
Proceedings Paper:
Su, J., Anderson, S. and Mihaylova, L. orcid.org/0000-0001-5856-2223 (2022) A deep learning method with cross dropout focal loss function for imbalanced semantic segmentation. In: 2022 Sensor Data Fusion: Trends, Solutions, Applications (SDF) Proceedings. IEEE Sensor Data Fusion Workshop, 12-14 Oct 2022, Bonn, Germany. Institute of Electrical and Electronics Engineers (IEEE) . ISBN 9781665486736

https://doi.org/10.1109/SDF55338.2022.9931700

A Deep Learning Method with Cross Dropout Focal Loss Function for Imbalanced Semantic Segmentation

Jingxuan Su
Dept. of Automatic Control
& Systems Engineering
University of Sheffield, UK
jsu14@sheffield.ac.uk

Sean Anderson
Dept. of Automatic Control
& Systems Engineering
University of Sheffield, UK
s.anderson@sheffield.ac.uk

Lyudmila S. Mihaylova
Dept. of Automatic Control
& Systems Engineering
University of Sheffield, UK
l.s.mihaylova@sheffield.ac.uk

***Abstract*—Deep learning methods have proven their potential in semantic segmentation. However, they depend on the data quality and training process. Often, the data corresponding to the objects to be segmented are of different sizes and this creates difficulties for the segmentation method. Objects are segmented and associated with categories during the training process. Data imbalance is a challenging problem, which often results in unsatisfactory segmentation performance. This paper proposes a solution to this task based on a novel cross dropout focal loss (CDFL) function, which represents well the change between the cross-entropy and other state-of-the-art loss functions providing a balance between the precision and accuracy of segmentation. The performance of the considered fully convolutional network (FCN) with different loss functions is considered and carefully evaluated. The proposed loss function improves efficiently the semantic segmentation performance over other well-known loss functions. It is demonstrated on Cityscapes and PASCAL VOC 2010 publicly available datasets. The implementation is over relatively large data sets. The achieved mean accuracy of the proposed CDFL network on Cityscapes dataset is 76.41% and on PASCAL VOC 2010 dataset is 79.63% which is with approximately 2.5% improvement compared with the same network implemented with the cross-entropy loss function.**

***Keywords*—Deep learning, Semantic segmentation, Loss function, Imbalanced class dataset.**

I. Introduction

Semantic segmentation is defined as the pixel-level classification [1]–[3]. The results depend to a large extent on the dataset balance. When one label data is in the minority category while millions of labels are in the majority category, it can lead to a slight bias or severe imbalance in the predictive results [4]. This means data imbalance is a fundamental problem in semantic segmentation tasks, which restricts the accuracy and precision of the image segmentation. The imbalanced dataset poses a challenge for prediction since many semantic segmentation algorithms assume an equal number of each category. However, in semantic segmentation datasets, the category imbalance is inevitable. For instance, in Cityscapes datasets [5], a traffic sign and a person are considered as the minority of the segmentation categories, while a large number of categories correspond to buildings, roads and the sky. Commonly, models need to pay sufficient attention to the minority of categories for safety reasons. However, models naturally have a bias towards the majority categories in the training process, which leads to low accuracy and precision results, especially on a small number of categories. The choice of the loss function and how it is linked to the labels plays an important role in improving the image segmentation performance.

A series of significant works show how to alleviate the impact of the data imbalance on the segmentation results [2], [6]. The majority of the methods focus on the design of loss functions that consider well both the minority and majority classes. There are three types of loss functions [7]–[9]. Firstly, the region-based loss function directly optimizes the intersection-over-union (IoU) [10]. This type of loss function mainly applies to medical segmentation. Secondly, the statistics-balanced loss function adjusts the weight of category distribution based on its margin or size, i.e. class-balanced loss [11] and a label-distribution-aware margin (LDAM) loss function [12]. It encourages overfull false positives in the small number of categories. However, this approach could undermine the learning capability in feature extraction [13]. Thirdly, the performance-balanced loss function adds factors to weight the distribution of each category, i.e. as it is in the focal loss function [14]. However, its applications face challenges sometimes [11] since it cannot balance between the small and large number of categories that up-weight the minority category [13].

