电子科技大学
University of Electronic Science and Technology of China

Capsule Neural Network and Discussion on Sync NN

Wei Han

Data Mining Lab,
Big Data Research Center, UESTC
Email：weihan@std.uestc.edu.cn

Outline

- Capsule Neural Network
  - Background and Motivation
  - Structure of CapsNet and a demo
  - Experiments

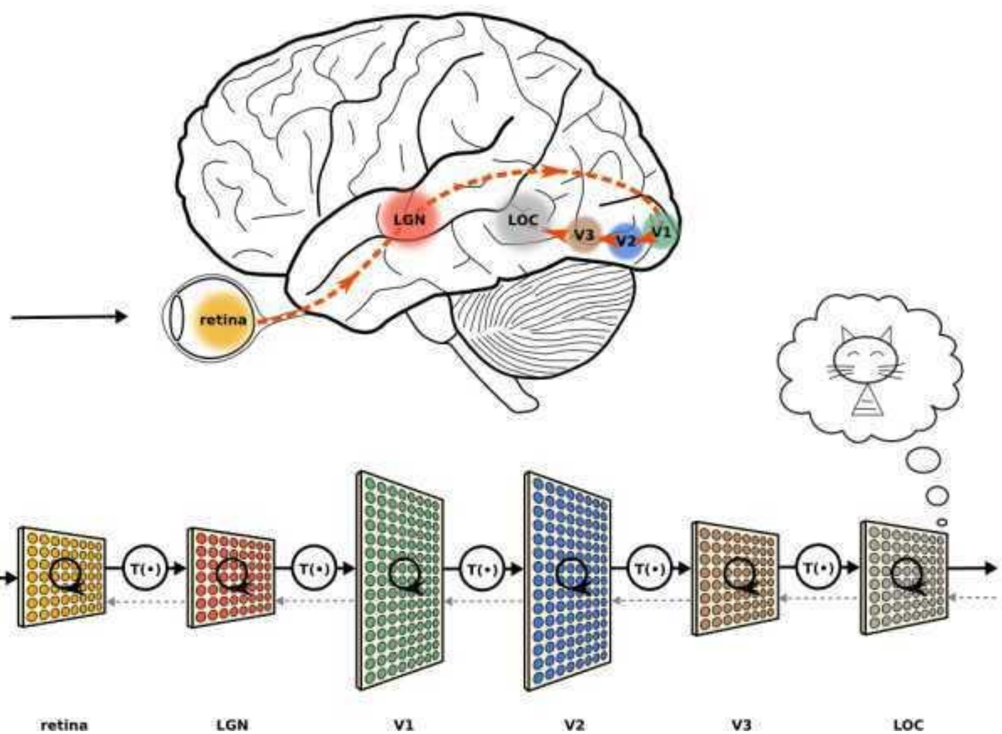
- Synchronization Neural Network
  - Background and Motivation
  - Potential ways and Related works

Capsule Neural Network

Hinton's Thinking:

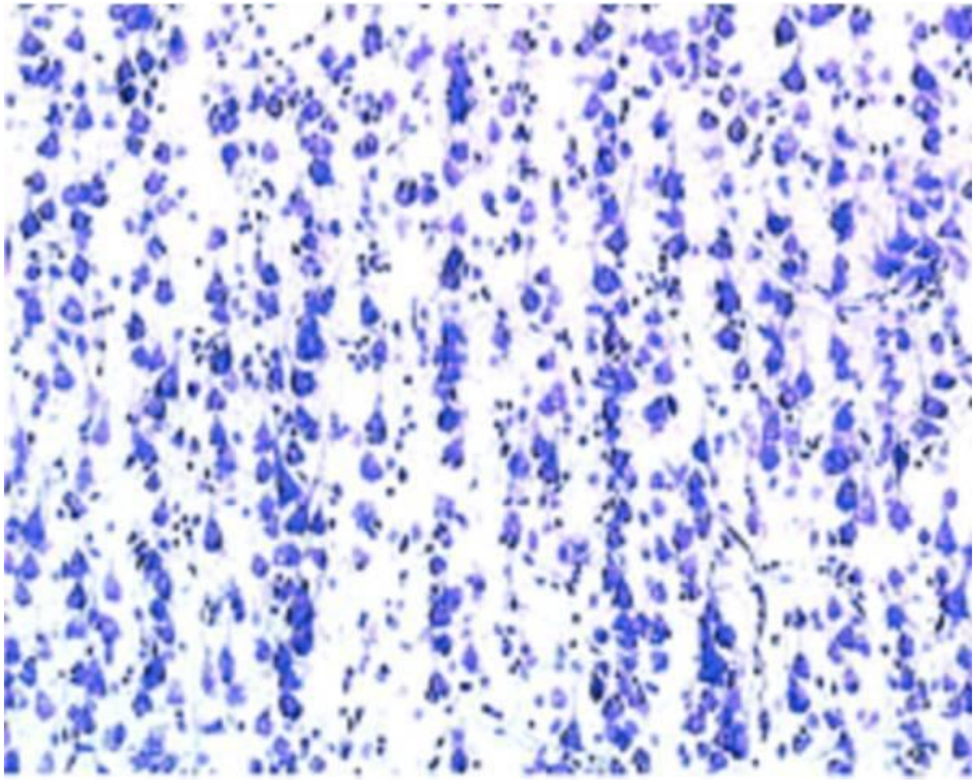
- No BP in the brain

- Fewer layers and slower computing



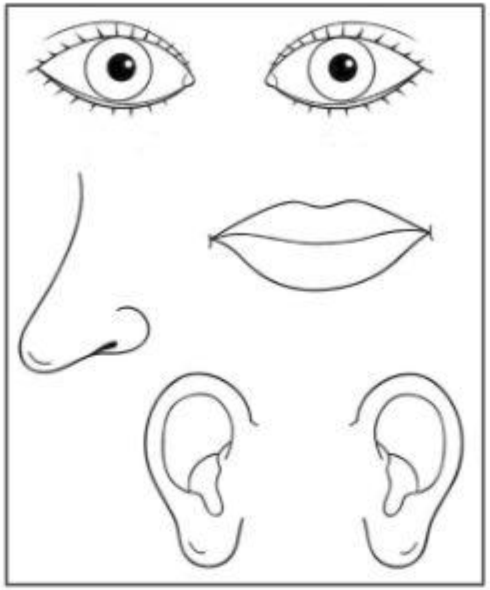
1.1. Background and Motivation

- Cortical minicolumn in the cortex
  - Include hundreds of neuros
  - Internal stratification

