Loss Functions: Evaluating Their Performance and the Need For An Adaptive Approach

Haroon Haider Khan and Majid Iqbal Khan

October 31, 2021

Loss Functions: Evaluating their Performance and the Need For an Adaptive Approach

Haroon Haider Khan
Department of Computer Science
Commission of Science and Technology for Sustainable Develeopment in the South (COMSATS)
Islamabad, Pakistan
hhaider@hpsoft.co.uk

Dr Majid Iqbal
Department of Computer Science
Commission of Science and Technology for Sustainable Develeopment in the South (COMSATS)
Islamabad, Pakistan
majid_iqbal@comsats.edu.pk

Abstract— **Mainstream machine learning is dominated by semi-supervised learning. Developments in this field has permitted scholars to harness large amounts of unlabeled data with typically smaller sets of labelled data. This study focuses on the need for an adaptive loss function which automatically adjusts itself by training the model on various datasets. Once semantic segmentation is embedded in the architecture of any model, deeper layers are needed to extract features from images, causing the gradient to be too small for training the network during the learning process, particularly when pixelwise cross entropy loss function is in high dimensional settings, with large number of classes larger objects often overlap with smaller objects causing inaccurate detection. The need is to overcome the impact of super imposed objects on accuracy of classification caused by model confusion owing to the large number of classes. Our research endeavors to deal with the imbalanced data set problem in neural networks by experimenting on various loss functions. The experiments conducted on two different data sets show that different loss functions produce varying results. We present results on Indian driving dataset (IDD) and Cityscapes.**

Keywords—neural networks, image process, loss function, semantic segmentation, cross entropy, focal loss, combo loss

I. Introduction

Recently, there have been significant improvements in deep learning algorithms used for object detection, recognition and semantic segmentation. Semantic segmentation has been embedded in various applications linking images to a class label. In this regard various loss functions have evolved and successfully applied for classification tasks in state-of-the-art models. Among the plethora of loss functions available, it can be challenging to choose a suitable one for training a neural network. The loss function of the model plays a key role in the optimization of the process and determines how accurate the estimate will be. Loss functions measure the difference between a predicted value and its true value and check the performance of a model. Sometimes the functions are not fixed and can manually be altered depending upon the task they are required to handle.

Majority of the state-of-art deep learning models like GANs (Generative Adversial Models) and FCN (Fully Convolutional Network) are embedded with the de facto Cross Entropy loss function. In semantic segmentation since each pixel is trained independently using the standard cross entropy loss, the model fails to perform where faced with tasks with sparse labels. To deal with convoluted variations, it is important to extract multi-scale strong features and abundant context information [1,2].

Various datasets employed for assessing the effectiveness of semantic segmentation in road scenario use a limited number of classes and assume a well structured environment with clear road boundaries and less variation in background, like in Europe and North America. They have definite lanes, meagre traffic concentration, minimal deviations in the objects and background and traffic regulations are meticulously followed. However, such ideal circumstances are not found in most of the other parts of the world like Asia and Africa where the variety of traffic participants is greater, comprising innovative and greater classes such as autorickshaws or animals which behave contrary to vehicles. For orthodox classes like cars, the appearance deviations are complex because of wear and tear. The progress of smart automobiles in such an unstructured environment is an exceedingly challenging job.

When the labeled image is extremely sparse there exists foreground-background imbalances because only a very small amount of pixels are labeled as foreground class. Sparsely labeled pixels will be punished heavily during training because the probability for the real label is too small. Where labels are evenly distributed this does not present a problem. However, since each pixel is trained independently in standard cross entropy loss, the model fails to perform where faced with tasks with sparse labels.

The cost function is an important element for adjusting the weight of a neural network during the training process for creating an affective machine learning model. To alleviate the imbalance, a lot of segmentation models employ weighted cross-entropy as their loss function. The weights are determined on the basis of statistics or experience. Focal Loss, an extension of Cross entropy loss proved effective for multiclass classification where some classes are easy and others are difficult to classify. In [3], Focal Loss is used for overcoming the disparity between foreground and background by down-weighting the loss of simple objects thereby aiming on training the hard negative objects. Class imbalance is an acute problem in semantic segmentation, because of hard objects. Pixel wise classification lies at the heart of semantic segmentation which naturally results in a large number ratio among objects. Additionally majority of the pixels become easy examples as the training stages advance, specially when mIoU between prediction and ground truth is more than 70%. This is the step where most pixels' predictions to ground truth are greater than the estimations to other classes. Some portion of the main loss consists of these pixels which are the easy examples which are mainly objects appearing regularly in the Cityscapes dataset [4] like sky and road.

