Technical Report on Kappa Loss

Jing Zhang

October 2019

Abstract

Skin melanoma represents a major health issue. Today, diagnosis and follow-up maybe performed thanks to computer-aided diagnosis tool, to help dermatologists segment and quantitatively describe the image content. In particular, deep convolutional neural networks (CNN) have lately been become the state-of-the-art in medical image segmentation. The loss function plays an important role in CNN in the process of backpropagation. In this work, we propose a metric-inspired loss function, based on the Kappa index. Unlike the Dice loss, the Kappa loss takes into account all the pixels in the image, including the true negative, which we believe can improve the accuracy of the evaluation process between prediction and ground truth. We demonstrate the differentiability of the Kappa loss and present some results on six public datasets of skin lesion. Experiments have shown promising results in skin lesion segmentation.

1 Introduction

Melanocytes are cells located in the bottom layer of the skin's epidermis. They can be the starting point for melanoma, a type of skin cancer. If melanoma is not detected and treated within limited time, this disease may be fatal, since it can spread to other organs quickly. According to a survey [1], in 2019, about 7230 people will lose their lives because of melanoma in the USA. Dermatologists establish their diagnosis by visual inspections of moles, and by extracting texture, size and shape analysis information. One first step is often the segmentation of the mole. Manual segmentation by dermatologists as described in [9] is a time-consuming process, hardly compatible with the usual workload of medical experts, and that can be subjective. Today, computer-aided diagnosis tool may help to segment the images, that can rely on thresholding, active contours and region merging [10, 9]. However, these computer vision based methods are sensitive to the images quality and characteristics, such as illumination and noise. Thus automated approaches have been proposed in last decades, based on machine learning, and more recently deep learning, with the use of deep convolutional neural networks (CNN). CNN nowadays are popular topics in the field of computer vision. They are widely used in the image classification, object detection, image segmentation and so on. Because CNN are

trained in the way of end-to-end, the features of objects can be learned automatically. Thus, using this method can have better results than the traditional ones [4]. In image segmentation, each single pixel of an image has to be classified into specific classes. One key component of CNN is the loss function, that drives the backpropagation of the error between the predicted value and and the reference label. Cross-entropy is a widely used, standard loss function. In order to handle class imbalance, weights can be assigned to samples of different classes. The Dice loss function [12], derived from the well-known Dice metric, is specifically designed for image segmentation. However, the Dice loss only considers foreground (i.e. object) pixels, and does not take into account the background pixels in the image.

In this work, we propose a loss function called Kappa loss based on Kappa index that can not only deal with class imbalance problems in medical image segmentation, but also considers the whole information of an image. The motivating factor spurring our approach is rooted in the fact that all pixels should be taken into account, since a large part of the image is occupied by object (or in our case melanoma) pixels. We believe that by using the Kappa loss, we will enforce the constrain on the true negative pixels in addition to the true positive ones, just reaching a better balance between the two classes.

Our paper is organized as follows. In Section 2, we present related works in image segmentation. We define the Kappa loss and show its derivability in Section 3. Experimental settings and results are presented in Section 4, and we conclude in Section 5.

2 Related works

Below we present some related works in skin lesion image segmentation, then we focus on the varieties of loss functions that have been proposed in the literature of deep learning based image segmentation.

Skin lesion image segmentation. Computer vision based image segmentation methods have been thoroughly investigated before the advent of deep learning. [26] uses active contour to detect border of skin lesion. In [28], the authors combine gradient vector flow with mean shift to segment the skin lesion images. Prepossessing is performed in [19] to tackle the problem of low contrast and color between background and object. [15] presents Delaunay Triangulation to extract the contour of skin lesion image. Region merging based approach is used in [24]. In [10], authors apply image processing tools to extract features (Asymmetry, Border, Color, Diameter) of skin lesion in order to classify the image as melanoma or not. For the skin lesion segmentation with deep learning methods, the early work of [9] includes 3 steps: preprocessing (image filtering), CNN, post processing (selection). In [2], the authors spend effort on aggregating convolutional and recurrent neural networks. In the work of [27], authors use CNN with a proposed loss function based on Jaccard distance that can deal with class imbalance problem.

