Tversky-Kahneman: A New Loss Function For Skin Lesion Image Segmentation

Do-Hai-Ninh Nham, Minh-Nhat Trinh, Van-Truong Pham, Thi-Thao Tran

Abstract This paper proposes a novel loss function inspired by the Tversky-Kahneman probability-weighting function to effectively deal with medical image segmentation tasks. The proposed loss, which is called Tversky-Kahneman loss function, is assessed on the official Skin Lesion datasets of the ISIC 2017 and ISIC 2018 Challenge. To evaluate our new loss function, we propose modified U-Net-based model to obtain quantitative results in Dice score and Jaccard index. Various experiments indicate that our new loss function provides a more promising and time-saving performance than other loss functions.

Key words: Medical Image Segmentation, Tversky-Kahneman, Skin Lesion datasets

1 Introduction

Image segmentation is an important and difficult aspect of image processing as well as computer vision [1]. It has become a hotspot in the field of medical image analysis. A common task in medical image analysis is detecting and segmenting pathological regions that typically occupy a small area of the full image. In recent years, dominant research area in image segmentation has developed with many successful strategies.

Convolutional neural network (CNN) has been a popular method for pixel-wise semantic segmentation. In the field of skin lesion segmentation, CNN has been applied to produce a segmentation mask that specifies the lesion region [2]. A relatively new segmentation method by CNNs is the application of dilated convolutions. In the convolutional layers, the weights are sparsely distributed over a larger receptive

Do-Hai-Ninh Nham
School of Applied Mathematics and Informatics, Hanoi University of Science and Technology

Minh-Nhat Trinh, Van-Truong Pham, Thi-Thao Tran
School of Electrical Engineering, Hanoi University of Science and Technology
e-mail: thao.tranthi@hust.edu.vn

field without losing coverage on the input image [3, 4]. Dilated CNNs are therefore efficient to gain a large receptive field with the usage of a limited number of trainable weights and a limited number of convolutional layers without subsampling layers.

CNNs have also achieved remarkable successes as in [5], Long *et al.* have proposed a fully convolutional networks (FCNs) for semantic segmentation tasks. Conventional fully connected layers in CNNs are replaced with convolutional layers to achieve a coarse label map, before upsampling it with deconvolutional layers to leverage temporal coherence, learn discriminative features of target objects then obtain pixel classification results. Moreover, Conditional Random Fields (CRFs) and CNNs have been combined by Lin *et al.* [6] for a better exploration on spatial correlation between pixels, although they still need to implement a dense CRF to refine the CNN output.

U-Net [7] is another famous image segmentation architecture trained end-to-end for pixel-wise prediction. Skip connections operation is used, which allows propagating dense feature maps from the encoder to the corresponding layers in the decoder. Thus, the output segmentation map is proved to be more accurate, as the spatial information is applied to the deeper layer.

Proof has been indicated that deep learning model efficiency could be enhanced by applying various loss functions. Regarding to classification problems, the $L2$ norm, which is known as mean squared error (MSE) and cross-entropy (CE), are widespreadly used as loss functions [8]. Also, cross-entropy and the Dice loss (DC) have typically been used for segmentation tasks [7], with the main role on extracting features from specific regions [9]. Though experiments show that this could result in good classification and segmentation performance, CE and DC have certain drawbacks in highly-unbalanced class training, due to their assumption on identical importance of distribution of labels.

In recent years, there has been an increasing interest in exploiting the active contour models as loss functions for training the neural networks. Basically, in active contour models, Mumford-Shah functional [10], which was proposed by Mumford and Shah has been inspiring to many developed methods such as active contour model and level-set methods [11], and proximal methods [12] for classical image segmentation. However, in the traditional active contour-based level set approach, the results highly depend on the contour initialization, despite the level set functions could enable the contour to change its topology. Meanwhile, the deep learning based approach can overcome this drawback by eliminating initialization of contour. Progressively encouraging the effectiveness of Mumford-Shah functional and the AC loss with some modifications, we obtain the LMS loss provided by Trinh *et al.* [13]. Even though this loss function has been designed for CMR image dataset segmentation, it is proved to gain quite an auspicious performance on different kinds of clinical images. However, LMS performance depends significantly on the fine tuning of parameters α and β in the level-set formulation and slowly converges.

