Land Use Land Cover Scene classification with Focal Loss optimization of Convolutional Networks using Sentinel -2 EuroSAT Dataset

Neha Kumari[1], Sonajharia Minz[1]*

[1] School of Computer and Systems Sciences,

Jawaharlal Nehru University,

New Delhi, India

Abstract

Remote sensing scene classification aims to label images acquired by sensors with a set of semantic labels. With the increasing spatial resolution of sensors, scene classification has gained immense importance over pixel classification in Land Use Land Cover Classification. Deep learning techniques in remote sensing image scene classification have drawn remarkable attention with a looming paradigm shift towards data availability and high computational GPUs. However, inadequate training samples and imbalanced class distribution in land cover classes are the significant challenges that question classification accuracy. In this context, saving time for human annotation at a minimal loss with sufficient accuracy is imperative. High-resolution satellite image scene classification has an inherent class imbalance problem. Class imbalance inherits bias in the model. The model gets more confident in predicting the majority of scenes.

Further, the impact of class imbalance on classification performance is detrimental as the probability of misclassifications increases manifold. Hence, this research elucidates the novel idea of using Class Focal Loss Optimization with differential weight decay assignment to classes in an imbalanced dataset. Also, supervised classifiers lead to extensive overfitting in the dense layers of the network due to excessive parameters in case of inadequate samples. Hence, pre-trained Residual Nets with deeper connections have been proposed to extract pre-trained activations for training the model. For experimentation, we have used Sentinel – 2 satellite images. The experiments were performed on the RGB bands of the EuroSAT dataset. This dataset is benchmarked for the EuroSAT dataset using Convolutional Neural Networks. This novel satellite image dataset comprises 27,000 labeled images with ten different land use and land cover classes. Results corroborated by the experiments illustrated improved class accuracy and fewer misclassifications per class. Accurate predictions have been calculated per class to indicate improvement in class accuracy. To the best of author's knowledge, no prior literature addressed the issue of class imbalance and misclassification in remote sensing scene multiclass classification using Class Focal Loss Optimization.

1. Introduction

Supervised remote sensing scene classification in high-spatial-resolution images is an important research topic. Remote sensing scene classification maps images acquired by sensors with a set of semantic labels [1][2]. Inadequate training samples and imbalanced class distribution in land cover classes are significant challenges that question classification accuracy in scene classification[1-4]. Also, it is costly to manually label the satellite images because of the huge size of the imagery. The analysis is further challenging due to distortions caused by the sensors and atmospheric conditions [1]. Most existing satellite mapping efforts are manual, time-consuming, and give erroneous results. It requires 6-8 hours to map the disasters after satellite images availability manually. Further spatial and spectral variability in resolution makes the automated process more complex. In this context, saving time for human annotation at a minimal loss with sufficient accuracy is imperative [1-4].

Deep learning techniques in remote sensing image scene classification have drawn remarkable attention with a looming paradigm shift towards data availability and high computational GPUs [2-4]. Satellite scene images are rich in spectral and spatial information hence, the choice of feature representation is essential. Satellite image scene understanding has gained momentum, focusing on learning spectral, spatial, and hierarchical features for thematic classification. Supervised classifiers overfit at dense layers due to numerous parameters when representative samples are insufficient. In this context, deep Pretrained Convolutional Networks can leverage feature representations from data. These networks have scalable architectures and can learn high discriminative features contributing to classification accuracy [1-4].

Literature demonstrated many exemplary works on pre-trained deep networks and their importance in classifying satellite images [1-11]. The initial layers of deep networks consist of generic representations. The pre-activations obtained from initial network layers trained on millions of images can be used to learn a model and generalized to cross domains. This concept is called transfer learning [2-4]. Research findings in the literature corroborated increased accuracy and reduced computational parameters with exceptional feature extraction capabilities using transfer learning techniques [1-11]. In [5], the author described the fusion of features from different layers of pre-trained networks and illustrated the extraction of weights from the last layers. Naushad et al. [6] illustrated that Pretrained models like VGG 16 and Residual Networks (Resnet) enhance classification accuracy with fine tuning and data augmentation techniques. In [7], the author compared different transfer learning techniques on sentinel 2 EuroSAT dataset. In [14], author addressed the issue of class imbalance using focal loss optimization in high resolution satellite images.

1.1 Class Imbalance

High-resolution remote sensing image scene classification has an inherent class imbalance problem. Class imbalance can be illustrated as an imbalance in gradient norm distribution. Class imbalance inherits bias in the model. In minority classes, the learned representations are forgotten when weights are updated. So, the scenes in minority and majority classes have different degrees of attenuation [13-16].

The majority class instances dominate gradient descent. The model gets more confident in predicting the majority scenes instead of the minority classes. Further, the impact of class imbalance on classification performance is detrimental as the probability of misclassifications increases manifold [12]. Cross Entropy is the significant loss function used in deep learning algorithms. But this loss function fails to give more attention to minority examples [12]. Focal loss optimization has been used in this research to alleviate this issue. State of the art primarily focused on increasing overall accuracy, and less strenuous efforts were made for misclassifications and model calibration. As per the insights from the literature, no prior work has tried to alleviate imbalance and misclassification in remote sensing scene classification with Focal Loss Optimization [13-16].

