Non-Iterative Cluster Routing

Subjects: Computer Science, Artificial Intelligence

Contributor: Huong Pham , Samuel Cheng

In conventional routing, a capsule network employs routing algorithms for bidirectional information flow between layers through iterative processes.

data-dependent routing   capsnet   capsule network

1. Introduction

A Capsule Network, often referred to as CapsNet, is an advanced type of neural network that employs neuron clusters known as "capsules". Unlike traditional Convolutional Neural Networks (CNNs), which output scalar values, these capsules produce vector outputs. These vectors represent not only the probability of a feature's existence but also its instantiation parameters, such as pose (position, size, orientation), deformation, texture, and so on. This richer representation allows the network to capture and maintain spatial relationships between features more effectively than traditional methods. CapsNet introduces a unique mechanism, known as "routing-by-agreement", which replaces the pooling layers found in conventional CNNs. This routing process enables capsules at one level to send their outputs to higher-level capsules only if there is a strong agreement (i.e., high probability) that the higher-level capsule's entity is present in the input. This agreement is determined according to the dot product between the output of a lower-level capsule and the predicted output of a higher-level capsule, iteratively refined through a routing process. This architecture ensures that, during forward propagation, information flows through the network in a way that preserves spatial relationships, making it inherently more capable of handling variations in viewpoint, scale, and rotation without the need for extensive data augmentation [1].

Capsule networks aim to address some fundamental limitations of CNNs, especially in terms of preserving spatial hierarchies between features within an image. As CNNs rely on the scalar output of neurons within layers for feature detection and representation, they sometimes fail to recognize objects captured from different viewpoints if they are not covered in the training data [2][3]. CapsNets, in contrast, can better preserve the pose information (position and orientation) of features, thus making them more robust to variations in the input data [1][4][5].

There are several main distinctions between CapsNets and CNNs. First, CapsNets use a vector or group of neurons as a basic unit, whereas CNNs use a single neuron. These vectors, called capsules, can potentially represent different parts of an object. Second, while CNNs capture hierarchical features through the depth of the network, with each layer learning different levels of abstract and complex features, CapsNets model hierarchical relationships between parts and whole objects, providing more interpretable representations of learned features. Third, unlike CNNs—which lack a dedicated mechanism for routing information between layers—CapsNets use

data-dependent routing to determine the flow of information between capsules, allowing for better modeling of part–whole relationships.

In classic routing procedures, a CapsNet begins with a set of primary capsules that represent low-level features extracted from the input data. Each primary capsule makes predictions (or votes) by computing an affine transformation of its output and sending their votes to the capsules of the next layer. These capsules at the higher level compute a weighted sum of the predictions received from capsules at the lower level. Routing weights are normalized through layer normalization [6] or a squashing function [1]. An iterative routing process is used to determine an agreement on how capsules at one level should connect to capsules in the next level by updating the weights [1][4]. In contrast, a non-iterative routing procedure computes the routing weights and information of capsules only once [5][7]. By simplifying the process into a single forward pass, non-iterative routing methods alleviate the computational load associated with iteration.

Unlike conventional routing methods [1][4], the cluster routing paradigm involves capsules generating vote clusters (instead of individual votes) for the subsequent layer's capsules [7]. Each vote cluster consists of multiple votes, with each vote potentially originating from a distinct capsule in the previous layer. The proximity of votes within a cluster signifies the information extracted from the same part of an object from capsules in previous layers. Consequently, the variance within a vote cluster can serve as an indicator of the confidence in the encoded information that the vote clusters represent. This suggests that vote clusters with lower variance are more reliable in encoding information related to a specific part of an object. Thus, greater weights are assigned to centroids originating from clusters with lower variance.