While Mumford-Shah loss, AC loss and LMS loss focus on edge-preserving filtering method, region-based Tversky loss [14] and Focal Tversky loss [15] prefer to control the information flow implicitly through pixel-level affinity and tackle class-imbalanced problem. However, their convergence rate is witnessed to be not good enough. Hence, we propose a novel regional Tversky-Kahneman loss to not

only address non-lesion tissue, but also progressively encourage the convergence rate during end-to-end training.

In this paper, the main contributions of our work are:

- Proposing a novel loss called Tversky-Kahneman to improve the overall segmentation performance.
- Experimenting on multiple datasets are prerequisites for showing the efficiency of the proposed loss function over other loss functions applying in various models. Due to this respect, experiments are executed on offcial skin lesion datasets from Lesion Boundary Segmentation challenge of ISIC 2017 and ISIC 2018, without external data usage. Tversky-Kahneman loss is proved to express better performances in almost cases.
- Making some modifications based on some U-Net-based models in [20, 24, 26] to customize a new U-Net [7] architecture and to evaluate the performance of our proposed loss as well as provide state-of-the-art results.

2 Related Work

FCN-based semantic segmentation. FCNs [5] frequently generate pixel-wise labelling results with usage of encoder-decoder architecture for high-resolution images. For multi-scale feature fusion, feature pyramid plays as an efficient approach. Pyramid Pooling module has been introduced in PSP-Net [21], which encourages more represensative context information extraction. Furthermore, as Atrous Spatial Pyramid Pooling [22] (ASPP) makes the use of atrous convolution filters [23] at several dilation rates to capture small image information, we utilize it into the model structure for the sake of higher segmentation accuracy.
Channel Attention. Attention mechanisms prefers highlighting the most salient features, as well as avoiding meaningless features that provide favorable segmentation results. Residual Attention Network [16] is proposed by using self-attention mechanisms to effectively seize long-range dependencies. Wang *et al.* [17] create attention maps integrating local and global features by a attention module; while Dual Attention network [18] has been successfully applied in various segmentation contexts. Also, Attention U-Net [19] uses gating signal as cross-attention in the skip connection. Attention Up and Concate block [20] helps feeding attention modules with more essential features, before these modules are concatenated with the max-pooling outputs of decoder layers.
Tversky Loss. Lacking of labels balance, the training process might be led to unwanted local minima convergence, thus predictions might susceptibly face with non-lesion tissue; which results in high-precision and low-recall segmentation outcome. In order to weigh the false-negative (FNs) during segmentation process for imbalanced data, Tversky loss [14] has become useful as the additional level of control over the FNs yields better small scale lesions than other losses.

3 Methodology

3.1 Proposed Network Architecture

As shown in Fig. 1, an input image first goes through a customized U-Net structure, which includes an encoder and a decoder, to generate feature maps for the subsequent segmentation mask prediction. The input images are resized to 192×288, then normalized by rescaling to $[0, 1]$ before being fed into the network. The encoder, which is well-known for extracting salient features, is divided into four downsample deep blocks. Each downsample block contains two sub-blocks, each includes a 2D convolution layer before Mean-Variance normalization layer (MVN) [9] and ReLU activation. While BatchNorm could calculate windowed statistics and switch between accumulating or using fixed statistics, MVN primarily centers and standardizes a single batch at a time. However, MVN is utilized since it is a simple but effective technique that substantially strengthen the learning capacity of the model.

Instead of applying conventional skip connection paths, we replace them by residual skip connections [25], which could provide additional challenging spatial information. In fact, it is experimentally validated that this kind of connection are noticeably advantageous in model convergence, as it skips some layer in the neural network before feeding the output of one layer as the input to the next layer. In our model structure, dataflow passes through a chain of two downsample sub-blocks, before reaching 2×2 Max Pooling layer then being delivered into residual connection. Residual connection is skip connection via addition [27], which is short skip connection; associating with long skip connections of U-Net basement, our model helps access stronger gradient signals thus keeping as much fine-grained details as possible.

