3D-UCaps: 3D Capsules Unet for Volumetric Image Segmentation

Tan Nguyen[1], Binh-Son Hua[1], and Ngan Le[2]

[1] VinAI Research, Vietnam
{v.tannh10,v.sonhb}@vinai.io
[2] Department of Computer Science and Computer Engineering, University of Arkansas, Fayetteville USA 72701
thile@uark.edu

Abstract. Medical image segmentation has been so far achieving promising results with Convolutional Neural Networks (CNNs). However, it is arguable that in traditional CNNs, its pooling layer tends to discard important information such as positions. Moreover, CNNs are sensitive to rotation and affine transformation. Capsule network is a data-efficient network design proposed to overcome such limitations by replacing pooling layers with dynamic routing and convolutional strides, which aims to preserve the part-whole relationships. Capsule network has shown a great performance in image recognition and natural language processing, but applications for medical image segmentation, particularly volumetric image segmentation, has been limited. In this work, we propose 3D-UCaps, a 3D voxel-based Capsule network for medical volumetric image segmentation. We build the concept of capsules into a CNN by designing a network with two pathways: the first pathway is encoded by 3D Capsule blocks, whereas the second pathway is decoded by 3D CNNs blocks. 3D-UCaps, therefore inherits the merits from both Capsule network to preserve the spatial relationship and CNNs to learn visual representation. We conducted experiments on various datasets to demonstrate the robustness of 3D-UCaps including iSeg-2017, LUNA16, Hippocampus, and Cardiac, where our method outperforms previous Capsule networks and 3D-Unets. Our code is available at https://github.com/VinAIResearch/3D-UCaps.

Keywords: Capsule network · CapsNet · Medical image segmentation.

1 Introduction

Medical image segmentation (MIS) is a visual task that aims to identify the pixels of organs or lesions from background medical images. It plays a key role in medical analysis, computer-aided diagnosis, and smart medicine due to the great improvement in diagnostic efficiency and accuracy. Thanks to recent advances of deep learning, convolutional neural networks (CNNs) can be used to extract hierarchical feature representation for segmentation, which is robust to image degradation such as noise, blur, contrast, etc. Among many CNNs-based segmentation approaches, FCN [19], Unet [5], and Auto-encoder-like

architecture have become the desired models for MIS. Particularly, such methods achieved impressive performance in brain tumor [15,16], liver tumor [2,18], optic disc [23,28], retina [17], lung [26,12], and cell [8,20]. However, CNNs are limited in their mechanism of aggregating data at pooling layers. Notably, pooling summarizes features in a local window and discards important information such as pose and object location. Therefore, CNNs with consecutive pooling layers are unable to perverse the spatial dependencies between objects parts and wholes. Moreover, the activation layer plays an important role in CNNs; however, it is not interpretable and has often been used as a black box. MIS with CNNs is thus prone to performance degradation when data undergoes some transformations such as rotations. A practical example is during an MRI scan, subject motion causes transformations to appear in a subset of slices, which is a hard case for CNNs [29].

To overcome such limitations by CNNs, Sabour et al. [24] developed a novel network architecture called Capsule Network (CapsNet). The basic idea of CapsNet is to encode the part-whole relationships (e.g., scale, locations, orientations, brightnesses) between various entities, i.e., objects, parts of objects, to achieve viewpoint equivariance. Unlike CNNs which learn all part features of the objects, CapsNet learns the relationship between these features through dynamically calculated weights in each forward pass. This optimization mechanism, i.e., dynamic routing, allows weighting the contributions of parts to a whole object differently at both training and inference. CapsNet has been mainly applied to image recognition; its performance is still limited compared to the state-of-the-art by CNNs-based approaches. Adapting CapsNet for semantic segmentation, e.g., SegCaps [13,14], receives even less attention. In this work, we propose an effective 3D Capsules network for volumetric image segmentation, named 3D-UCaps. Our 3D-UCaps is built on both 3D Capsule blocks, which take temporal relations between volumetric slices into consideration, and 3D CNNs blocks, which extract contextual visual representation. Our 3D-UCaps contains two pathways, i.e., encoder and decoder. Whereas encoder is built upon 3D Capsule blocks, the decoder is built upon 3D CNNs blocks. We argue and show empirically that using deconvolutional Capsules in the decoder pathway not only reduces segmentation accuracy but also increases model complexity.

