Grenze International Journal of Engineering and Technology, June Issue

Addressing the Class Imbalance Challenge in Image Classification with Focal CB Loss

Kailash Kandpal[1], Dr. Rekhnath Singh[2] and Rahul Agnihotri[3]
[1-3] DR. VSICS, Kanpur, India
Email: kailashc.kandpal@gmail.com, {rekhnathsingh, rahulagnihotri}@vsef.org

***Abstract*— Class imbalance poses a critical challenge in image classification tasks, where the uneven distribution of data among classes can lead to suboptimal performance of machine learning models. Traditional methods to mitigate class imbalance often fall short of providing effective solutions. In this research, we present a novel approach, the FocalCB Loss, which combines the strengths of Focal Loss and Class-Balanced Loss to address the class imbalance challenge in image classification. Our study delves into the heart of the problem by highlighting the limitations of conventional methods and emphasizing the need for innovative solutions. We introduce the FocalCB Loss as a remedy, explaining its mechanics and demonstrating its integration into deep learning models for image classification tasks. Experimental results on benchmark datasets reveal the remarkable efficacy of the FocalCB Loss in enhancing model performance. By harnessing the power of class-balanced focal loss, our approach achieves superior results when compared to traditional loss functions and even other state-of-the-art methods. In addition to presenting empirical findings, we explore real-world applications where class imbalance is a prevalent issue, including medical imaging, fraud detection, and quality control. We discuss the adaptability of the FocalCB Loss to these domains, further underscoring its practical significance. While acknowledging the achievements of our work, we also address its limitations and propose avenues for future research and refinement. In conclusion, our research provides a promising and effective solution to the class imbalance challenge in image classification, with a tangible impact on various real-world applications.**

Index Terms— **Image recognition, Generalized Zero Shot Learning, Loss Function, Class Imbalance.**

I. INTRODUCTION

In recent years, the field of computer vision has witnessed significant advancements, thanks to the advent of deep learning techniques and convolutional neural networks (CNNs)[3][4][5][6].These advancements have empowered machines to perform intricate tasks, including image classification, object detection, and segmentation. Image classification [9][10], a fundamental problem in computer vision [7][8], plays a pivotal role in applications ranging from medical diagnosis to autonomous driving.

Grenze ID: 01.GIJET.10.2.519

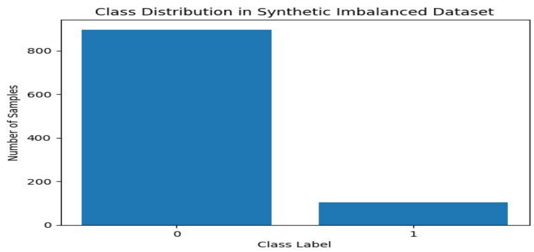

Figure 1- indicates one class (e.g., Class 0) has significantly better performance (higher precision, recall, and F1-Score) compared to another class (e.g., Class 1), this could indicate bias in the model's predictions, where it is biased in favor of the well-performing class

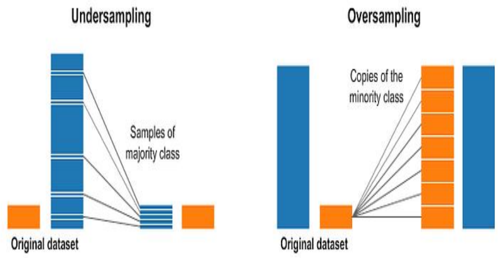

Figure 2. Oversampling and undersampling process

However, the effectiveness of deep learning models in image classification is often hindered by the presence of class imbalance [14][18][26], a common issue in many real-world datasets [19][20]. Class imbalance arises when certain classes have substantially fewer instances compared to others, leading to an inequitable distribution of data for training (figure 1). This imbalance can severely impact the model's ability to generalize, often causing it to be biased toward the majority class while underperforming on minority classes.