This work develops a novel data-balanced driven semantic segmentation solution consisting of a fully connected convolutional neural network and a cross dropout focal loss function. The cross dropout focal loss function down-weights, respectively up-weighs a category based on the output for this category. Unlike the statistics-balanced losses, the cross dropout focal loss has dynamic weight components based on per-category network outputs, compared to the statistics-balanced losses. In our experiments, the cross dropout focal loss can effectively address data imbalance and improves the accuracy and IoU.

The main contributions of this work are as following.

- A novel loss function is introduced. The cross dropout

focal loss not only depends on the weights to adjust the loss but keeps the statistic capability of the cross-entropy loss function.

- The cross dropout focal loss updates weights based on the segmentation output per category after T dropout times.
- The proposed loss function with FCN improves the segmentation performance, which is demonstrated over two popular semantic segmentation datasets, Cityscapes [5] and PASCAL [15]. The results show that the cross dropout focal loss achieves better performance than the well-known loss functions such as entropy and the focal loss.

The rest of this paper is organized as follows. Section II summarizes the related works, Section III describes the main proposed methodology. Section IV shows the performance validation and evaluation of the proposed algorithm. Finally, Section V concludes this work and summarizes future directions.

II. Related Work

Deep learning methods have become increasingly popular for semantic segmentation tasks [2], [16], [17] but still challenges in pixel-level image segmentation remain due to imbalanced datasets. The fully convolutional network [16] as a generic forerunner of the state-of-the-art algorithms and is often chosen to be in the heart of many deep learning approaches [18], [19]. Fully convolutional networks [16] learn their representations by using skip layers. Fully convolutional networks provide efficient inference and learning which can be extended to other well-known networks, such as the U-Net [20], and SegNet [21], [22]. These architectures [23], [24] have a number of attractive properties. They provide smooth predictions and easy visualisations of the feature activation in the pixel label space. Thus, we choose the fully convolutional network architecture as the deep learning backbone to balance the computational time and accuracy. Despite the power of the described networks, they still cannot solve the data imbalance problem.

The quality of the data sets is of significant importance when it trains semantic segmentation models. Deep learning semantic segmentation methods such as the FCN [16], U-Net [20], and recently developed methods, such as Deeplab [17] face the category imbalance problem. When training of neural networks is based on easy examples this could lead to insufficient learning and as a whole to inefficiently accurate results. A common solution is to increase the hard examples [25]–[27]. In contrast to these works, we propose the cross dropout focal loss function that solves the imbalance issue without complex computation and huge sampling.

Different loss functions have been used. The cross-entropy [28] is one of them and it is defined to measure the difference between two probability distributions. The cross-entropy has been widely applied to semantic segmentation. The weighted cross-entropy [29] proposes to weight positive and negative examples, which leads to better results than with the cross-entropy in an imbalanced class. The balanced cross entropy [30] is motivated by the weighted cross-entropy which leads to an efficient use of the number of samples in each class. The focal loss function [14] affords training on a sparse set of hard examples and can improve the accuracy when applied to object detection tasks.

These loss functions were proposed, driven by the motivation of improving the weighting of the class labels [4], [31]. However, these could introduce excessive false positives and adversarial results [7]. The Intersection over Union (IoU) is the most commonly used evaluation indicator in segmentation networks. Lovasz Softmax [32] directly optimizes the IoU on the Lovasz convex extension. In [33] the Dice similarity coefficient is leveraged and the trade-off between false positives and negatives in image segmentation is controlled. In contrast, the proposed cross dropout focal loss function does not rely fully on label weights and considers the balance between different classes.

III. Methodology

This section describes the known cross entropy and focal loss functions and the proposed cross dropout focal loss (CDFL) function for multi-category segmentation. Their performance is compared next in Section IV.

A. Cross Entropy for Multi-category Segmentation

The cross-entropy [4] has been widely applied in many semantic segmentation tasks [4], [34]. It uses the number of pixels for each category to optimize the geometric mean confidence of each weighted category. The formula for the cross-entropy CE is the following:

$$CE = -\sum_{c=1}^{M} y_c \log\left(p_c\right), \tag{1}$$

where M denotes the category number, p_c represents the corresponding value of the c-th category in the output of the softmax activation function, y_c denotes the value of true predictions in the category c. If the category of prediction and label are the same, then the value of 1 is assigned, otherwise it is 0. However, this approach with the cross-entropy has an obvious drawback that it applies to a balanced dataset. When the number of pixels in the minority category is much smaller than the number of pixels in the majority category for the same image, the $y_c = 0$ in the function will dominate. Thus, the number of pixels influences on the value of y_c. In other words, if the number of $y_c = 0$ is much larger than the number of $y_c = 1$, this situation will make the model heavily biased towards the main label which results in poor results.