Loss functions for image segmentation. The loss function plays a key

role in CNN, which is to measure the consistency between the predicted value and and the reference label. It influences the convergence speed and the accuracy of segmented results. Cross-entropy and Weighted Cross Entropy [13] have been widely used. WCE is used to balance the classes, in case of class imbalance. However, these per-pixel losses have been shown to be outperformed by other, "metric-sensitive" losses, based on metrics such as the Dice or the Jaccard index [3], for image segmentation, especially in the medical imaging field. The Dice loss [12] has been established as a popular choice for binary image segmentation. The Generalised Dice loss is introduced in [20] to handle multiple classes. The Tversky loss [18] has been proposed to achieve a better tradeoff between precision and recall. Another metric inspired loss function has recently been proposed, relying on the Hausdorff distance, the maximum point-to-point distance between two contours, called the "weighted Hausdorff distance" [16], but for object localization. A loss function called "contour loss" that takes into account distance information via the distance map of the ground truth, has shown interesting smoothing effect in a 3D segmentation setting [11].

Fine tuning of loss functions. It has been shown that changing the loss during the training process, can help to make the backpropagation more stable. Dynamic losses have thus been proposed: In the work of [23], the authors improve the Focal loss by converting two hype parameters into one parameter that can automatically adapt during the training progress. Another idea of dynamic loss in [29] is that they initialize the training with Cross entropy loss, and then training with Focal loss in the application of medical image segmentation. Similarly, in the work of [21], in which they combined Cross entropy and Weighted cross entropy.

3 Methodology

3.1 From metrics to loss

The Dice index (DI) is widely known as an overlap measure in binary image segmentation, defined as the ratio between twice the intersection of two regions over their union. The DI values range from 0 to 1, 0 means that there is no intersection while 1 indicates perfect match.

The Dice Index was originally designed to be an inter-rater agreement [7], independently from the pixel labeling problem. Let us define as a the number of counts where raters agree positively, d the number of counts where raters agree negatively, and b and c where raters disagree, N is the sum of a, b, c and d (see Tab. 1). Fig. 1 shows the interpretation of number of counts with respect to the segmentation problem. In the image segmentation, we regard the two raters as ground truth and the predicted image, as shown in Fig.1. More specifically, the element a is foreground (represented by positive "+") shared by both of ground truth and predicted area, b is foreground of predicted image and background of ground truth, c is foreground of ground truth and background of predicted image, d is background (represented by negative "-") shared by both of ground

Table 1: Counts of agreement and disagreement from two raters. $a + d$ is the number of targets for which two raters agreed, $b + c$ is the number of targets for which they disagreed, $N = a + b + c + d$.

		Ground Truth (Rater 1)		
		+	-	Total
Predicted	+	a	b	$a + b$
Results	-	c	d	$c + d$
(Rater 2)	Total	$a + c$	$b + d$	N

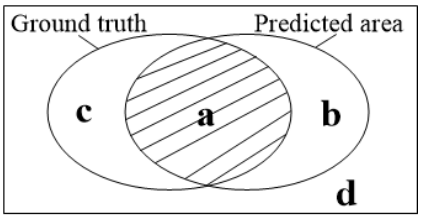

Figure 1: Ground truth contour vs predicted contour, with a, b, c the number of pixels included in both contours, only in the predicted area, only in the ground truth, respectively.

truth and predicted area. One can write the DI as:

$$DI = \frac{2a}{2a + b + c}$$

Note that the d counts where raters disagree negatively are not taken into account in this definition. Other agreement rates can take into account these true negative. In particular, the κ coefficient, a chance-corrected measure of agreement voted by two raters, is defined as:

$$\kappa = \frac{2(ad - bc)}{(a + b)(b + d) + (a + c)(c + d)} \tag{1}$$

As recalled in the pioneering paper [30] that first uses the Dice index as a metric to evaluate segmentation quality, the Dice index is a limit case of the Kappa index κ when $d \gg a, b, c$:

$$\lim_{d \to \infty} \kappa = \frac{2a}{2a + b + c}, \tag{2}$$

The Dice index only considers the foreground pixels to compute the overlap of the predicted region and the ground truth, based on the assumption that region or object pixels are small compared to the background area.