1.2 Problem Formulation

Given, a satellite image, I, consisting of scenes s_i, as a tensor.
Let a set of discrete labels $Y = \{y_1, y_2 \ldots y_k\}$, here k is the total number of discrete labels,
Let D be the training set where, $D_{train} = S[i] = \{s_0, s_1 \ldots s_n\}$ and d be the dimensional feature vector $[a_0, a_1 \ldots \ldots a_n]$, $\forall s_i \in D$, then the classification problem can be defined as

$$f: s \to y \tag{1}$$

2. Proposed Model

2.1 State of the Art Focal Loss Function

Tsung-Yi et al. [13] proposed Focal Loss for model calibration in the dense object detection approach in imbalanced classes. Focal Loss optimization gives weight to minority instances to improves the class's accuracy. Differential weight decay assigned to classes can improve accuracy in imbalanced datasets [12]. This work has extended this concept for multiclass classification.

If the predicted output of the model for the classes is $z = [z_1, z_2 \ldots z_C]^T$,where C is the number of classes. SoftMax function calculates the probability distribution on all classes with the assumption of mutually exclusive classes [12].

$$pi = \frac{exp(z_i)}{\sum_{j=1}^{C} (exp\, z_j)} \tag{2}$$

where, $\forall i \in \{1,2 \ldots C\}$

Given, a scene with class label y, then SoftMax Cross-Entropy (CE) Loss is written as

$$CE\ (z, y) = -\, log\left(\frac{exp(z_y)}{\sum_{j=1}^{c} exp\ (z_J)}\right) \tag{3}$$

Cross Entropy is the significant loss function used in deep learning algorithms. But this loss function fails to give more attention to minority examples.

Since the Cross-Entropy Loss function does not work well with imbalanced data, a modulating factor $(1-\rho)^\gamma$ is added to the CE Loss for focusing on misclassified or difficult samples [12-16].

2.2 Class Focal Loss

Focal Loss gives weight to minority and misclassified instances and improves the class's accuracy [12-16]. Class focal loss emphasizes that the weight decay coefficients for classes should be different. The underlying concept accentuates assigning greater weight decay to easy classified samples and smaller to easy examples of minority classes [12].

Class Focal Loss Function is

$$\text{CFL}(z, y) = -(1-p_i)^{\gamma\beta y} \log(p_i) \tag{4}$$

where, γ, is a modulating factor, β_y, represents the proportion of samples with the class label, *y* [12].

$$\beta_y = \frac{Ny}{\sum_{j=1}^{c} N_j} \quad ; \sum_{c=1}^{C} \beta_c = 1, \tag{5}$$

Here, N_j denotes the quantity of samples in category j

2.3 Architecture of proposed Model Residual Nets with Class Focal Loss Optimization

For spatial learning of hierarchical feature vectors, Resnet 18 [13] has been considered. The probabilities assigned by the SoftMAX Function is then optimized through Focal Loss [15].