The balanced cross entropy (BCE) [4] adds a weight parameter for each category to solve data imbalance problem. The balanced cross entropy BCE for muti-segmentation is represented with the equation:

$$BCE = -\sum_{c=1}^{M} w_c y_c \log\left(p_c\right). \tag{2}$$

The weight parameter w_c calculation formula is $w_c = \frac{N-N_c}{N}$, where N denotes the total number of pixels, and N_c shows the number of pixels in the ground truth per category.

The variable y_c still has the same meaning as in the cross-entropy expression. In this way, the balanced cross entropy can represent well the different categories with different weights for small or large categories. However, it did not consider the easy-hard imbalance in per category. The balanced cross entropy cannot address the data imbalance issue effectively when facing a big semantic segmentation dataset.

B. Focal loss for Multi-category Segmentation

In [14] the focal loss function is proposed for binary segmentation. The idea for the focal loss is inspired by the cross-entropy. The focal loss has two hyperparameters, γ and α that are introduced for balancing between the easy and hard examples. In this paper, we extend the focal loss to the multi-segmentation task. The activation function can only be the softmax [14], [35] function. The multi-focal loss with the softmax function FL_{softmax} is defined as:

$$FL_{\text{softmax}} = -\sum_{c=1}^{M} \alpha_{\text{c}} \left(1-\text{p}_{\text{c}}\right)^{\gamma} \log\left(\text{p}_{\text{c}}\right), \tag{3}$$

where α_{c} indicates the weight of the c-th category label, p_c denotes the output of the c-th category after softmax function. The value of p_c can reflect the degree of difficulty of the sample in segmentation. When $p_c > 0.5$, it belongs to easy-segmented region, otherwise is a hard-segmented region. If the value of p_c is big, the prediction results will be more accurate. The parameter γ adjusts the rate of easy label down-weighted labels. The parameter α represents the adjustment weight of the corresponding positive sample. However, this loss function only considers the easy-hard imbalance, without considering the imbalance category.

C. Cross Dropout Focal Loss Function

The proposed dropout cross focal loss function aims to improve the model performance and weight well the balance per category for easy-hard segmentation. In the proposed approach the input data are considered with T dropout times into the segmentation architecture. Thanks to the Monte Carlo dropout procedure [36], deep neural network output $\hat{y}_t$ will be different after dropout at each time. Then an indicator variable $u(\hat{y}$ is introduced which depends on the network predicted output $\hat{y}_t$ and can be expressed by the following equation:

$$u(\hat{y}) \approx \frac{1}{T}\sum_{t=1}^{T}\left(\hat{y}_t\right)^2 - \left(\frac{1}{T}\sum_{t=1}^{T}\hat{y}_t\right)^2. \tag{4}$$

The value of the indicator variable $u(\hat{y})$ represents the easy-hard degree of segmentation from the dropout output per category perspective. Motivated by the focal loss, the value of the indicator variable $u(\hat{y})$ can replace the modulating factor $(1-p_t)$ from equation (3). Thus, we update the focal loss to the dropout focal (DF) loss shown in the following equation:

$$DF = -\frac{1}{N}\sum_{i=1}^{N} \alpha_{\text{i}} \left(u_i(\hat{y})\right)^{\gamma} \log\left(\hat{y}_i\right). \tag{5}$$

When the value of $u(\hat{y})$ is close to 0, the value of the dropout focal loss function will reduce. That means the easy-segmented labels are down-weighted. In this paper, the T value was set equal to 5 to keep an efficient computational time and sufficient accuracy.

The cross-entropy and focal loss functions face challenges with the imbalanced dataset. Thus, we propose a novel loss function, the cross dropout focal loss (CDFL). Based on the cross-entropy we add the dropout focal loss with a weighted index ω as a modulating factor to solve the data imbalance problem. The cross dropout focal loss $CDFL$ is represented with the following equation:

$$\begin{aligned} CDFL &= CE + \omega DF \\ &= -\sum_{i=1}^{N} y_i \log\left(p_i\right) - \omega[\frac{1}{N}\sum_{i=1}^{N} \alpha_{\text{i}} \left(u_i(\hat{y})\right)^{\gamma} \log\left(\hat{y}_i\right)]. \end{aligned} \tag{6}$$