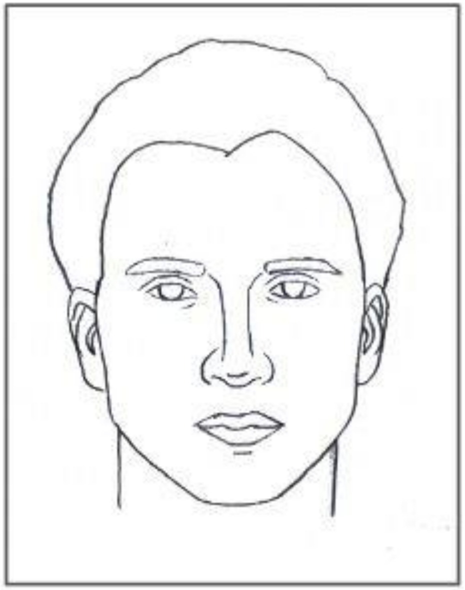


1.1. Background and Motivation

- Invariance & Equivariance



Not Face

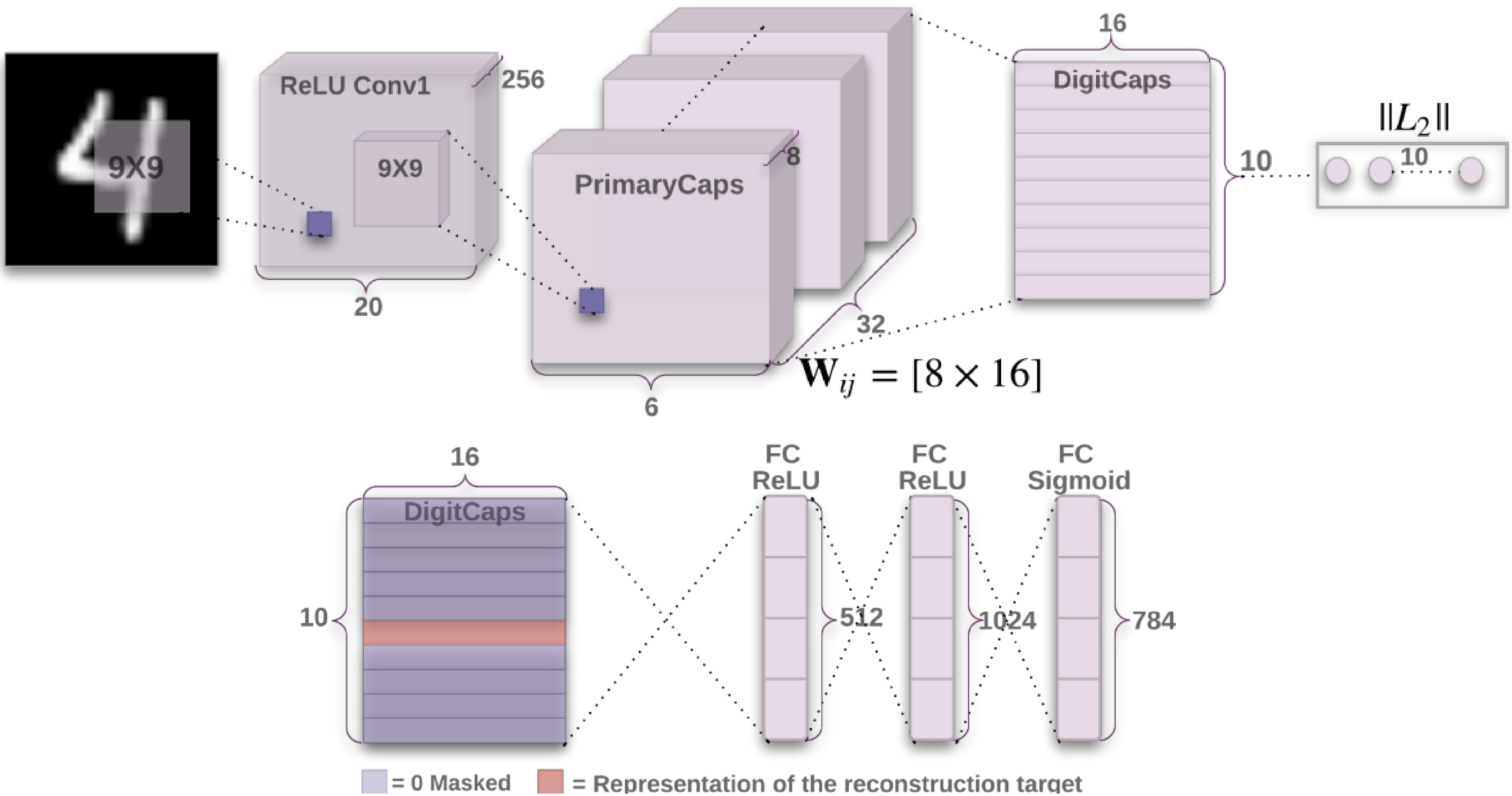


Face

- ‘Routing’ mechanism

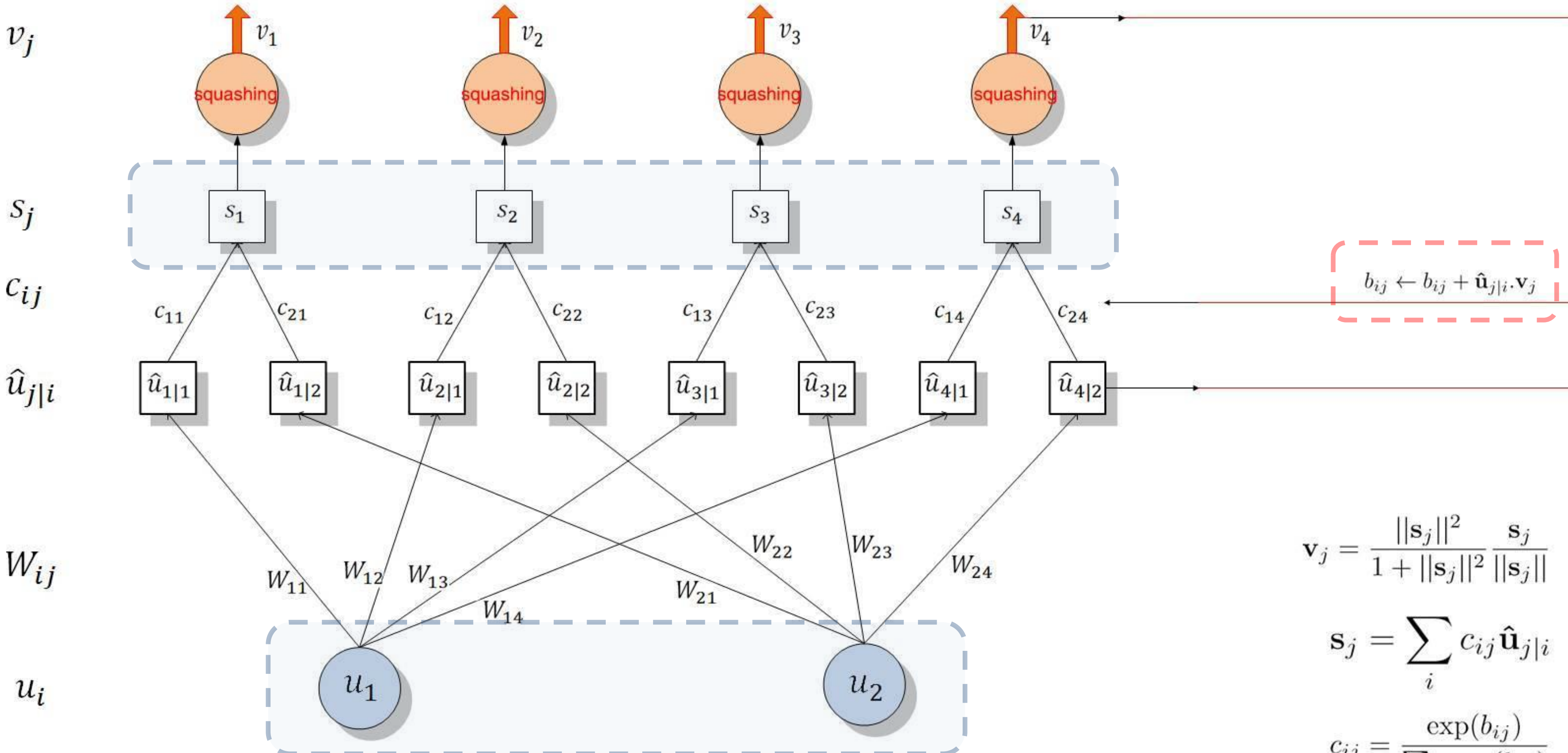
1.2. Structure of CapsNet

- Capsule: a group of neurons



1.2. Structure of CapsNet

- Layer connection



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

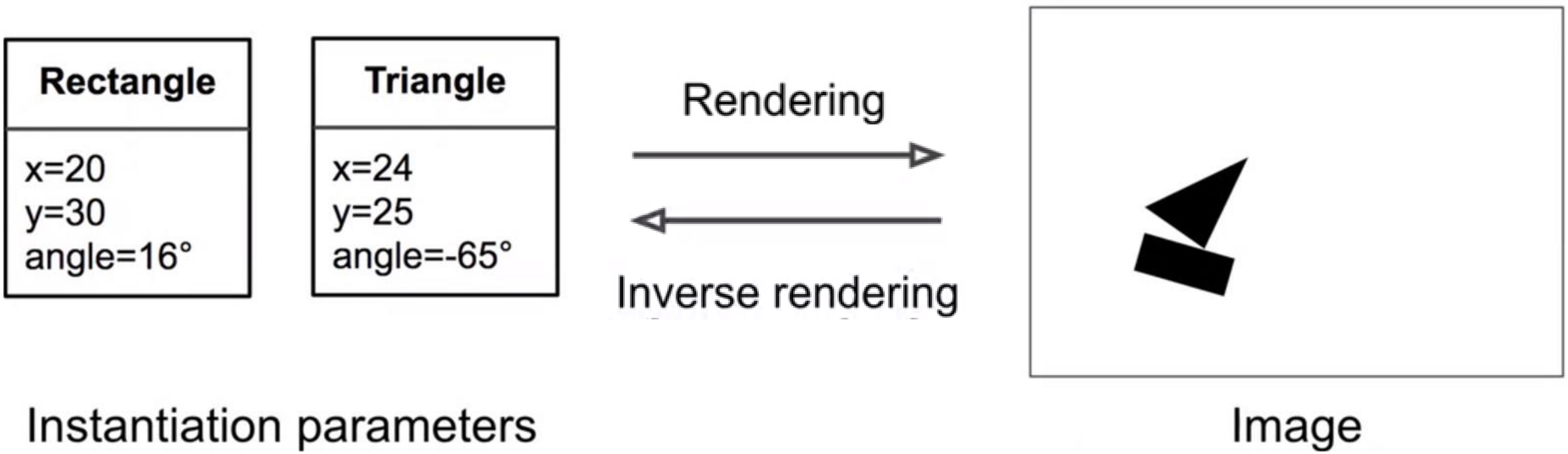
$$\mathbf{s}_j = \sum_i c_{ij} \hat{\mathbf{u}}_{j|i}$$