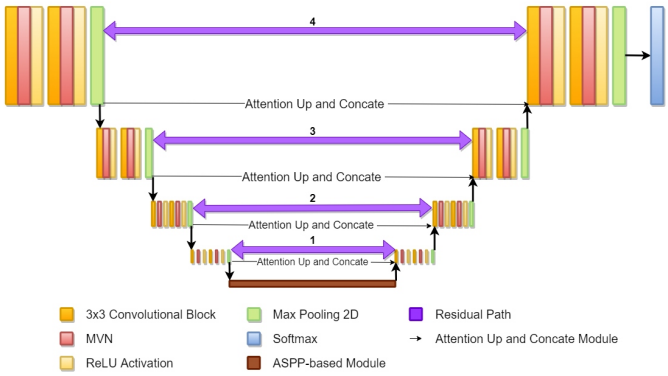

Fig. 1 Our proposed customized U-Net-based Model

With reference to the decoder architecture, it contains four upsample deep blocks and a bottle-neck block including ASPP-based structure. Our version of ASPP contains five parallel operations, these are a series of 1×1 convolution and four

3×3 convolutions with dilation rates set $(1, 6, 12, 18)$. These rates are arranged to capture multi-scale context; also, incorporated image-level features via Global Average Pooling combination. As above, both training and testing images have been cropped to 192×288, by the usage of the feature maps' nominal stride 16, we receive feature vectors of size 12×18. This output feature maps are delivered into a 1×1 convolution with 1024 filters, before being bilinearly upsampled to the initial dimensions and combined into a single vector via concatenation. We also have an adjustment on ASPP structure, that we displace Batch Norm layer by Mean-Variance normalization layer. Accordingly, the feature map outputs of the encoder path, which are symbolized by:

$$E_1, E_2, E_3, E_4 \quad \text{which} \quad E_i \in \mathbb{R}^{H_i \times W_i \times C_i}$$

$$H_i = \frac{1}{2} H_{i-1};\ W_i = \frac{1}{2} W_{i-1};\ C_i = 2^{5+i}; H_i, W_i, C_i \in \mathbb{N}^* \quad i \in [1, 4] \tag{1}$$

, are then passed into skip-connected Attention Up and Concate block at the first stage of each upsample deep block, before going deeper into that upsample deep block to eliminate gradient vanishing and assemble more features. Consequently, the outputs of upsample blocks and ASPP-based block are:

$$D_1, D_2, D_3, D_4, D_5 \quad \text{which} \quad D_i \in \mathbb{R}^{H_i \times W_i \times C_i}$$

$$H_i = \frac{1}{2} H_{i-1};\ W_i = \frac{1}{2} W_{i-1};\ C_i = 2^{5+i} \quad i \in [1, 5] \tag{2}$$

D_{i+1} and E_i are fed into Attention Up and Concate block that:

$$D_{i+1}, E_i \xrightarrow{attention} D'_{i+1}, E'_i \xrightarrow{concatenation} [D'_{i+1}, E'_i] \quad i \in [1, 4]$$

Denote D_0 as the final output of model, providing that D_0 is convolutional output of D_1 that $D_0 \in \mathbb{R}^{H_i \times W_i \times 2}$ with 2 is the number of predicted labels. D_0 is aimed for further comparison with the ground truth.

3.2 Proposed Tversky-Kahneman Loss

In model structure, Softmax activation is applied to form the loss function from the output layer, which includes c planes. c stands for the total number of labels for each pixel to be classified into; and in skin lesion segmentation $c = 2$. Here we imagine that there are totally N pixels for prediction as well as ground truth labels, denote P and L be the predicted set and the ground truth set correspondingly such that $|P| = |L| = N$. Denote p_{ic} and l_{ic} be orderly the element of P and L, with condition that $i \in \{1, 2, ..., N\}$ and $c \in \{0, 1\}$; $p_{ic} \in [0, 1]$; $l_{ic} \in \{0, 1\}$ representing predicted label probability and ground truth labels respectively.