In summary, our contributions are: (1) An effective 3D Capsules network for volumetric image segmentation. Our 3D-UCaps inherits the merits from both 3D Capsule block to preserve spatial relationship and 3D CNNs block to learn better visual representation. (2) Extensive experiments on various datasets and ablation studies that showcase the effectiveness and robustness of 3D-UCaps for MIS.

2 Background

In CNNs, each filter of convolutional layers works like a feature detector in a small region of the input features and as we go deeper in a network, the detected low-level features are aggregated and become high-level features that can be used to distinguish between different objects. However, by doing so, each feature map

only contains information about the presence of the feature, and the network relies on fixed learned weight matrix to link features between layers. It leads to the problem that the model cannot generalize well to unseen changes in the input image and usually perform poorly in that case.

CapsNet [24] is a new concept that strengthens feature learning by retaining more information at aggregation layer for pose reasoning and learning the part-whole relationship, which makes it a potential solution for semantic segmentation and object detection tasks. Each layer in CapsNet aims to learn a set of entities (i.e., parts or objects) with their various properties and represent them in a high-dimensional form, particularly vector in [24]. The length of this vector indicates the presence of the entity in the input while its orientation encodes different properties of that entity. An important assumption in CapsNet is the entity in previous layer are simple objects and based on an agreement in their votes, complex objects in next layer will be activated or not. This setting helps CapsNet reflect the changes in input through the activation of properties in the entity and still recognize the object successfully based on a dynamic voting between layers. Let $\{c_1^l, c_2^l, \ldots, c_n^l\}$ be the set of capsules in layer l, $\{c_1^{l+1}, c_2^{l+1}, \ldots, c_m^{l+1}\}$ be the set of capsule in layer $l+1$, the overall procedure will be:

$$c_j^{l+1} = \text{squash}\left(\sum_i r_{ij} v_{j|i}\right), \qquad v_{j|i} = W_{ij} c_i^l \tag{1}$$

where W_{ij} is the learned weight matrix to linear mapping features of capsule c_i^l in layer l to feature space of capsule c_j^{l+1} in layer $l+1$. The r_{ij} are coupling coefficients between capsule i and j that are dynamically assigned by a routing algorithm in each forward pass such that $\sum_j r_{ij} = 1$.

SegCaps [13,14], a state-of-the-art Capsule-based image segmentation, has made a great improvement to expand the use of CapsNet to the task of object segmentation. This method functions by treating an MRI image as a collection of slices, each of which is then encoded and decoded by capsules to output the segmentation. However, SegCaps is mainly designed for 2D still images, and it performs poorly when being applied to volumetric data because of missing temporal information. Our work differs in that we build the CapsNet to consume 3D data directly so that both spatial and temporal information can be fully used for learning. Furthermore, our 3D-UCaps is able to take both advantages of CapsNet and 3D CNNs into consideration.

3 Our Proposed 3D-UCaps Network

In this work, we propose a hybrid 3D-UCaps network, which inherits the merits from both CapsNet and 3D CNNs. Our proposed 3D-UCaps follows Unet-like architecture [5] and contains three main components as follows.

Visual Feature Extractor: We use a set of dilated convolutional layers to convert the input to high-dimensional features that can be further processed by capsules. It contains three convolution layers with the number of channels