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})}$$

$$\hat{u}_{j|i} = W_{ij} u_i$$

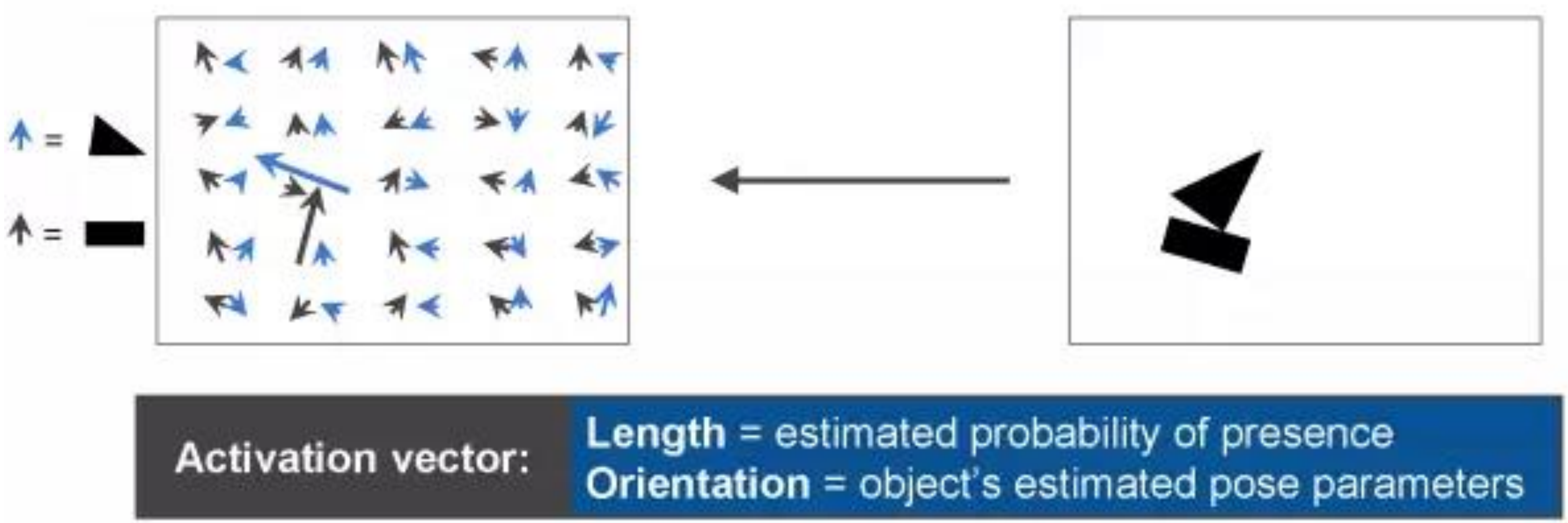
1.2. Demo

- Representation



1.2. Demo

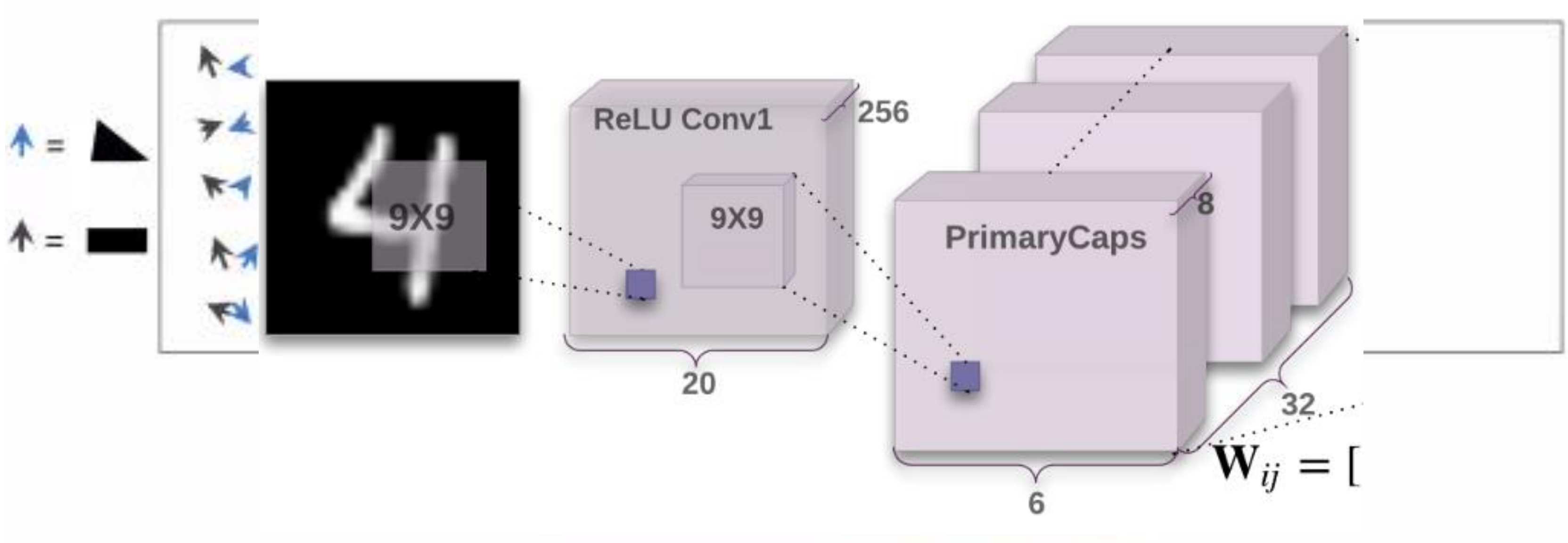
- Capsule



1.2. Demo

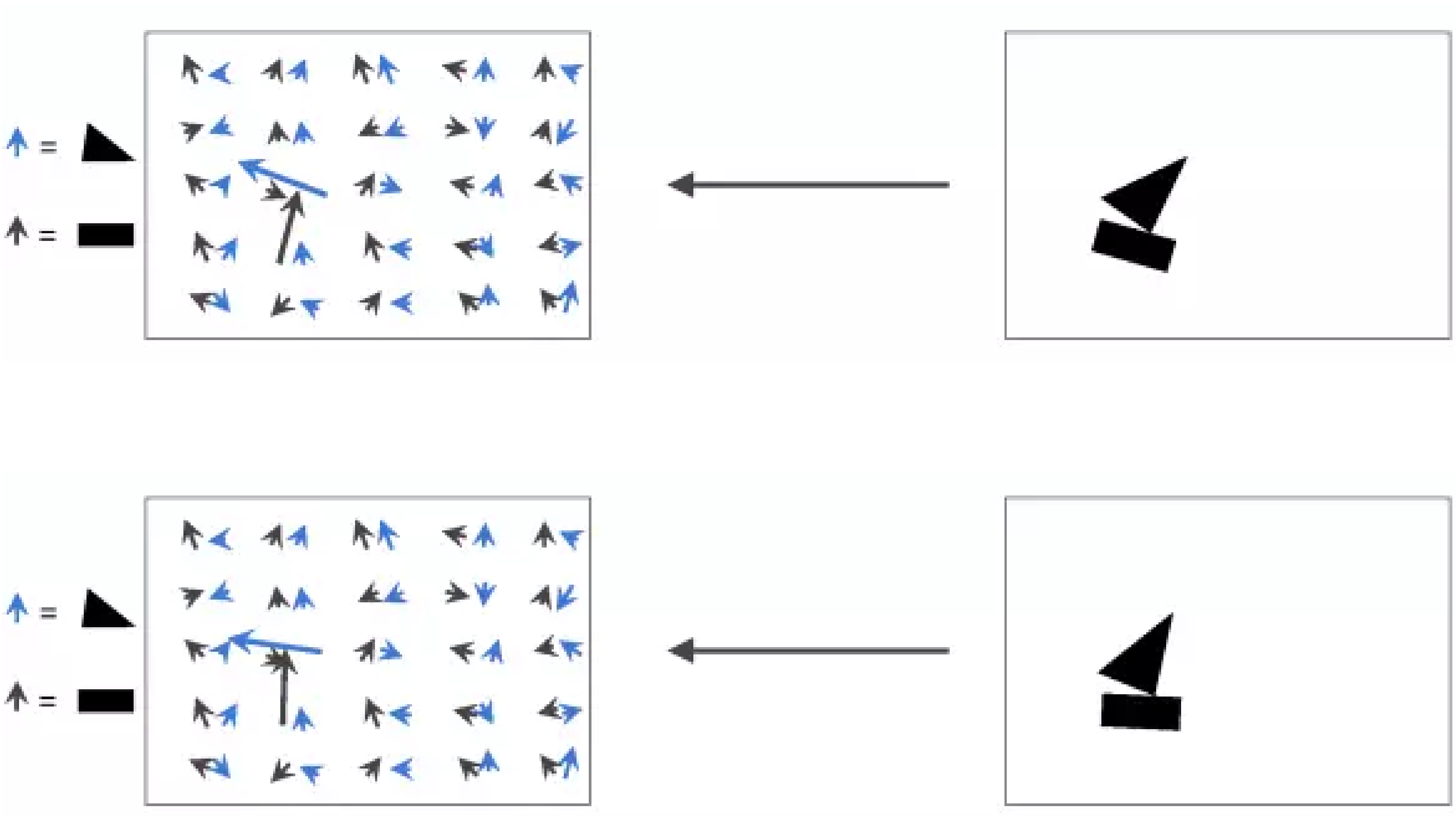
数据挖掘实验室
Data Mining Lab

- Capsule



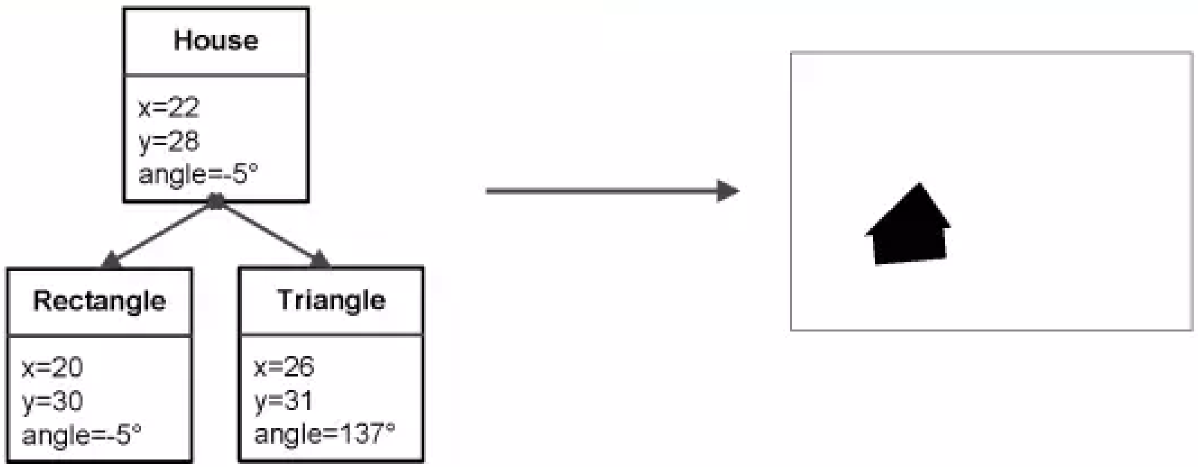
1.2. Demo

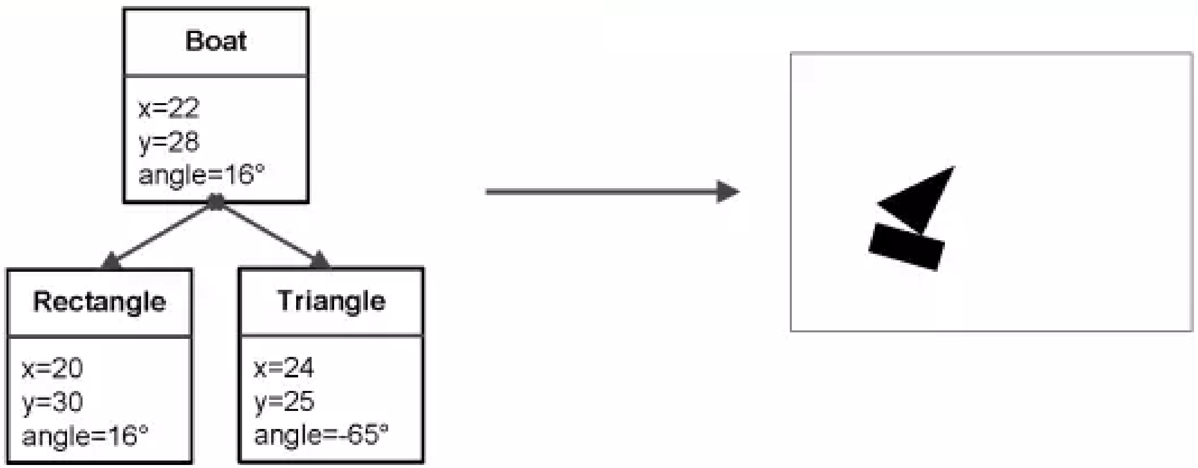
- Equivariance



1.2. Demo

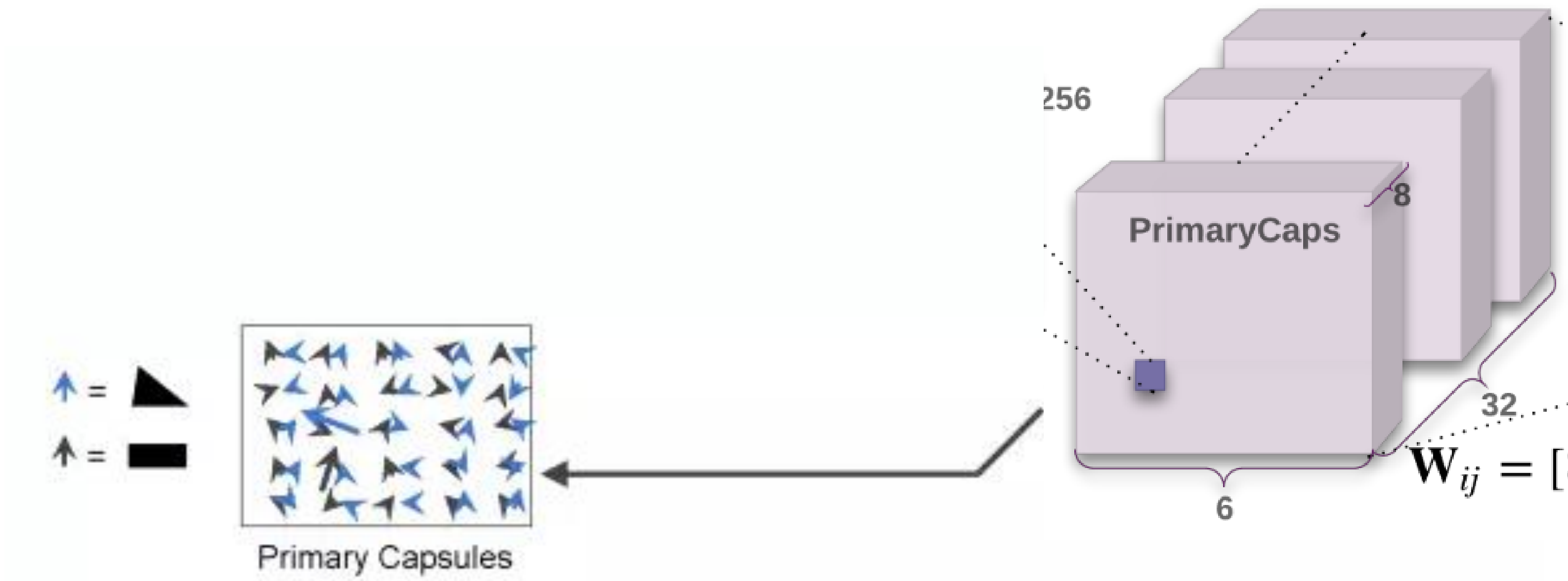
- Hierarchy of parts





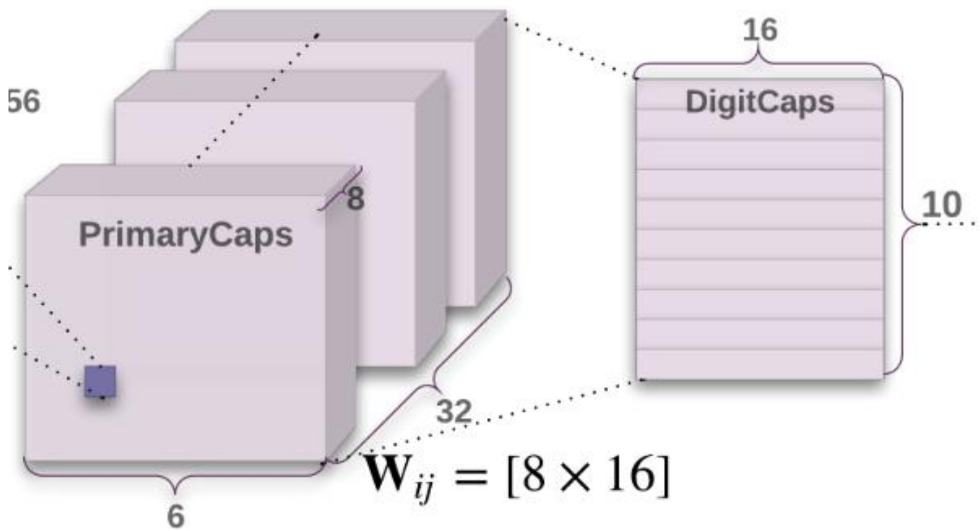
1.2. Demo

- Primary Capsules



1.2. Demo

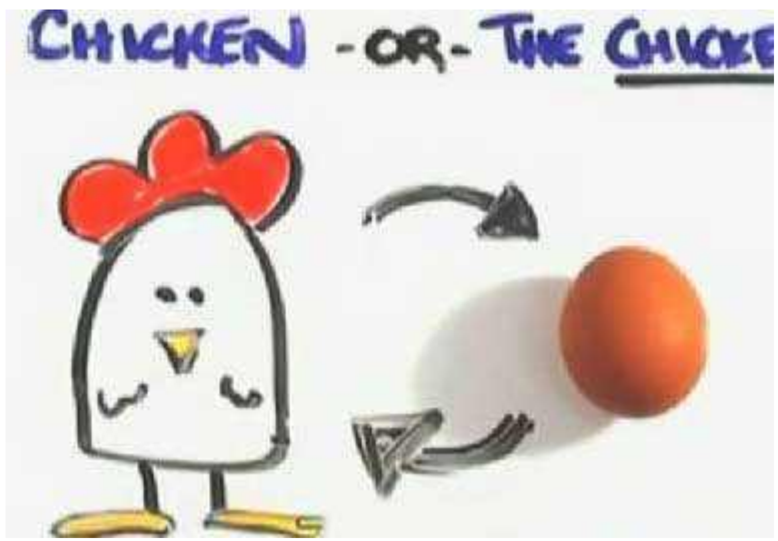
- Predict Next Layer's Output



- Chicken & Egg

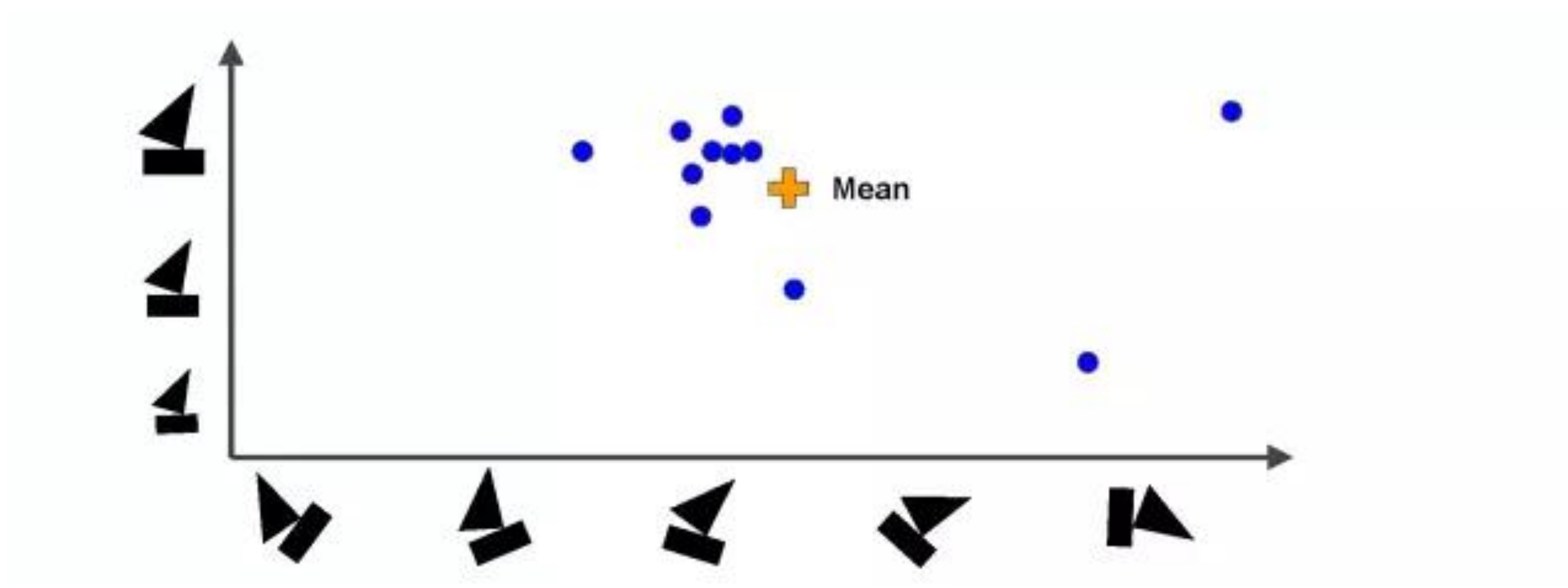
$$\mathbf{s}_j = \sum_i c_{ij} \hat{\mathbf{u}}_{j|i} \,, \qquad \hat{\mathbf{u}}_{j|i} = \mathbf{W}_{ij} \mathbf{u}_i$$