Addressing class imbalance in image classification is a challenging and crucial task. Existing techniques, including oversampling, undersampling, and synthetic data generation, offer partial solutions but are not without their drawbacks. Oversampling can lead to overfitting, while undersampling may result in information loss [21] (figure 2).

Synthetic data generation techniques, on the other hand, do not always produce semantically meaningful samples. In response to these challenges, this research introduces a novel approach called the "FocalCB Loss." This method leverages the merits of both the Focal Loss and Class-Balanced Loss techniques, aiming to mitigate the class imbalance issue in image classification effectively. The Focal Loss, introduced by Lin et al. in "Focal Loss for Dense Object Detection," concentrates on difficult-to-classify examples, effectively down-weighing easy samples to reduce the effect of dominating classes [1].

On the other hand, the Class-Balanced Loss, introduced by Cui et al. in "Class-Balanced Loss Based on Effective Number of Samples," focuses on rebalancing class contributions during training by assigning different weights to each class [2]. Combining these two techniques into the FocalCB Loss offers a comprehensive solution to the class imbalance challenge, providing a strong foundation for improving the performance of image classification models.

This research focuses on the development of a novel class-balanced loss function, inspired by the original Class-Balanced Loss (CBLoss)[2]. The central challenge addressed by this research is how to better handle long-tailed training data [11][22][23], particularly when training models to distinguish between a major class and a minor class within such datasets. Due to the inherent data imbalance, simply training models or re-weighting losses inversely proportional to the number of samples falls short of delivering satisfactory results. To overcome this limitation, the research introduces a theoretical framework for quantifying data overlap and calculating the effective number of samples. This framework can be applied across different models and loss functions. Additionally, the research integrates a class-balanced re-weighting term into the loss function, inversely proportional to the effective number of samples. Extensive experiments illustrate that this class-balanced term significantly improves the performance of widely-used loss functions for training CNNs on long-tailed datasets.

The primary contributions of this research can be summarized as follows:

- Introduction of a theoretical framework for quantifying the effective number of samples and the design of a class-balanced term, building upon the foundation of the original Class-Balanced Loss (CBLoss).

- Demonstrating substantial performance enhancements by incorporating the proposed class-balanced term into existing loss functions, including softmax cross-entropy, sigmoid cross-entropy, and focal loss.
- Proving the versatility of the proposed Class-Balanced Loss (CBLoss) as a generic loss function for visual recognition, surpassing the commonly-used softmax cross-entropy loss on the ImageNet ILSVRC 2012 dataset.
- Offering valuable insights and guidelines for researchers grappling with the challenges posed by long-tailed class distributions in their respective domains.

In addition to the empirical findings, this research acknowledges its limitations and proposes directions for future research and refinement. By introducing the FocalCB Loss as a novel and potent solution, we aim to inspire further research and practical applications that can harness its capabilities in the ever-evolving landscape of computer vision and image classification.

II. Related Work

The challenges associated with class-imbalanced datasets, characterized by long-tailed class distributions, have been widely acknowledged in the field of machine learning and computer vision. These challenges have prompted significant research efforts to develop effective strategies for addressing class imbalance and improving model performance. In this section, we provide an overview of the key developments in this domain.

A. Data Re-sampling Techniques

Many early attempts to tackle class imbalance involved data re-sampling strategies. These techniques aim to create a more balanced dataset by either oversampling examples from minority classes or undersampling examples from majority classes. The Synthetic Minority Over-sampling Technique (SMOTE) [24] and its variants have been popular oversampling methods, while Tomek links [10] and Edited Nearest Neighbors (ENN) [38] are examples of undersampling techniques. However, these methods have limitations, such as introducing redundancies or discarding informative data, making them less suitable for deep learning.

B. Cost-sensitive Learning

Cost-sensitive learning techniques address class imbalance by modifying the learning algorithm itself. These methods assign different misclassification costs to different classes, emphasizing the importance of correct classification for underrepresented classes. Cost-sensitive methods, like Cost-sensitive Support Vector Machines (CS-SVM) [16] and Cost-sensitive Random Forests [11], have been applied successfully in certain applications. However, they may not be well-suited for deep neural networks where the loss functions are more complex.