The standard Focal Loss function is as follows:

$$\mathrm{FL}(pt) = -(\lambda - pt)\,\gamma\ \log(pt) \tag{1}$$

Here, pt is the probability of ground-truth and λ and γ are hyper parameters. If $\gamma = 0$ then Focal Loss function works as standard cross-entropy. This equation tends to ignore the hard to detect objects and simply focuses on the easy ones. Therefore Equation (1) is considered to be a weighted cross-entropy. The weight $(\lambda - \mathrm{pt}\ )\ \gamma$ is negatively related to pt , which means that the weights can be adjusted in accordance with the prediction to the ground truth. To ensure that the easy, negative and majority classes do not super impose the difficult, minority and hard classes, Focal Loss includes a regulating element to the standard cross entropy loss.

Many research papers have attempted to manually alter and set the value of γ in order to detect the easy as well as hard objects. When the γ is set to 5, it moves its attention from easy to hard objects but tends to ignore the easy objects. This study, focuses on the imbalance issue which arise in various data sets based on static loss functions. Our objective is to train any given model using any number of classes avoiding model confusion which arise due to class imbalance in the dataset to detect objects efficiently.

II. Related Work

Under semantic segmentation semi-supervised learning not only make use of the few pixel-wise annotated samples but also leverages added annotation-free images [5]. Some models use two network branches referred to as the Semi-Supervised Semantic Segmentation GAN (s4GANs) and Multi-Label Mean Teacher (MTML) that connects classification with segmentation in a semi-supervised environment in combination with self-training. The function of generator is performed by the segmentation network which uses a standard loss function like cross entropy to train a discriminator. The output from the discriminator is used as a quality measure for determining the finest outputs which are further engaged for self-training. On the Cityscape dataset s4GANs performance is restricted to 19 classes only.

Since the origination of GANs, researchers have endeavored to refine GANs in multiple ways. [6] proposed a conditional GAN model (CGAN), in which road shape prediction network is trained end-to-end as a part of generator. In the study the author has proposed a semi-supervised learning (SSL) road detection method based on generative adversarial networks (GANs) and a weakly supervised learning (WSL) method based on conditional GANs. The discriminator is able to lead an untapped process of annotation on the unlabeled data by training under these frameworks. As a result, the network is able to exploit unlabelled as well as labelled data using improved images of road scene using the basic cross entropy loss function.

In semantic segmentation, object detection suffer greatly from imbalanced data and hard examples. Weighted cross-entropy is used to mitigate the imbalance of objects in many semantic segmentation tasks. The weights are determined on the basis of statistics or experience [7]. In [8], small batches in each epoch are collected to depict all the classes in a uniform manner rather than erratically dis-organizing the dataset and taking advantage of training crops out of random positions.

Some semantic segmentation methods use feature extractor of a real-time object detection model [9] to overcome the huge imbalance of objects in Cityscapes dataset and (IDD) [10]. A loss function similar to Focal Loss was recommended the aim of emphasizing on the difficult hard pixels. This was achieved by refining the weight of the difficult negatives as the same time maintaining the weight of simple examples. Majority of the pixels' predictions to ground truth in semantic segmentation, range from 0.4 to 0.6 (difficult example). Similar classes in Cityscapes such as roads and sidewalks, have predictions close to each other causing the pixels' predictions of ground truth to fall between 0.4 and 0.6. Moreover, it reduces the edge pixels' predictions to ground truth to the range of 0.4 and 0.6. The pixels' loss stated above will be very less when $\lambda = 1$. In this research experiments, to enlarge the weight of hard examples and preserve the weight of easy examples, λ is manually adjusted at 2. Darknet53 was used as the backbone of DeepLabv3+ and tested on Cityscapes test set.

The performance of Dice based loss function for binary-class segmentation problems has so far displayed better performance for solving the imbalance problem [11]. The potential of this loss function to trade off between false positives and false negatives (i.e., output imbalance) needs further research.