2. Agreement Routing in Capsule Network

CapsNets play a crucial role in feature encoding. They transform the extracted features into capsules, which are vectors capable of characterizing parts of an object, unlike the singular units used in conventional CNNs. These capsules possess the ability to represent more intricate aspects of an object, such as its pose, orientation, and texture style. The capsule system proposed in the dynamic routing paper titled "Dynamic Routing Between Capsules" [1] employs a routing-by-agreement mechanism. This mechanism iteratively determines its predictions based on the agreement between lower- and higher-level capsules, which is achieved through adjusting the weights connecting them. These endeavors have led to diverse approaches and innovations in the field. The matrix capsules with EM routing approach [4] uses high-dimensional coincidence filtering and a fast iterative process to determine the routing weights, resulting in better performance and higher robustness to adversarial attacks than baseline CNNs. RS-CapsNet [8] integrates *Res2Net* and *Squeeze-and-Excitation* blocks to extract multi-scale features, emphasizes useful information, and employs linear combinations between capsules in order to enhance object representation while reducing the capsule count. Cluster routing [7] uses variance as the fundamental indicator of agreement and confidence in the information encoded within the vote cluster. Group normalization [9] also utilizes the mean and standard deviation of a group. It splits the output channels of a convolutional layer into several groups and normalizes the features within each group according to the mean and standard deviation.

CapsNets can outperform conventional CNNs in various applications, utilizing fewer parameters [10][11][12]. This advantage is particularly notable in medical image processing, where the main obstacles include detecting small lesions and overcoming class imbalances. Unlike CNNs, which often require substantial amounts of labeled data (that may be scarce in medical contexts), CapsNets can achieve similar levels of performance with a smaller data set [13][14]. CapsNets employ capsules that encapsulate richer feature vectors, unlike CNNs that use scalar neurons, making them more effective in addressing these challenges [13]. Furthermore, CapsNets are superior in maintaining part–whole relationships and geometric details, significantly improving their performance in medical segmentation tasks [11][12][15][16].

AI-generated deepfakes, including face-swapping videos and images, have been proliferating across the internet, driven by significant advancements in graphics processing units and AI algorithms. These technologies enable individuals to effortlessly create manipulated and unethical media. In areas such as deepfake detection, where novel and unforeseeable attacks are frequent, the strong generalization capability of capsules becomes vital. The nature of deepfake technology allows attackers to continuously develop new methods to bypass detection systems, making it imperative for defense mechanisms to possess the ability to generalize from known attacks to novel ones effectively. CapsNets are particularly suited to this task, due to their ability to understand the underlying structure of the data in a way that mirrors human visual perception. This understanding includes recognizing when an image or video deviates from the norm in a manner that suggests manipulation, even if the specific technique used for manipulation has not been encountered by the system before. A notable application of CapsNets in this context is Capsule-Forensic, which has been shown to be effective in identifying altered or synthetically produced images and videos [17][18]. The efficacy of CapsNets in this application stems from their unique ability to encode hierarchical relationships between objects and their components, including detailed pose information. This makes it an invaluable tool in the fight against the unethical use of AI for media manipulation [19].

3. Attention Routing

The ideas behind CapsNets share similarities with those of attention mechanisms, initially introduced in transformers [20]. Attention between capsules [21] replaces dynamic routing with a convolutional transform and attention routing, thus utilizing fewer parameters. The inclusion of a dual attention mechanism after the convolution layer and primary capsules in [22] enhanced the performance of the CapsNets. The use of a self-attention mechanism in [23] allowed for alternative non-iterative routing, efficiently reducing the number of parameters while maintaining effectiveness.

References

1. Sabour, S.; Frosst, N.; Hinton, G.E. Dynamic routing between capsules. In Proceedings of the 31st Conferenceon Neural Information Processing Systems (NIPS2017), Long Beach, CA, USA, 4–9 December 2017; pp. 3856–3866.

2. Engstrom, L.; Tsipras, D.; Schmidt, L.; Madry, A. A rotation and a translation suffice: Fooling cnns with simple transformations. arXiv 2017, arXiv:1712.02779.

3. Alcorn, M.A.; Li, Q.; Gong, Z.; Wang, C.; Mai, L.; Ku, W.S.; Nguyen, A. Strike (with) a pose: Neural networks are easily fooled by strange poses of familiar objects. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 4845–4854.