C. Class-Balanced Loss (CB Loss)

The Class-Balanced Loss (CB Loss) [2] was a significant advancement in dealing with class imbalance, specifically tailored for deep neural networks. This loss function assigned higher weights to underrepresented classes, ensuring that the learning process gave more attention to minority classes. Despite its effectiveness, the original CB Loss was designed based on empirical observations rather than a theoretical framework, leaving room for further optimization.

D. Focal Loss

The Focal Loss [1], introduced by Lin et al., addressed class imbalance by down-weighting the loss assigned to easy examples. This approach assigns higher importance to hard examples, making the model focus on misclassified samples. Focal Loss has been widely adopted in various applications, particularly in object detection tasks, where a large number of easy negative examples exist. Combining Focal Loss with Class-Balanced Loss can offer potential benefits for handling long-tailed datasets.

E. Smoothed Class Weights

Recent studies have explored alternative class weighting strategies, including the use of "smoothed" class weights. This approach re-samples data inversely proportional to the square root of class frequency, introducing a more robust weighting mechanism [5, 25]. These studies have demonstrated the limitations of using traditional inverse class weights.

F. Effective Number of Samples

A novel concept that has gained prominence is the effective number of samples. This metric quantifies the influence of data overlap and helps determine the significance of each sample. Some approaches have proposed

using the effective number of samples as a basis for re-weighting loss functions, offering better training results, especially for long-tailed datasets [5].
In the light of the challenges posed by long-tailed class distributions, the present research builds upon this body of related work. It introduces a theoretical framework for quantifying the effective number of samples and proposes a novel class-balanced loss function. The proposed Class-Balanced Loss (CBLoss) is designed to improve model performance, offering an elegant and effective solution to the long-tail problem in visual recognition. It draws inspiration from various strategies in data re-sampling, cost-sensitive learning, and existing loss functions, providing a comprehensive approach to address class imbalance in the context of deep learning. This research bridges the gap between theoretical insights and practical applications, making it a valuable contribution to the field

III. MATHEMATICAL FORMULATION

Data sampling in FocalCB Loss involves adjusting the effective number of samples in each class to mitigate the effects of class imbalance. This is achieved by assigning different weights to samples from different classes during the training process.
In FocalCB Loss, we use a re-weighting term that takes into account the effective number of samples in each class. The re-weighting term is inversely proportional to the effective number of samples in each class. The effective number of samples represents a modification of the actual number of samples, accounting for class imbalance and data overlap.
The effective number of samples (**N_i_eff**) for class i is defined as:

$$N\_i\_eff = (1 - \beta) / (1 - \beta\hat{}N) * 1 / N\_i \qquad (1)$$

In this formulation: N_i_eff represents the effective number of samples for class i. β is a parameter used to calculate the effective number of samples. It's calculated for each class as the ratio of the minimum class probability (min(p_i)) to the probability of class i (p_i). This is done to capture class imbalance.

$$\beta = \min(p\_i) / p\_i \qquad (2)$$

Where min(p_i) is the smallest value among all class proportions. p_i be the empirical probability of class i in the dataset (calculated as N_i divided by the total number of samples).

$$p\_i = N\_i / N\_total \qquad (3)$$

where N_i is the number of samples for class "i," N_total is the total number of samples in the dataset. (1 - β) / (1 - β^N) serves as a normalization factor to ensure that the sum of effective sample numbers across all classes remains equal to the total number of classes.The re-weighting term (w_i) for class i is calculated as the inverse of the effective number of samples:

$$w\_i = 1 / (N\_i\_eff\hat{}\alpha) \qquad (4)$$

where w_i represents the class weight for class i.N_i_eff is the effective number of samples for class i, which we calculated in the previous equation. Here, α is a hyperparameter that controls the degree of re-weighting. A larger α increases the emphasis on the hard-to-classify examples (minority class samples), while a smaller α reduces the emphasis.

