TGNET: A TASK-GUIDED NETWORK ARCHITECTURE FOR MULTI-ORGAN AND TUMOUR SEGMENTATION FROM PARTIALLY LABELLED DATASETS

Hao Wu *Shuchao Pang* *Arcot Sowmya*

School of Computer Science and Engineering, University of New South Wales, Sydney, NSW, Australia

ABSTRACT

The cost of labour and expertise makes it challenging to collect a large amount of medical data and annotate 3D medical images at a voxel level. Most public medical datasets are only labelled with one type of organ or tumour. For example, in the liver segmentation task, only the liver is labelled while all the other organs, even when present, as well as irrelevant parts are annotated as background, resulting in partially labelled datasets. Current popular methods usually build multiple neural network models for different tasks that lead to great model redundancy, or design a unified network for all tasks which suffers from low ability to extract task-related features. In this paper, we propose a unified and task-guided network architecture to efficiently learn task-related features and avoid mixing representations of different organs and tumours from different tasks. Specifically, a novel residual block and attention module are devised in a task-guided way to fuse image features and task encoding constraints. Moreover, both designs significantly suppress task-unrelated features and highlight features related to the specific segmentation task. Experiments conducted on seven benchmark datasets illustrate that our task-guided model achieves more competitive performance compared with state-of-the-art approaches in segmenting multiple organs and tumours from partially labelled data.

Index Terms— task-guided attention module, multiple organs and tumours, partially-labelled datasets, unified model

1. INTRODUCTION

Nowadays, artificial intelligence is being widely used in autonomous vehicles, quality tests as well as medical data analysis [1][2]. In particular, deep neural networks are often utilised for processing medical images to help doctors segment human organs and tumours [3]. Due to the high cost of labour and expertise, it is challenging to annotate 3D medical images at voxel level. Most public medical datasets are only labelled with one type of organ or tumour. This issue is known as the partially labelled dataset problem. Considering the memory and time consumed by training multiple neural networks for different segmentation tasks and the scarcity of accessible 3D medical image datasets, multi-organ and tumour segmentation from partially labelled datasets remains a worthy problem to solve. Building a unified model using partially labelled datasets has attracted researchers' interest recently [3][4].

Traditionally, neural networks [5] are designed for one specific task, for example, liver and liver tumour extraction [6]. Because of the high computational cost and limited datasets, these networks only use abdomen CT scans for single organ and tumour segmentation, and are not able to make use of abdomen CT scans with other labelled organs, making abdomen medical analysis progress slowly. There are many efforts to integrate these partially labelled datasets into one model to increase training samples, even though they have labels of different organs and tumours. Chen et al. [7] gathered data from different datasets and trained a shared-encoder neural network with different decoders for different segmentation tasks, which showed that even partially labelled datasets can improve the segmentation accuracy of different tasks compared with training on a single dataset. However, this kind of multi-head model enlarges the model size and increase the training and inference time. Very recently, Zhang et al. [4] proposed a simple unified model with a shared encoder-decoder network for all segmentation tasks. And a dynamic controller was designed to generate weights of dynamic filters for different tasks then these dynamic filters were used to segment different organs and tumours. Although the unified method simplifies the network and achieves good segmentation performance, it easily fails in distinguishing different organs and tumours in different segmentation tasks.

To address these problems, in this paper we propose a task-guided network architecture (namely TGNet) to accurately segment multi-organs and tumours from partially labelled datasets. We first propose a task-guided attention module placed at skip connections, which consists of a GAP (global average pooling) module followed by task-encoding concatenation and a convolutional layer with kernel size $1 \times 1 \times 1$. Then a sigmoid activation function is used to generate channel-wise attention for different feature maps, which aims to highlight task-related features and suppress task-unrelated features according to different segmentation tasks. During the encoding path, we devise novel task-guided residual blocks where the task-guided attention mechanism is also implemented on the residual from the input, which can

Corresponding author: Shuchao Pang.