where c_{ij} is the similarity of s_j and $u_{j|i}$



1.2. Demo

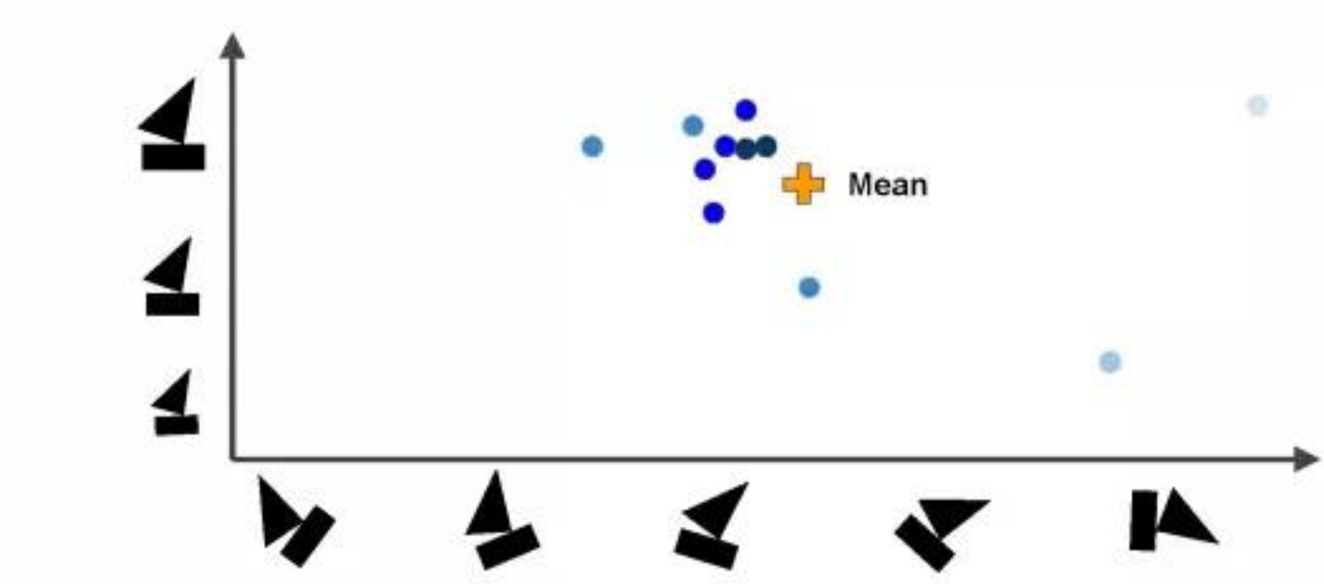
- Clusters of Agreement



1.2. Demo

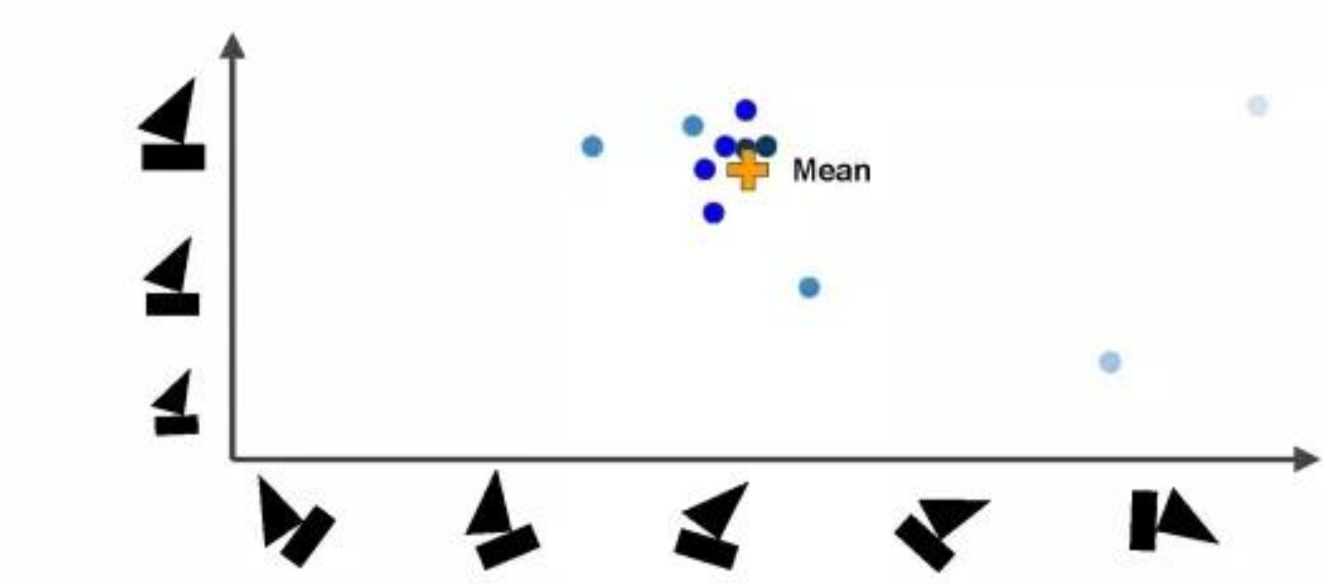
数据挖掘实验室
Data Mining Lab

- Clusters of Agreement



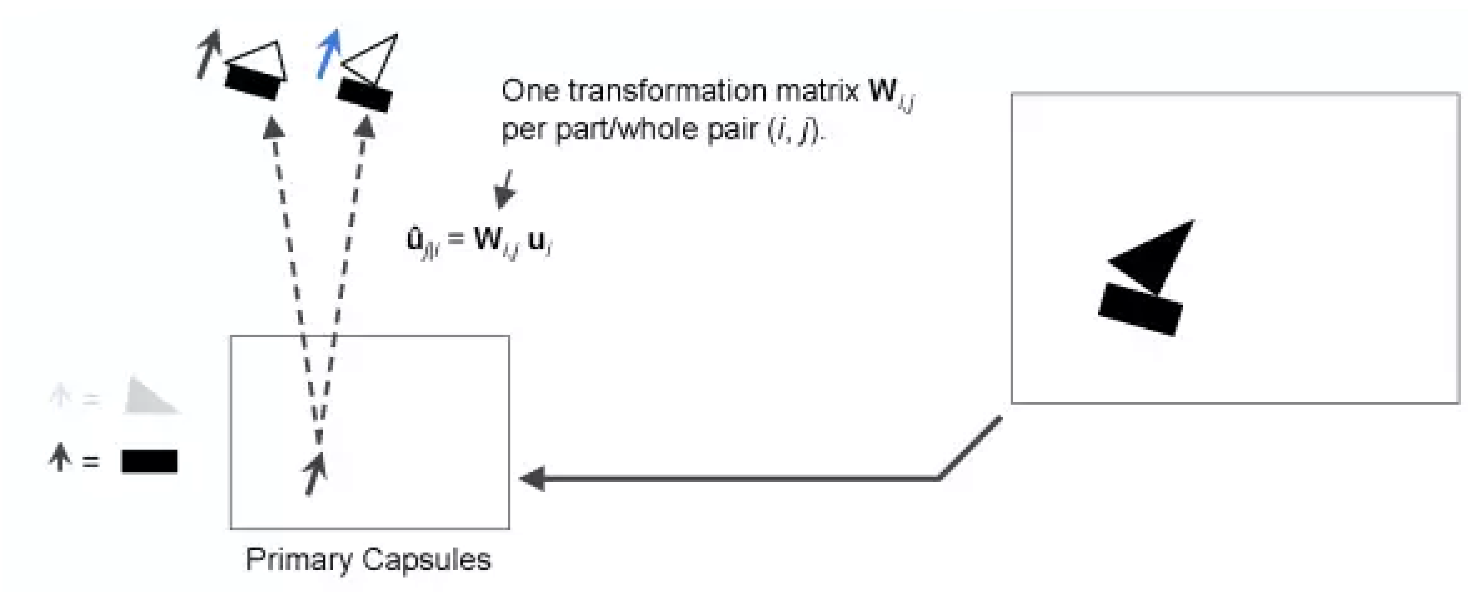
1.2. Demo

- Clusters of Agreement



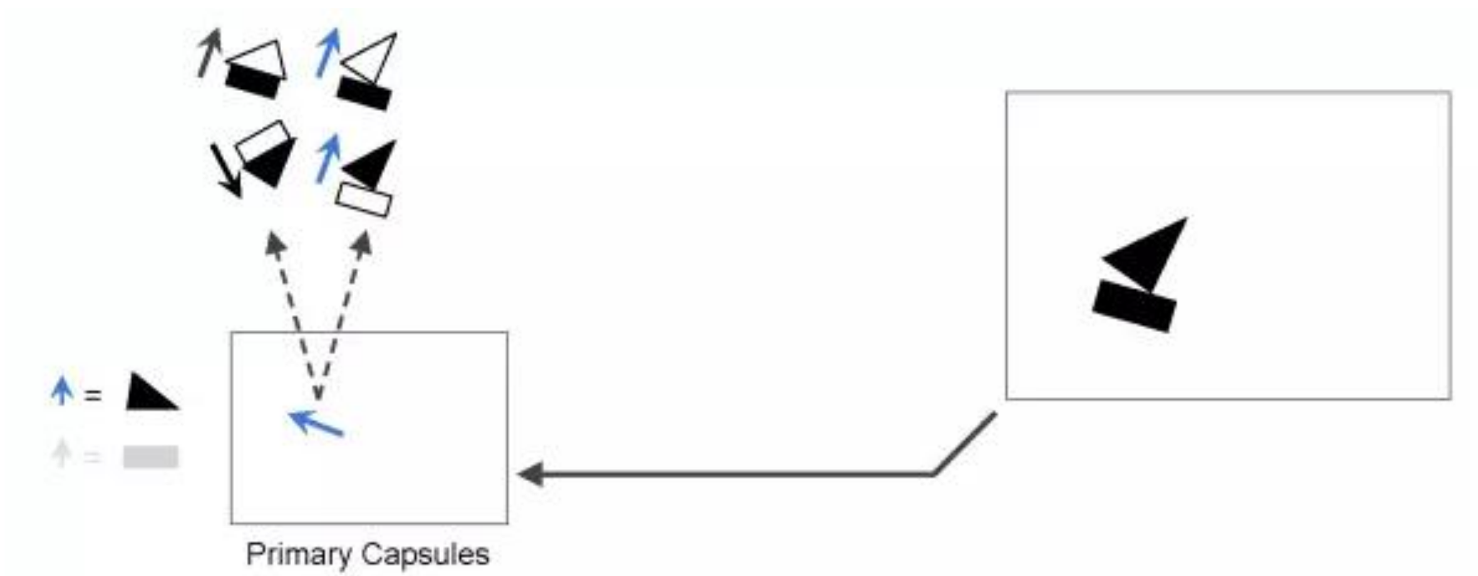
1.2. Demo

- Predict Next Layer's Output



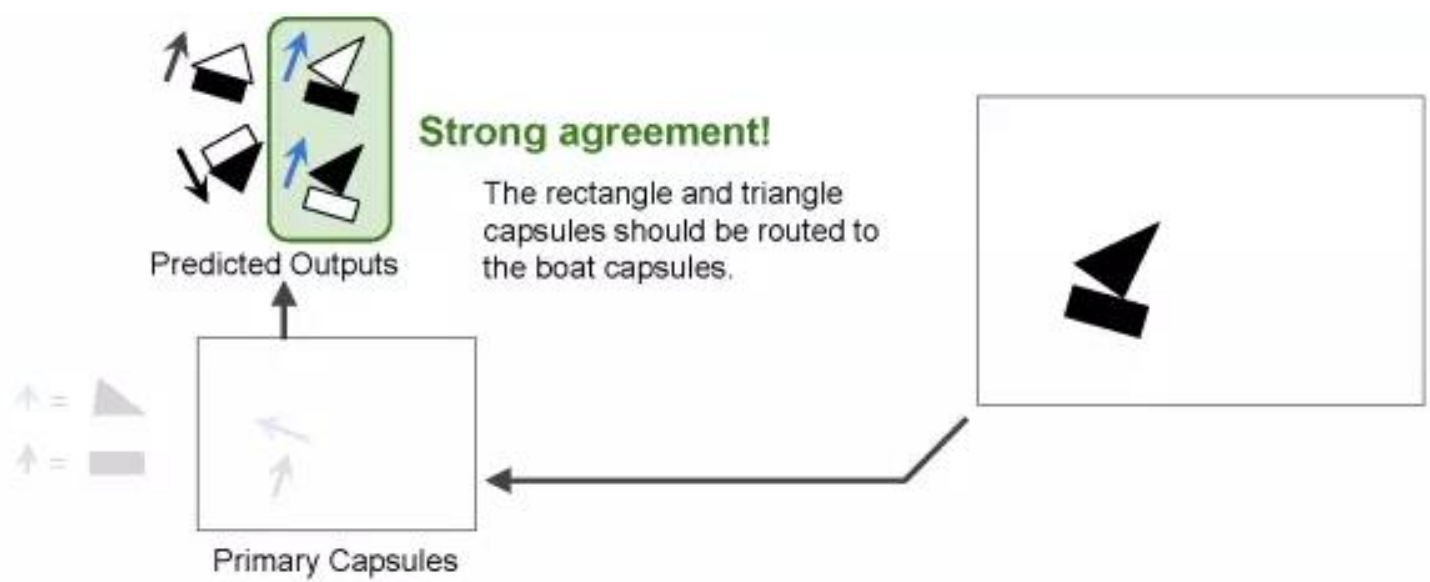
1.2. Demo

- Predict Next Layer's Output



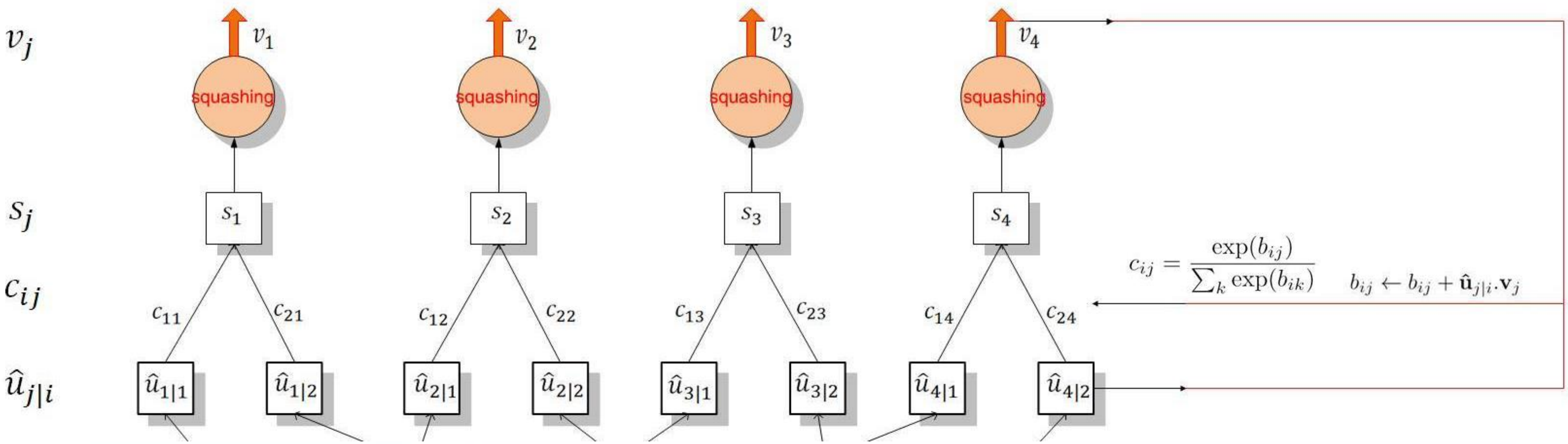
