Balanced Affinity Loss for Highly Imbalanced Baggage Threat Contour-Driven Instance Segmentation

IEEE ICIP 2022

Supplementary Material

1. Proposed Network Architecture:

This section discusses the architectural details of the proposed instance segmentation framework driven via SegNet [1]. As reported in Table 1, we can observe that the network has one input layer, 48 convolution layers, 47 batch normalization layers, 41 ReLU activations, one max pooling, 5 zero padding, 13 additions, 1 reshape, 3 upsampling, and one clustering layer having different kernel sizes. Moreover, the network has 14.86M parameters in which 14.83M parameters are trainable, and 32K are non-trainable. Apart from this, the size of the latent space representation within the proposed framework is 1x884736. The detailed model summary is also provided in the source code repository[1].

2. Additional Ablation Experiments:

2.1. Determining the optimal $\boldsymbol{\beta}$ value:

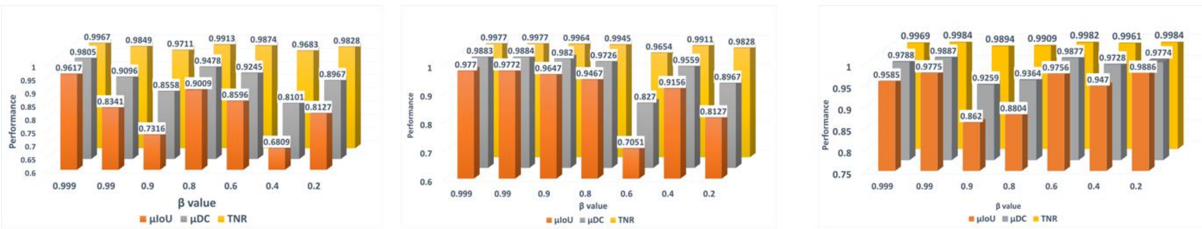
$\beta \in [0, 1]$ is a hyperparameter within the proposed balanced affinity loss function that controls how fast the effective number grows as the number of samples grows, as evident from Eq. 2 of the main manuscript. In this ablation study, we determined the optimal value of β that allows the proposed framework to produce best threat detection performance under extreme class imbalance. For this purpose, we plugged different values of β within the proposed loss function, measured the performance of the proposed framework, in terms of mean intersection-over-union (μIoU), mean dice coefficient (μDC), and true negative rate (TNR), at the inference stage across COMPASS-XP [2], OPIXray [3] and SIXray [4] datasets. From Fig. 1 (a), we can observe that for COMPASS-XP, the value of the proposed framework, in terms of μDC, μIoU, and TNR has reached up to 0.9805, 0.9617, and 0.9967, respectively, when β = 0.999. Moreover, on the OPIXray dataset, the best detection performance of the proposed framework is achieved when β =0.990, i.e., the μDC, μIoU, and TNR scores of

[1] Source code repository: https://github.com/taimurhassan/balanced_affinity

0.9884, 0.9772, and 0.9977, respectively (see Fig. 1-b). Similarly, the optimal value of β on SIXray is 0.990, yielding the score of 0.9887, 0.9775, and 0.9984 in terms of μDC, μIoU, and TNR, respectively (see Fig. 1-c). Therefore, in the rest of the experimentation, we used β = 0.999 for COMPASS-XP, and β = 0.990 for OPIXray and SIXray datasets (along with SIXray subsets) towards detecting imbalanced baggage threats.

Table 1: Proposed Network Architecture

Layers	Number of Layers	Parameters	Kernel Size
Convolution	48	14,799,302	1x1, 3x3, 7x7
Batch Normalization	47	65,024	-
ReLU	41	0	-
Max Pooling	1	0	3x3
Zero Padding	5	0	1x1, 3x3
Input	1	0	-
Addition	13	0	-
Reshape	1	0	-
UpSampling	3	0	2x2
Clustering	1	0	-
Total Parameters:	14,864,326		
Trainable Parameters:	14,831,814		
Non-trainable Parameters	32,512		



(a) (b) (c)

Fig. 1 Performance comparison of balanced affinity loss function on the (a) COMPASS-XP; (b) OPIXray; and (c) SIXray datasets with different β values.