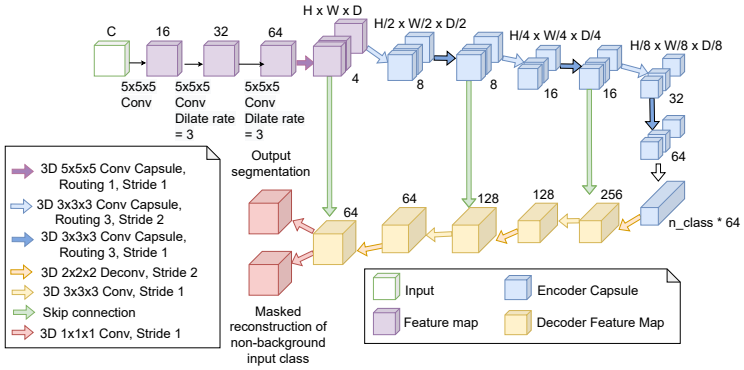

Fig. 1. Our proposed 3D-UCaps architecture with three components: visual feature extraction; capsule encoder, and convolution decoder. Number on the blocks indicates number of channels in convolution layer and dimension of capsules in capsule layers.

increased from 16 to 32 then 64, kernel size $5 \times 5 \times 5$ and dilate rate set to 1, 3, and 3, respectively. The output of this part is a feature map of size $H \times W \times D \times 64$.

Capsule Encoder: The visual feature from the previous component can be cast (reshaped) into a grid of $H \times W \times D$ capsules, each represented as a single 64-dimensional vector. Our capsule layer is a 3D convolutional capsule, which consumes volumetric data directly instead of treating it as separate slices as in SegCaps [13]. The advantage of our 3D capsule layer is that contextual information along the temporal axis can be included in the feature extraction. In additional to increasing the dimensionality of capsules as we ascend the hierarchy [24], we suggest to use more capsule types in low-level layers and less capsule types in high-level layers. This is due to the fact that low-level layers represent simple object while high-level layers represent complex object and the clustering nature of routing algorithm [9]. The number of capsule types in the encoder path of our network are set to $(16, 16, 16, 8, 8, 8)$, respectively. This is in contrast to the design in SegCaps, where the numbers of capsules are increasing $(1, 2, 4, 4, 8, 8)$ along the encoder path. We make sure that the number of capsule types in the last convolutional capsule layer is equal to the number of categories in the segmentation, which can be further supervised by a margin loss [24]. The output from a convolution capsule layer has the shape $H \times W \times D \times C \times A$, where C is the number of capsule types and A is the dimension of each capsule.

Convolutional Decoder: We use the decoder of 3D Unet [5] which includes deconvolution, skip connection, convolution and BatchNorm layers [10] to generate the segmentation from features learned by capsule layers. Particularly, we reshape the features to $H \times W \times D \times (C \star A)$ before passing them to the next convolution layer or concatenating with skip connections. The overall architecture can be seen in Fig. 1. Note that in our design, we only use capsule layers in the contracting path but not expanding path in the network. Sabour et al. [24] point out that "routing-by-agreement" should be far more effective than max-pooling, and max-pooling layers only exist in the contracting path of Unet.

This contradicts to the design by LaLonde et al. [13], where capsules are used in the expanding path in the network as well. We empirically show that using capsules in the expanding path has negligible effects compared to the traditional design while incurring high computational cost due to routing between capsule layers.

Training Procedure. We supervise our network with ground truth segmentation as follows. The margin loss is applied at the capsule encoder with downsampled ground truth segmentation. The weighted cross entropy loss is applied at the decoder head to optimize the entire network. To regularize the training, we also use an additional branch to output the reconstruction of the original input image as in previous work [24,13]. We use masked mean-squared error for the reconstruction. The total loss is the weighted sum of the three losses.

4 Experimental Results

Evaluation Setup

We perform experiments on various MIS datasets to validate our method. Specifically, we experiment with iSeg-2017 [29], LUNA16 [1], Hippocampus, and Cardiac [25]. iSeg is a MRI dataset of infant brains that requires to be segmented into white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF). A recent analysis [29] shows that previous methods tend to perform poorly on subjects with movement and unusual poses. We follow the experiment setup by 3D-SkipDenseSeg [3] to conduct the report on this dataset where 9 subjects are used for training and 1 subject (subject #9) is for testing.

Additionally, we experiment on LUNA16, Hippocampus, and Cardiac [25] to compare with other capsule-based networks [13,27]. We follow a similar experiment setup in SegCaps [13] to conduct the results on LUNA16. We also use 4-fold cross validation on training set to conduct the experiments on Hippocampus and Cardiac.