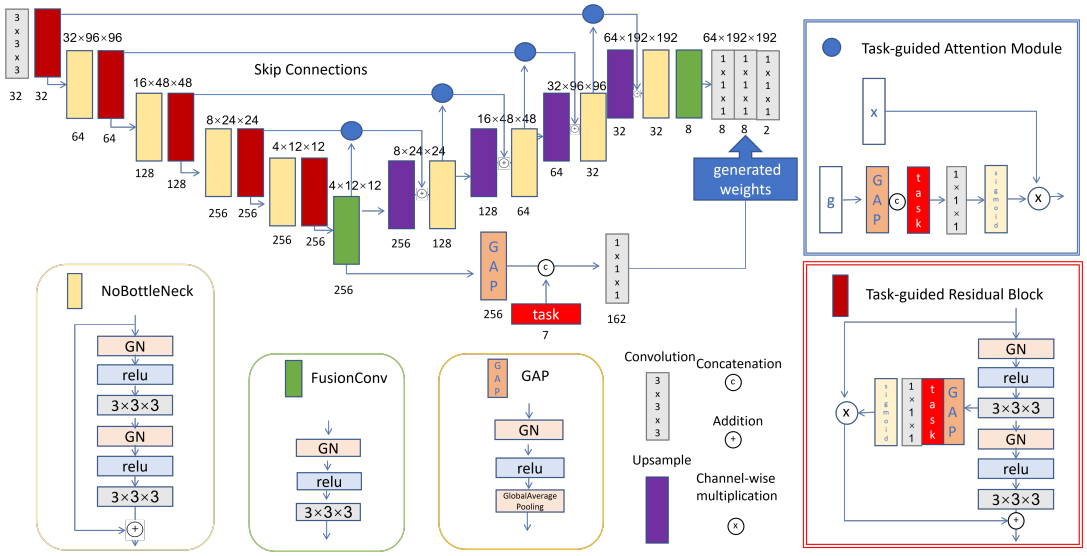

Fig. 1. Proposed task-guided network architecture to segment multi-organs and tumours from partially labelled datasets

further extract organ and tumour features related to the specific task through the encoder. Experimental results show that the proposed network outperforms existing network models and achieves more competitive segmentation results on public benchmark datasets.

2. METHOD

The architecture of the proposed TGNet is shown in **Fig.1.**. It consists of two novel modules: task-guided attention modules and residual blocks that can help to learn task-specific features from multiple partially labelled segmentation datasets.

2.1. Task-guided Attention Module

In the widely used medical image segmentation network UNet [10], the decoder restores the feature map back to the image using the information provided by the encoder. Traditional attention modules [11] use either spatial attention, channel attention or combined attention to capture the salient parts of the input through skip connections. However, when dealing with the skip connection in a unified model for different tasks, these attention mechanisms increase the segmentation accuracy on one or two tasks, but significantly decrease the accuracy on other tasks, therefore the overall accuracy declines. When exploring the reason behind it, we found that traditional attention modules mistakenly mix organs or tumours of other tasks with the one we want. This is likely due to some tasks sharing mutual features, therefore focussing on these features and suppressing other features likely makes it difficult for the network to distinguish between organs and tumours that share mutual features.

Therefore, the decoder should use task-focused encoder information to help build accurate segmentation. We need a special mechanism to guide the model to pay more attention to task-related information and suppress unrelated information in order to process the encoder information more efficiently and accurately. To address this, we propose an innovative attention module, namely the task-guided attention module.

Let x be the input feature map of the encoder and g the descriptor from the decoder. We first apply group normalization and ReLU on g, which is widely used to accelerate convergence and obtain non-linear activation. In order to associate the descriptor with specific tasks and decrease the number of parameters, global average pooling is used afterwards to obtain the feature vector. This feature vector is then concatenated with the task-encoding to acquire the task-guided feature vector. A convolutional layer with kernel size $1 \times 1 \times 1$ and a $sigmoid$ activation function are implemented to obtain the task-guided channel-wise attention. Thus we can use channel-wise multiplication with x in skip connections which enhances the channels that are significantly related to the task and suppresses unrelated channels. The final output of the skip connection is: $x\prime = x \odot \sigma(f(GAP(g)))$ where $\odot$ means channel-wise multiplication, σ represents sigmoid activation function, f denotes the $1 \times 1 \times 1$ convolution and GAP is global average pooling.

