Supplementary Material for "A Receptor Skeleton for Capsule Neural Networks"

Jintai Chen [1] **Hongyun Yu** [1] **Chengde Qian** [2] **Danny Z. Chen** [3] **Jian Wu** [4]

1. Model Specifications

To fairly compare our approach and various known routing algorithms, we unify the main structures of the CapsNets in the classification experiments. Since the CapsNets with Inverted Dot-Product Attention Routing (IDPA-Routing) are slightly different from the CapsNets with other routing algorithms and our approach, we separately show the model specifications. The model specifications of CapsNets with the simple backbone are provided in Table 1 and Table 2, and the CapsNets with the ResNet-18 backbone are provided in Table 3 and Table 4. Similar to the CapsNets with IDPA-Routing (Tsai et al., 2020), in the CapsNets with our approach, we use the convolution-based transformation function (Conv Transformation) to deal with the representations of matrix capsules and use the fully connection based transformation function (Linear Transformation) to deal with the representations of vector capsules. We use the original transformation functions in the CapsNets with the other routing algorithms[1]. In all the tables, "S", "SN", "AN", "F-M", "M-M", "C10", and "C100" denote "SVHN", "SmallNorb", "AffNIST", "Fashion-MNIST", "Multi-MNIST", "CIFAR-10", and "CIFAR-100", respectively. The "Routing" in Table 1 (rows (4)—(6)) and Table 3 (rows (6)—(8)) indicates the corresponding routing algorithms or our approach[2]. We report the shapes of the representations that are output by the convolutional layers following "channel size * width * height". The representation shapes of the matrix capsules are reported following "capsule number (capsule type) * capsule size * width * height".

[1]College of Computer Science and Technology, Zhejiang University, Hangzhou, China; [2]School of Statistics and Data Science, Nankai University, China; [3]Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN 46556, USA; [4]The First Affiliated Hospital, and Department of Public Health, Zhejiang University School of Medicine, Hangzhou, China;. Correspondence to: Jian Wu <wujian2000@zju.edu.cn>.

Proceedings of the 38^{th} International Conference on Machine Learning, PMLR 139, 2021.

[1]The transformation function is the same as described in Sec. 2.1 of the main paper.

[2]As mentioned in the main paper, our approach plays the same role as the known routing algorithms.

References

Tsai, Y.-H. H., Srivastava, N., Goh, H., and Salakhutdinov, R. Capsules with inverted dot-product attention routing. In *ICLR*, 2020.

Table 1. The model specifications of the CapsNets using the simple backbone with various routing algorithms or our approach, except for the Inverted Dot-Product Attention Routing. "BN" and "LN" denote "Batch Normalization" and "Layer Normalization", respectively.

Operations	Output Shapes of Representations			
	S & M-M & SN & C10	C100	F-M	AN
(1) 3 * 3 Conv, stride=2, pad=1, BN, RELU	128 * 16 * 16	256 * 16 * 16	128 * 14 * 14	128 * 20 * 20
(2) 3 * 3 Conv, stride=2, pad=1, BN, RELU	640 * 8 * 8	1440 * 8 * 8	640 * 7 * 7	640 * 10 * 10
(3) Capsule reshape	40 * (4 * 4) * 8 * 8	40 * (6 * 6) * 8 * 8	40 * (4 * 4) * 7 * 7	40 * (4 * 4) * 10 * 10
(4) Transformation, Routing, Squashing or LN	40 * (4 * 4) * 8 * 8	40 * (6 * 6) * 8 * 8	40 * (4 * 4) * 7 * 7	40 * (4 * 4) * 10 * 10
(5) Transformation, Routing, Squashing or LN	20 * (4 * 4) * 4 * 4	20 * (6 * 6) * 4 * 4	20 * (4 * 4) * 4 * 4	20 * (4 * 4) * 5 * 5
(6) Transformation, Routing, Squashing or LN	10 * (4 * 4) * 2 * 2	10 * (6 * 6) * 2 * 2	10 * (4 * 4) * 2 * 2	10 * (4 * 4) * 2 * 2
(7) Reshape	10 * 64	10 * 144	10 * 64	10 * 64
(8) Transformation, Routing, Normalization	10 (or 5) * 16	100 * 36	10 * 16	10 * 16

Table 2. The model specifications of the CapsNet using the simple backbone with Inverted Dot-Product Attention Routing (IDPA-Routing). "BN" and "LN" denote "Batch Normalization" and "Layer Normalization", respectively.

