Self-Routing Capsule Networks

Taeyoung Hahn   Myeongjang Pyeon   Gunhee Kim

Seoul National University

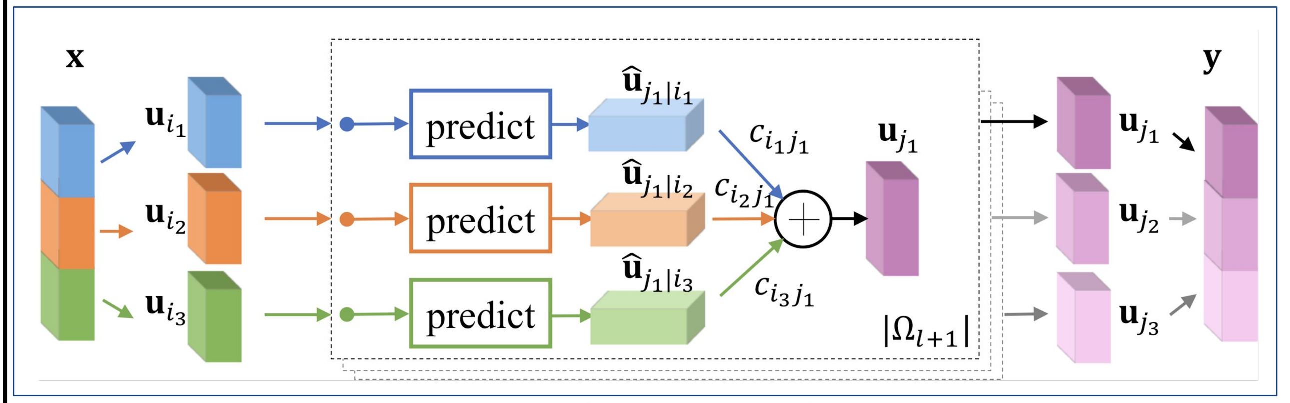
SEOUL NATIONAL UNIV. VISION & LEARNING

Source code is available at http://vision.snu.ac.kr/projects/self-routing

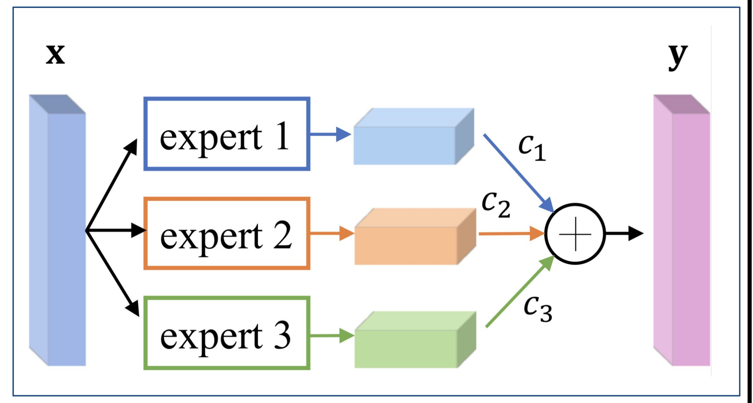
NeurIPS 2019

Capsule Networks

1. Structure that encourages ensemble of weak submodules.



Capsule Networks



Mixture-of-Experts

2. Routing-by-Agreement coordinates connections between capsules in adjacent layers.

Routing-by-Agreement

- Dynamic Routing

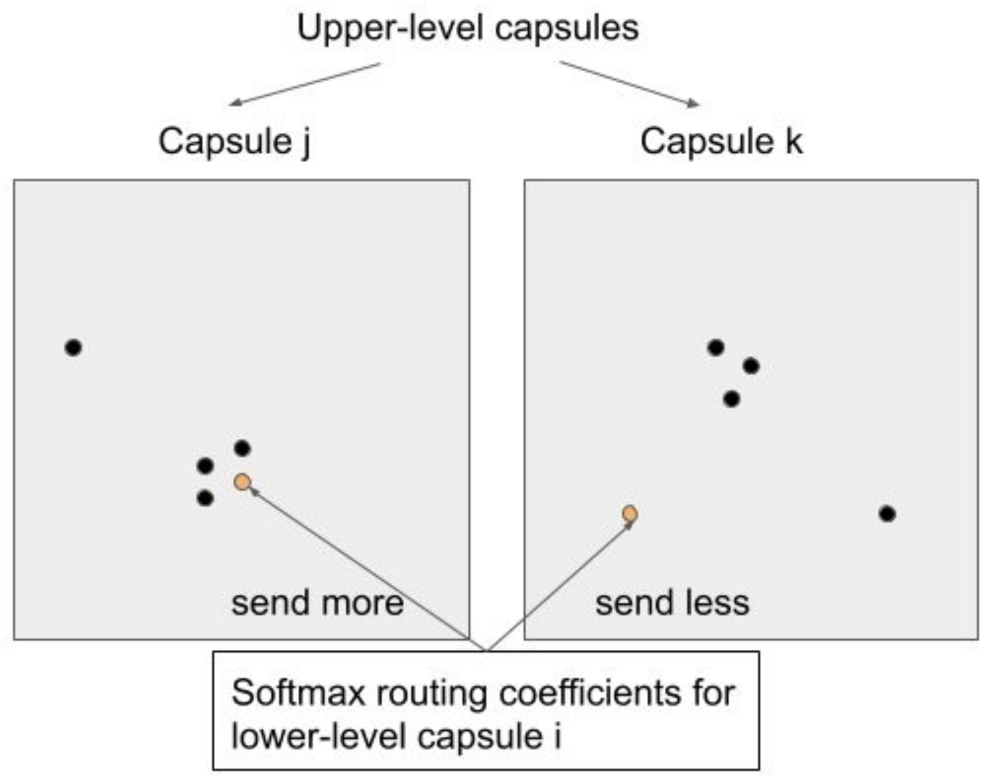
$$b_{ij}^{(t+1)} \leftarrow b_{ij}^{(t)} + \hat{\mathbf{u}}_{j|i} \cdot \mathbf{u}_j^{(t)}$$