2.2. Task-guided Residual Block

Residual Deep Learning [12] was proposed to utilize residuals to solve the Gradient Descent issue, making it possible to build much deeper and wider neural networks. To further improve the accuracy of segmentation of multi-organs and tumours from partially labelled datasets in a unified model, we

Table 1. Comparison with state-of-the-art methods for multi-organ and tumour segmentation on seven partially labelled datasets. Best result in each column is shown in red and second best in blue.

Methods	Task1:Liver				Task2: Kidney				Task3: Hepatic Vessel				Task4: Pancreas				Task5: Colon		Task6: Lung		Task7: Spleen		Average Score	
	Dice		HD		Dice		HD		Dice		HD		Dice		HD		Dice	HD	Dice	HD	Dice	HD	Dice	HD
	Organ	tumour	Organ	tumour	Organ	tumour	Organ	tumour	Organ	tumour	Organ	tumour	Organ	tumour	Organ	tumour	tumour	tumour	tumour	tumour	Organ	Organ		
Multi-Nets	96.61	61.65	4.25	41.16	96.52	74.89	1.79	11.19	63.04	72.19	13.73	50.70	82.53	58.36	9.23	26.13	34.33	103.91	54.51	53.68	93.76	2.65	71.67	28.95
Multi-Head [7]	96.75	64.08	3.67	45.68	96.60	79.16	4.69	13.28	59.49	69.64	19.28	79.66	83.49	61.22	6.40	18.66	50.89	59.00	64.75	34.22	94.01	3.86	74.55	26.22
TAL [3]	96.18	60.82	5.99	38.87	95.95	75.87	1.98	15.36	61.90	72.68	13.86	43.57	81.35	59.15	9.02	21.07	48.08	66.42	61.85	39.92	93.01	3.10	73.35	23.56
Cond-NO	69.38	47.38	37.79	109.65	93.32	70.40	8.68	24.37	42.27	69.86	93.35	70.34	65.31	46.24	36.06	76.26	42.55	76.14	57.67	102.92	59.68	38.11	60.37	61.24
Cond-Input [8]	96.68	65.26	6.21	47.61	96.82	78.41	1.32	10.10	62.17	73.17	13.61	43.32	82.53	61.20	8.09	31.53	51.43	44.18	60.29	58.02	93.51	4.32	74.68	24.39
Cond-Dec [9]	95.27	63.86	5.49	36.04	95.07	79.27	7.21	8.02	61.29	72.46	14.05	65.57	77.24	55.69	17.60	48.47	51.80	63.67	57.68	53.27	90.14	6.52	72.71	29.63
DoDNet* [4]	96.76	64.97	3.73	42.81	96.07	78.14	2.99	8.59	62.35	75.02	13.52	51.49	80.55	56.82	13.23	38.03	49.14	60.00	56.32	27.80	83.90	34.37	72.73	26,96
TGNet	96.83	63.84	3.57	40.57	96.17	81.05	4.67	7.03	62.59	74.65	13.27	44.67	83.30	60.71	7.00	21.21	47.37	65.32	61.69	34.48	94.18	2.51	74.76	22.21

propose a novel task-guided residual block which inserts task-guided attention block into the regular convolutional block, which is shown in **Fig.1.**. Let x be the input of this block. We first implement a $3 \times 3 \times 3$ convolutional layer and obtain $f(x)$. There are two branches after this convolutional layer. One branch is followed by GAP (global average pooling) and the task-encoding is concatenated with the output of GAP. After it, a convolutional layer with kernel size $1 \times 1 \times 1$ follows it. Sigmoid activation is used in this branch to generate the channel-wise attention to perform channel-wise multiplication with the residual x. For the other branch, group normalization, ReLU activation function and another $3 \times 3 \times 3$ convolutional layer are implemented. The final output of this attention residual block is $R(x) = x \odot \sigma(g(f(x)||task)) + f\prime(f(x))$ where $\odot$ is channel-wise multiplication, f and $f\prime$ are $3 \times 3 \times 3$ convolutional layers with group normalisation and ReLU while g is a $1 \times 1 \times 1$ convolutional layer, σ represents the sigmoid activation function and $||$ means concatenation. From experimental results, this task-guided residual mechanism was found to help encoder blocks focus on the more task-related features in the passed residual, enabling computation of a more accurate task-specific feature map. Therefore, the information passed in both skip connections and the encoder can be more task-specific. This special residual block, combined with task-guided attention modules in skip connections, enhances the segmentation accuracy on several tasks, especially those with fewer samples.

