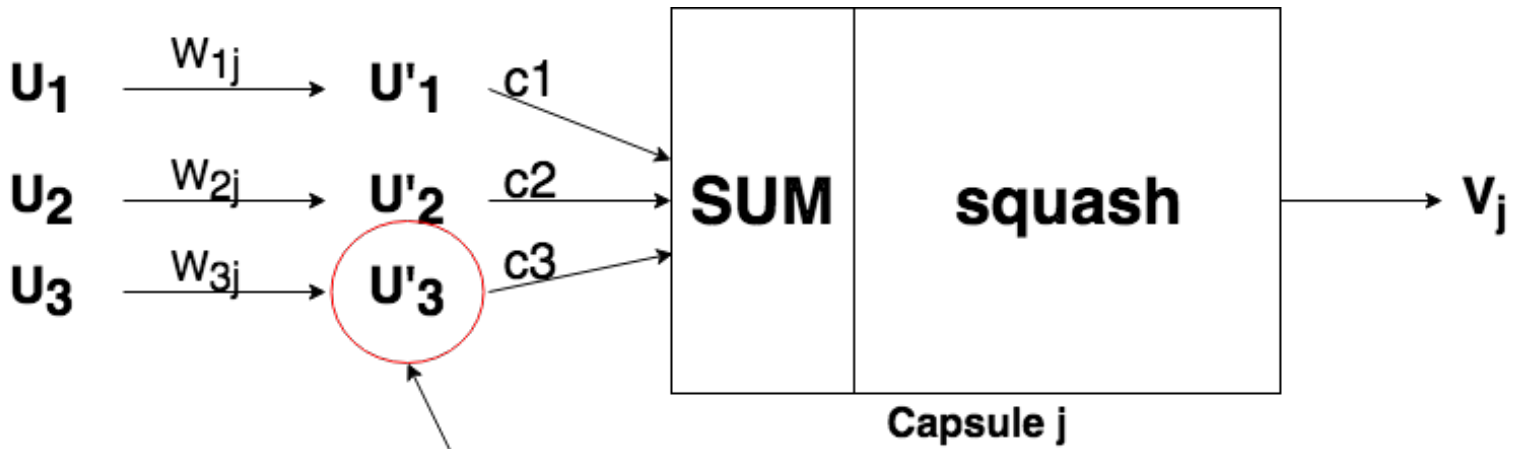
Capsule



a learnable weight matrix that encodes spatial relationship

Santa Clara University

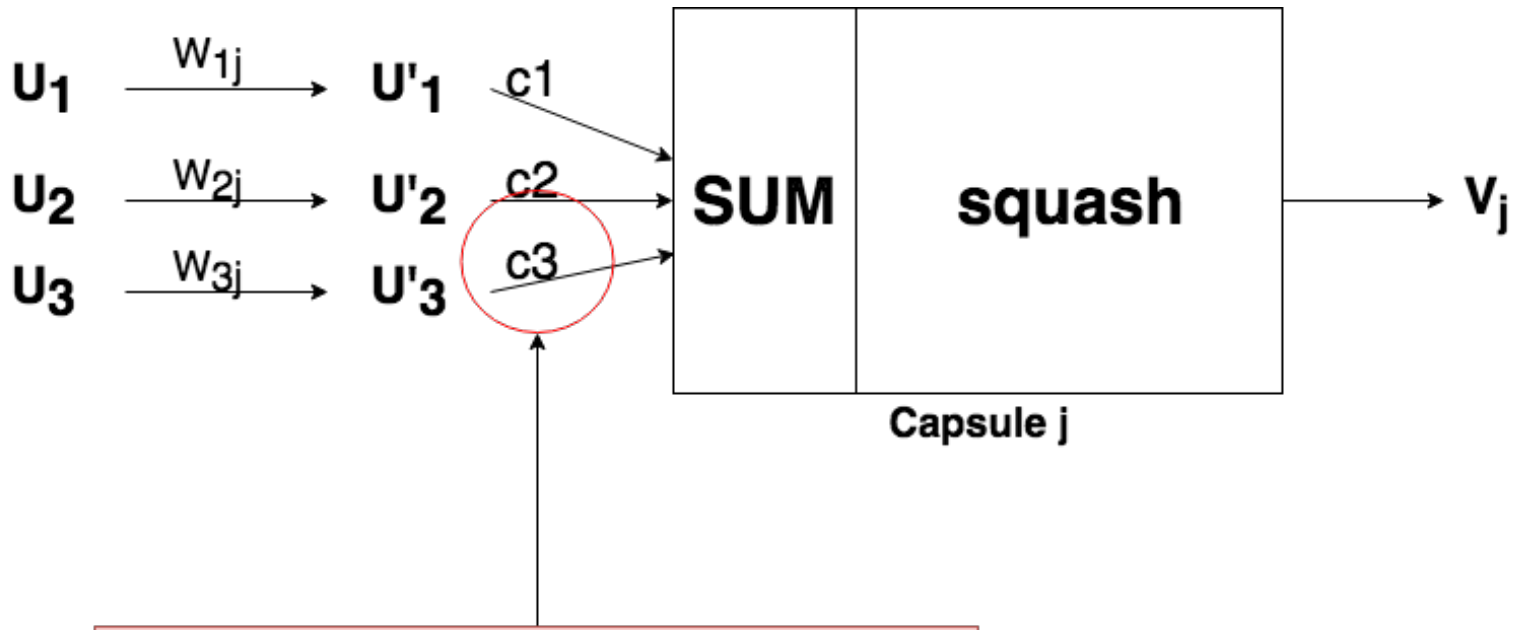