数据挖掘实验室
Data Mining Lab

- Routing by Agreement

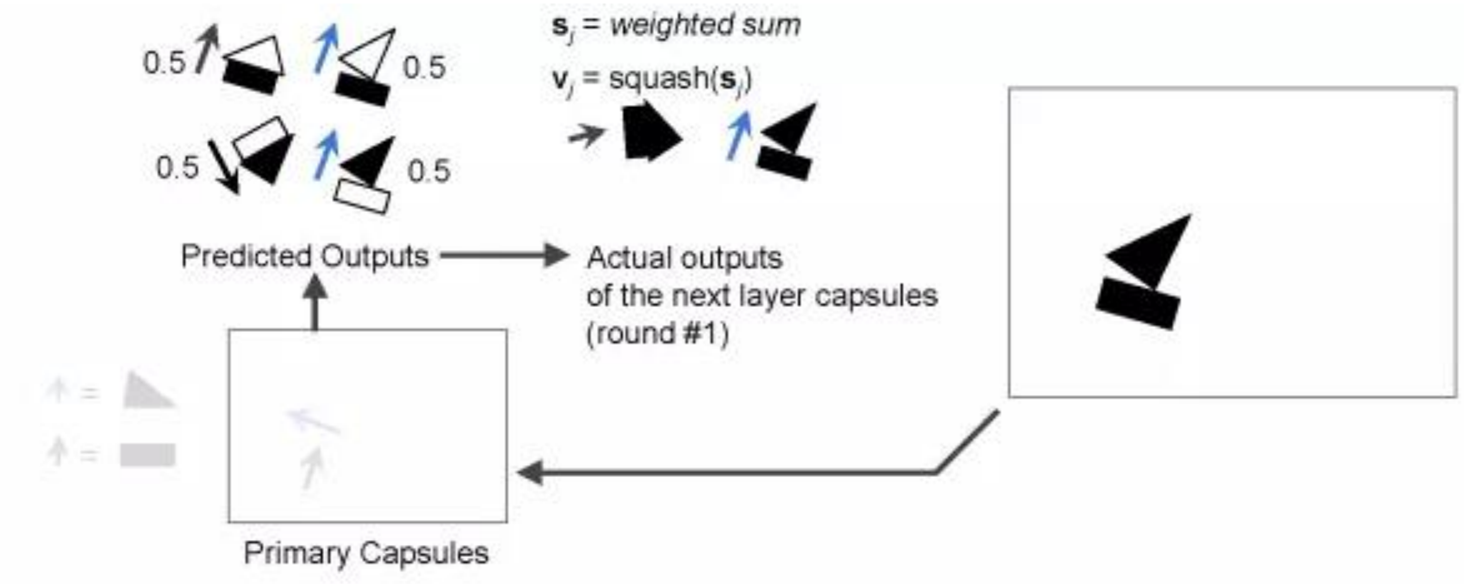


数据挖掘实验室
Data Mining Lab

- Routing Weights



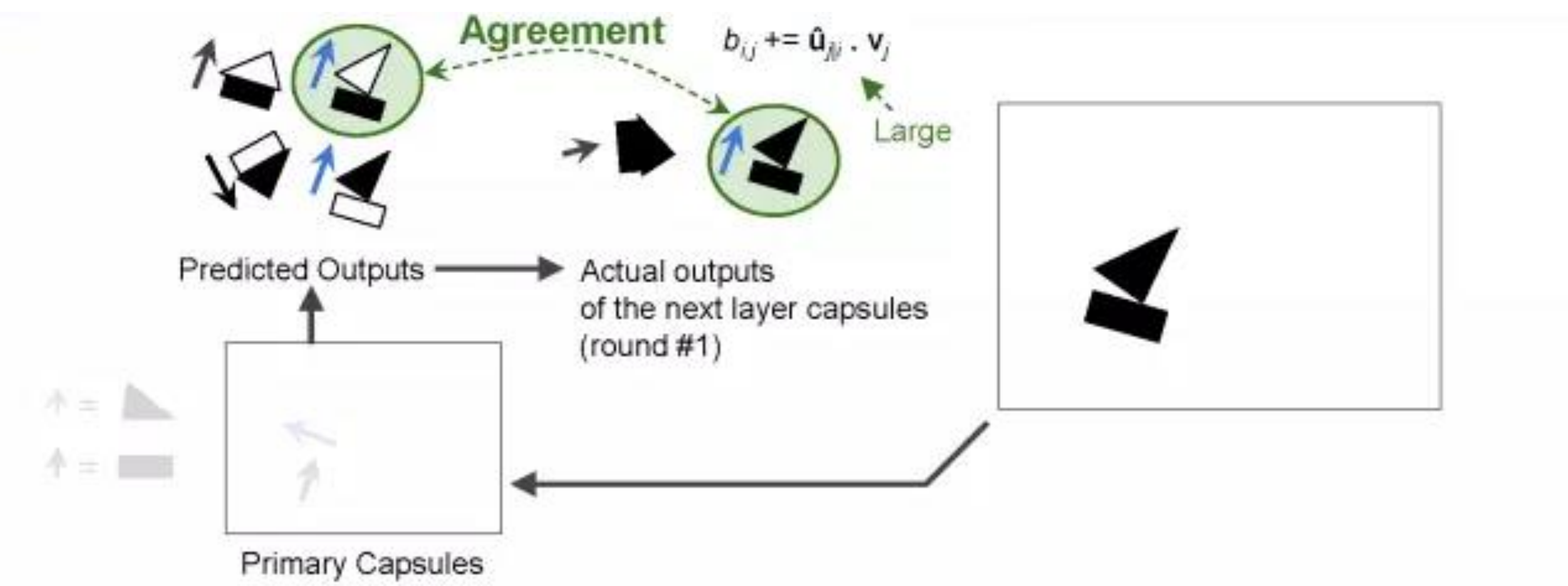
- Compute Next Layer's Output



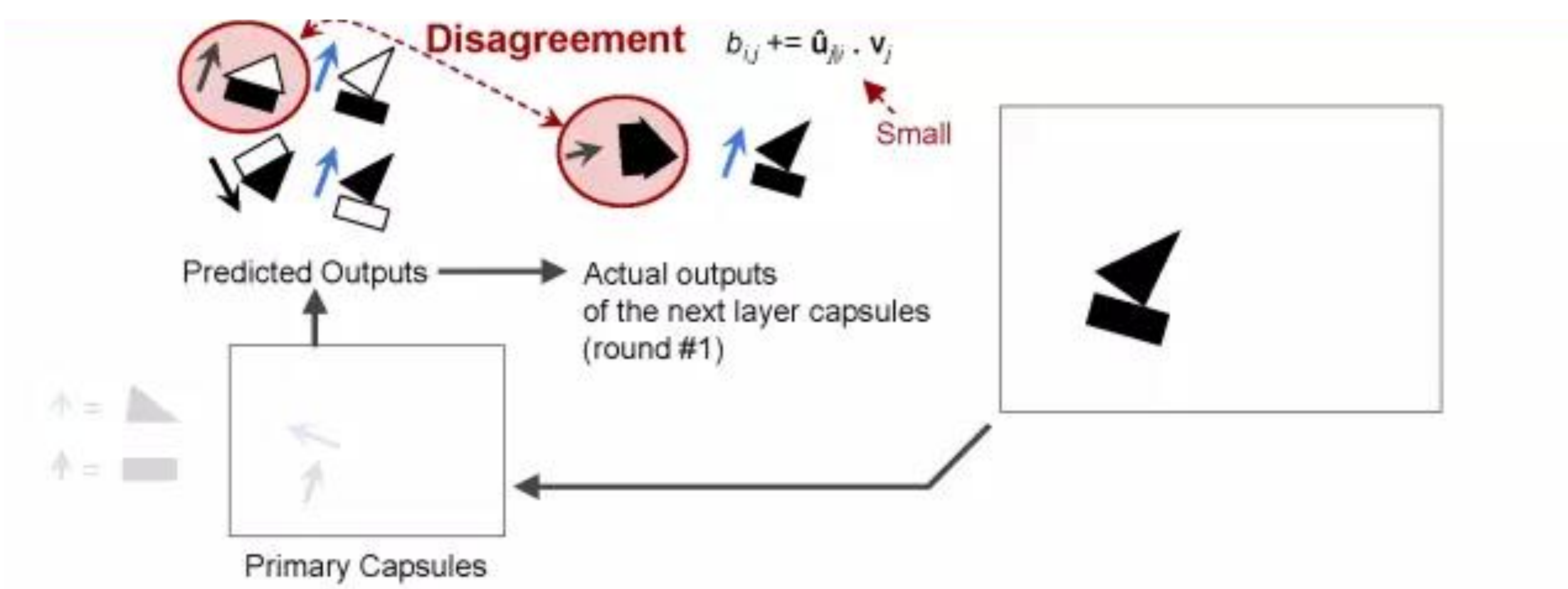
数据挖掘实验室

Data Mining Lab

- Update Routing Weights

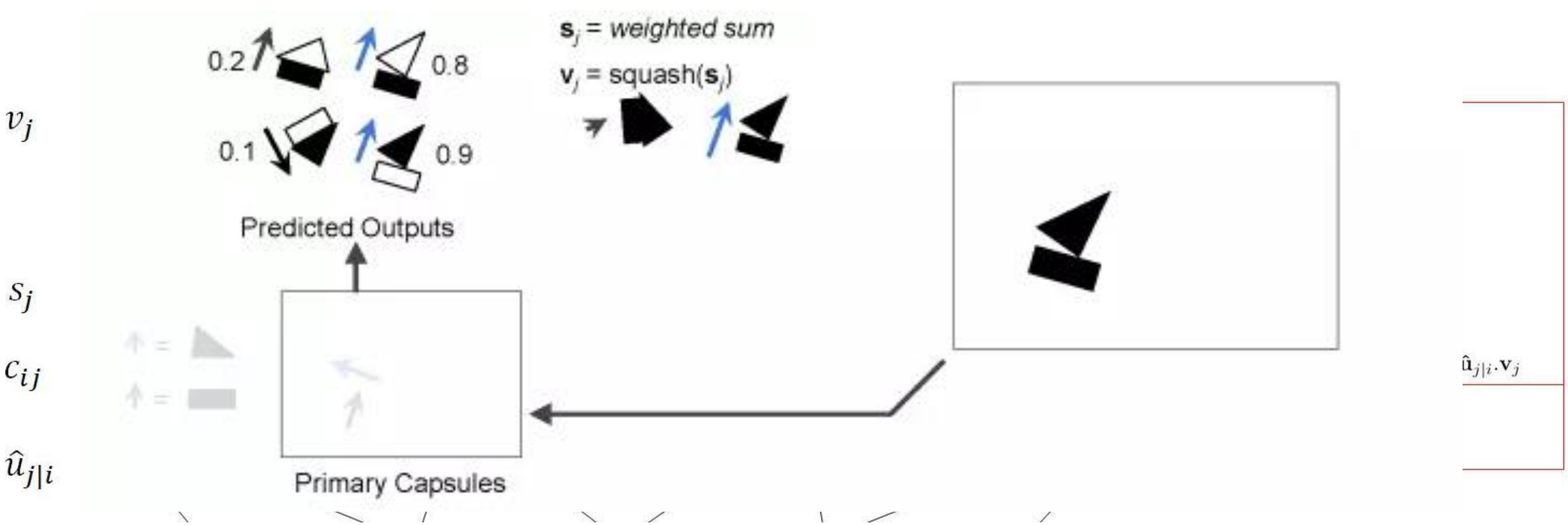


- Update Routing Weights

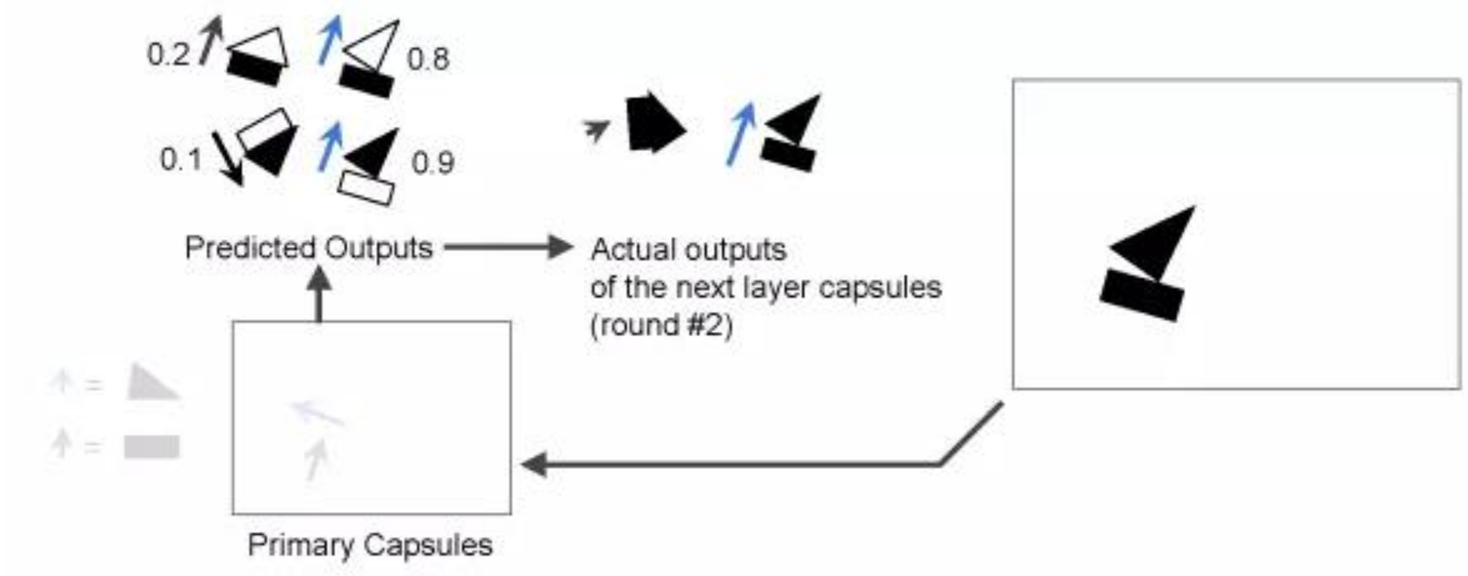


1.2. Demo

- Compute Next Layer's Output



- Compute Next Layer’s Output