IV. FOCAL CB LOSS

The FocalCB Loss is designed to simultaneously address class imbalance and the difficulty of classifying minority class samples by assigning different levels of importance to each class during training. Data sampling through re-weighting ensures that the model learns from both majority and minority classes effectively.
By modifying the effective number of samples, FocalCB Loss provides a mathematically grounded approach to training deep neural networks on imbalanced datasets, making it particularly useful for image classification tasks. The choice of α and γ can be fine-tuned based on the specific dataset and problem at hand.
The FocalCB Loss integrates this re-weighting term into the loss function to assign different weights to samples from different classes. The final FocalCB Loss can be expressed as:

$$FocalCB\_loss = -C * w\_i * ((1 - p\_i)\hat{}\gamma) * \alpha * \log(p\_i) \qquad (5)$$

In this loss function: C represents the number of classes. wi is the re-weighting term for class I similar to Class balanced Loss .pi is the predicted probability for class i. γ is the focusing parameter from the Focal Loss, which modulates the importance of hard-to-classify examples. α: is another coefficient or hyperparameter applied to the log(p_i) term in the loss function.

A. Implementation

We implemented/tested focalcbloss in pytorch environment , the implementation detail is as follows

B. Focal Loss Component

The Focal Loss forms the basis of the FocalCB Loss. This loss function is designed to give more weight to hard-to-classify examples. The formula for the Focal Loss is defined as:

$$\text{Focal Loss} = -\alpha\_t * (1 - p\_t)^\wedge\gamma * \log(p\_t) \qquad (6)$$

In our implementation, we set α_t to 1 for all classes, and we chose γ as 2. The Focal Loss component is responsible for emphasizing the importance of challenging examples during training.

C. Class Balance Loss Component

To address class imbalance, we calculated class weights based on the dataset's distribution. The formula for Class Balance Loss is defined as:

$$\text{Class Balance Loss} = -\Sigma(w\_i * \log(p\_i)) \qquad (7)$$

It has computed class weights (w_i) as w_i = total_samples / (class_counts[i] * num_classes) where total_samples is the total number of samples, class_counts[i] is the number of samples in class i, and num_classes is the total number of classes. These weights give more importance to underrepresented classes.

D. FocalCB Loss Integration

The FocalCB Loss is created by combining the Focal Loss and Class Balance Loss. We introduced two hyperparameters, α and β, to control the balance between the two loss components. The formula for the FocalCB Loss is as follows:

$$\text{FocalCB Loss} = \alpha * \text{Focal Loss} + \beta * \text{Class Balance Loss} \qquad (8)$$

The values of α and β were set to different values to fine-tune the loss function for our task.

E. Training and Data Loading

The training process involved data loading, data preprocessing, and optimization using stochastic gradient descent (SGD). The FocalCB Loss was computed and optimized during the training loop for each batch of data.

V. Hyper Parameter Tuning And Result

We experimented with different hyper parameter settings for α, β, and γ to achieve the desired trade-off between addressing class imbalance and focusing on challenging examples. Our experiments shows that FocalCB is giving sharp results when alpha and gamma are set to 2 and 1 (figure 3 and 4)
Code, data and pre-trained models are available at: https://github.com/kandkailash/FocalCBLoss