Capsule



a transformed state of the feature.

Santa Clara University

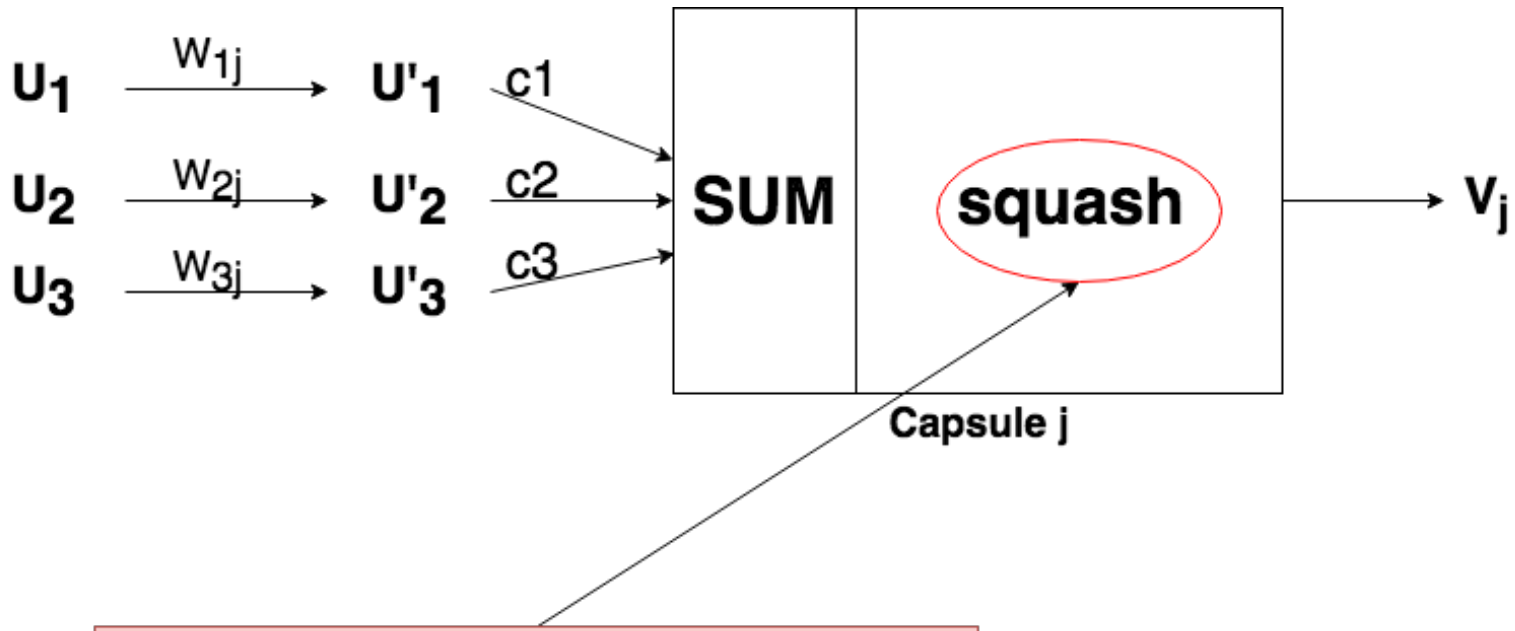
Capsule



a non-negative scalar weight determined by routing algorithm.