1.2. Demo

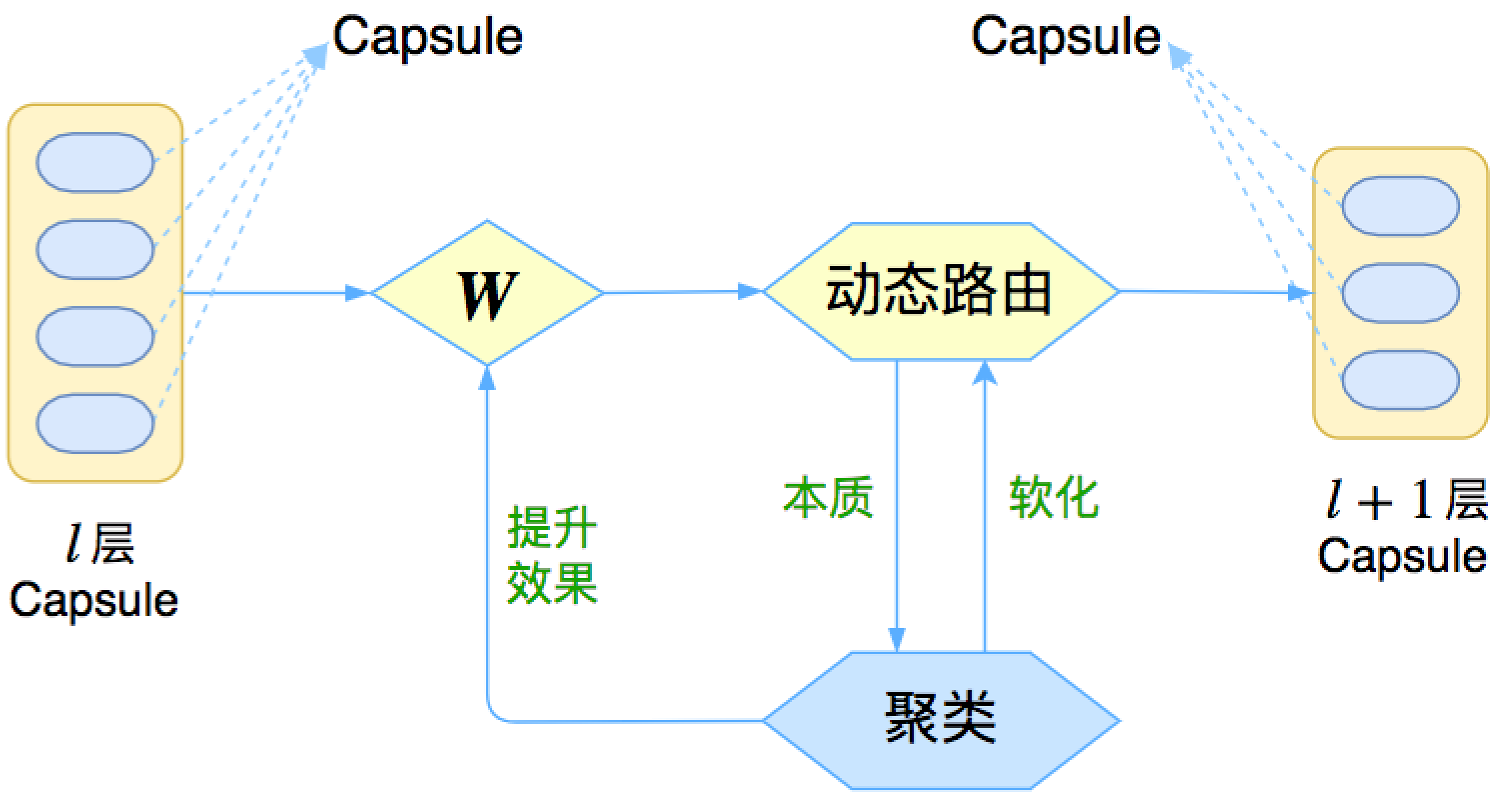
- Routing algorithm

Procedure 1 Routing algorithm.

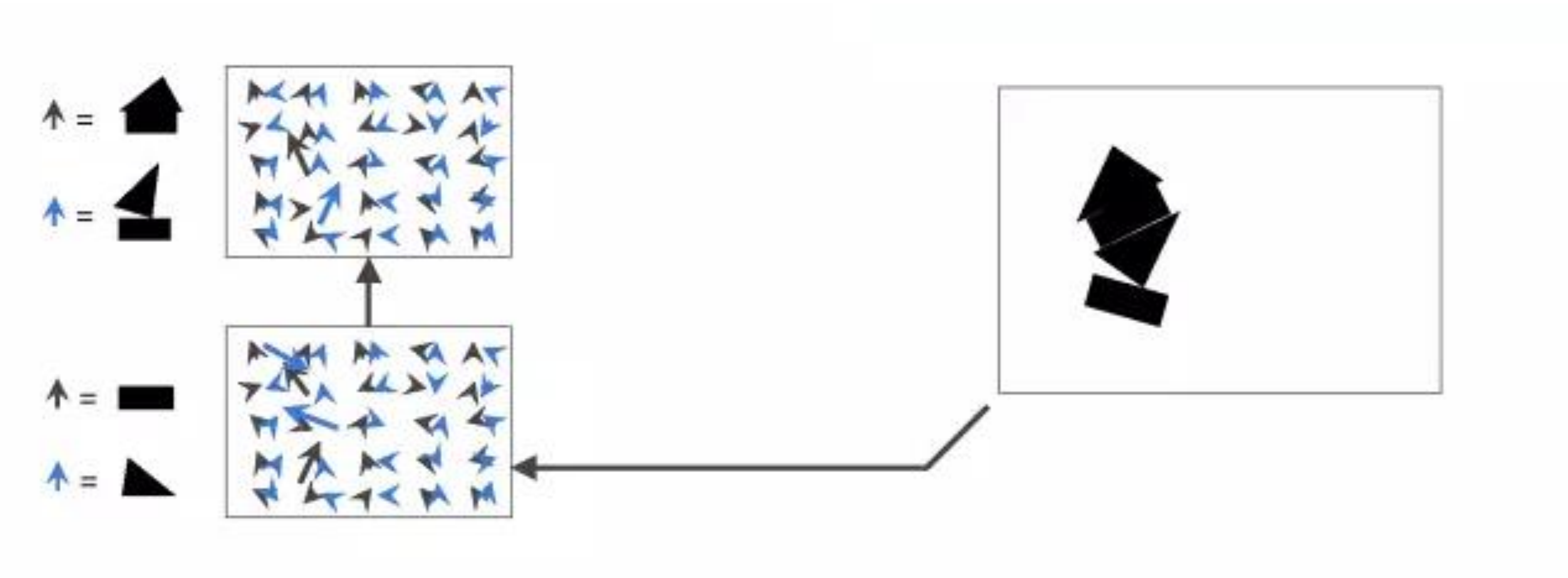
procedure ROUTING($\hat{\boldsymbol{u}}_{j|i}$, r, l)
    for all capsule i in layer l and capsule j in layer $(l+1)$: $b_{ij} \leftarrow 0$.
    for r iterations **do**
        for all capsule i in layer l: $\mathbf{c}_i \leftarrow \texttt{softmax}(\mathbf{b}_i)$ ▷ softmax computes Eq. 3
        for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