The Dice function is a popular mechanism for assessing the precision of image segmentation. To exploit the Dice loss function which addresses the input class-imbalance issues, some researchers have combined the weighted sum of Dice loss and the cross entropy into a new loss called the Combo Loss. Using the binary version of the cross entropy loss function helps to enforce the intended switch over between false negatives and positives [12].

III. problem statement

Present day loss functions cannot uniformly detect the features of road. There are differences in terms of undetected objects and areas, overlapping and blurry boundaries leading to incomplete predictions. Moreover some loss functions focus on bigger objects only overlooking the smaller objects on road.

Cross entropy is prone to class imbalance, as the algorithm focuses on the total calculated loss of the minor classes (small objects) like stone, road divider, animals, motorbikes etc and major classes (bigger objects) e.g road, sky, side wall etc. In the process, the major classes have a more total calculated loss as compared to minor total calculated loss. This causes the algorithm to focus majorly on major classes instead of minor classes which are sparsely labelled causing the classifier to overwhelm smaller objects by the bigger ones and hinder the training process.

During training and learning most of the contribution is coming from bigger object classes loss as the loss is greater than the total smaller object classes loss. Hence the relative weight given by the algorithm to small or sparsely labelled objects in a road scene is very less. This hinders the training process of the classifier which fails to effectively identify small objects and/or results in larger objects engulfing smaller/sparse ones.

Our objective is to propose a adaptive loss function which can assign a greater probability to minor classes in road scenario so that sparsely labeled and small objects on the road are focused more during the training process. Assigning a greater value to such objects will help to identify and detected them and prevent the bigger objects from overwhelming smaller objects ensuring the safety of the autonomous vehicle, the passengers on board and the surroundings.

IV. EXPERIMENTS

We present our results on two separate datasets; IDD (19 classes only) and Cityscapes. Our results and analysis are divided into three parts. Firstly we compare the performance of Cross Entropy loss function trained on Cityscapes and validated on IDD as well as Cityscapes. In the second part, we show a comparison of Cross Entropy Loss, Combo Loss and Focal Loss trained on IDD. Finally we depict results of Cross Entropy Loss, Combo Loss and Focal Loss trained on Cityscapes.

A. Experiments using Cross Entropy loss function trained and validated on Cityscapes and IDD datasets

We conducted different experiments to provide an insight into Cross Entropy loss functions used on two different datasets. A graphical comparison is shown in Fig. 1 and tabular comparison can be seen in table 1.

In our first experiment under this section, the original s4GAN [5] model is trained on original Cityscape dataset using de facto loss function and that is cross entropy loss function as a main loss function. This trained s4GANs model is then validated on the same Cityscapes dataset showing a mean IOU 0.57 accuracy. Next, the already trained s4GANs model which was trained on Cityscapes dataset while using cross entropy loss function is validated on IDD dataset having 30 classes, the validation results show mean IOU 0.047. Lastly the s4GANs model is trained on IDD dataset while using cross entropy loss function comprising of 30 classes and validated on 19 classes of IDD dataset showing mean IOU 0.48 accuracy.

TABLE 1. SHOWING THE IOU AND MEANIOU OF S4GAN TRAINED ON CITYSCAPE AND INDIAN DRIVABLE DATASET AND ALSO VALIDATED ON CITYSCAPE AND INDIAN DRIVABLE DATASET WHILE USING CROSS ENTROPY FUNCTION

Classes	s4GAN [5] trained and validation on Cityscape data set using cross entropy loss function (IOU)	s4GAN [5] trained on cityscape and validation on IDD using cross entropy loss function (IOU)	s4GAN [5] trained and validation on IDD using cross entropy loss function (IOU)
Road	0.95	0.01	0.89
Sidewalk	0.66	0.1	0.42
Building	0.84	0.2	0.66
Wall	0.37	0.01	0.36
Fence	0.37	0.02	0.32
Pole	0.41	0.01	0.39
Traffic_light	0.39	0.20	0.00
Traffic_sign	0.47	0.08	0.15
Vegetation	0.86	0.01	0.83
Terrain	0.46	0.03	0.38
Sky	0.88	0.01	0.95
Person	0.65	0.04	0.60
Rider	0.35	0.02	0.54
Car	0.88	0.03	0.77
Truck	0.35	0.00	0.56
Bus	0.46	0.02	0.58
Train	0.42	0.00	0.01
Motorcycle	0.48	0.02	0.65
Bicycle	0.58	0.01	0.18
Mean IOU	0.57	0.047	0.486