3. EXPERIMENTS AND ANALYSIS

3.1. Image Datasets and Evaluation Metrics

Experiments were conducted on seven partially labelled organ and/or tumour datasets, including LiTS, KiTS and MSD [13]. There were 1155 3D abdominal CT scans in total, including 131, 210, 303, 281, 126, 63 ,41 of liver, kidney, hepatic vessel, pancreas, colon, lung and spleen respectively. We followed Zhang et al. [4] to split the training set into 920 scans for training and 235 for testing, where the ratio was the same for all tasks.

Dice score and Hausdorff distance 95 were used as the metrics for measuring segmentation accuracy. We trained the propsoed network using Pytorch and two NVIDIA TESLA V100. For training, we used the Dice loss function plus Cross Entropy loss as the loss function, defined as below:

$$L = 1 - \frac{2\sum_{i=1}^{V} p_i y_i}{\sum_{i=1}^{V}(p_i + y_i + \epsilon)} - \sum_{i=1}^{V}(y_i \log p_i + (1 - y_i)\log(1 - p_i)) \quad (1)$$

where p_i and y_i are the prediction and ground truth of the voxel i, V means the number of voxels in total, and ϵ represents the smoothing factor. We used SGD optimizer with the learning rate initialized to 0.01 and a decay formula $lr = lr_{init} \times (1 - k/K)^{0.9}$,where k is the current epoch and K the maximal epochs which was set to 1000. Images were randomly cropped to the size of $64 \times 192 \times 192$ during training and a sliding window of size of $64 \times 192 \times 192$ was used to predict the maps. The results were post-processed by retaining connected areas corresponding to different tasks.

3.2. Comparison with State-of-the-art Methods

We compared the proposed TGNet with other state-of-the-art methods that deal with partially labelled datasets: (1) seven networks that were separately trained for seven specific tasks using the corresponding datasets (i.e., Multi-Net); (2) two networks using multiple heads (i.e., Multi-Head [7] and TAL [3]); (3) a single network without task constraints (i.e., Cond-NO); (4) two single networks with task constraints (i.e., Cond-Input [8] and Cond-Dec [9]); and (5) a single network with dynamic filters and task constraints (i.e., DoDNet [4]). Following the same reported setting[4], DoDNet was trained as our baseline on the same machine for fair comparison.

3.3. Results

Our proposed network achieved competitive results compared with state-of-the-art methods. In **Table 1**, a quantitative comparison between state-of-the-art methods is shown, and we make the following observations:

1. Compared to Multi-Nets, Multi-Head and TAL, our unified network achieves either similar or better segmentation accuracy, but with significantly fewer parameters and lower computational cost;

Table 2. Comparison between TGNet and TGNet-A, and channel-wise attention, spatial attention, serial combination of two attention modules, parallel combination of two attention modules