However in some cases, especially in medical skin images, this assumption does not hold. We show examples of such cases, in the 3 images on the right in

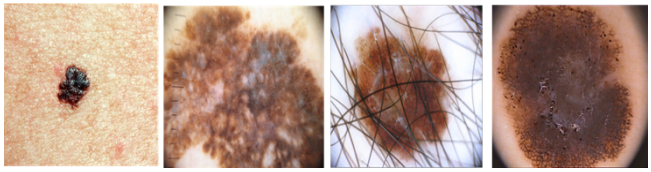

Figure 2: Samples of skin images: 1st column is from dataset Skin-Cancer-Detection, 2nd column from ISIC 2017, 3rd&4th column from ISIC 2018. In the skin image of 2-4th column, the object is large with respect to the image.

Fig 2. This is the rationale behind the use of the kappa index as loss function: all pixels in the image are taken into account, and not only the foreground pixels.

3.2 Kappa loss

In order for the Kappa coefficient to be used as a loss function in a CNN, it has to be differentiable so that its gradient may be computed. Because the Kappa loss has to be differentiable in order to be able to compute the gradient of this loss with respect to the parameters of the network. Thus the probabilities (i.e. output values of the last layer) have to be used, instead of hard labels, in the definition of the Kappa loss. We rewrite a, b, c and d by taking into account the predicted segmentation (or probability) at pixel i, denoted p_i, and the ground truth at this same pixel, denoted g_i:

$$\begin{aligned} a &= \sum_{i=1}^{N} (p_i g_i), \\ b &= \sum_{i=1}^{N} (1 - p_i) g_i, \\ c &= \sum_{i=1}^{N} (1 - g_i) p_i, \\ d &= \sum_{i=1}^{N} (1 - p_i)(1 - g_i) \end{aligned} \tag{3}$$

where N is the number of pixels in the image. We obtain the soft approximation of the Kappa loss by replacing the affectations from Eq. (3) in Eq. (1):

$$\text{Kappa loss} = 1 - \frac{2 \sum_{i=1}^{N} p_i g_i - \sum_{i=1}^{N} p_i \cdot \sum_{i=1}^{N} g_i / N}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} g_i - 2 \sum_{i=1}^{N} p_i g_i / N} \tag{4}$$

Deriving the Kappa loss with respect to predicted probabilities at pixel j,

the gradient of Kappa loss is:

$$\begin{aligned}\frac{\partial K}{\partial p_j} = &-2\frac{g_j(\sum g_i + \sum p_i - 2\sum p_i \sum g_i/N)}{(\sum p_i + \sum g_i - 2\sum p_i g_i/N)^2} \\ &+ \frac{\sum p_i g_i(1 - 2\sum g_i/N^2) + (\sum g_i)^2/N)}{(\sum p_i + \sum g_i - 2\sum p_i g_i/N)^2}\end{aligned} \tag{5}$$

In the case of Dice loss, Eq (4) of Kappa loss simplifies down to [12]:

$$\text{Dice loss} = 1 - \frac{2\sum_{i=1}^{N} p_i g_i}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} g_i} \tag{6}$$

For this Dice loss, variants of this definition may have $p_i^2 + g_i^2$ instead of $p_i + g_i$ in the denominator, or include a smoothness term (a small value) added to the denominator and the numerator [20, 25, 14], but that only helps in case of missing labels and is not critical.

3.3 CNN for image segmentation

We have used the well-known U-Net architecture, one of the most popular CNN for medical image segmentation [17], to implement the Kappa loss function. The U-Net is fully convolutional network with an encoder-decoder architecture and skip connections. Compared to the original architecture, we simplified the network, given the limited amount of images. The original U-Net includes 64 filters at the first level, for a total of 31,031,685 parameters. We set the initial number of filters to be 16 (3×3), so the number of parameters is 1,946,449. We also add a batch normalization operation after each ReLU activation function to avoid gradient vanishing. In order to avoid overfitting, we use drop out in the U-Net with a rate of 0.5.