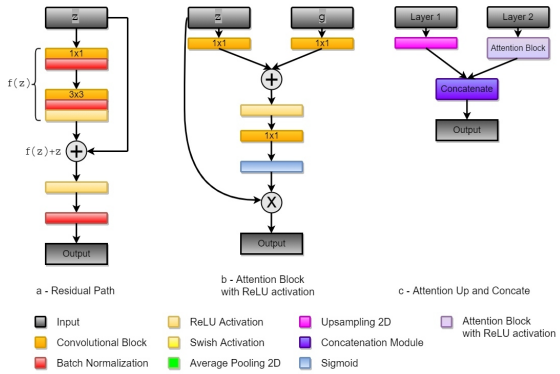

Fig. 2 Sub-Modules applied for proposed architecture

In Expected Utility Theory, Cumulative Prospect Theory accounts for anomalies in the observed behavior of economic agents [28]. Several different probability weighting functions have been suggested to be applied to the cumulative probability of the outcomes, one of them is the Tversky-Kahneman probability weighting function:

$$\omega(z) = \frac{z^{\gamma}}{[z^{\gamma} + (1-z)^{\gamma}]^{\frac{1}{\gamma}}} \tag{3}$$

where $z \in [0, 1]$ is the cumulative probability distribution of gains or losses in a number of economical fields, $\gamma \in (0, 1)$ is a parameter. To be more detailed, z would be the probability distribution of gain or loss if there is only one economical field; otherwise, z is the cumulative probability distribution of gains or losses in several economical fields.

Inspired by this kind of function, we propose a new loss for medical image segmentation, which is also named as Tversky-Kahneman:

$$\Omega(x) = \frac{x^{\gamma}}{[x^{\gamma} + (1-x)^{\gamma}]^{\frac{1}{\gamma}}} \tag{4}$$

subject to

$$x = \frac{\alpha \sum_{i=1}^{N} p_{i1} l_{i0} + \beta \sum_{i=1}^{N} p_{i0} l_{i1}}{0.5 \sum_{i=1}^{N} (p_{i0} l_{i0} + p_{i1} l_{i1}) + \alpha \sum_{i=1}^{N} p_{i1} l_{i0} + \beta \sum_{i=1}^{N} p_{i0} l_{i1}} \qquad i \in \{1, 2, ..., N\} \tag{5}$$

and α, β regulate the penalty amplitude of $p_{i1} l_{i0}$ as the false positive (FP) and $p_{i0} l_{i1}$ as the false negative (FN), respectively. It is noteworthy that these two parameters are tuned for balancing the trade-off between FP and FN, as [14] has proved, with the condition that $\alpha + \beta = 1$. The larger β, the higher weight of recall compared to precision. When $\alpha = \beta = 0.5$, x is simplified as the accuracy loss. And it could be easily seen that $x \in [0, 1]$.

Noticably, $\gamma = 1$ shortens the loss function to $\Omega(x) = x$. Additionally, we observe that $\gamma > 1$ the loss function focuses more on low-level accuracy predictions that have been mis-classified. Furthermore, we also evidence that suppression of the the loss when the class accuracy stays high, usually as the model is swiftly convergent and $\gamma > 2$. This tendency is visualized in Fig. 3 as the increasing trend of x value is mapped to less flatter regions of the Tversky-Kahneman curve with increasing values of γ. Experiments have been conducted with high values of γ till it indicates that the overall result displays the best when $\gamma \in (1, 2)$ and in addition, the best performance is confirmed with $\gamma = \frac{4}{3}$. Thus we train all experiments in case of $\gamma = \frac{4}{3}$.