We set up the values of the γ and α parameters respectively equal to 2 and 0.75, which is the same choice as in [14]. The weight factor ω balances the impact of the cross-entropy and of the dropout focal loss. For the purposes of performance validation, different values of ω are used and these are $0.1, 0.01$ and 0.001. The best performance has been obtained with the 0.01 value of ω. As a result, the cross dropout focal loss function avoids excessive weighting of hard-segmented examples or of the minority categories which could cause undesirable results. Meanwhile, the $CDFL$ provides a suitable training weight for the different inputs. Therefore, the cross dropout focal loss achieves a higher balancing ability than the cross-entropy. It can handle both category and easy-hard segmentation imbalance situations.

IV. Experiments and Analysis

The experiments are performed with the Ubuntu 20.04 system. The server environment uses Python 3.7, Pytorch 1.12.1 and CUDA 10.1.

A. Datasets and Implementation Details

In order to evaluate the proposed loss function, the performance of a FCN with different loss functions is evaluated over two popular semantic segmentation datasets, Cityscapes for outdoor driving and PASCAL VOC 2010.

1) Cityscapes: Cityscapes [5] is a popular data set for semantic segmentation, which comprises urban street scenes. It is a large-scale driving database that contains fine annotated data and coarse annotated data of around 25000. There are 8 groups with 30 categories. Data was captured from 50 cities under different environmental conditions. In this paper, the dataset adopts 3475 fine annotations images for train and validation sets and 1525 dummy annotations for the test set. It has 19 classes shown in Fig. 2.

2) PASCAL VOC 2010: Pattern Analysis, Statical Modeling and Computational Learning (PASCAL) Visual Object Classes (VOC) [15] is a computer vision challenge for five different competitions and provides ground truth annotated datasets. This paper only focuses on the PASCAL VOC

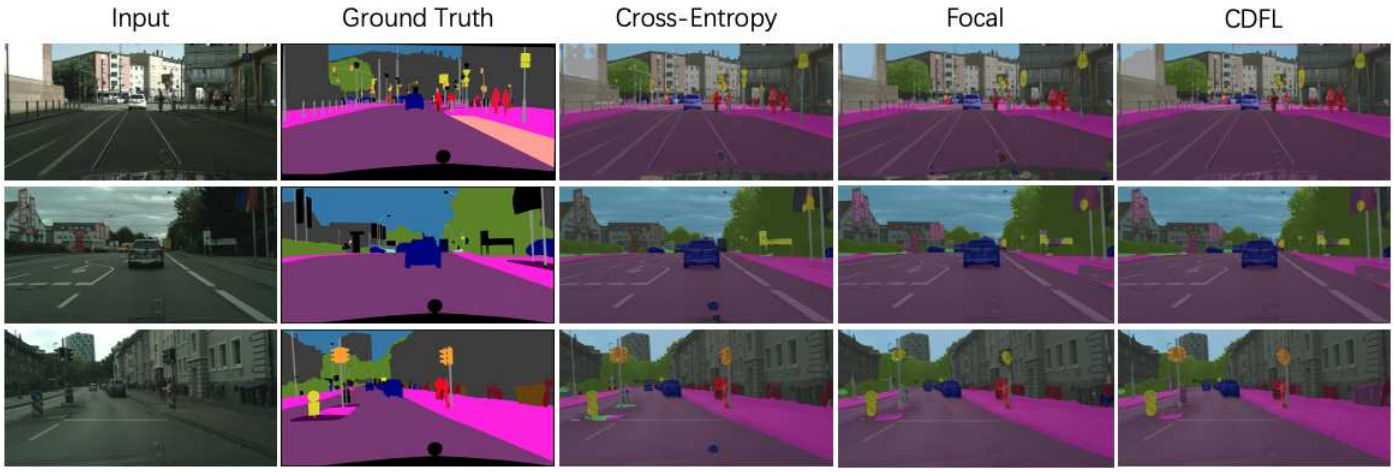

Fig. 1: Visualization of segmentation results on Cityscapes with FCN.

2010, which is a two dimensional (2D) segmentation dataset. Especially, the dataset supports pixel-level segmentation. It contains 540 classes grouped into 3 categories (objects, stuff, and hybrids). The dataset contains 4998 images for training and 1550 for validation. It has 20 classes: aeroplane, bag, bed, bedclothes, bench, bicycle, bird, boat, book, bottle, building, bus, cabinet, car, cat, ceiling, chair, cloth, computer, cow and others.