4 Experiments

4.1 Datasets

We use 6 publicly available datasets of skin images with mole or melanoma, to assess the proposed Kappa-based loss function: Skin-Cancer-Detection (SCD, 206 images, supplied by Vision and Image Processing Lab, University of Waterloo), split into two subsets which are melanoma (Mel, 119 images) and not-melanoma (Non-mel, 87 images), and the datasets from International Skin Imaging Collaboration: ISIC 2018 [6, 22], 2594 images; ISIC 2017 [5], 2000 images; ISIC 2016 [8], 900 images. In the latter, images not only include a lesion part but also present noise such as hair, which increases the difficulty of segmentation. Moreover, the object (lesion area) in the image of dataset SCD or ISIC 2018 varies from one to another in size. We split each dataset into training set, validation set and test set, respectively, with the same amount of images in each set. Images are resized to 256 × 256.

Table 2: Averaged Dice index (DI) and Hausdorff distance (HD) values (± standard deviation), for Dice loss and Kappa loss on 6 different datasets (87, 119, 206, 900, 2000, 2594 images respectively).

	Dice loss		Kappa loss		p-value	
dataset	DI	HD	DI	HD	DI	HD
Non-mel	0.65±0.11	5.06±1.79	0.73±0.11	4.70±2.02	0.006	0.012
Mel	0.80±0.06	6.70±1.93	0.81±0.03	6.59±1.88	0.289	0.424
SCD	0.82±0.04	7.94±1.72	0.83±0.03	7.91±1.68	0.273	0.735
ISIC-16	0.80±0.05	8.42±2.19	0.84±0.01	8.41±2.25	0.000	0.875
ISIC-17	0.80±0.05	8.07±1.93	0.84±0.05	8.03±1.94	0.067	0.223
ISIC-18	0.81±0.03	7.59±2.60	0.82±0.04	7.52±2.66	0.027	0.066

4.2 Experimental settings

Data augmentation including rotation, shifting, shearing, zooming and flipping is used in training. 3-fold cross validation is performed. The optimizer is Adam with a learning rate of $1e^{-4}$. The batch size is 4. Evaluation metrics are the Dice index (DI) and the Hausdorff distance (HD), which is the maximum point-to-point distance between two contours. The implementation tool is based on Keras and Tensorflow 1.0. Tesla P100 GPU server is used in the experiments.

4.3 Results

4.3.1 The results of Kappa loss

We trained the U-Net described in the previous section from scratch on the 6 datasets independently, with two different loss functions, the Dice loss functions that will be the baseline, and the Kappa loss functions. Results are shown in Table 2. The Hausdorff distances between Kappa loss and Dice loss are similar, which means that Kappa is not correcting distant, false positive pixels, except for the first dataset (Non-mel), where HD drops by 7%. We performed p-value of t-test on Dice index and Hausdorff distance respectively. The t-test can be used to check if the means of two sets of data are significantly different from each other with given significant level $\alpha = 0.05$.

However, substantial improvement on the Dice Index (DI) is obtained for the Kappa loss, in comparison to the Dice loss, for several datasets. Looking at Fig. 4, we can also observe that the Kappa loss converges faster than Dice loss under the same U-Net model settings. In addition, we set 200 epochs to see the loss and Dice index changing tendency (Fig. 5), we can know that the results are similar to the curve when epochs is 100. Some segmentation results are shown in Fig.3, which shows that in some cases, the Kappa loss can help to make the segmentation more accurate.

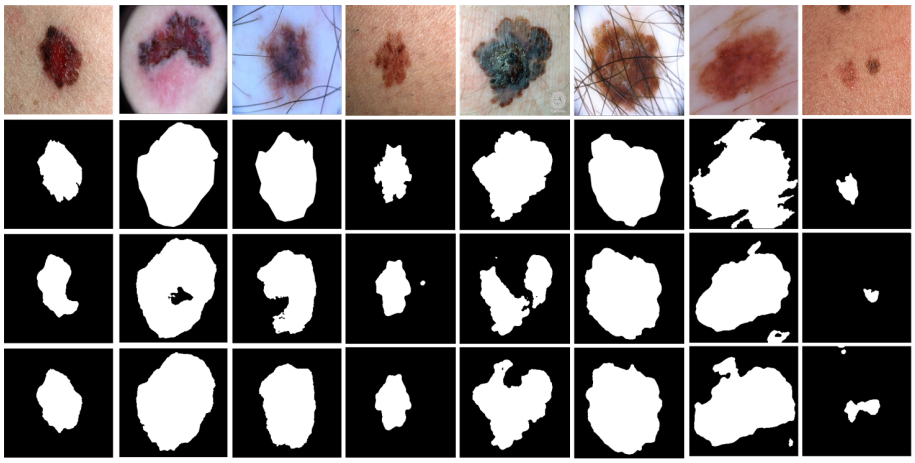

Figure 3: Examples of segmentation result. For each row from top to bottom: skin lesion image, ground truth, segmentation result with Dice loss, and result with Kappa loss.