Methods	Task1:Liver				Task2: Kidney				Task3: Hepatic Vessel				Task4: Pancreas				Task5: Colon		Task6: Lung		Task7: Spleen	
	Dice		HD		Dice		HD		Dice		HD		Dice		HD		Dice	HD	Dice	HD	Dice	HD
	Organ	tumour	Organ	tumour	Organ	tumour	Organ	tumour	Organ	tumour	Organ	tumour	Organ	tumour	Organ	tumour	tumour	tumour	tumour	tumour	Organ	Organ
TGNet	**96.83**	**63.84**	**3.57**	40.57	96.17	81.05	4.67	7.03	**62.59**	**74.65**	13.27	44.67	83.30	**60.71**	7.00	**21.21**	**47.37**	**65.32**	**61.69**	**34.48**	**94.18**	**2.51**
TGNet-A	96.71	63.52	3.84	**35.69**	**96.29**	**81.73**	**4.56**	**6.91**	62.40	74.45	**13.05**	**36.25**	**84.02**	59.25	**6.86**	37.07	46.10	67.65	54.04	52.01	83.77	34.06
channel	**96.71**	**63.52**	**3.84**	35.69	**96.29**	**81.73**	4.56	**6.91**	**62.40**	**74.45**	**13.05**	**36.25**	**84.02**	**59.25**	**6.86**	37.07	**46.10**	**67.65**	54.04	52.01	**83.77**	**34.06**
spatial	95.76	58.03	7.79	48.02	95.24	74.29	4.36	22.08	62.24	68.47	15.59	63.78	78.40	58.47	18.49	**24.96**	39.73	80.96	65.61	49.46	49.68	123.71
serial	95.86	59.69	9.00	**34.90**	95.75	74.34	**3.12**	21.11	61.88	68.71	17.87	70.28	77.88	53.68	16.12	37.41	38.07	81.24	64.24	35.27	80.70	38.90
parallel	95.48	58.49	11.03	42.80	92.72	66.75	14.16	43.20	60.83	67.00	25.27	81.22	77.49	52.23	18.46	34.87	34.67	93.40	**74.33**	**9.92**	42.23	123.53

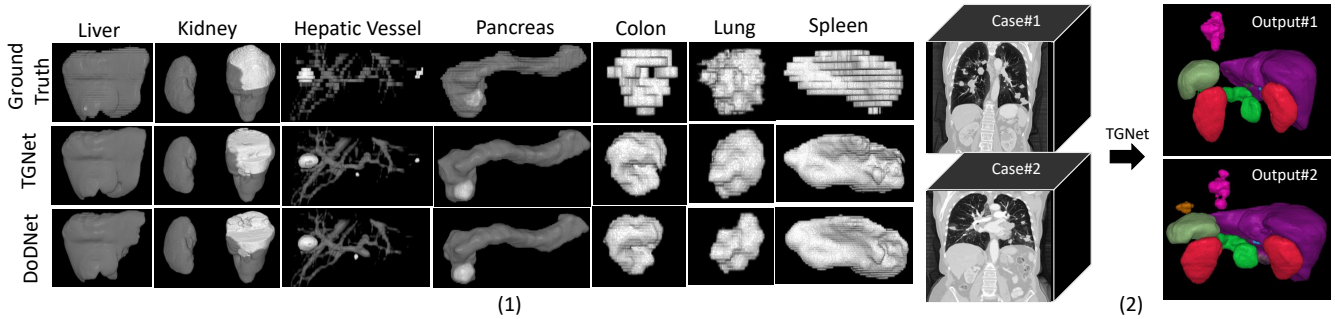

Fig. 2. (1) Visualization comparison between TGNet and DoDNet as unified models on multiple partially labelled segmentation tasks; (2) Segmentation results of multi-organs and tumours on any two cases from partially labelled datasets using TGNet

2. Compared to single networks (Cond-NO, Cond-Input and Cond-Dec), our network generally outperformed them;
3. Compared to DoDNet, our novel task-guided attention modules and residual blocks with only 1.45% increment in parameters produced better results on most tasks. For kidney, our network achieved the best result with respect to tumour segmentation, where dice score increased by 2.91 and HD decreased by 1.56 compared to DoDNet. For hepatic vessel, organ segmentation results were similar but HD for tumour dropped by 6.82. For pancreas, dice score rose by 2.75 and 3.89 for organ and tumour respectively and HD was improved by a big margin (6.23 and 16.82). With respect to colon and lung, two networks produced competitive accuracy. Lastly, dice score increased by 10.28 and HD decreased by 31.86 for spleen;
4. In terms of average scores among all tasks, our network achieved the best result compared to others.