3) Evaluation metrics: The cross dropout focal loss based on FCN [16] is implemented for the segmentation task on the two above mentioned datasets. In semantic segmentation, the mean accuracy (mACC) and the mean IoU (mIoU) [12], [13], [37], [38] are important metrics. Here we apply them to evaluate the image semantic segmentation performance.

The mean Intersection over Union (IoU) [39] characterizes the balance between precision and recall performance measures. We also show both precision and recall results and demonstrate that the performance of the FCN with the cross dropout focal loss function gives very good segmentation results.

- **Mean Intersection over Union (mIoU)**, $mIoU$: In semantic segmentation, this evaluation metric calculates the intersection ratio of two sets. These two sets are annotated data and predicted outputs [2]. It is computed by

$$mIoU = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}+\sum_{j=0}^{k}p_{ji}-p_{ii}}, \tag{7}$$

  where p_{ij} and p_{ji} represents false positive and false negatives for category i and category j respectively. The value of p_{ii} is the number of true positives. The value of k is the total number of categories.

- **Mean Accuracy**: It computes two sets, which are the number of the correct pixels p_{ii} and the total number of pixels per category [2]. After getting per-category accuracy, the mean accuracy $mAcc$ averages the total $k+1$ categories:

$$mAcc = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}}. \tag{8}$$

- **Precision** It [34] refers to the proportion of the total number of true positives (TPs) divided by the sum of all TPs and false positives (FPs):

$$Precision = \frac{TP}{TP+FP}. \tag{9}$$

- **Recall** It [34] defines the number of correct positive predictions, which are achieved from all the positive predictions. False negatives and true positives denote total samples:

$$Recall = \frac{TP}{TP+FN}. \tag{10}$$

The next subsection presents results over the considered public data and evaluates the segmentation results.

B. Validation Results and Analysis

We compare the fully convolutional network with three well-known loss functions for imbalanced semantic segmentation, namely the cross-entropy, focal loss [14] and Lovasz Softmax loss function [32]. Our novel loss function shows the best results compared with the other loss functions over the two datasets.

The segmentation results of the Cityscapes outdoor driving dataset are presented in Table I. The cross-entropy, focal loss and cross dropout focal loss all improved the mean accuracy and IoU compared to the Lovasz softmax loss. The performance of the cross dropout focal loss (CDFL) is very similar to the performance of the cross-entropy. However, the

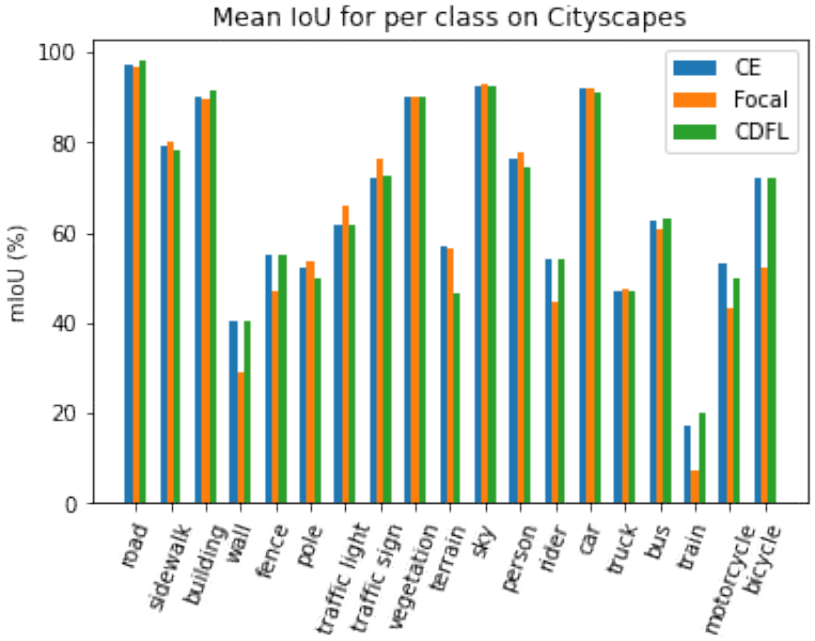

Fig. 2: Mean IoU per category on Cityscapes with FCN

TABLE I: Quantitative FCN performance with different losses on Cityscapes

loss	mIoU (%)	mAcc (%)	mPre (%)	mRec (%)
Cross-entropy	66.51	76.33	80.78	77.86
Focal loss	62.1	74.47	79.25	72.75
Lovasz softmax loss	57.14	70.51	75.22	70.51
CDFL	66.62	76.41	**81.23**	**78.11**