$$c_{ij} = \mathrm{softmax}(\mathbf{b}_i)$$

- EM Routing

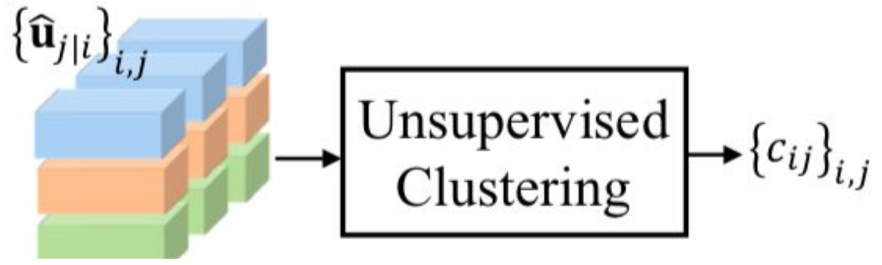
$$a_j^{(t)}, \boldsymbol{\mu}_j^{(t)}, \boldsymbol{\sigma}_j^{(t)} \leftarrow \mathtt{M-step}(a_i, c_{ij}^{(t)}, \hat{\mathbf{u}}_{j|i})$$

$$c_{ij}^{(t+1)} \leftarrow \mathtt{E-step}(a_j^{(t)}, \boldsymbol{\mu}_j^{(t)}, \boldsymbol{\sigma}_j^{(t)})$$



Motivation

- The unsupervised clustering part in Routing-by-Agreement brings two inherent weaknesses
  - High computational cost
  - Cluster assumption that may not hold when input noise is present.



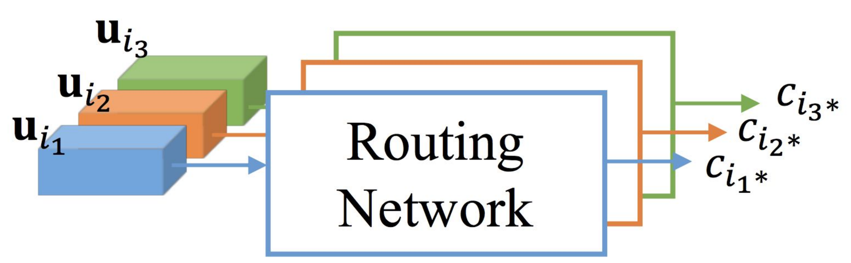
- But is Routing-by-Agreement really necessary?
  - Would there be simpler alternative that works as effectively?

Self-Routing

- A routing method without notion of agreement.
- Each capsule relies on its subordinate routing network.

$$c_{ij} = \mathrm{softmax}(\mathbf{W}_i^{route}\mathbf{u}_i)_j, \quad a_j = \frac{\sum_{i\in\Omega_l} c_{ij}a_i}{\sum_{i\in\Omega_l} a_i}.$$

$$\hat{\mathbf{u}}_{j|i} = \mathbf{W}_{ij}^{pose}\mathbf{u}_i, \quad \mathbf{u}_j = \frac{\sum_{i\in\Omega_l} c_{ij}a_i\hat{\mathbf{u}}_{j|i}}{\sum_{i\in\Omega_l} c_{ij}a_i}.$$



- Our method is more similar to Mixture-of-Experts (MoE).

Experiment Settings

- CIFAR-10 and SVHN
  - Append capsule layers to ResNet-20 after Conv layers
  - Test robustness against adversarial attacks
- SmallNORB
  - Append capsule layers to 7-layer CNN after Conv layers
  - Test robustness against viewpoint changes.