2.2. Determining the optimal backbone:

The next ablation experiment is about determining the optimal backbone model within the proposed framework that gives the best threat detection performance on the three datasets. For this purpose, we employed different encoder-decoder, scene parsing and fully convolutional networks, such as SegNet [1], UNet [5], PSPNet [6], and FCN-8 [7], within the proposed framework and measured the detection performance in terms of μIoU and μDC scores. The results are reported in Table 2 in which we can observe that across all the datasets the SegNet gives the best threat detection performance when constrained using the proposed balanced affinity loss function. For example: it achieved 4.64%, 5.63%, 7.25%, 3.67%, and

1.02% improvement in terms of μIoU over second-best UNet across SIXray10, SIXray100, SIXray1000, OPIXray, and COMPASS-XP datasets, respectively. Similarly, in terms of μDC, the proposed framework (backboned through SegNet) achieved 2.35%, 2.89%, 4.36%, 1.85%, and 0.52% improvements over other variants across SIXray10, SIXray100, SIXray1000, OPIXray, and COMPASS-XP datasets, respectively. Therefore, in the rest of experimentation, we chose SegNet as a candidate backbone model towards detecting imbalanced baggage threats.

Table 2: Performance evaluation of proposed framework with different backbone models. To ensure fairness, all models are driven through ResNet-50 [8] and constrained using proposed balanced affinity loss function. Moreover, bold indicates the best performance and the second-best scores are underlined.

Metric	Dataset	SegNet	PSPNet	UNet	FCN-8
μIoU	SIXray10	**0.9775**	0.9051	<u>0.9341</u>	0.8761
	SIXray100	**0.9498**	0.8673	<u>0.8991</u>	0.8257
	SIXray1000	**0.6611**	0.5738	<u>0.6164</u>	0.5248
	OPIXray	**0.9772**	0.9239	<u>0.9426</u>	0.8863
	COMPASS-XP	**0.9617**	0.9406	<u>0.9519</u>	0.9012
μDC	SIXray10	**0.9886**	0.9502	<u>0.9659</u>	0.9340
	SIXray100	**0.9743**	0.9289	<u>0.9469</u>	0.9045
	SIXray1000	**0.7960**	0.7292	<u>0.7627</u>	0.6884
	OPIXray	**0.9885**	0.9604	<u>0.9705</u>	0.9397
	COMPASS-XP	**0.9805**	0.9694	<u>0.9754</u>	0.9480

Fig. 2: Failure cases of the proposed framework. (A) and (C) shows the original scan from SIXray and OPIXray datasets, whereas (B) and (D) shows the results of the proposed framework on (A) and (C).

3. Failure Cases:

This section discusses the limitation of the proposed framework towards recognizing the imbalanced baggage threats. From Fig. 2, we can see that although the proposed framework can recognize the challenging instances of the contraband data depicted within the complex scans of SIXray (see Fig. 2-A, B) and OPIXray datasets (see Fig. 2-C, D). However, the masks of the items are not very accurate. For example: see the mask of *gun* in (B), and the *scissor* in (D), pointed out by red arrow. Although seen rarely on some of the complex X-ray scans, this limitation of the proposed framework emanates from the fact that it is trained on an extremely imbalanced datasets (e.g., SIXray1000), where the balanced affinity loss function could not constrain the proposed framework enough to accurately segment the extremely cluttered objects with the imbalanced training. However, this limitation can be easily addressed by employing morphological operations as a postprocessing step. Also, considering the fact that the proposed framework, when constrained with balanced affinity loss function, can accurately recognize the threatening objects from such complex X-ray scans with imbalanced training, the performance of the proposed framework is appreciable.

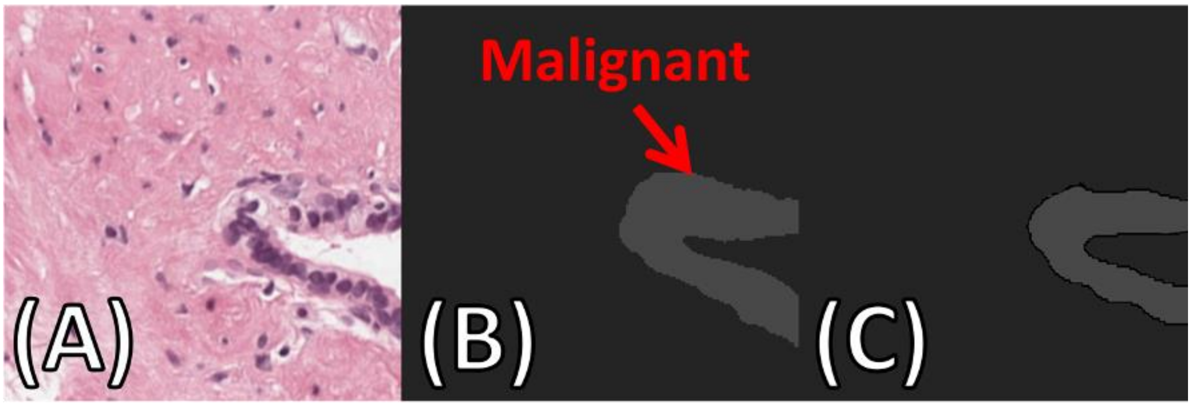


Fig. 3: (A) Prostate WSI patch with majority benign and minority malignant tissues, (B) the ground truth where malignant region is highlighted with red arrow and the remaining region is benign, (C) regions extracted by the proposed framework.