Implementation Details

We implemented both 3D-SegCaps and 3D-UCaps in Pytorch. The input volumes are normalized to [0, 1]. We used patch size set as $64 \times 64 \times 64$ for iSeg and Hippocampus whereas patch size set as $128 \times 128 \times 128$ on LUNA16 and Cardiac. Both 3D-SegCaps and 3D-UCaps networks were trained without any data augmentation methods. We used Adam optimization with an initial learning rate of 0.0001. The learning rate is decayed by a factor of 0.05 if Dice score on validation set not increased for 50,000 iterations and early stopping is performed with a patience of 250,000 iterations as in [13]. Our models were trained on NVIDIA Tesla V100 with 32GB RAM, and it takes from 2-4 days depends on the size of the dataset.

Performance and Comparison

In this section, we compare our 3D-UCaps with both SOTA 3D CNNs-based segmentation approaches and other existing SegCaps methods. Furthermore, we have implemented 3D-SegCaps which is an extension version of 2D-SegCaps [13] on volumetric data to prove the effectiveness of incorporating deconvolution

Method	Depth	Dice Score			
		WM	GM	CSF	Average
Qamar et al. [22]	82	90.50	**92.05**	**95.80**	**92.77**
3D-SkipDenseSeg [3]	47	**91.02**	91.64	94.88	92.51
VoxResNet [4]	25	89.87	90.64	94.28	91.60
3D-Unet [5]	18	89.83	90.55	94.39	91.59
CC-3D-FCN [21]	34	89.19	90.74	92.40	90.79
DenseVoxNet [11]	32	85.46	88.51	91.26	89.24
SegCaps (2D) [13]	16	82.80	84.19	90.19	85.73
Our 3D-SegCaps	16	86.49	88.53	93.62	89.55
Our 3D-UCaps	17	**90.95**	**91.34**	**94.21**	**92.17**

Table 1. Comparison on iSeg-2017 dataset. The first group is 3D CNN-based networks. The second group is Capsule-based networks. The best performance is in **bold**.

layers into 3D-UCaps. Our 3D-SegCaps share similar network architecture with 2D-SegCaps [13] and implemented with 3D convolution layers. This section is structured as follows: We first provide a detailed analysis on iSeg with different criteria such as segmentation accuracy, network configurations, motion artifact, and rotation invariance capability. We then report segmentation accuracy on various datasets, including LUNA16, Hippocampus, and Cardiac.

Accuracy: The comparison between our proposed 3D-SegCaps, 3D-UCaps with SOTA segmentation approaches on iSeg dataset [29] is given in Table 1. Thanks to taking both spatial and temporal into account, both 3D-SegCaps, 3D-UCaps outperforms 2D-SegCaps with large margin on iSeg dataset. Moreover, our 3D-UCaps consisting of Capsule encoder and Deconvolution decoder obtains better results than 3D-SegCaps, which contains both Capsule encoder and Capsule decoder. Compare to SOTA 3D CNNs networks our 3D-UCaps achieves compatible performance while our network is much shallower i.e our 3D-UCaps contains only 17 layers compare to 82 layers in [22]. Compare to SOTA 3D CNNs networks which has similar depth, i.e. our 3D-UCaps with 18 layers, our 3D-UCaps obtains higher Dice score at individual class and on average.

Method	Dice Score			
	WM	GM	CSF	Average
change number of capsule (set to 4)	89.02	89.78	89.95	89.58
without feature extractor	89.15	89.66	90.82	89.88
without margin loss	87.62	88.85	92.06	89.51
without reconstruction loss	88.50	88.96	90.18	89.22
3D-UCaps	**90.95**	**91.34**	**94.21**	**92.17**

Table 2. Performance of 3D-UCaps on iSeg with different network configurations

Network configuration: To prove the effectiveness of the entire network architecture, we trained 3D-UCaps under various settings. The results are given in Table 2. We provide a baseline where we change the number of capsules at the first layer from 16 capsules (our setting in Section 3) to 4 capsules (similar to SegCaps). We also examine the contribution of each component by removing feature extraction layer, margin loss, reconstruction loss, respectively. The result shows that each change results in accuracy drop, which validates the competence of our network model.