数据挖掘实验室
Data Mining Lab

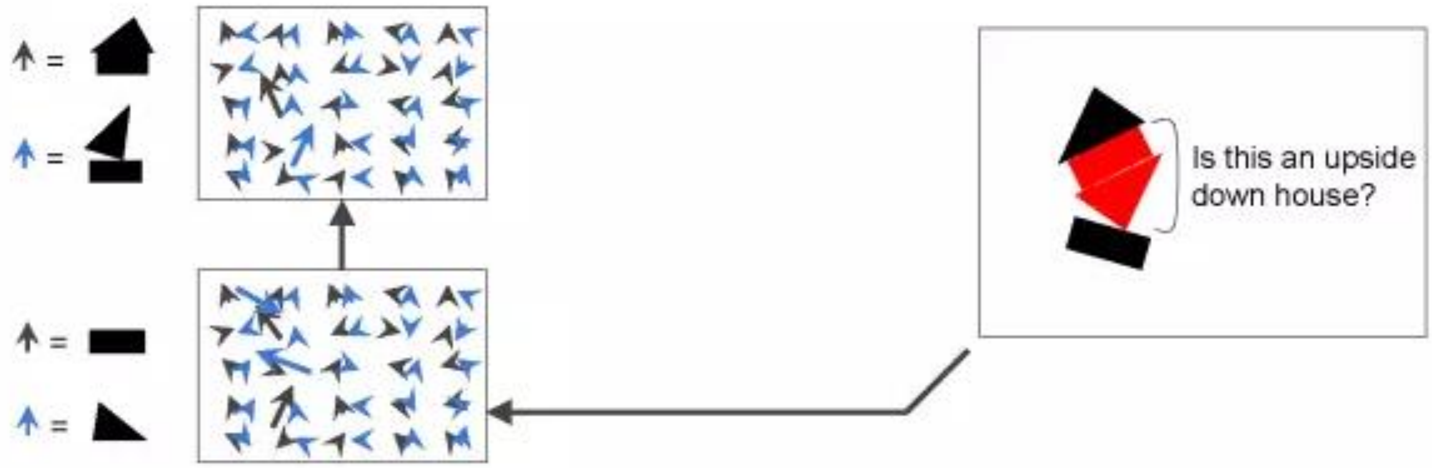
- Routing algorithm



- Handling Crowded Scenes

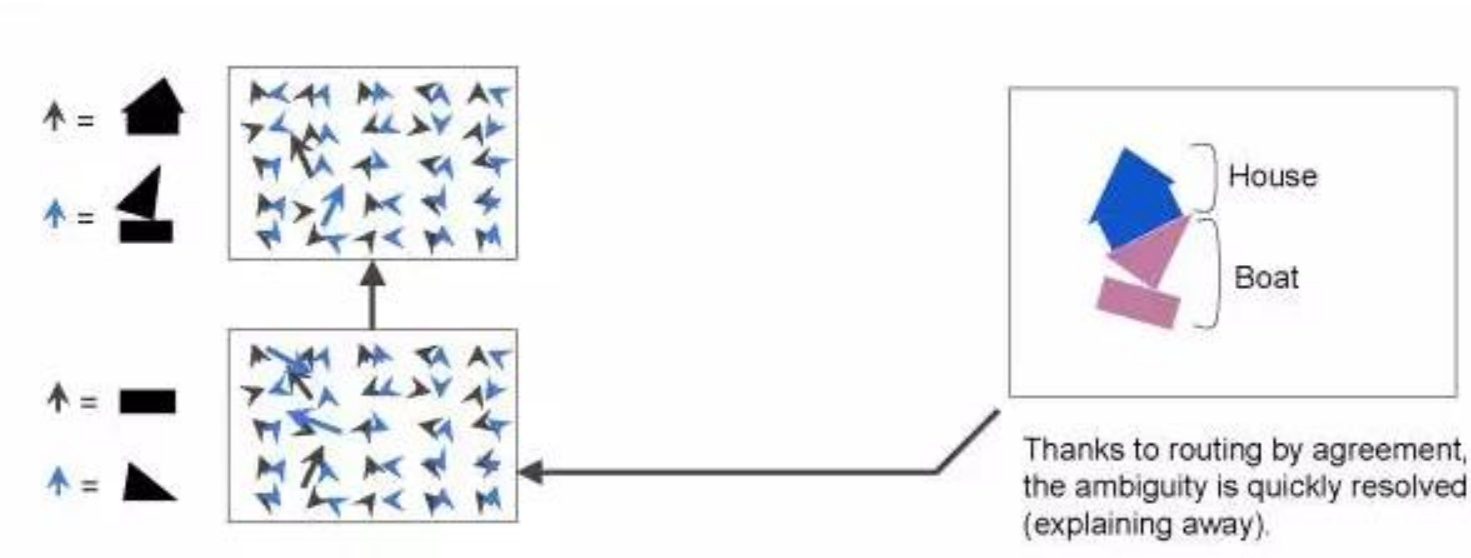


- Handling Crowded Scenes



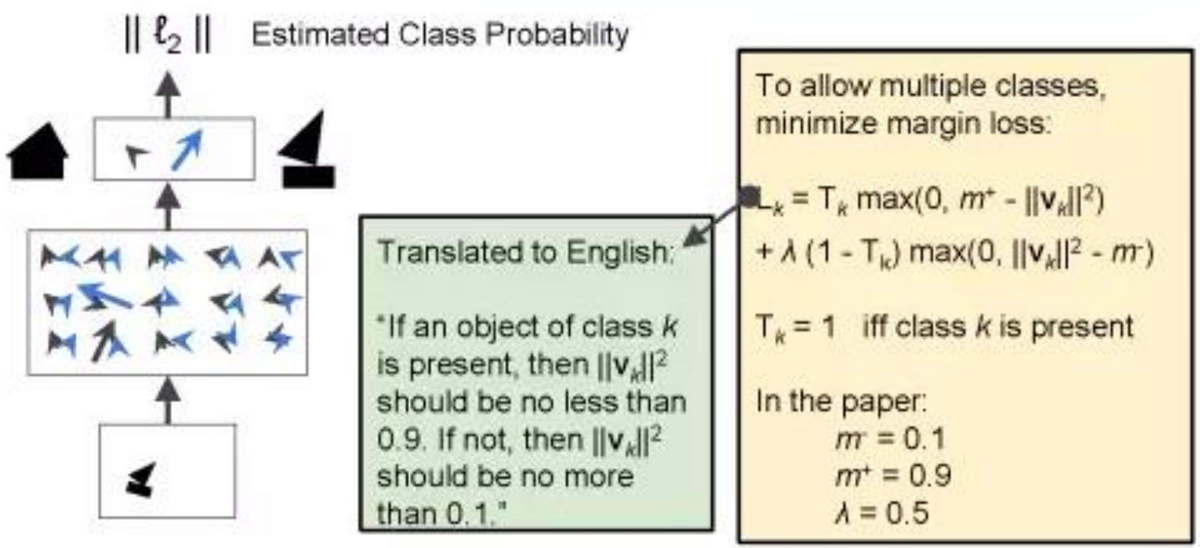
1.2. Demo

- Handling Crowded Scenes



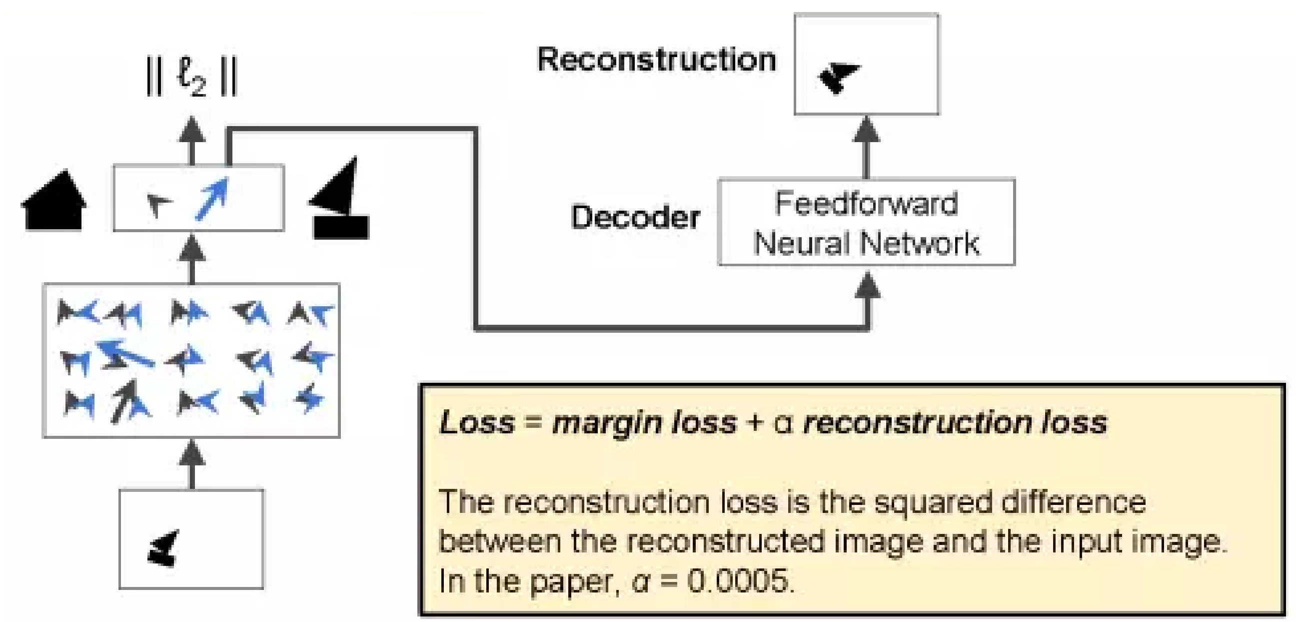
数据挖掘实验室
Data Mining Lab

- Training



1.2. Demo

- Regularization by Reconstruction



- Classification accuracy on MNIST

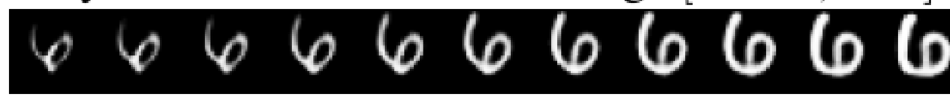
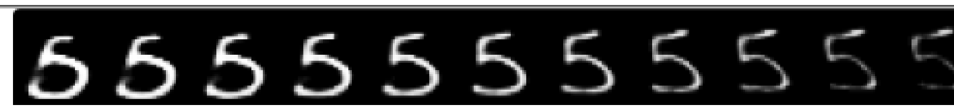
Table 1: CapsNet classification test accuracy. The MNIST average and standard deviation results are reported from 3 trials.

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**

1.3. Experiments

- Reconstruction with perturbations

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 $[-0.25, 0.25]$.

Scale and thickness	
Localized part	
Stroke thickness	
Localized skew	
Width and translation	
Localized part	

1.3. Experiments

- Robustness to Affine Transformations

CNN:	99.22% -> 66%
CapsNet:	99.23% -> 79%

1.3. Experiments

- MultiMNIST result

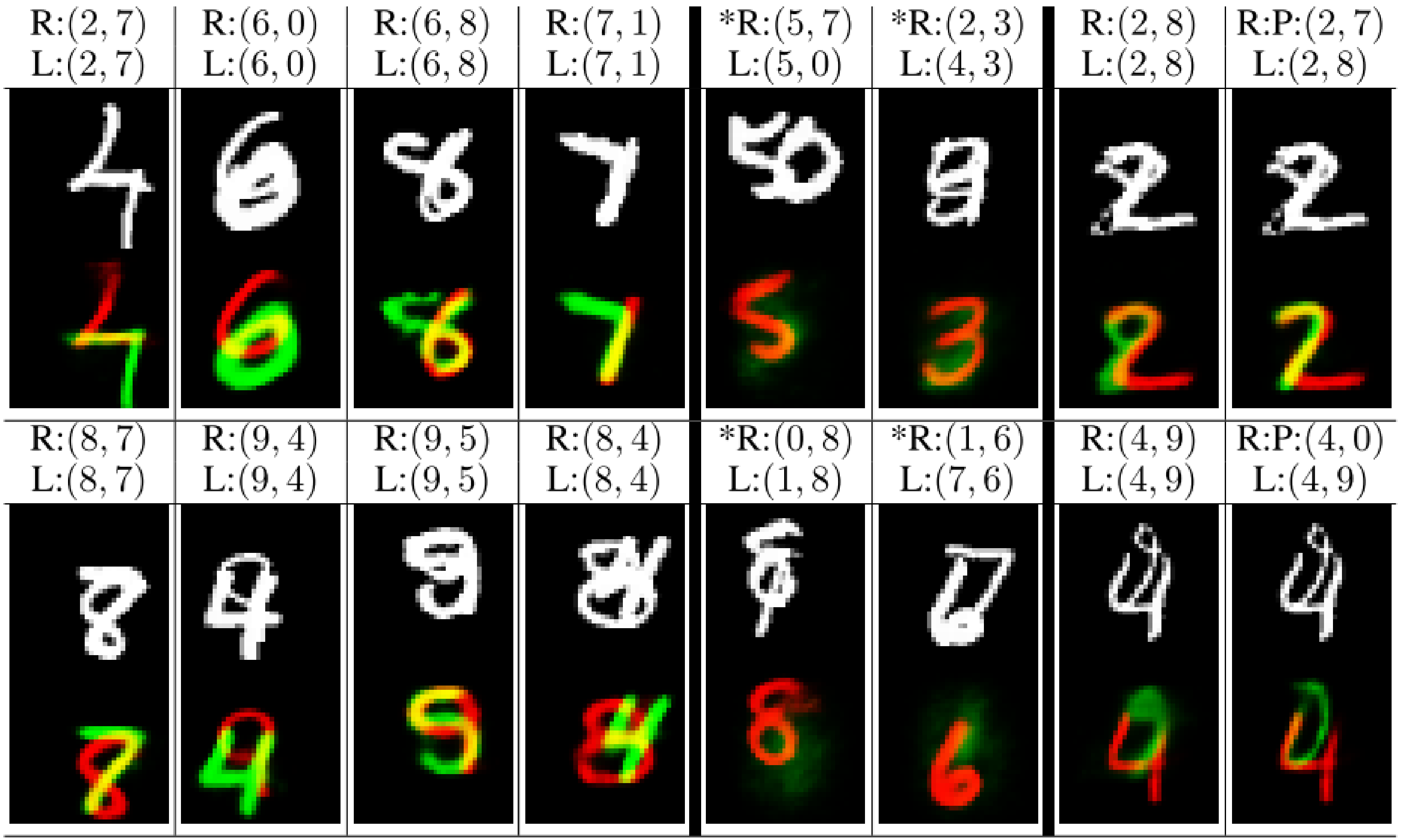
CNN: 5% error rate -> in <4% overlapping

CapsNet: 5% error rate -> in 80% overlapping

1.3. Experiments

数据挖掘实验室
Data Mining Lab

- Sample reconstructions on MultiMNIST



1.4. Conclusion

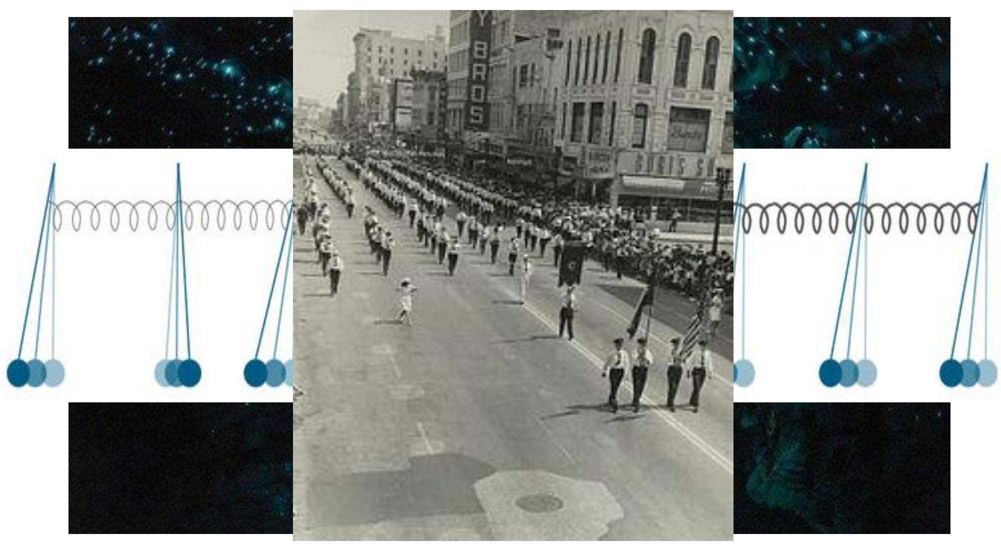
- Pros
  - Preserve position and pose information (equivariance)
  - Require less training data
  - Routing by agreement is great for overlapping objects
  - Robust to affine transformations
  - Interpretability