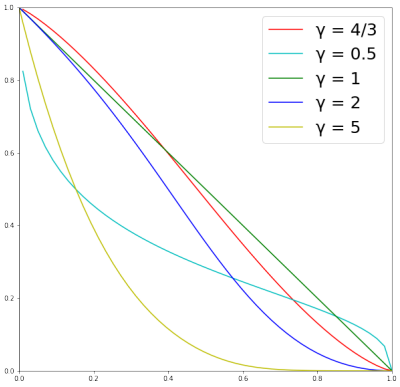

Fig. 3 Tversky-Kahneman Graph Function on different γ cases

3.3 Evaluation Metrics

In medical image analysis, Dice coefficient is a statistical tool for measuring similarity between segmentation maps. The Dice coefficient is calculated by:

$$dice(P, L) = \frac{\sum_{i=1}^{N} 2p_{ic}l_{ic} + \epsilon}{\sum_{i=1}^{N}(p_{ic} + l_{ic}) + \epsilon} \tag{6}$$

The smooth coefficient ϵ is provided for preventing zero division, in experiment we suppose ϵ to be $1e - 10$.

The Jaccard Index is also a statistical tool to gauge the similarity and diversity of sample pixel sets. It is determined by the formula:

$$jaccard(P, L) = \frac{\sum_{i=1}^{N} p_{ic}l_{ic} + \epsilon}{\sum_{i=1}^{N}(p_{ic} + l_{ic} - p_{ic}l_{ic}) + \epsilon} \tag{7}$$

4 Experimental Results

4.1 Datasets

We experiment our loss on the official ISIC Skin Lesion 2017 and 2018 Dataset. In the ISIC 2017, there are 2000 images for training data and blind held-out testing data with 600 images. While in ISIC 2018, there are 2594 images officially with 80-20 train-test split in our experiments.

4.2 Training

We have performed our customized U-Nets with Tversky-Kahneman loss layer to segment multiple skin lesions. Our model is trained end-to-end, with cost minimization on several epochs (base on different cases) is performed by using NADAM optimizer [29] with an initial learning rate of 0.001. Learning rate is multiplied by 0.5 every 10 epochs, before reaching 0.00001 and being constantly kept through the remainder training period. All images are all rigidly pre-processed by cropping center and normalization, before being augmented by randomly flipping horizontally and vertically, and rotated within $\frac{\pi}{6}$ radians. The training time for each of these networks is approximately 4 hours at maximum on a workstation with NVIDIA Tesla P100 16GB GPU.

4.3 Representative Results

In this section, we provide quantitative results produced by our proposed model and Tversky-Kahneman loss function for some difficult input cases. As Fig. 4 displays, our loss and model have obtained quite a high-quality performance on inputs with complexity.

4.4 Evaluation

To evaluate the efficiency of our new Tversky-Kahneman loss function before making comparison with other loss functions in skin lesion segmentation, we trained our U-Net model with several parameter pairs of α and β. The performance outcomes are displayed in Table 1. From Table 1, it is obviously that according to all combined test measures, the best performance is observed from the customized U-Net trained with $\alpha = \beta = 0.5$.

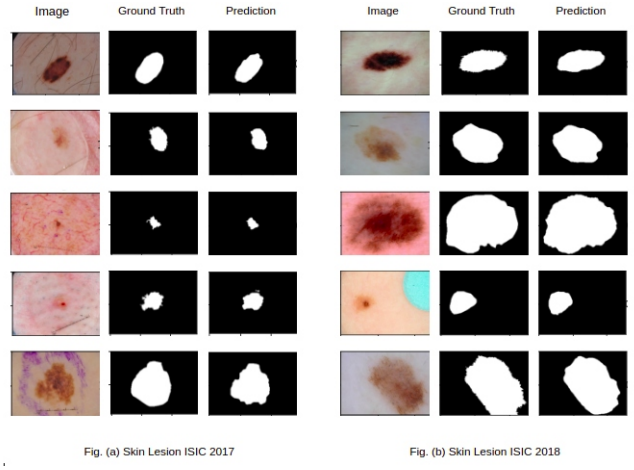

Fig. 4 Representative segmentation performances of our proposed model and Tversky-Kahneman loss for some difficult input cases of (a) Skin Lesion ISIC 2017, and (b) Skin Lesion ISIC 2018 datasets