Method	x-axis			y-axis			z-axis		
	CSF	GM	WM	CSF	GM	WM	CSF	GM	WM
3D-SkipDenseSeg [3]	83.93	88.76	88.52	78.98	87.80	87.89	82.88	88.38	88.27
SegCaps (2D) [13]	88.11	83.01	82.01	86.43	81.80	80.91	89.36	83.99	82.76
Our 3D-SegsCaps	90.70	86.15	84.24	87.75	84.21	82.76	89.77	85.54	83.92
Our 3D-UCaps	**91.04**	**88.87**	**88.62**	**90.31**	**88.21**	**88.12**	**90.86**	**88.65**	**88.55**

Table 3. Performance on iSeg with motion artifact on different axis. The experiment was conducted 5 times and report average number to minimize the effect of randomization

Moving artifact: Motion artifact caused by patient moving when scanning was reported as a hard case in [29]. We examine the influence of motion artifact to our 3D-UCaps in Table 3. In this table, motion artifact at each axis was simulated by randomly rotating 20% number of slices along the axis with an angle between -5 and 5 degree. As can be seen, 3D-based capsules (3D-SegCaps and 3D-UCaps) both outperforms SegCaps in all classes in all rotations.

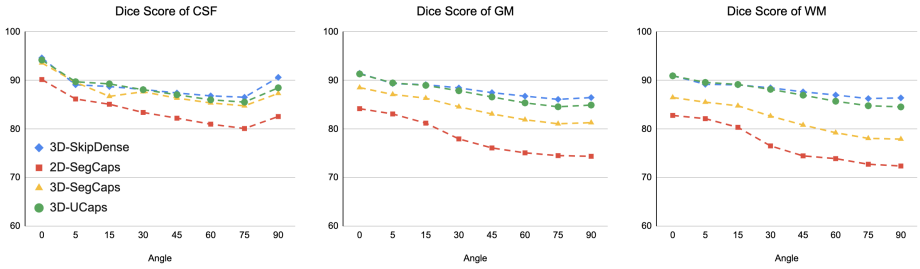

Fig. 2. Comparison on iSeg with test object rotated about z-axis from zero to 90 degree. Best view in zoom.

Rotation invariance: To further study rotation equivariance and invariance properties in our 3D-UCaps, we trained our network without any rotation data augmentation. During testing, we choose an axis to rotate the volume, and apply the rotation with angle values fixed to 5, 10, 15, .., 90 degrees. Here we conduct the experiment on iSeg and choose z-axis as the rotation axis. We choose 3D

SkipDense [3] as 3D CNNs-based segmentation method and compare robustness to rotations between our 3D-UCaps, our 3D-SegCaps, 2D-SegCaps, and 3D SkipDense [3]. The segmentation accuracy of rotation transformation on each target class is reported in Figure 2. We found that the performance tends to drop slightly when the rotation angles increases. Except 2D-SegCaps, there is no significant difference in performance between 3D CNNs-based network and Capsule-based networks even though traditional 3D CNN-based network is not equipped with learning rotation invariance property. This could be explained by that the networks perform segmentation on a local patch of the volume at a time, making them resistant to local changes. Further analysis of the robustness of capsule network on the segmentation task would be necessary, following some recent analysis on the classification task [6,7].

Results on other datasets: Besides iSeg, we continue benchmarking our 3D-UCaps on other datasets. The performance of 3D-UCaps on LUNA16, Hippocampus, and Cardiac is reported in Table 4, 5, 6. Different from other datasets, LUNA16 was annotated by an automated algorithm instead of a radiologist. When conducting the report on LUNA16, SegCaps [13] removed 10 scans with exceedingly poor annotations. In Table 4, we compare our performance in two cases: full dataset and remove 10 exceedingly poor annotations. The results show that our 3D-UCaps outperforms previous methods and our 3D-SegCaps baseline, respectively.