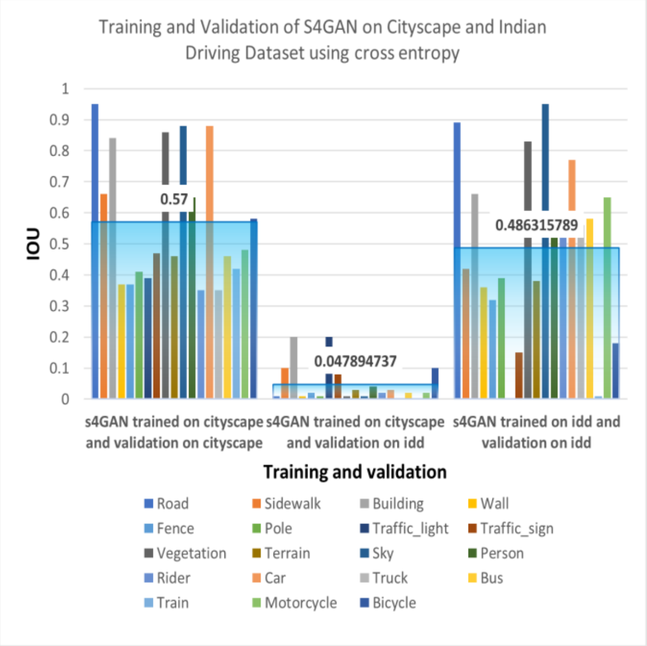

Fig. 1. Bar graph showing IOU and MeanIOU performance of s4GAN firstly trained on cityscape then validated on cityscape and Indian driving dataset secondly s4GAN

B. Result comparison of Combo loss, Cross Entropy loss and Focal loss function on IDD

In our next set of experiments, s4GANs model is being trained on IDD and validation on 19 classes by using IDD dataset while using focal loss function embedded with Gamma value 2 . A comparison is shown of the same model trained on same dataset but using two different loss functions.

A typical cross entropy focuses on those class whose samples numbers are more in data set, enabling it to predict those classes more accurately which carried more weight in the dataset. In order to overcome this, the basic cross entropy loss is extended. Focal loss also focuses on less accurate predicted classes. Firstly cross entropy is taken out and then less accurate predicted classes are given attention in order to predict them more accurately. This training is being performed on 30 classes of the data set IDD. In the computation of Focal loss function, we have set gamma value to 2 which is suitable gamma value that gives an appropriate attention to in corrected predicted classes. The X-axis or probability of ground truth class is denoted by ‘pt’.

Table 2 below show the results of experiments conducted and its IOU against each class leading to a total of 19 classes and Mean IOU. The second column depicts results of s4GAN trained and validated on IDD using cross entropy loss, the third column shows results of s4GAN trained and validated on IDD and using Focal Loss and the fourth column shows results of s4GAN trained and validated on IDD using combo loss function. A graphical representation can also been seen in Fig. 2 showing different Mean IOU against each loss function along with some visualization results in Fig. 3. The first column displays the input image whereas the ground truth is depicted in the second column. The next column displays the predicted output of s4GANS using standard cross entropy, in the fourth column the predicted output of s4GANS using combo loss function and lastly in the fifth column the predicted output of s4GANs under focal loss function can be seen.

As can be seen in Fig. 3, training on a single IDD dataset leads to variations in results of different loss functions. Small objects like speed breakers, cow and road separator are masked and labeled differently using the three loss functions under consideration. Hence there lacks conformity in the results of loss functions.

TABLE 2. SHOWING THE IOU AND MEANIOU OF S4GAN TRAINED AND VALIDATED ON INDIAN DRIVABLE DATASET WHILE USING CROSS ENTROPY, FOCAL LOSS AND COMBO LOSS FUNCTION.