- Cons
  - Not effective in large datasets
  - Slow training

Discussion on Synchronization Neural Network

2.1. Background

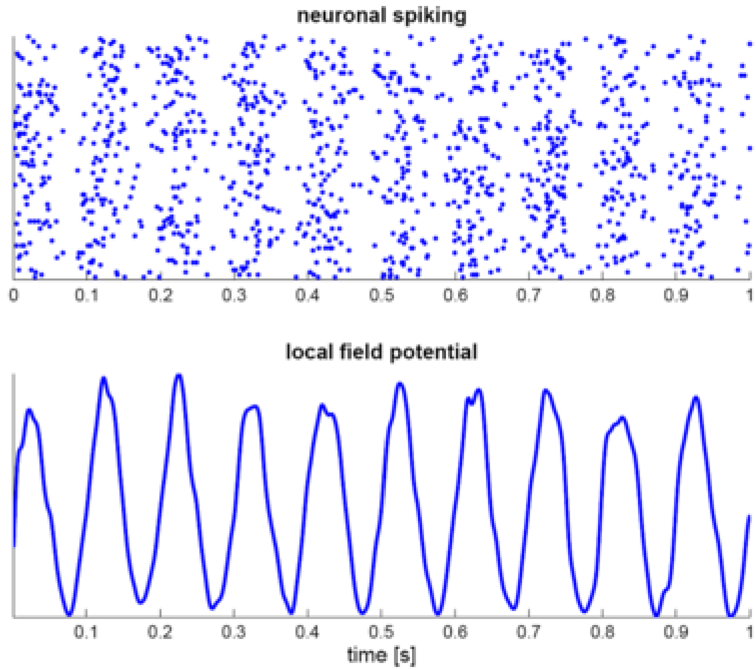
- Oscillations and Amazing Synchronization
  - Animate oscillator
  - Inanimate oscillator
  - Complex system



2.1. Background

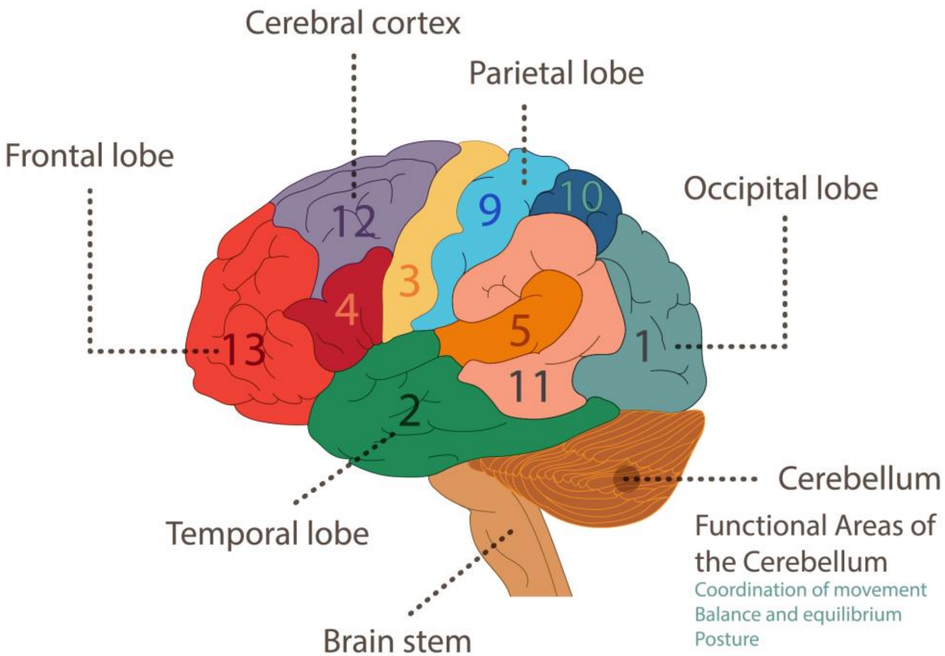
- Neural oscillation / synchronization
  - Spike trains
  - Local field potentials
  - Large-scale oscillations

- Property
  - Frequency
  - Amplitude
  - Phase



2.1. Motivation

- What's wrong on ANN model (comparing to human brain)?
  - Overfitting / Flexible
  - Intricate representation
  - Task-dependent representation
  - Data hungry
  - Catastrophic forgetting



2.1. Motivation

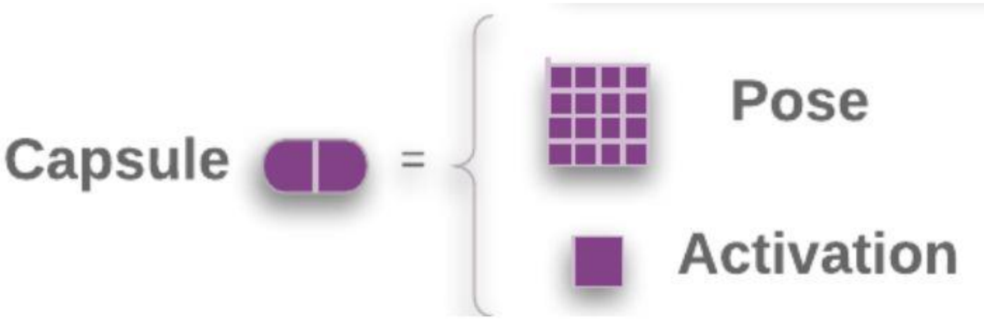
数据挖掘实验室
Data Mining Lab

- Interpretability
- Hierarchy
  - Protocol oriented (multi-source) transfer learning
  - Life-long learning (besides Online learning)
  - Super-parallel (Task hierarchy)
- Disentangled representation
  - Robustness
  - Image segmentation / description
  - ……
- Additive decision (Multi-view)

2.2. Potential ways

数据挖掘实验室
Data Mining Lab

- What to sync?
  - Vector/Matrix representation [1][2]
  - Complex Neural Network [3-6]
  - Amplitude and phase [7]
  - Reusable module [8]



[1] Sabour, S., Frosst, N., & Hinton, G. E. (2017). Dynamic routing between capsules. In *Advances in Neural Information Processing Systems* (pp. 3856-3866).

[2] Hinton, G. E., Sabour, S., & Frosst, N. (2018). Matrix capsules with EM routing.

2.2. Potential ways

数据挖掘实验室
Data Mining Lab

- What to sync?
  - Vector/Matrix representation [1][2]
  - Complex Neural Network [3-6]
  - Amplitude and phase [7]
  - Reusable module [8]

[3] Jose, C., Cisse, M., & Fleuret, F. (2017). Kronecker Recurrent Units. *arXiv preprint arXiv:1705.10142*.

[4] Trabelsi, C., Bilaniuk, O., Zhang, Y., Serdyuk, D., Subramanian, S., Santos, J. F., ... & Pal, C. J. (2017). Deep complex networks. *arXiv preprint arXiv:1705.09792*.

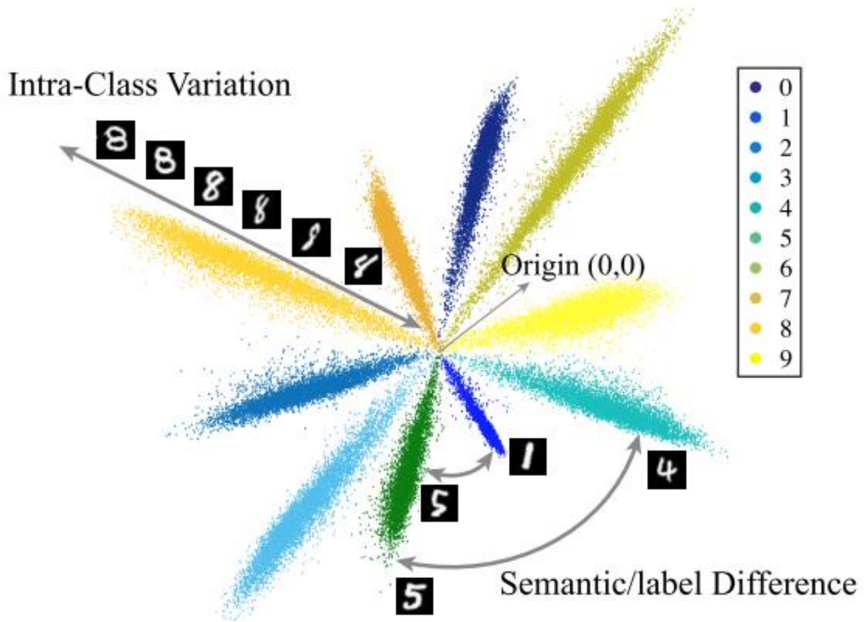
[5] Mescheder, L., Nowozin, S., & Geiger, A. (2017). The numerics of GANs. In *Advances in Neural Information Processing Systems* (pp. 1825-1835).

[6] Rahman, M., & Geiger, D. (2016). Quantum Clustering and Gaussian Mixtures. *arXiv preprint arXiv:1612.09199*.

2.2. Potential ways

数据挖掘实验室
Data Mining Lab

- What to sync?
  - Vector/Matrix representation [1][2]
  - Complex Neural Network [3-6]
  - Amplitude and phase [7]
  - Reusable module [8]

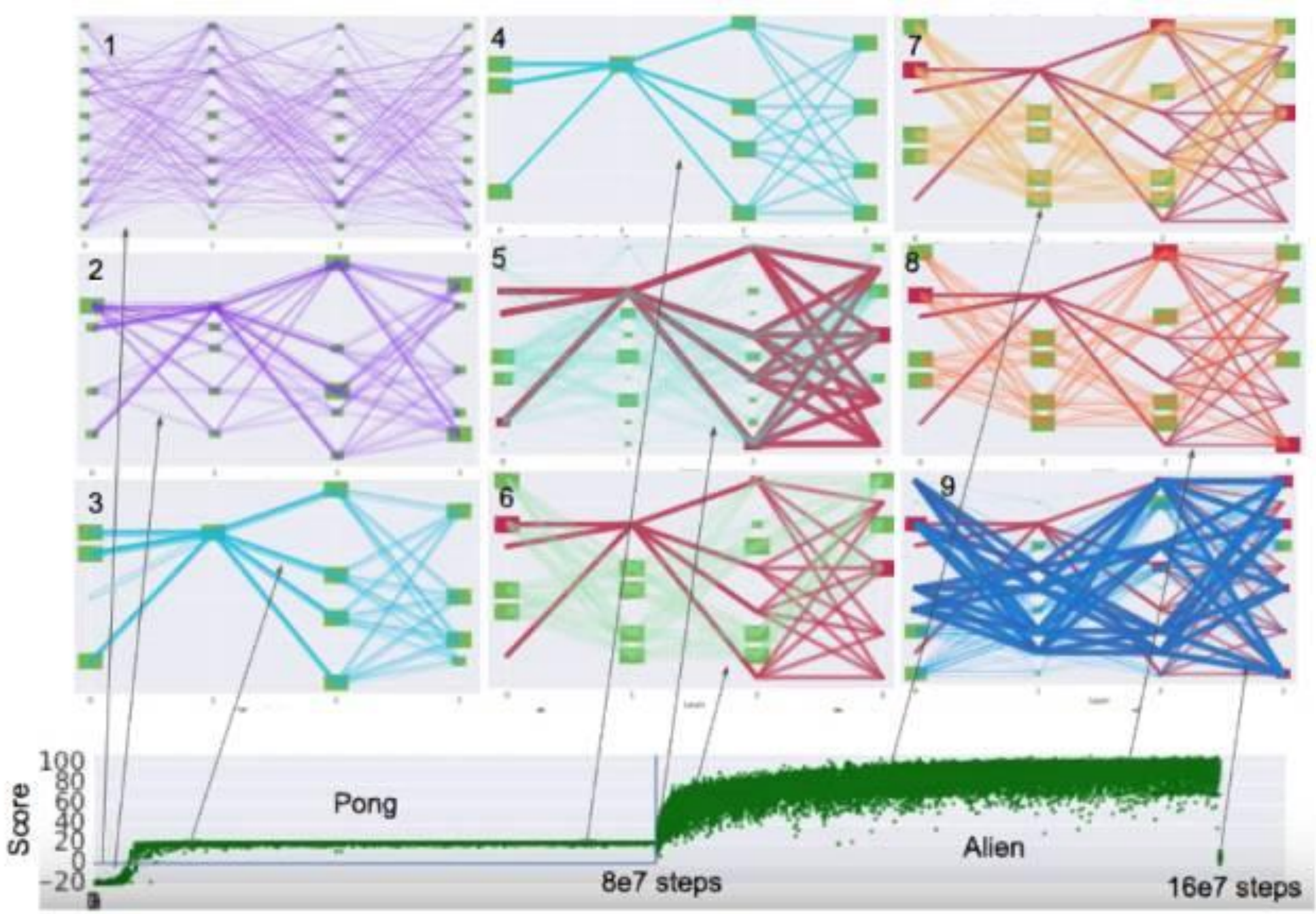


$$f(\boldsymbol{w},\boldsymbol{x}) = \langle \boldsymbol{w},\boldsymbol{x}\rangle = \boldsymbol{w}^\top \boldsymbol{x},$$

$$f(\boldsymbol{w},\boldsymbol{x}) = \|\boldsymbol{w}\|\,\|\boldsymbol{x}\|\, cos(\theta_{(\boldsymbol{w},\boldsymbol{x})}) = h(\|\boldsymbol{w}\|,\|\boldsymbol{x}\|)\cdot g(\theta_{(\boldsymbol{w},\boldsymbol{x})})$$

[7] Liu, W., Liu, Z., Yu, Z., Dai, B., Lin, R., Wang, Y., ... & Song, L. (2018, April). Decoupled Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2771-2779).

2.2. Potential ways

- 

[8] Fernando, C., Banarse, D., Blundell, C., Zwols, Y., Ha, D., Rusu, A. A., ... & Wierstra, D. (2017). Pathnet: Evolution channels gradient descent in super neural networks. *arXiv preprint arXiv:1701.08734*.

Thank you