Method	Split-0	Split-1	Split-2	Split-3	Average
SegCaps (2D) [13]	**98.50**	98.52	98.46	98.47	98.48
Our 3D-UCaps	98.49	**98.61**	**98.72**	**98.76**	**98.65**
SegCaps* (2D) [27]	98.47	98.19	98.07	98.24	98.24
Our 3D-UCaps*	**98.48**	**98.60**	**98.70**	**98.76**	**98.64**

Table 4. Comparison on LUNA16 in two cases where * indicates full dataset. The best score is in **bold**.

Method	Anterior			Posterior		
	Recall	Precision	Dice	Recall	Precision	Dice
Multi-SegCaps (2D) [13]	80.76	65.65	72.42	84.46	60.49	70.49
EM-SegCaps (2D) [27]	17.51	20.01	18.67	19.00	34.55	24.52
Our 3D-SegCaps	94.70	75.41	83.64	93.09	73.20	81.67
Our 3D-UCaps	**94.88**	**77.48**	**85.07**	**93.59**	**74.03**	**82.49**

Table 5. Comparison on Hippocampus dataset with 4-fold cross validation.

Method	Recall	Precision	Dice
SegCaps (2D) [13]	**96.35**	43.96	60.38
Multi-SegCaps (2D) [27]	86.89	54.47	66.96
Our 3D-SegCaps	88.35	56.40	67.20
Our 3D-UCaps	92.69	**89.45**	**90.82**

Table 6. Comparison on Cardiac dataset with 4-fold cross validation.

5 Conclusion

In this work, we proposed a novel network architecture that can both utilize 3D capsules for learning features for volumetric segmentation while retaining the advantage of traditional convolutions in decoding the segmentation results. Even though we use capsules with dynamic routing [24,13] only in the encoder of a simple Unet like architecture, we can achieve competitive result with the state-of-the-art models on iSeg-2017 challenge while outperforming SegCaps [13] on different complex datasets. Exploring hybrid architecture between Capsule-based and traditional neural network is therefore a promising approach to medical image analysis while keeping model complexity and computation cost plausible.

Acknowledgment: This material is based upon work supported by the National Science Foundation under Award No. OIA-1946391.

Disclaimer: Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

1. Armato III, S.G., McLennan, G., Bidaut, L., McNitt-Gray, M.F., Meyer, C.R., Reeves, A.P., Zhao, B., Aberle, D.R., Henschke, C.I., Hoffman, E.A., et al.: The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans. Medical physics **38**(2), 915–931 (2011)
2. Bilic, P., Christ, P.F., Vorontsov, E., Chlebus, G., Chen, H., Dou, Q., Fu, C.W., Han, X., Heng, P.A., Hesser, J., et al.: The liver tumor segmentation benchmark (lits). arXiv preprint arXiv:1901.04056 (2019)
3. Bui, T.D., Shin, J., Moon, T.: Skip-connected 3d densenet for volumetric infant brain mri segmentation. Biomedical Signal Processing and Control **54**, 101613 (2019)
4. Chen, H., Dou, Q., Yu, L., Qin, J., Heng, P.A.: Voxresnet: Deep voxelwise residual networks for brain segmentation from 3d mr images. NeuroImage **170**, 446–455 (2018)
5. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: International conference on medical image computing and computer-assisted intervention. pp. 424–432. Springer (2016)