4. Validation of the Proposed L_{cba} Function in Other Domains:

Apart from validating the applicability of the proposed balanced affinity loss function (L_{cba}) for learning the highly imbalanced baggage threat detection tasks, we also used it in another domain to extract benign and malignant tissues from the prostate whole slide images (WSIs). The dataset which we used for this purpose is called the PANDA dataset [9] and is publicly available online. Moreover, PANDA dataset contains around 11,000 WSIs [9], which are decomposed into 79.4M non-overlapping WSI patches of 350x350x3. Moreover, these patches are marked with the labels of background, healthy stroma, benign tissues, and Gleason Patterns (GP-3, 4, and 5), as per the clinical standards. However, since the purpose of these experiments is to verify the imbalanced learning of benign and malignant tissues using L_{cba}, where the benign and malignant tissues are extremely imbalanced in nature [10-11], i.e., we have majority of pixel-level regions for benign tissues and minority of pixel-level regions for malignant tissues (see Fig. 3), we combined the GP-3, 4, and 5 categories as malignant, and healthy stroma and benign categories as benign, whereas the background class remains the same. The results are reported in Table 3 in which we can observe that when the state-of-the-art DRN [10] model is trained using the proposed balanced affinity loss function (L_{cba}) on the PANDA dataset, it achieved 3.78% and 2.64% improvements in terms of μIoU and μDC scores, respectively, for extracting the benign tissues, and 6.64% and 6.94% improvements in terms of in terms of μIoU and μDC scores, respectively, for extracting the malignant tissues, as compared to the original DRN model [10]. Moreover, we also trained here the TST [12] using the L_{ba} loss function to extract benign and malignant prostate tissues. However, since TST leverages contour information to segment the region-of-interest [12], it does not perform well on the WSIs as compared to the state-of-the-art. Apart from this, these experiments also evidence the superiority of the proposed L_{cba} function towards accurately extracting the imbalanced region-of-interest across different applications.

Table 3: Performance evaluation of the proposed framework on the PANDA dataset towards extracting the imbalanced benign and malignant prostate tissues. $DRN_{L_{cba}}$ denotes the DRN framework trained using proposed balanced affinity loss function (L_{cba}), $TST_{L_{cba}}$ denotes the TST framework trained using proposed L_{cba} function. Moreover, bold indicates the best score while the second-best score is underlined.

Metric	Categories	$DRN_{L_{cba}}$	DRN [10]	$TST_{L_{cba}}$
μIoU	Benign	0.7126	0.6748	0.6104
	Malignant	0.4173	0.3509	0.2462
μDC	Benign	0.8322	0.8058	0.7581
	Malignant	0.5889	0.5195	0.3951

References:

[1] V. Badrinarayanan, A. Kendall and R. Cipolla, “SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017.

[2] L. D. Griffin, M. Caldwell, and J. T. A. Andrews, “COMPASS-XP Dataset”, Computational Security Science Group, UCL, 2019.

[3] Y. Wei, R. Tao, Z. Wu, Y. Ma, L. Zhang, and X. Liu, “Occluded Prohibited Items Detection: An X-ray Security Inspection Benchmark and De-occlusion Attention Module”, ACM Multimedia, 2020.

[4] C. Miao, L. Xie, F. Wan, C. Su, H. Liu, J. Jiao, and Q. Ye, “SIXray: A Large-scale Security Inspection X-ray Benchmark for Prohibited Item Discovery in Overlapping Images”, IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[5] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation”, International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015.

[6] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid Scene Parsing Network”, IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[7] J. Long, E. Shelhamer, and T. Darrell, “Fully Convolutional Networks for Semantic Segmentation”, IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[8] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[9] W. Bulten et al., “Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge”, Nature Medicine, January 2022.

[10] T. Hassan, B. Hassan, A. El-Baz, and N. Werghi, “A Dilated Residual Hierarchically Fashioned Segmentation Framework for Extracting Gleason Tissues and Grading Prostate Cancer from Whole Slide Images”, IEEE Sensors Applications Symposium (SAS), 2021.

[11] S. Graham, Q. D. Vu, S. E. A. Raza, A. Azam, Y. W. Tsang, J. T. Kwak, N. Rajpoot, “HoVer-Net: Simultaneous Segmentation and Classification of Nuclei in Multi-Tissue Histology Images”, Medical Image Analysis, 2019

[12] T. Hassan, N. Werghi, “Trainable Structure Tensors for Autonomous Baggage Threat Detection Under Extreme Occlusion”, Asian Conference on Computer Vision (ACCV), 2020.