Dataset	Penalties	Dice Coefficient	Jaccard Index
Skin Lesion ISIC 2017	$\alpha = 0.5, \beta = 0.5$	**0.854**	**0.767**
	$\alpha = 0.4, \beta = 0.6$	0.847	0.763
	$\alpha = 0.3, \beta = 0.7$	0.843	0.757
	$\alpha = 0.2, \beta = 0.8$	0.844	0.757
	$\alpha = 0.1, \beta = 0.9$	0.842	0.756
Skin Lesion ISIC 2018	$\alpha = 0.5, \beta = 0.5$	**0.886**	**0.813**
	$\alpha = 0.4, \beta = 0.6$	0.853	0.771
	$\alpha = 0.3, \beta = 0.7$	0.881	0.808
	$\alpha = 0.2, \beta = 0.8$	0.864	0.789
	$\alpha = 0.1, \beta = 0.9$	0.830	0.751

Table 1 Performances under variety of α and β parameters

To present a fairer evaluation of our proposed U-Net model and the proposed Tversky-Kahneman loss and to prove our results being state-of-the-art, we validate 8 different cases of variations each skin lesion dataset. Ablation test results are recorded in Table 2. As Table 2 demonstrates, it is observable that training with Tversky-Kahneman could obtain much effective and accurate segmentation results. To be more detailed, in the ISIC 2017 Skin Lesion Dataset the best performance exceeds others by 0.3% to 1.5%, while in the ISIC 2018 Skin Lesion Dataset the best outpaces others by 0.3% to 3.9%.

From each dataset, top 5 loss functions with best accuracy are selected to observe training processes through loss function graphs. Obiviously, the ISIC 2017 Dataset observes the best performances by Tversky-Kahneman, Tversky, Accuracy, Mumford-Shah and Active Contour loss functions. Whilst with the ISIC 2018 Dataset, top 5 are Tversky-Kahneman, LMS, Active Contour, Mumford-Shah and BCE Dice loss functions. Therefore, we display the experimental training process as Fig. 5 follows.

Datasets	Performance Metrics	Dice Coefficient		Jaccard Index
		Result	Epoch	
Skin Lesion ISIC 2017	Binary Cross Entropy (BCE) + Dice loss	0.840	19-22	0.756
	Accuracy loss	0.849	29-32	0.766
	Tversky loss ($\alpha = 0.3, \beta = 0.7$)	0.851	40-43	0.766
	Focal Tversky loss ($\alpha = 0.3, \beta = 0.7$)	0.839	38-40	0.752
	Mumford-Shah loss	0.848	18-20	0.763
	Active Contour loss	0.847	19-21	0.762
	LMS loss	0.844	30-32	0.758
	Tversky-Kahneman loss (ours)	**0.854**	46-48	**0.767**
Skin Lesion ISIC 2018	Binary Cross Entropy (BCE) + Dice loss	0.861	18-20	0.781
	Accuracy loss	0.860	20-23	0.779
	Tversky loss ($\alpha = 0.3, \beta = 0.7$)	0.847	32-35	0.764
	Focal Tversky loss ($\alpha = 0.3, \beta = 0.7$)	0.854	59-62	0.773
	Mumford-Shah loss	0.874	46-48	0.802
	Active Contour loss	0.876	46-49	0.803
	LMS loss	0.883	40-42	0.809
	Tversky-Kahneman loss (ours)	**0.886**	46-48	**0.813**

Table 2 Skin Lesion segmentation performances on different types of loss functions. Here the "Epoch" column indicates the range epoch number for the Dice Coefficient and the Jaccard Index reaching their peaks

As Fig. 5 evidences, the Tversky-Kahneman line is proved to be the most effective, the smoothest and the most fastly-convergent, out of all the other loss functions.