Classes	s4GAN trained and validation on IDD (IOU) using Cross Entropy loss	s4GAN trained and validation on IDD (IOU) using Focal Loss	s4GAN trained and validation on IDD (IOU) using Combo Loss
Road	0.89	0.89	0.89
Sidewalk	0.42	0.40	0.44
Building	0.66	0.64	0.66
Wall	0.36	0.36	0.38
Fence	0.32	0.29	0.30
Pole	0.39	0.36	0.39
Traffic_light	0.00	0.00	0.00
Traffic_sign	0.15	0.18	0.27
Vegetation	0.83	0.82	0.83
Terrain	0.38	0.36	0.38
Sky	0.95	0.94	0.95
Person	0.6	0.58	0.60
Rider	0.54	0.51	0.54
Car	0.77	0.75	0.76
Truck	0.56	0.57	0.57
Bus	0.58	0.53	0.56
Train	0.01	0.01	0.33
Motorcycle	0.65	0.63	0.64
Bicycle	0.18	0.15	0.25
MeanIOU	0.486	0.472	0.512

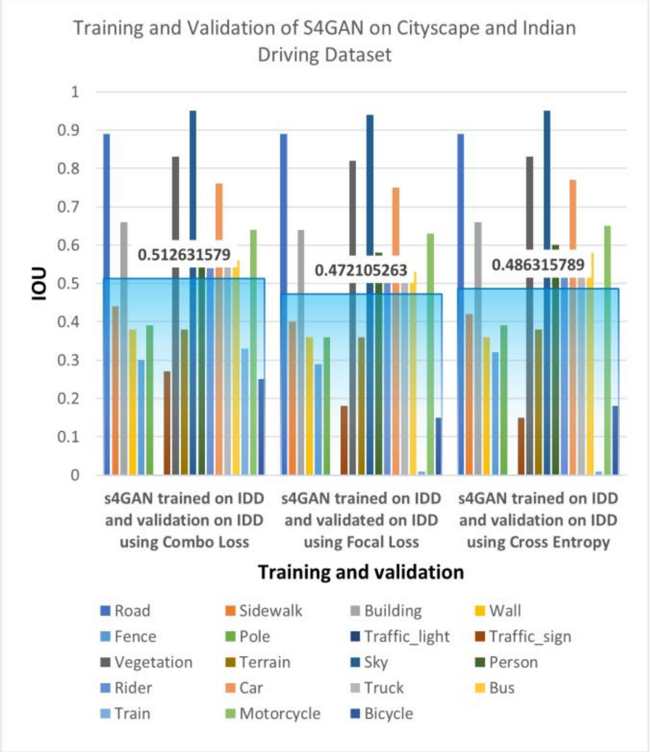

Fig. 2. Bar graph showing IOU and MeanIOU performance of s4GAN trained and validated on Indian driving dataset using combo loss, focal loss and cross entropy loss function

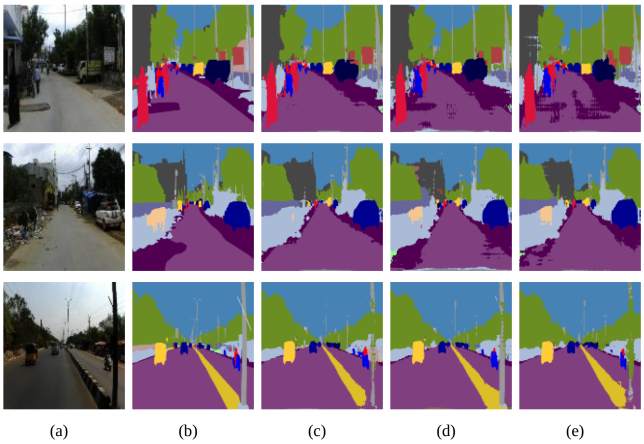

Fig. 3. (a) input; (b) ground truth; (c) s4GAN model trained with cross entropy on IDD dataset; (d) s4GAN model trained with combo loss on IDD dataset; (e) s4GAN model trained with focal loss on IDD dataset

C. *Result comparison of Combo loss, Cross Entropy loss and Focal loss Function on Cityscapes*

Table 3 below show the results of experiments conducted and its IOU against each class leading to a total of 19 classes and Mean IOU. The second column depicts results of s4GAN trained and validated on Cityscapes using cross entropy loss, the third column shows results of s4GAN trained and validated on Cityscapes using Focal Loss and the fourth column shows results of s4GAN trained and validated on Cityscapes using combo loss function. A graphical representation is shown in Fig. 4 showing different Mean IOU against each loss function along with visualization results in Fig. 5. When the s4GANs model is trained on Cityscapes its results show different predicted outputs against each loss function.