Operations	Output Shapes of Representations			
	S & M-M & SN & C10	C100	F-M	AN
(1) 3 * 3 Conv, stride=1, pad=1, BN, RELU	128 * 32 * 32	256 * 32 * 32	128 * 28 * 28	128 * 40 * 40
(2) 3 * 3 Conv, stride=2, pad=1, BN, RELU	640 * 16 * 16	1440 * 16 * 16	640 * 14 * 14	640 * 20 * 20
(3) Capsule reshape	40 * (4 * 4) * 16 * 16	40 * (6 * 6) * 16 * 16	40 * (4 * 4) * 16 * 16	40 * (4 * 4) * 20 * 20
(4) 3 * 3 Conv Transformation, IDPA-Routing, LN	40 * (4 * 4) * 7 * 7	40 * (6 * 6) * 7 * 7	40 * (4 * 4) * 7 * 7	40 * (4 * 4) * 10 * 10
(5) 3 * 3 Conv Transformation, IDPA-Routing, LN	20 * (4 * 4) * 5 * 5	40 * (6 * 6) * 5 * 5	20 * (4 * 4) * 5 * 5	20 * (4 * 4) * 5 * 5
(6) Flatten	500 * (4 * 4)	1000 * (6 * 6)	500 * (4 * 4)	500 * (4 * 4)
(7) Linear Transformation, IDPA-Routing, LN	10 * 16	20 * 36	10 * 16	10 * 16
(8) Linear Transformation, IDPA-Routing, LN	10 (or 5) * 16	100 * 36	10 * 16	10 * 16

Table 3. The model specifications of the CapsNets using the ResNet-18 backbone with various routing algorithms or our approach, except for the Inverted Dot-Product Attention Routing. "BN" and "LN" denote "Batch Normalization" and "Layer Normalization", respectively.

Operations	Output Shapes of Representations	
	C10	C100
(1) 3 * 3 Conv, stride=1, pad=1, BN, RELU	64 * 32 * 32	64 * 32 * 32
(2) Layer 1 in the ResNet-18 (stride=2)	64 * 16 * 16	64 * 16 * 16
(3) Layer 2 in the ResNet-18 (stride=2)	128 * 8 * 8	128 * 8 * 8
(4) 1 * 1 Conv, stride=1, pad=1	640 * 8 * 8	1440 * 8 * 8
(5) Capsule reshape	40 * (4 * 4) * 8 * 8	40 * (6 * 6) * 8 * 8
(6) Transformation, Routing, Squashing or LN	40 * (4 * 4) * 8 * 8	40 * (6 * 6) * 8 * 8
(7) Transformation, Routing, Squashing or LN	20 * (4 * 4) * 4 * 4	20 * (6 * 6) * 4 * 4
(8) Transformation, Routing, Squashing or LN	10 * (4 * 4) * 2 * 2	10 * (6 * 6) * 2 * 2
(9) Reshape	10 * 64	10 * 144
(10) Transformation, Mean, Normalization	10 * 16	100 * 36

Table 4. The model specifications of the CapsNet using the ResNet-18 backbone with Inverted Dot-Product Attention Routing (IDPA-Routing). “BN” and “LN” denote “Batch Normalization” and “Layer Normalization”, respectively.

Operations	Output Shapes of the Representations	
	C10	C100
(1) 3 * 3 Conv, stride=1, pad=1, BN, RELU	64 * 32 * 32	64 * 32 * 32
(2) Layer 1 in the ResNet-18 (stride=1)	64 * 32 * 32	64 * 32 * 32
(3) Layer 2 in the ResNet-18 (stride=2)	128 * 16 * 16	128 * 16 * 16
(4) 1 * 1 Conv, stride=1, pad=1	640 * 16 * 16	1440 * 16 * 16
(5) Capsule reshape	40 * (4 * 4) * 16 * 16	40 * (6 * 6) * 16 * 16
(6) 3 * 3 Conv Transformation, IDPA-Routing, LN	40 * (4 * 4) * 7 * 7	40 * (6 * 6) * 7 * 7
(7) 3 * 3 Conv Transformation, IDPA-Routing, LN	20 * (4 * 4) * 5 * 5	40 * (6 * 6) * 5 * 5
(8) Flatten	500 * 4 * 4	1000 * 6 * 6
(9) Linear Transformation, IDPA-Routing, LN	10 * 16	20 * 36
(10) Linear Transformation, IDPA-Routing, LN	10 * 16	100 * 36