Santa Clara University

Capsule



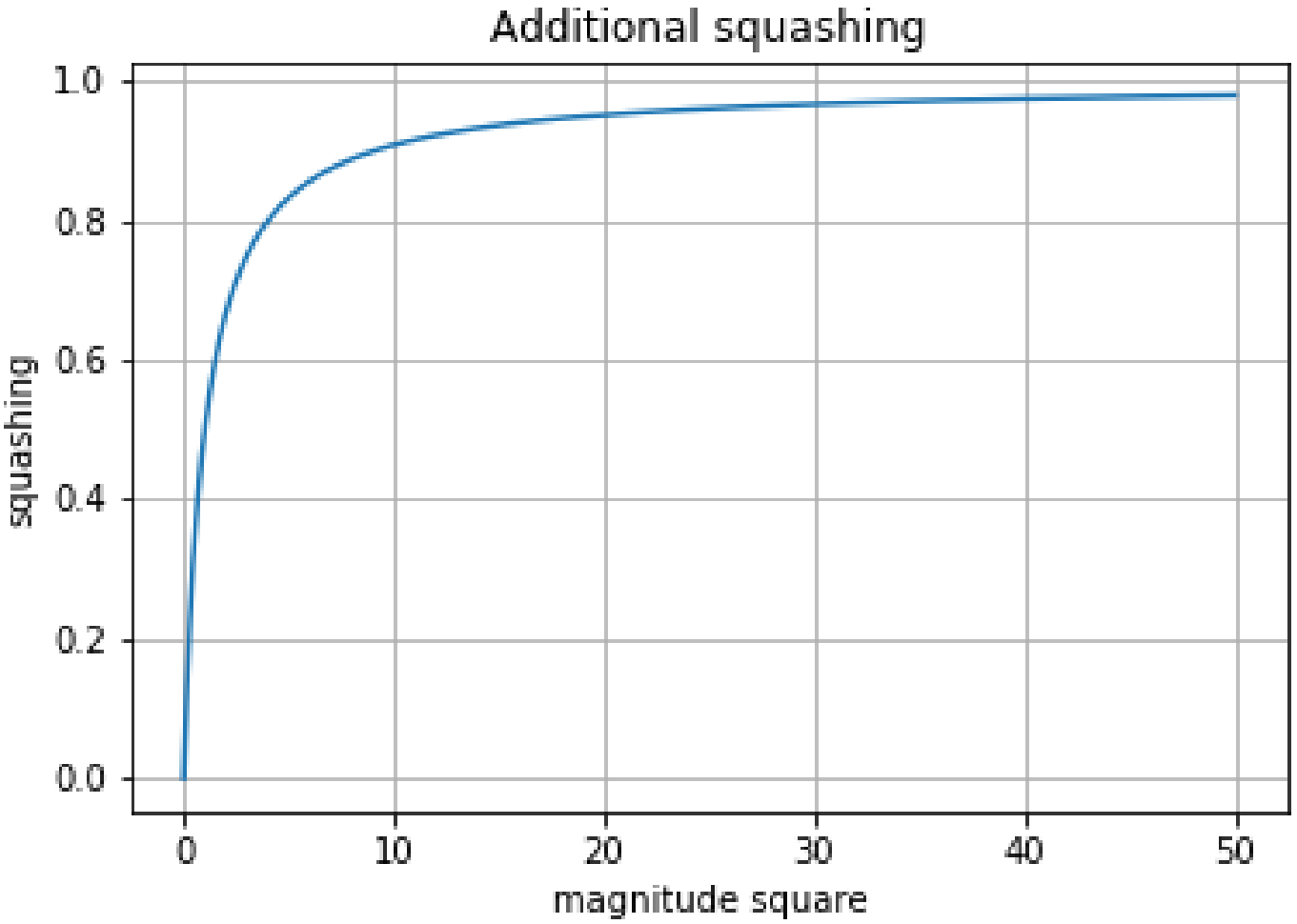
a non-linear function that takes a vector as an input and outputs another vector.

Squash Function

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

Santa Clara University

Squash Function



Santa Clara University

Dynamic Routing

- Each low-level capsule i, it has a weight $c_{i,j}$ as a probability of its output belong to each high level capsule j.
- $\sum_j c_{i,j} = 1$
- $\forall i, j \qquad c_{i,j} >= 0$

Dynamic Routing

1. for all capsule i in layer l and capsule j in layer $l+1$: $b_{ij} = 0$
2. for r iterations do:
   1. for all capsule i in layer l : $c_i = \text{softmax}(b_i)$
   2. for all capsule j in layer $l+1$: $s_j = \sum_i c_{ij} \hat{u}_{j|i}$
   3. for all capsule j in layer $l+1$: $v_j = \text{squash}(s_j)$
   4. for all capsule i in layer l and capsule j in layer $l+1$: $b_{ij} = b_{ij} + \hat{u}_{j|i} \cdot v_j$
3. return v_j