6. Gu, J., Tresp, V.: Improving the robustness of capsule networks to image affine transformations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7285–7293 (2020)
7. Gu, J., Tresp, V., Hu, H.: Capsule network is not more robust than convolutional network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14309–14317 (2021)
8. Hatipoglu, N., Bilgin, G.: Cell segmentation in histopathological images with deep learning algorithms by utilizing spatial relationships. Medical & biological engineering & computing **55**(10), 1829–1848 (2017)
9. Hinton, G.E., Sabour, S., Frosst, N.: Matrix capsules with em routing. In: International conference on learning representations (2018)
10. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International conference on machine learning. pp. 448–456. PMLR (2015)
11. Jégou, S., Drozdzal, M., Vazquez, D., Romero, A., Bengio, Y.: The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 11–19 (2017)
12. Jin, D., Xu, Z., Tang, Y., Harrison, A.P., Mollura, D.J.: Ct-realistic lung nodule simulation from 3d conditional generative adversarial networks for robust lung segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 732–740. Springer (2018)
13. LaLonde, R., Bagci, U.: Capsules for object segmentation. arXiv preprint arXiv:1804.04241 (2018)
14. LaLonde, R., Xu, Z., Irmakci, I., Jain, S., Bagci, U.: Capsules for biomedical image segmentation. Medical image analysis **68**, 101889 (2021)
15. Le, N., Gummadi, R., Savvides, M.: Deep recurrent level set for segmenting brain tumors. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) pp. 646–653 (2018)
16. Le, N., Le, T., Yamazaki, K., Bui, T.D., Luu, K., Savides, M.: Offset curves loss for imbalanced problem in medical segmentation. 25th International Conference on Pattern Recognition (ICPR) pp. 6189–6195 (2021)
17. Le, N., Yamazaki, K., Gia, Q.K., Truong, T., Savvides, M.: A multi-task contextual atrous residual network for brain tumor detection & segmentation. 25th International Conference on Pattern Recognition (ICPR) pp. 5943–5950 (2021)
18. Li, X., Chen, H., Qi, X., Dou, Q., Fu, C.W., Heng, P.A.: H-denseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes. IEEE transactions on medical imaging **37**(12), 2663–2674 (2018)
19. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3431–3440 (2015)
20. Moshkov, N., Mathe, B., Kertesz-Farkas, A., Hollandi, R., Horvath, P.: Test-time augmentation for deep learning-based cell segmentation on microscopy images. Scientific reports **10**(1), 1–7 (2020)
21. Nie, D., Wang, L., Adeli, E., Lao, C., Lin, W., Shen, D.: 3-d fully convolutional networks for multimodal isointense infant brain image segmentation. IEEE transactions on cybernetics **49**(3), 1123–1136 (2018)
22. Qamar, S., Jin, H., Zheng, R., Ahmad, P., Usama, M.: A variant form of 3d-unet for infant brain segmentation. Future Generation Computer Systems **108**, 613–623 (2020)

23. Ramani, R.G., Shanthamalar, J.J.: Improved image processing techniques for optic disc segmentation in retinal fundus images. Biomedical Signal Processing and Control **58**, 101832 (2020)
24. Sabour, S., Frosst, N., Hinton, G.E.: Dynamic routing between capsules. arXiv preprint arXiv:1710.09829 (2017)
25. Simpson, A.L., Antonelli, M., Bakas, S., Bilello, M., Farahani, K., Van Ginneken, B., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., et al.: A large annotated medical image dataset for the development and evaluation of segmentation algorithms. arXiv preprint arXiv:1902.09063 (2019)
26. Souza, J.C., Diniz, J.O.B., Ferreira, J.L., da Silva, G.L.F., Silva, A.C., de Paiva, A.C.: An automatic method for lung segmentation and reconstruction in chest x-ray using deep neural networks. Computer methods and programs in biomedicine **177**, 285–296 (2019)
27. Survarachakan, S., Johansen, J.S., Aarseth, M., Pedersen, M.A., Lindseth, F.: Capsule nets for complex medical image segmentation tasks. Colour and Visual Computing Symposium (2020)
28. Veena, H., Muruganandham, A., Kumaran, T.S.: A review on the optic disc and optic cup segmentation and classification approaches over retinal fundus images for detection of glaucoma. SN Applied Sciences **2**(9), 1–15 (2020)
29. Wang, L., Nie, D., Li, G., Puybareau, É., Dolz, J., Zhang, Q., Wang, F., Xia, J., Wu, Z., Chen, J.W., et al.: Benchmark on automatic six-month-old infant brain segmentation algorithms: the iseg-2017 challenge. IEEE transactions on medical imaging **38**(9), 2219–2230 (2019)