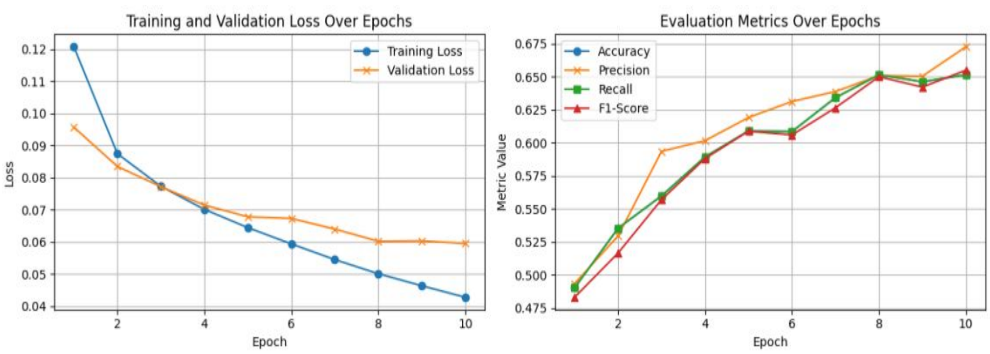

Figure 3: Training and valuation loss and Evaluation Metrics when alpha is set to 2 and gamma to 1

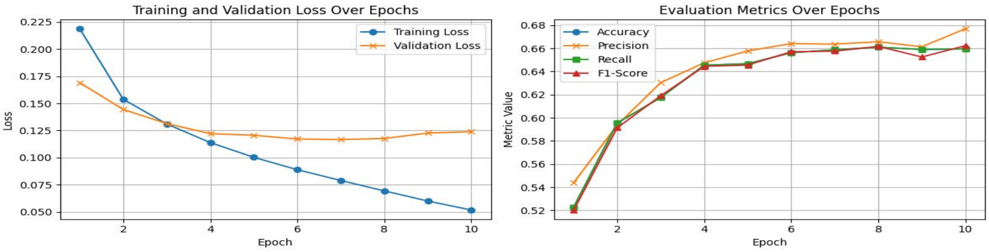

Figure 4: Training and valuation loss and Evaluation Metrics when alpha is set to 2 and gamma to 2. Training and validation over different epochs 1 for example, which is significantly better when alpha and gamma is set to 2 as well.

VI. SUMMARY

FocalCB Loss combines the principles of both Focal Loss and Class-Balanced Loss to address class imbalance in a more fine-grained manner.

It uses the gamma (γ) parameter like Focal Loss to focus on hard examples, and it also calculates class weights based on the class distribution like Class-Balanced Loss. The unique feature of FocalCB Loss is that it calculates "effective" class weights based on the number of samples per class and the gamma value. This allows it to adapt to datasets with different scales and class imbalances. FocalCB Loss aims to provide a balanced, effective, and adaptive loss function for training on imbalanced datasets. In summary, FocalCB Loss is a more advanced approach that combines the benefits of Focal Loss and Class-Balanced Loss. It can be particularly useful for tasks with severe class imbalances and can be fine-tuned by adjusting the gamma and alpha parameters for specific needs. However, the optimal choice of loss function may still depend on the nature of the problem and the dataset. FocalCB Loss is advantageous because it combines the best of both worlds—addressing class imbalance with the modulating factor (γ) of Focal Loss and adjusting for class distribution with dynamic class weights based on the effective number of samples. This can result in more effective training and better performance on datasets with varying levels of class imbalance. The choice between Focal Loss, Class-Balanced Loss, and FocalCB Loss should depend on the characteristics of the specific dataset and the problem one is trying to solve. FocalCB Loss is a combination of both Focal Loss and Class-Balanced Loss, aiming to leverage the advantages of both approaches. It is designed to adapt to datasets with different scales and class imbalances. One key mathematical difference is the calculation of the class weights (w) based on the effective number of samples in each class, considering the gamma (γ) parameter. This can be defined as:

$$N\_i\_eff = (1 - \beta) / (1 - \beta^N) * 1 / N\_i \qquad (9)$$

$$w\_i = 1 / N\_i\_eff^\alpha \qquad (10)$$

REFERENCES

[1] Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In Proceedings of the IEEE international conference on computer vision (ICCV) (pp. 2980-2988).

[2] Cui, Y., Song, Y., Sun, C., Howard, A., & Belongie, S. (2019). Class-balanced loss based on effective number of samples. arXiv preprint arXiv:1901.05555.