From visualization results (**Fig.2(1).**), as a unified model our network outperformed DoDNet with more accurate segmentation. At the same time, our method could achieve more realistic and precise segmentation of patients' multi-organs and tumours, as shown in **Fig.2(2).**.

3.4. Ablation Study

In **Table 2**, a comparison between our network without the task-guided residual block, namely TGNet-A), and the full TGNet is shown. As we can see, task-guided residual blocks generally increase the accuracy, especially on difficult tasks. In **Table 2** the segmentation results are also shown using different kinds of attention modules in skip connections. All these attention modules were compared without task-guided residual blocks. Spatial attention modules were also guided by tasks. Experimental results showed that task-guided channel-wise attention modules achieved the best overall result and traditional spatial attention could improve the accuracy of one or two tasks but generally decreased the accuracy. Combined serial and parallel attention were tried, however both performed badly compared to task-guided channel attention.

4. CONCLUSION

In this paper, we propose a task-guided network architecture (TGNet) to segment multi-organs and tumours from partially labelled datasets. The newly-devised task-guided attention module and residual block can help to highlight task-related features and suppress task-unrelated features. Extensive experiments conducted on seven benchmark datasets show that these two new mechanisms make our network achieve state-of-the-art performance with only 1.45% increment in the backbone parameters. Future work will focus on integrating task-encoding with attention module in a more efficient and effective way and exploring better use of dynamic filters and task-related information through the encoder and the decoder.

5. COMPLIANCE WITH ETHICAL STANDARDS

This is a retrospective research study on public benchmark medical datasets for which no ethical approval was required.

6. ACKNOWLEDGMENTS

The authors declare no conflict of interest.

7. REFERENCES

[1] Qianfei Zhao, Huan Wang, and Guotai Wang, "Lcov-net: A lightweight neural network for covid-19 pneumonia lesion segmentation from 3d ct images," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2021, pp. 42–45.

[2] Shuchao Pang, Juan Jose Del Coz, Zhezhou Yu, Oscar Luaces, and Jorge Díez, "Combining deep learning and preference learning for object tracking," in *International Conference on Neural Information Processing*. Springer, 2016, pp. 70–77.

[3] Xi Fang and Pingkun Yan, "Multi-organ segmentation over partially labeled datasets with multi-scale feature abstraction," *IEEE Transactions on Medical Imaging*, vol. 39, no. 11, pp. 3619–3629, 2020.

[4] Jianpeng Zhang, Yutong Xie, Yong Xia, and Chunhua Shen, "Dodnet: Learning to segment multi-organ and tumors from multiple partially labeled datasets," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 1195–1204.

[5] Shuchao Pang, Matthew Field, Jason Dowling, Shalini Vinod, Lois Holloway, and Arcot Sowmya, "Training radiomics-based cnns for clinical outcome prediction: Challenges, strategies and findings," *Artificial Intelligence in Medicine*, vol. 123, pp. 102230, 2022.

[6] Hyunseok Seo, Charles Huang, Maxime Bassenne, Ruoxiu Xiao, and Lei Xing, "Modified u-net (mu-net) with incorporation of object-dependent high level features for improved liver and liver-tumor segmentation in ct images," *IEEE transactions on medical imaging*, vol. 39, no. 5, pp. 1316–1325, 2019.

[7] Sihong Chen, Kai Ma, and Yefeng Zheng, "Med3d: Transfer learning for 3d medical image analysis," *arXiv preprint arXiv:1904.00625*, 2019.

[8] Qifeng Chen, Jia Xu, and Vladlen Koltun, "Fast image processing with fully-convolutional networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2497–2506.

[9] Konstantin Dmitriev and Arie E Kaufman, "Learning multi-class segmentations from single-class datasets," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9501–9511.

[10] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[11] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al., "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[13] Amber L Simpson, Michela Antonelli, Spyridon Bakas, Michel Bilello, Keyvan Farahani, Bram Van Ginneken, Annette Kopp-Schneider, Bennett A Landman, Geert Litjens, Bjoern Menze, et al., "A large annotated medical image dataset for the development and evaluation of segmentation algorithms," *arXiv preprint arXiv:1902.09063*, 2019.