4. Hinton, G.E.; Sabour, S.; Frosst, N. Matrix capsules with EM routing. In Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018.

5. Hinton, G.E.; Krizhevsky, A.; Wang, S.D. Transforming auto-encoders. In Proceedings of the International Conference on Artificial Neural Networks, Espoo, Finland, 14–17 June 2011; Springer: Berlin/Heidelberg, Germany, 2011; pp. 44–51.

6. Ba, J.L.; Kiros, J.R.; Hinton, G.E. Layer normalization. arXiv 2016, arXiv:1607.06450.

7. Zhao, Z.; Cheng, S. Capsule networks with non-iterative cluster routing. Neural Netw. 2021, 143, 690–697.

8. Yang, S.; Lee, F.; Miao, R.; Cai, J.; Chen, L.; Yao, W.; Kotani, K.; Chen, Q. RS-CapsNet: An advanced capsule network. IEEE Access 2020, 8, 85007–85018.

9. Wu, Y.; He, K. Group normalization. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.

10. Ribeiro, F.D.S.; Duarte, K.; Everett, M.; Leontidis, G.; Shah, M. Learning with capsules: A survey. arXiv 2022, arXiv:2206.02664.

11. Zade, A.A.T.; Aziz, M.J.; Masoudnia, S.; Mirbagheri, A.; Ahmadian, A. An improved capsule network for glioma segmentation on MRI images: A curriculum learning approach. Comput. Biol. Med. 2022, 148, 105917.

12. Wang, L.; Tang, M.; Hu, X. Evaluation of grouped capsule network for intracranial hemorrhage segmentation in CT scans. Sci. Rep. 2023, 13, 3440.

13. Zhang, Z.; Ye, S.; Liao, P.; Liu, Y.; Su, G.; Sun, Y. Enhanced capsule network for medical image classification. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, BC, Canada, 20–24 June 2020; IEEE: New York, NY, USA, 2020; pp. 1544–1547.

14. Akinyelu, A.A.; Zaccagna, F.; Grist, J.T.; Castelli, M.; Rundo, L. Brain tumor diagnosis using machine learning, convolutional neural networks, capsule neural networks and vision transformers, applied to MRI: A survey. J. Imaging 2022, 8, 205.

15. Tran, M.; Vo-Ho, V.K.; Quinn, K.; Nguyen, H.; Luu, K.; Le, N. CapsNet for medical image segmentation. In Deep Learning for Medical Image Analysis; Elsevier: Amsterdam, The Netherlands, 2024; pp. 75–97.

16. Avesta, A.; Hui, Y.; Aboian, M.; Duncan, J.; Krumholz, H.; Aneja, S. 3D capsule networks for brain image segmentation. Am. J. Neuroradiol. 2023, 44, 562–568.

17. Bhagtani, K.; Yadav, A.K.S.; Bartusiak, E.R.; Xiang, Z.; Shao, R.; Baireddy, S.; Delp, E.J. An overview of recent work in media forensics: Methods and threats. arXiv 2022, arXiv:2204.12067.

18. Nguyen, H.; Yamagishi, J.; Echizen, I. Use of a capsule network to detect fake images and videos. arXiv 2019, arXiv:1910.12467.

19. Mehra, A.; Spreeuwers, L.J.; Strisciuglio, N. Deepfake Detection using Capsule Networks and Long Short-Term Memory Networks. In Proceedings of the VISIGRAPP, Virtual Event, 8–10 February 2021.

20. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is All You Need. In Proceedings of the Advances in Neural Information Processing Systems 30 (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.

21. Choi, J.; Seo, H.; Im, S.; Kang, M. Attention routing between capsules. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Seoul, Republic of Korea, 27–28 October 2019.

22. Huang, W.; Zhou, F. DA-CapsNet: Dual attention mechanism capsule network. Sci. Rep. 2020, 10, 11383.

23. Mazzia, V.; Salvetti, F.; Chiaberge, M. Efficient-capsnet: Capsule network with self-attention routing. Sci. Rep. 2021, 11, 14634.