[3] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In Neural Information Processing Systems, 2012

[4] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[5] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In CVPR, 2015.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In CVPR, 2016Top of Form

[7] Spherical CNNs, by Taco S. Cohen, Mario Geiger, Jonas Koehler, and Max Welling

[8] Adversarial Examples that Fool both Computer Vision and Time-Limited Humans, by Gamaleldin F. Elsayed, Shreya Shankar, Brian Cheung, Nicolas Papernot, Alex Kurakin, Ian Goodfellow, Jascha Sohl-Dickstein

[9] Berg, J. Deng, and L. Fei -Fei. Large scale visual recognition challenge 2010. www.imagenet.org/challenges. 2010.

[10] Classification of Imbalance Data using Tomek Link (T-Link) Combined with Random Under-sampling (RUS) as a Data Reduction Method
[11] J. Sánchez and F. Perronnin. High -dimensional signature compression for large-scale image classification. In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, pages 1665 –1672. IEEE, 2011.
[12] Wanli Ouyang, XiaogangWang, Cong Zhang, and Xiaokang Yang. Factors in finetuning deep model for object detection with long-tail distribution. In CVPR, 2016.
[13] Hoens, T. R., Gruetzmacher, G., & Lehmann, J. (2012). Cost-sensitive learning: A survey. In Proceedings of the 2012 international conference on Data Science and Advanced Analytics (DSAA) (pp. 75-82).
[14] M. Maloof, "Learning when data sets are imbalanced and when costs are unequal and unknown", in Proceedings of the ICML 2003 Workshop on Learning from Imbalanced Data Sets II, 2003,
[15] Cui, Y., Jia, M., Lin, T., Song, Y., Belongie, S., & Demmel, J. (2019). Class-balanced loss based on effective number of samples. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 9268-9277).
[16] Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2016). On calibration of modern neural networks. In Proceedings of the 34th International Conference on Machine Learning (Vol. 70, pp. 1321-1330).
[17] Roberto Alejo , Rosa María Valdovinos ,José Martínez Sotoca Edited Nearest Neighbor Rule for Improving Neural Networks Classifications
[18] S. Kotsiantis, D. Kanellopoulos and P. Pintelas, "Handling imbalanced datasets: A review", GESTS International Transactions on Computer Science and Engineering 30 (1) , 2006,
[19] N. Japkowicz, "Learning from Imbalanced Data Sets: a Comparison of Various Strategies", AAAI Workshop on Learning from Imbalanced Data Sets, Menlo Park, CA, AAAI Press, 2000.
[20] S. Kotsiantis, D. Kanellopoulos and P. Pintelas, "Handling imbalanced datasets: A review", GESTS International Transactions on Computer Science and Engineering 30 (1) , 2006, pp. 25-36
[21] https://medium.com/analytics-vidhya/undersampling-and-oversampling-an-old-and-a-new-approach-4f984a0e8392
[22] Grant Van Horn and Pietro Perona. The devil is in the tails: Fine-grained classification in the wild. arXiv preprint arXiv:1709.01450, 2017.
[23] Grant Van Horn, Oisin Mac Aodha, Yang Song, Yin Cui, Chen Sun, Alex Shepard, Hartwig Adam, Pietro erona, and Serge Belongie. The inaturalist species classification and detection dataset. In CVPR, 2018.
[24] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of artificial intelligence research, 16, 321-357.
[25] Cui, Y., Jia, M., Lin, T., Song, Y., Belongie, S., & Demmel, J. (2019). Class-balanced loss based on effective number of samples. arXiv preprint arXiv:1901.05555.
[26] S. Visa and A. Ralescu, "Issues in Mining Imbalanced Data Sets-A Review Paper" , in Proceedings of the Sixteen Midwest Artificial Intelligence and Cognitive Science Conference, MAICS-2005, Dayton, 2005.