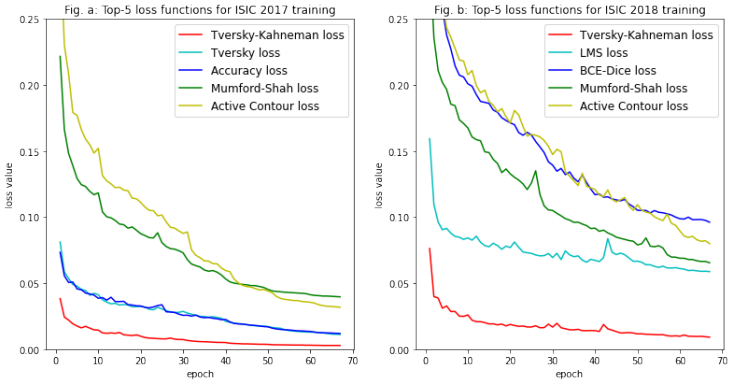

Fig. 5 Experimental Training Process with top 5 loss functions in each dataset

5 Conclusion

We have introduced a new loss function based on the Tversky-Kahneman probability weighting function, to achieve a more balanced trade-off between precision and recall also a better convergence rate in segmentation. Moreover, our proposed loss layer has been added to a customized U-Net network to produce state-of-the-art performances (0.854 Dice Coefficient Score and 0.767 Jaccard Index Score on ISIC 2017 Dataset and 0.886 Dice Coefficient Score and 0.813 Jaccard Index Score on ISIC 2018 Dataset). Our experimental results in skin lesion segmentation evidently indicate that all performance evaluation metrics on the test data using the Tversky-Kahneman loss layer further exceed other performances provided by other kinds of loss layer. Even though parameters α and β have been weighted equally, our proposed approach still obtains better outcomes when comparing to the latest results in skin lesion segmentation. In the future, we will discover more about the relationship between parameters α, β on different models, as well as work on complex segmentation datasets to compare against state-of-the-art results using relevant criteria.

Acknowledgement

This research is funded by the Hanoi University of Science and Technology (HUST) under project number T2021-PC-005.

References

1. Szeliski, Richard: Computer Vision: Algorithms and Applications, Springer-Verlag (2010)
2. Jafari, M.H. and Karimi, N. and Nasr-Esfahani, E. and Samavi, S. and Soroushmehr, S.M.R. and Ward, K. and Najarian, K.: Skin lesion segmentation in clinical images using deep learning (2016), *2016 23rd International Conference on Pattern Recognition (ICPR)* doi: 10.1109/ICPR.2016.7899656
3. Wolterink, Jelmer M. and Leiner, Tim and Viergever, Max A. and Išgum, Ivana: Dilated Convolutional Neural Networks for Cardiovascular MR Segmentation in Congenital Heart Disease, *Lecture Notes in Computer Science* doi: 10.1007/978-3-319-52280-7-9
4. Fisher Yu and Vladlen Koltun: Multi-Scale Context Aggregation by Dilated Convolutions, *arXiv* (2016)
5. Jonathan Long and Evan Shelhamer and Trevor Darrell: Fully Convolutional Networks for Semantic Segmentation, *arXiv* (2015)
6. Guosheng Lin and Chunhua Shen and Anton van dan Hengel and Ian Reid: Efficient piecewise training of deep structured models for semantic segmentation, *arXiv* (2016)
7. Olaf Ronneberger and Philipp Fischer and Thomas Brox: U-Net: Convolutional Networks for Biomedical Image Segmentation, *arXiv* (2015)
8. Xiao-Yun Zhou and Mali Shen and Celia V. Riga and Guang-Zhong Yang and Su-Lin Lee: Focal FCN: Towards Small Object Segmentation with Limited Training Data, *CoRR* (2017)
9. Phi Vu Tran: A Fully Convolutional Neural Network for Cardiac Segmentation in Short-Axis MRI, *arXiv* (2017)