4.3.2 Fine-tuning the loss function

In the last part of our experiment, we fine-tune our model by combining another loss functions, we chose three data sets SCD, ISIC 2016 and ISIC 2018. In Table 3, there are 4 combinations of loss functions, Cross Entropy with Dice, Cross Entropy with Kappa, they are the combinations of two different natures of loss functions. Also Dice with Cross Entropy, Kappa with Cross Entropy, the opposite order of loss functions. We list the performance of loss function after fine tuning. The training epochs are 100. In two loss function combination, each model was trained for 50 epochs. We can see from the Tab. 3, the fine-tuning results of Dice loss after Cross entropy loss and Kappa loss after Cross entropy outperform the corresponding static loss function in baseline. This demonstrates that the dynamic loss functions in the form of Cross entropy loss + Dice loss and Cross entropy loss + Kappa loss have the promising performance.

5 Conclusion

In this work, we have proposed a new loss function, based on the Kappa index, to be used in CNN for medical image segmentation. Different from the Dice loss, this loss function considers all pixels (background pixels included) in the evaluation of the predicted segmentation. We believe that by enforcing constraint on both positive and negative pixels, segmentation accuray or converge will be improved. We have shown the Kappa loss diffentiability and have used the state-of-the-art U-Net architecture to implement it. We compared the Kappa loss quantitatively to the Dice loss on several public datasets of melanoma and skin segmentation. Promising results were obtained, showing the potential of

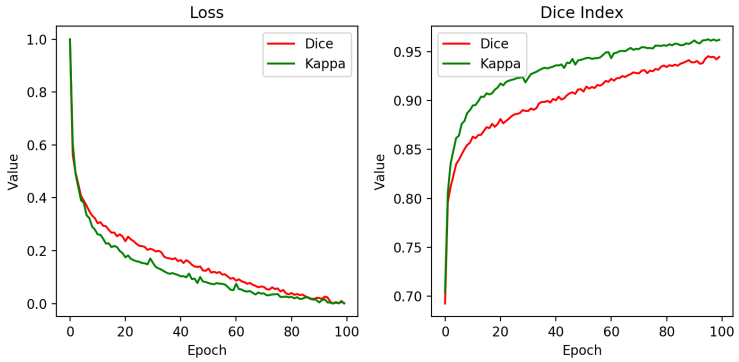

Figure 4: Loss function (left) and DI metric (right) during the 100 epochs of training process on the dataset ISIC 2016.

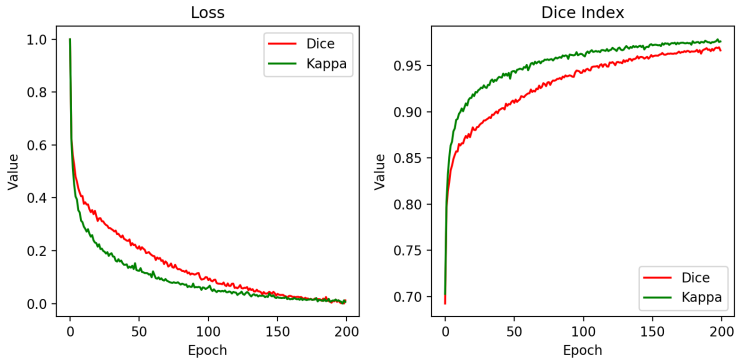

Figure 5: Loss function (left) and DI metric (right) during the 200 epochs of training process on the dataset ISIC 2016.

the Kappa loss. Future work involves extending our benchmarking experiments to other loss functions, to further investigate the behavior of Kappa loss with respect to other loss functions. and generalizing the Kappa loss to multilabel image segmentation as was proposed for the generalized Dice loss in [20]. Kappa loss could also be used in multiscale approaches, when segmentation is required inside a region of interest (e.g. bounding box), where there is a balance between positive and negative pixels. We also plan to investigate the CNN feature maps to understand how the Kappa loss acts to improve or speed up the process, when compared to the Dice loss.