cross dropout focal loss achieves distinguishable performance with respect to precision 81.23% and recall 78.11%. The cross dropout focal loss function outperforms the other loss functions while keeping a good mean IoU and accuracy. The mean IoU of separate categories are shown in Fig. 2. Specifically, the cross dropout focal loss improves the weight of small categories such as trains and building and maintains good precision and recall performance. The segmentation results are visualized in Fig. 1. The cross-entropy and focal loss give incorrect results about the black class as the blue category at the bottom. The focal loss and cross dropout focal loss can predict traffic signs precisely and this is shown in the second row of Fig. 1. Fig. 2 shows the histogram graph that demonstrates that the cross dropout focal loss encourages a correct prediction of the small categories, i.e. trains.

TABLE II: Quantitative FCN performance with different losses on PASCAL VOC 2010

loss	mIoU (%)	mAcc (%)	mPre (%)	mRec (%)
Cross-entropy	66.72	76.85	80.32	76.23
Focal loss	62.45	75.74	76.45	70.57
Lovasz	59.43	64.16	69.56	63.24
CDFL	67.74	79.63	**81.85**	**79.63**

We further show the performance of the proposed loss function and of other state-of-the-art loss functions on the PASCAL VOC 2010 segmentation dataset. Table II demonstrates that the cross dropout focal loss still achieves the best performance.

V. Conclusions

This paper presents a novel data-balanced cross dropout focal loss algorithm for semantic image segmentation. The proposed loss function with a FCN alleviates the category dataset imbalance problem and improves model performance. Specifically, this loss function is designed from an output perspective. It has dynamic weights to reflect relative categories and segments the corresponding objects in images. We demonstrated several advantages of the cross dropout focal loss function: 1) It addresses effectively the data imbalance problem and weights the dropout output per category dynamically. 2) It updates weights based on the result of dropout T times. The dropout results can reflect the easy-hard degree of segmentation and help generate suitable weights. 3) The performance validation on both Cityscapes and PASCAL datasets shows its outperformance compared with state-of-the-art loss functions with approximately 2.5% accuracy improvement. That also demonstrates the increased robustness of the proposed loss.

Acknowledgement

We acknowledge the support of the UK EPSRC project EP/V026747/1 (Trustworthy Autonomous Systems Node in Resilience).

References

[1] P. Wang, P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou, and G. Cottrell, "Understanding convolution for semantic segmentation," in *Proceedings of 2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. Ieee, 2018, pp. 1451–1460.

[2] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, "A review on deep learning techniques applied to semantic segmentation," *arXiv preprint arXiv:1704.06857*, 2017.

[3] Y. Guo, Y. Liu, T. Georgiou, and M. S. Lew, "A review of semantic segmentation using deep neural networks," *International Journal of Multimedia Information Retrieval*, vol. 7, no. 2, pp. 87–93, 2018.

[4] S. Jadon, "A survey of loss functions for semantic segmentation," in *Proceedings of 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*. IEEE, 2020, pp. 1–7.

[5] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The Cityscapes dataset for semantic urban scene understanding," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 3213–3223.

[6] E. Puyol-Antón, B. Ruijsink, S. K. Piechnik, S. Neubauer, S. E. Petersen, R. Razavi, and A. P. King, "Fairness in cardiac mr image analysis: an investigation of bias due to data imbalance in deep learning based segmentation," in *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2021, pp. 413–423.

[7] J. Tian, N. C. Mithun, Z. Seymour, H.-p. Chiu, and Z. Kira, "Recall loss for imbalanced image classification and semantic segmentation," in *In Proceedings of the International Conference on Learning Representations, 2021*. ICLR, 2021.

[8] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *Journal of Big Data*, vol. 6, no. 1, pp. 1–54, 2019.

[9] A. Singh and A. Purohit, "A survey on methods for solving data imbalance problem for classification," *International Journal of Computer Applications*, vol. 127, no. 15, pp. 37–41, 2015.

[10] M. A. Rahman and Y. Wang, "Optimizing intersection-over-union in deep neural networks for image segmentation," in *Proceedings of International Symposium on Visual Computing*. Springer, 2016, pp. 234–244.