10. Mumford, David and Shah, Jayant: Optimal approximations by piecewise smooth functions and associated variational problems, pp. 577-685, *Communications on Pure and Applied Mathematics* (1989)
11. Chan, T.F. and Vese, L.A.: Active contours without edges, pp. 266-277, *IEEE Transactions on Image Processing* doi: 10.1109/83.902291 (2001)
12. Chambolle, Antonin and Pock, Thomas: A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging, *Journal of Mathematical Imaging and Vision* doi: 10.1007/s10851-010-0251-1 (2011)
13. M.N. Trinh and N.T. Nguyen and T.T. Tran and V.T. Pham: A Semi-Supervised Deep Learning-based Approach with Multiphase Active Contour Loss for Left Ventricle Segmentation from CMR Images, *The 3rd International Conference on Sustainable Computing SUSCOM-2021* (2021)
14. Seyed Sadegh Mohseni Salehi and Deniz Erdogmus and Ali Gholipour: Tversky loss function for image segmentation using 3D fully convolutional deep networks, *arXiv* (2017)
15. Nabila Abraham and Naimul Mefraz Khan: A Novel Focal Tversky loss function with improved Attention U-Net for lesion segmentation, *arXiv* (2018)
16. Fei Wang, Mengqing Jiang, Chen Qian, Shuo Yang, Cheng Li, Honggang Zhang, Xiaogang Wang, Xiaoou Tang: Residual Attention Network for Image Classification, *CoRR* (2017)
17. Wang, Yi and Ni, Dong and Dou, Haoran and Hu, Xiaowei and Zhu, Lei and Yang, Xin and Xu, Ming and Qin, Jing and Heng, Pheng-Ann and Wang, Tianfu: Deep Attentive Features for Prostate Segmentation in 3D Transrectal Ultrasound, pp. 2768-2778, *IEEE Transactions on Medical Imaging* doi: 10.1109/tmi.2019.2913184 (2019)
18. Jun Fu and Jing Liu and Haijie Tian and Yong Li and Yongjun Bao and Zhiwei Fang and Hanqing Lu: Dual Attention Network for Scene Segmentation, *arXiv* (2019)
19. Jo Schlemper and Ozan Oktay and Michiel Schaap and Mattias Heinrich and Bernhard Kainz and Ben Glocker and Daniel Rueckert: Attention Gated Networks: Learning to Leverage Salient Regions in Medical Images, *arXiv* (2019)
20. Ninh, Quoc Cuong and Tran, Thi-Thao and Tran, Tien Thanh and Anh Xuan Tran, Thi and Pham, Van-Truong: Skin Lesion Segmentation Based on Modification of SegNet Neural Networks, pp. 575-578, *2019 6th NAFOSTED Conference on Information and Computer Science (NICS)* doi: 10.1109/NICS48868.2019.9023862 (2019)
21. Hengshuang Zhao and Jianping Shi and Xiaojuan Qi and Xiaogang Wang and Jiaya Jia: Pyramid Scene Parsing Network, *arXiv* (2017)
22. Liang-Chieh Chen and George Papandreou and Iasonas Kokkinos and Kevin Murphy and Alan L. Yuille: DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs, *arXiv* (2017)
23. Liang-Chieh Chen and George Papandreou and Florian Schroff and Hartwig Adam: Rethinking Atrous Convolution for Semantic Image Segmentation, *arXiv* (2017)
24. Baheti, Bhakti and Innani, Shubham and Gajre, Suhas and Talbar, Sanjay: Eff-UNet: A Novel Architecture for Semantic Segmentation in Unstructured Environment, pp: 1473-1481, *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* doi: 10.1109/CVPRW50498.2020.00187 (2020)
25. Ange Lou and Shuyue Guan and Murray Loew: DC-UNet: Rethinking the U-Net Architecture with Dual Channel Efficient CNN for Medical Images Segmentation, *arXiv* (2020)
26. M.N. Trinh and N.T. Nguyen and T.T. Tran and V.T. Pham: A Deep Learning-based Approach with Image-driven Active Contour Loss for Medical Image Segmentation, *The 2nd International Conference on Data Science and Applications ICDSA 2021*
27. Nikolas Adaloglou: Intuitive Explanation of Skip Connections in Deep Learning. `https://theaisummer.com/skip-connections/.Cited03.23.2020`
28. Ingersoll and Jonathan: Non-Monotonicity of the Tversky-Kahneman Probability-Weighting Function: A Cautionary Note, pp. 385-390, *European Financial Management* doi: 10.1111/j.1468-036X.2007.00439.x (2008)
29. Timothy Dozat: Incorporating Nesterov Momentum into Adam (2016)