Acknowledgments. The authors would like to thank the China Scholarship Council (CSC) for supporting Jing Zhang and acknowledge the CRIANN (Centre des Ressources Informatiques et Applications Numérique de Normandie,

Table 3: Average IoU (±standard deviation) on SCD (206 images), ISIC 2019 (900) and ISIC 2018 (2594) data set. Baseline: a single loss is used for training; Fine-tuning: a combination of 2 loss functions is used. Best results are in bold.

Loss functions	SCD	ISIC 2016	ISIC 2018
Baseline (single loss)			
Cross Entropy(CE)	0.78±0.01	0.79±0.03	0.77±0.07
Dice	0.82±0.04	0.80±0.04	0.81±0.03
Kappa	0.83±0.03	0.84±0.01	0.82±0.04
Fine-tuning (2 losses)			
Dice+CE	0.75±0.07	0.78±0.06	0.78±0.06
CE+Dice	**0.83±0.03**	0.78±0.02	0.81±0.03
CE+Kappa	**0.82±0.04**	0.77±0.06	0.82±0.04
Kappa+CE	0.78±0.06	0.76±0.07	0.77±0.07

France) for providing computational resources.

References

[1] C Halpern Allan, A Marghoob Ashfaq, and Reiter Ofer. *Melanoma Overview*, 2019 (accessed October 15, 2019). `https://www.skincancer.org/skin-cancer-information/melanoma/`.

[2] Mohamed Attia, Mohamed Hossny, Saeid Nahavandi, and Anousha Yazdabadi. Skin melanoma segmentation using recurrent and convolutional neural networks. In *IEEE ISBI*, pages 292–296, 2017.

[3] Jeroen Bertels, Tom Eelbode, Maxim Berman, Dirk Vandermeulen, Frederik Maes, Raf Bisschops, and Matthew B. Blaschko. Optimizing the dice score and jaccard index for medical image segmentation: Theory and practice. In *MICCAI (2)*, volume 11765 of *LNCS*, pages 92–100. Springer, 2019.

[4] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *CoRR*, abs/1606.00915, 2016.

[5] Noel C. F. Codella, David Gutman, M. Emre Celebi, Brian Helba, Michael A. Marchetti, Stephen W. Dusza, Aadi Kalloo, Konstantinos Liopyris, Nabin K. Mishra, Harald Kittler, and Allan Halpern. Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (ISIC). *CoRR*, abs/1710.05006, 2017.

[6] Noel CF Codella, David Gutman, M Emre Celebi, Brian Helba, Michael A Marchetti, Stephen W Dusza, Aadi Kalloo, Konstantinos Liopyris, Nabin Mishra, Harald Kittler, et al. Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In *IEEE ISBI 2018)*, pages 168–172, 2018.

[7] Lee R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.

[8] David Gutman, Noel C. F. Codella, M. Emre Celebi, Brian Helba, Michael A. Marchetti, Nabin K. Mishra, and Allan Halpern. Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC). *CoRR*, abs/1605.01397, 2016.

[9] Mohammad H Jafari, Nader Karimi, Ebrahim Nasr-Esfahani, Shadrokh Samavi, S Mohamad R Soroushmehr, K Ward, and Kayvan Najarian. Skin lesion segmentation in clinical images using deep learning. In *IEEE ICPR*, pages 337–342, 2016.

[10] Shivangi Jain, Nitin Pise, et al. Computer aided melanoma skin cancer detection using image processing. *Procedia Computer Science*, 48:735–740, 2015.

[11] Shuman Jia, Antoine Despinasse, Zihao Wang, Hervé Delingette, Xavier Pennec, Pierre Jaïs, Hubert Cochet, and Maxime Sermesant. Automatically segmenting the left atrium from cardiac images using successive 3d u-nets and a contour loss. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 221–229. Springer, 2018.

[12] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. *CoRR*, abs/1606.04797, 2016.