TABLE 3. SHOWING THE IOU AND MEANIOU OF S4GAN TRAINED AND VALIDATED ON CITYSCAPES DATASET WHILE USING CROSS ENTROPY, FOCAL LOSS AND COMBO LOSS FUNCTION.

Classes	s4GAN trained and validation on Cityscape (IOU) using Cross Entropy loss	s4GAN trained and validation on Cityscape (IOU) using Focal Loss	s4GAN trained and validation on Cityscape (IOU) using Combo Loss
Road	0.96	0.95	0.96
Sidewalk	0.67	0.62	0.67
Building	0.84	0.82	0.84
Wall	0.36	0.37	0.37
Fence	0.38	0.35	0.37
Pole	0.40	0.38	0.41
Traffic_light	0.38	0.32	0.38
Traffic_sign	0.47	0.40	0.47
Vegetation	0.86	0.85	0.86
Terrain	0.45	0.42	0.44
Sky	0.88	0.86	0.88
Person	0.66	0.61	0.66
Rider	0.35	0.31	0.37
Car	0.88	0.85	0.88
Truck	0.34	0.36	0.40
Bus	0.46	0.37	0.44
Train	0.42	0.43	0.49
Motorcycle	0.45	0.43	0.48
Bicycle	0.58	0.53	0.58
MeanIOU	0.567	0.538	0.576

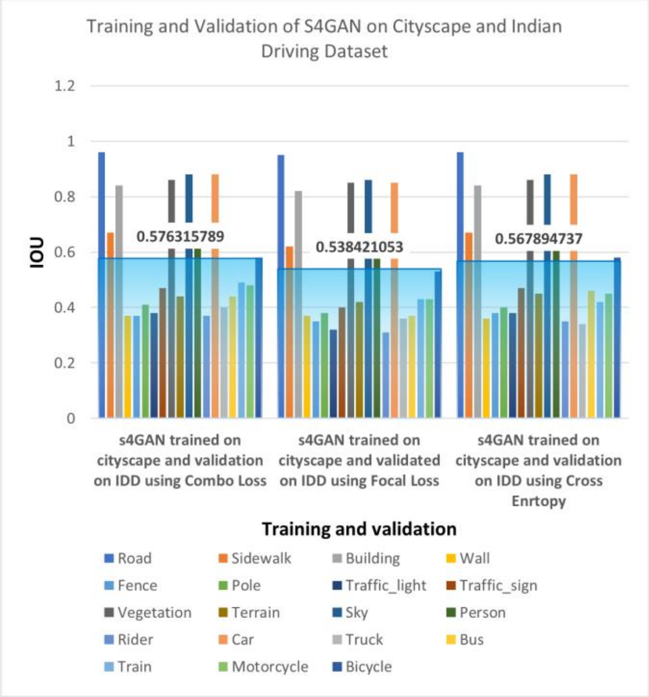

Fig. 4. Bar graph showing IOU and MeanIOU performance of s4GAN trained and validated on Cityscape dataset using combo loss, focal loss and cross entropy loss function.

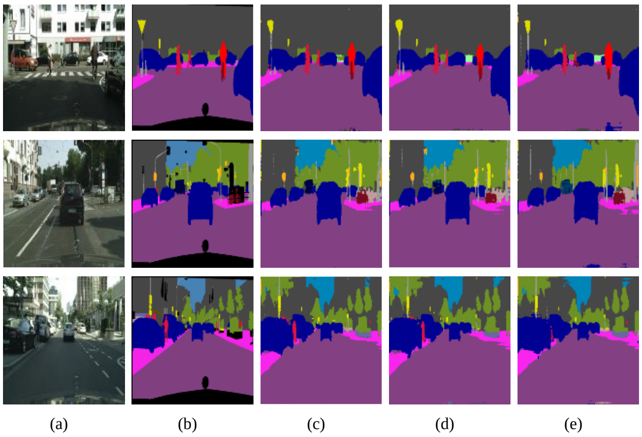
(a) (b) (c) (d) (e)

Fig. 5. (a) image; (b) ground truth; (c) s4GAN model trained with cross entropy on Cityscapes; (d) s4GAN model trained with combo loss on Cityscapes; (e) s4GAN model trained with focal loss on Cityscapes.