Santa Clara University

Architecture for MNIST

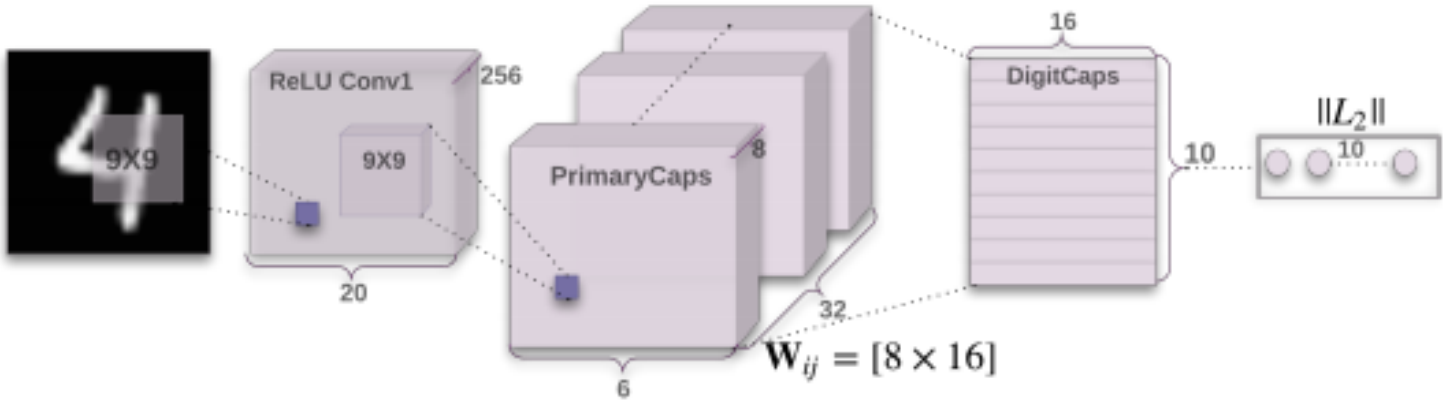


Figure: An encoder architecture[3]

[3]Image is taken from the [1]

Architecture for MNIST

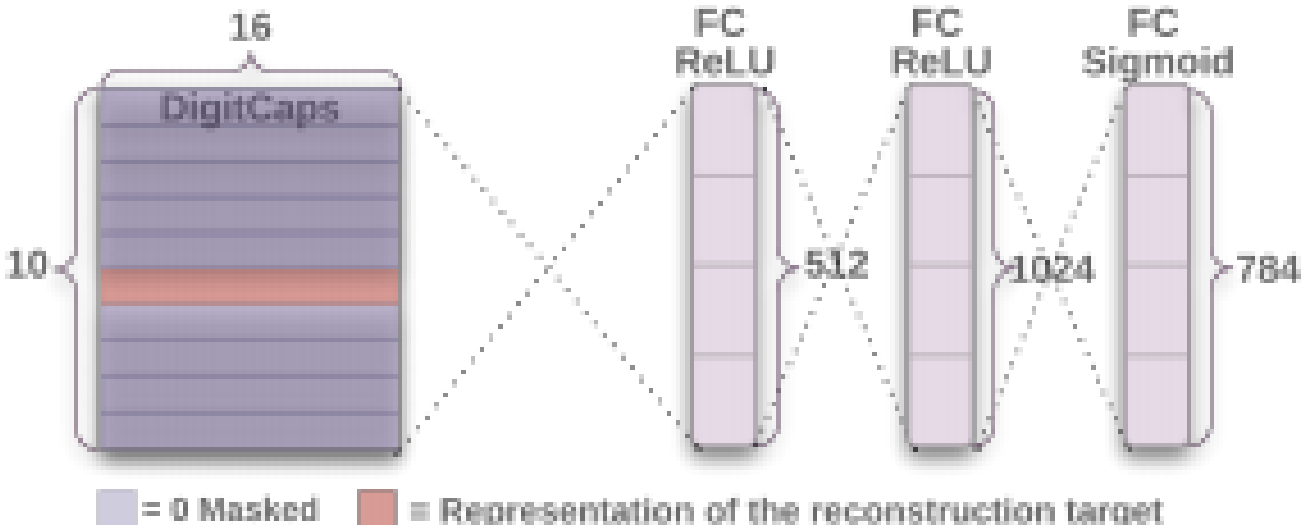


Figure: A decoder architecture[4]

[4] Image is taken from the [1]

Loss function

$$L_c = T_c \max(0, m^+ - \|V_c\|)^2 + \lambda(1 - T_c) \max(0, \|V_c\| - m^-)^2$$

References

1. https://medium.com/ai%C2%B3-theory-practice-business/understanding-hintons-capsule-networks-part-i-intuition-b
2. Dynamic Routing Between Capsules
   https://arxiv.org/abs/1710.09829

Santa Clara University

Questions

Questions?

schaidaroon@scu.edu

Santa Clara University