Train

Test

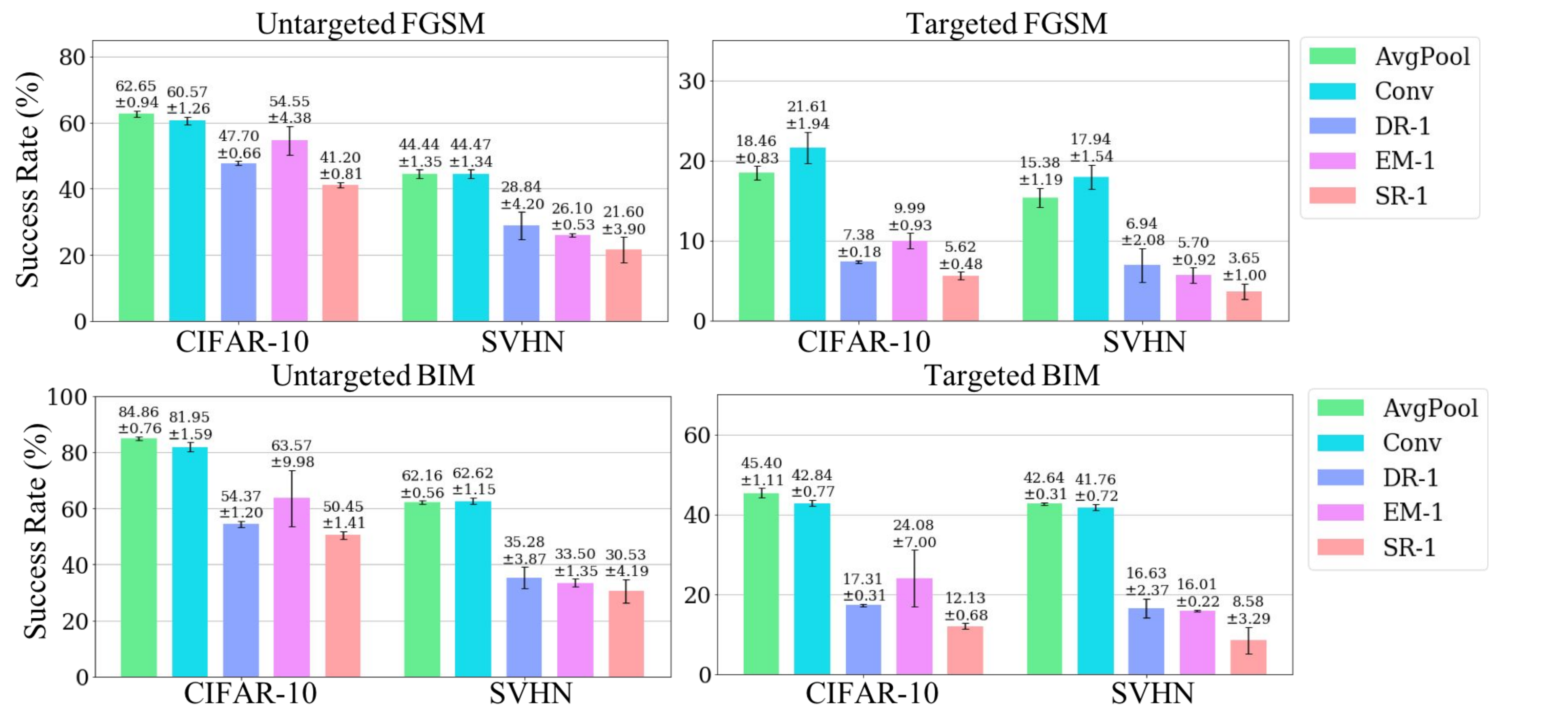
Computational Efficiency & Classification Error

Methods	# Param. (M)	# FLOPs (M)	CIFAR-10	SVHN
AvgPool	0.3	41.3	7.94±0.21	3.55 ±0.11
Conv	0.9	61.0	10.01±0.99	3.98 ±0.15
DR-1	5.8	73.5	8.46±0.27	3.49 ±0.69
DR-2	4.2	232.1	**7.86**±0.21	3.17 ±0.09
EM-1	0.9	76.6	10.25±0.45	3.85 ±0.13
EM-2	0.8	173.8	12.52±0.32	3.70 ±0.35
SR-1	0.9	62.2	8.17±0.18	3.34 ±0.08
SR-2	3.2	140.3	**7.86**±0.12	**3.12** ±0.13

- Self-Routing is computationally efficient as clustering part is removed.

Results on Adversarial Attacks

- All CapsNet variants are more robust than CNN baselines.
- Among CapsNets, Self-Routing is most robust.



Results on Viewpoint Generalization

- Self-Routing also shows better viewpoint generalization.

Methods	Azimuth		Elevation	
	Familiar	Novel	Familiar	Novel
AvgPool	8.49±0.45	21.76±1.18	5.68±0.72	17.72±0.30
Conv	8.39±0.56	22.07±1.02	7.51±1.09	18.78±0.67
DR-1	6.86±0.50	20.33±1.32	5.78±0.48	16.37±0.90
EM-1	7.36±0.89	20.16±0.96	5.97±0.98	17.51±1.52
SR-1	7.62±0.95	**19.86**±1.03	5.96±0.46	**15.91**±1.09

Scalability comparison

- Self-routing benefits from adding more capsules to layer.