[11] Y. Cui, M. Jia, T.-Y. Lin, Y. Song, and S. Belongie, "Class-balanced loss based on effective number of samples," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9268–9277.

[12] K. Cao, C. Wei, A. Gaidon, N. Arechiga, and T. Ma, "Learning imbalanced datasets with label-distribution-aware margin loss," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[13] B. Zhou, Q. Cui, X.-S. Wei, and Z.-M. Chen, "BBN: Bilateral-branch network with cumulative learning for long-tailed visual recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 9719–9728.

[14] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2980–2988.

[15] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2010 (VOC2010) Results," http://www.pascal-network.org/challenges/VOC/voc2010/workshop/index.html.

[16] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.

[17] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[18] J. Dai, Y. Li, K. He, and J. Sun, "R-fcn: Object detection via region-based fully convolutional networks," *Advances in Neural Information Processing Systems*, vol. 29, 2016.

[19] J. Sherrah, "Fully convolutional networks for dense semantic labelling of high-resolution aerial imagery," *arXiv preprint arXiv:1606.02585*, 2016.

[20] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proceedings of International Conference on Medical Image Computing and Computer-assisted Intervention*. Springer, 2015, pp. 234–241.

[21] 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*, vol. 39, no. 12, pp. 2481–2495, 2017.

[22] V. Badrinarayanan, A. Handa, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling," *arXiv preprint arXiv:1505.07293*, 2015.

[23] S. Hao, Y. Zhou, and Y. Guo, "A brief survey on semantic segmentation with deep learning," *Neurocomputing*, vol. 406, pp. 302–321, 2020.

[24] F. Lateef and Y. Ruichek, "Survey on semantic segmentation using deep learning techniques," *Neurocomputing*, vol. 338, pp. 321–348, 2019.

[25] P. F. Felzenszwalb, R. B. Girshick, and D. McAllester, "Cascade object detection with deformable part models," in *In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Ieee, 2010, pp. 2241–2248.

[26] A. Shrivastava, A. Gupta, and R. Girshick, "Training region-based object detectors with online hard example mining," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 761–769.

[27] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1. IEEE, 2001, pp. I–I.

[28] M. Yi-de, L. Qing, and Q. Zhi-Bai, "Automated image segmentation using improved pcnn model based on cross-entropy," in *Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, 2004.* IEEE, 2004, pp. 743–746.

[29] V. Pihur, S. Datta, and S. Datta, "Weighted rank aggregation of cluster validation measures: a Monte Carlo cross-entropy approach," *Bioinformatics*, vol. 23, no. 13, pp. 1607–1615, 2007.

[30] S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1395–1403.

[31] D. G. MacArthur, S. Balasubramanian, A. Frankish, N. Huang, J. Morris, K. Walter, L. Jostins, L. Habegger, J. K. Pickrell, S. B. Montgomery *et al.*, "A systematic survey of loss-of-function variants in human protein-coding genes," *Science*, vol. 335, no. 6070, pp. 823–828, 2012.

[32] M. Berman, A. R. Triki, and M. B. Blaschko, "The Lovász-Softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4413–4421.

[33] S. A. Taghanaki, Y. Zheng, S. K. Zhou, B. Georgescu, P. Sharma, D. Xu, D. Comaniciu, and G. Hamarneh, "Combo loss: Handling input and output imbalance in multi-organ segmentation," *Computerized Medical Imaging and Graphics*, vol. 75, pp. 24–33, 2019.

[34] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, pp. 3523–3542, 2022.

[35] K. Nemoto, R. Hamaguchi, T. Imaizumi, and S. Hikosaka, "Classification of rare building change using cnn with multi-class focal loss," in *Proceedings of IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2018, pp. 4663–4666.

[36] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *Proceedings of International Conference on Machine Learning*. PMLR, 2016, pp. 1050–1059.

[37] Z. Liu, Z. Miao, X. Zhan, J. Wang, B. Gong, and S. X. Yu, "Large-scale long-tailed recognition in an open world," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2537–2546.

[38] B. Kang, S. Xie, M. Rohrbach, Z. Yan, A. Gordo, J. Feng, and Y. Kalantidis, "Decoupling representation and classifier for long-tailed recognition," *arXiv preprint arXiv:1910.09217*, 2019.

[39] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 658–666.