V. Methodology

The purpose of our research is to evaluate the effectiveness of various state of the art loss functions. As can be seen from the empirical evidence above, the selection of an appropriate loss function has a direct impact on the accuracy of a semantic segmentation model. It is evident that no single static loss function is currently effective for solving the problem of image classification. While some loss functions emphasize on easy examples, other tend to ignore the simple examples and lay emphasis on the detection of hard examples. The issue is further affected by the type of dataset. The number of classes varies from one dataset to another. Cityscapes for instance consists of 19 classes whereas IDD is a 30 classes dataset.

Studying metadata can help us to gain significant insight into the dynamics and structure of various recognized datasets like IDD, Kitty and Cityscape. Instead of proposing a new loss function, we propose to work on a dynamic and adaptive loss function in future, for training the model without any manual parameter adjustment to classify an image regardless of any type of data set. Our model will be based on an adaptive loss function that can self adjust to incorporate the class imbalance present in any given dataset.

VI. Conclusion

Multiple loss functions have been introduced over the past few years to address the semantic segmentation issues. At the heart of all semantic segmentation models lie a static hard core loss function. Whereas some loss functions like cross entropy focus on major objects and tend to ignore the minor objects, others like focal loss tends to emphasize its learning on negative hard examples and down-weight easy examples. Experimental results on two datasets using various constant loss functions indicate that no single loss function is inclusive to tackle the challenge of object detection in varying scenarios. This signifies the need for a dynamic and self learning loss function which can take into consideration, the variances of different datasets.

Acknowledgment

This work is supported by COMSATS university and the experiments are being conducted on google co-lab.

References

[1] J. Liu, J. Fu, Z. Fang and H. Tian, “Dual attention network for scene segmentation”, CoRR, abs/1809.02983, 2018.

[2] J. Shen, H. Ling and W. Wang “A deep network solution for attention and aesthetics aware photo cropping”, IEEE Trans. Pattern Anal. Mach. Intell, 2019.

[3] T.Y. lin, P. Goyal, R. Girshick and K. He, “Focal loss for dense object detection”, In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2999–3007. IEEE, Piscataway 2017.

[4] S. Ramos, M. Cordts, T. Rehfeld, M. Omran , M. Enzweiler, R. Benenson, and B. Schiele, “The cityscapes dataset for semantic urban scene understanding”, In IEEE CVPR, 2016.

[5] T.Brox and S.Mittal, “Semi-Supervised Semantic Segmentation With High- and Low-Level Consistency”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 43, issue. 4, pp.1369–1379, Dec 2020.

[6] S. You, C. Zhao and H. Li, “Semisupervised and Weakly Supervised Road Detection Based on Generative Adversarial Networks”, IEEE SIGNAL PROCESSING LETTERS, VOL. 25, NO. 4, APRIL 2018.

[7] K. Yamamoto and T. Huy. “Resolving Class Imbalance in Object Detection with Weighted Cross Entropy Losses”, volume abs2006.01413, ArXiv, 2020.

[8] L. Porzi,P. Kontschieder and S. R. Bulo, “In-place activated BatchNorm for memory-optimized training of DNNs”, arXiv:1712.02616, 2017.

[9] J. Zhang, Z. Xu and W. Yang, “An efficient semantic segmentation method based on transfer learning from object detection” Iet Image Processing, DOI.org/10.1049/ipr2.12005, 2020.

[10] S. Subramanian, G. Varma, and A. Namboodiri, “IDD: A Dataset for Exploring Problems of Autonomous Navigation in Unconstrained Environments”, 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), DOI:10.1109/WACV.2019.00190, pp 1743-1751 2019.

[11] W. Li, T. Vercauteren, C. H. Sudre, M. J. Cardoso, and S. Ourselin “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, pp. 240–248, Springer, 2017.

[12] Y. Zheng , S. A. Taghanaki, S. K. Zhuo, “Combo Loss: Handling Input and Output Imbalance in Multi-Organ Segmentation”, Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society, vol 75, pp 24-33, 2019.