[13] Sankaran Panchapagesan, Ming Sun, Aparna Khare, Spyridon Matsoukas, Arindam Mandal, Björn Hoffmeister, and Shiv Vitaladevuni. Multi-task learning and weighted cross-entropy for dnn-based keyword spotting. In *INTERSPEECH*, 2016.

[14] Stefano Pedemonte, Bernardo Bizzo, Stuart Pomerantz, Neil Tenenholtz, Bradley Wright, Mark Walters, Sean Doyle, Adam McCarthy, Renata Rocha De Almeida, Katherine Andriole, et al. Detection and delineation of acute cerebral infarct on dwi using weakly supervised machine learning. In *MICCAI*, pages 81–88, 2018.

[15] Andrea Pennisi, Domenico D. Bloisi, Daniele Nardi, Anna Rita Giampetruzzi, Chiara Mondino, and Antonio Facchiano. Skin lesion image segmentation using delaunay triangulation for melanoma detection. *Computerized Medical Imaging and Graphics*, 52:89 – 103, 2016.

[16] Javier Ribera, David Güera, Yuhao Chen, and Edward Delp. Weighted hausdorff distance: A loss function for object localization. *CoRR*, abs/1806.07564, 2018.

[17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, pages 234–241. Springer, 2015.

[18] Seyed Sadegh Mohseni Salehi, Deniz Erdogmus, and Ali Gholipour. Tversky loss function for image segmentation using 3d fully convolutional deep networks. In *International Workshop on Machine Learning in Medical Imaging*, pages 379–387. Springer, 2017.

[19] Gerald Schaefer, Maher I Rajab, M Emre Celebi, and Hitoshi Iyatomi. Colour and contrast enhancement for improved skin lesion segmentation. *Computerized Medical Imaging and Graphics*, 35(2):99–104, 2011.

[20] Carole H Sudre, Wenqi Li, Tom Vercauteren, Sebastien Ourselin, and M Jorge Cardoso. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pages 240–248. Springer, 2017.

[21] Roger Trullo, Caroline Petitjean, Su Ruan, Bernard Dubray, Dong Nie, and Dinggang Shen. Segmentation of organs at risk in thoracic ct images using a sharpmask architecture and conditional random fields. In *IEEE ISBI*, pages 1003–1006, 2017.

[22] Philipp Tschandl, Cliff Rosendahl, and Harald Kittler. The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific data*, 5:180161, 2018.

[23] Michael Weber, Michael Fürst, and J. Marius Zöllner. Automated focal loss for image based object detection. *CoRR*, abs/1904.09048, 2019.

[24] Alexander Wong, Jacob Scharcanski, and Paul Fieguth. Automatic skin lesion segmentation via iterative stochastic region merging. *IEEE Transactions on Information Technology in Biomedicine*, 15(6):929–936, 2011.

[25] Ken CL Wong, Mehdi Moradi, Hui Tang, and Tanveer Syeda-Mahmood. 3d segmentation with exponential logarithmic loss for highly unbalanced object sizes. In *MICCAI*, pages 612–619. Springer, 2018.

[26] Xiaojing Yuan, Ning Situ, and George Zouridakis. A narrow band graph partitioning method for skin lesion segmentation. *Pattern Recognition*, 42(6):1017–1028, 2009.

[27] Yading Yuan, Ming Chao, and Yeh-Chi Lo. Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance. *IEEE transactions on medical imaging*, 36(9):1876–1886, 2017.

[28] Huiyu Zhou, Gerald Schaefer, M Emre Celebi, Faquan Lin, and Tangwei Liu. Gradient vector flow with mean shift for skin lesion segmentation. *Computerized Medical Imaging and Graphics*, 35(2):121–127, 2011.

[29] Xiao-Yun Zhou, Mali Shen, Celia Riga, Guang-Zhong Yang, and Su-Lin Lee. Focal fcn: towards small object segmentation with limited training data. *CoRR*, abs/1711.01506, 2017.

[30] Alex P Zijdenbos, Benoit M Dawant, Richard A Margolin, and Andrew C Palmer. Morphometric analysis of white matter lesions in mr images: method and validation. *IEEE transactions on medical imaging*, 13(4):716–724, 1994.