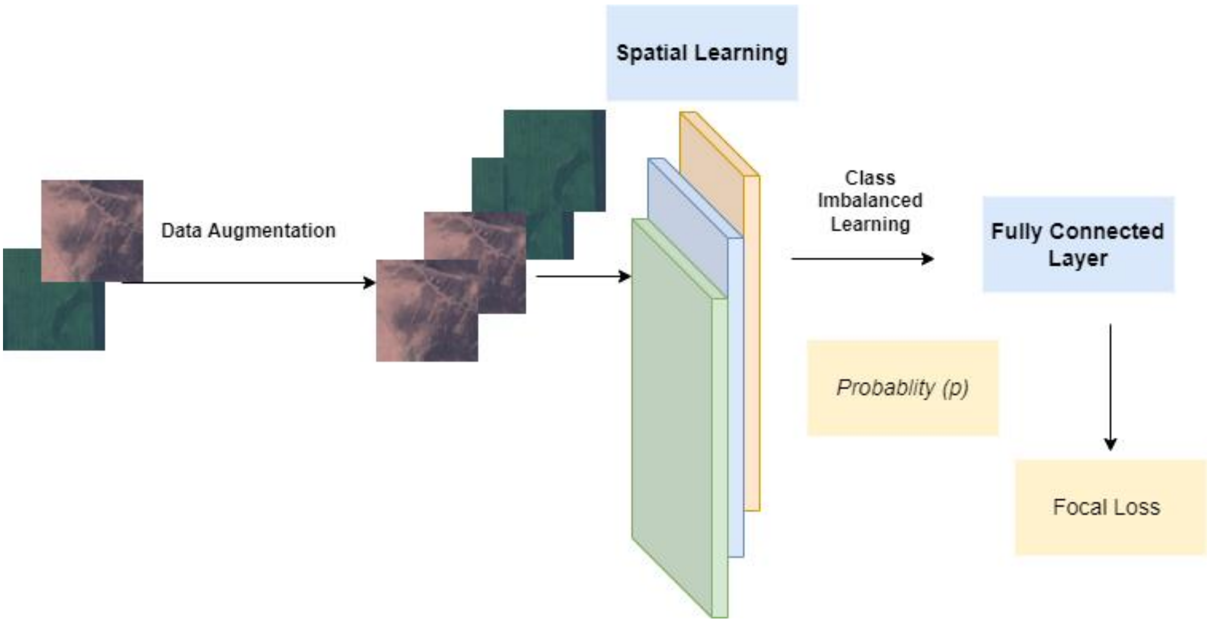

Figure 1. Framework for Scene Classification

3. Experimental Design

3.1 *Experimental Setup and Dataset Description*

Keras Framework has been used for model training. Experiments were performed in NVIDIA Tesla K80 GPU available with Google Colab Environment on publicly available Sentinel-2 satellite EuroSAT dataset. EuroSAT Dataset is the first patch-based LULC dataset. The experiments were performed on the RGB bands. It is benchmarked for the EuroSAT dataset using Convolutional Neural Networks [7-8][17]. This comprises 27,000 labeled images with ten different Land use and Land cover classes. The dataset is georeferenced and is based on open and accessible Earth observation data [17].

Table1 Class Distribution of Dataset

Class Number	Class Name	Number of Samples
1	Annual Crop	3000
2	Forest	3000
3	Herbaceous Vegetation	3000
4	Highway	2500
5	Industrial	2500
6	Pasture	2000
7	Permanent Crop	2500
8	Residential	3000
9	River	2500
10	Sea Lake	3000

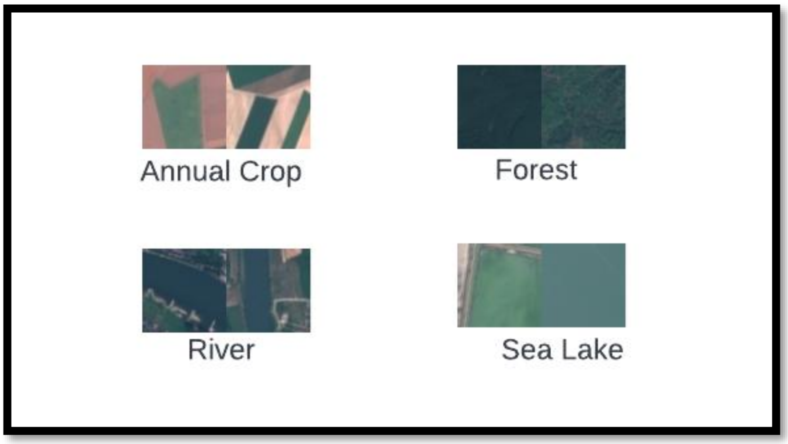

Figure 2. Sample image patches in the proposed EuroSAT Dataset

3.2 Experimental Parameters and Results

The proposed model is evaluated for performance metrics accuracy and F1 score. Model performance was further enhanced using data augmentation techniques, rotation, and scale. The proposed approach reduces misclassifications and improves class accuracy. The experimental settings of different hyperparameters are given in Table 2. Experiments were conducted with different hyperparameters, batch sizes, and learning rates and γ.

Table2 Hyperparameters for training the model

Hyperparameters	Values In Experiments
Batch size	64
Learning Rate	0.0006
Dropout Per layer	.25
γ	0.5(Optimal Value)
Data Augmentation	Rotate, Scale

Table3 Class Accuracy and True predictions with Focal Loss Function

Class	Y_true predictions	Label count	Class Accuracy
Annual Crop	1400	1500	0.9333
Forest	1498	1500	0.9986
Herbaceous Vegetation	1393	1500	0.9286
Highway	1200	1250	0.9600
Industrial	1185	1250	0.9480
Pasture	980	1000	0.9800
Permanent Crop	1199	1250	0.9592
Residential	1496	1500	0.9973
River	1190	1250	0.9520
Sea Lake	1400	1500	0.9333

Table 4 Class Accuracy and F1 Score on RGB Bands of EuroSAT Dataset with Cross Entropy Loss Function

Class Number	Class Name	Number of Samples	Class Accuracy	F1 Score
1	Annual Crop	3000	0.9000	0.8840
2	Forest	3000	0.9180	0.9130
3	Herbaceous Vegetation	3000	0.9550	0.9540

4	Highway	2500	0.9750	0.9659
5	Industrial	2500	0.9888	0.9780
6	Pasture	2000	0.9540	0.9630
7	Permanent Crop	2500	0.9528	0.9329
8	Residential'	3000	0.9986	0.9800
9	River	2500	0.9576	0.9476
10	Sea Lake	3000	0.9280	0.9180

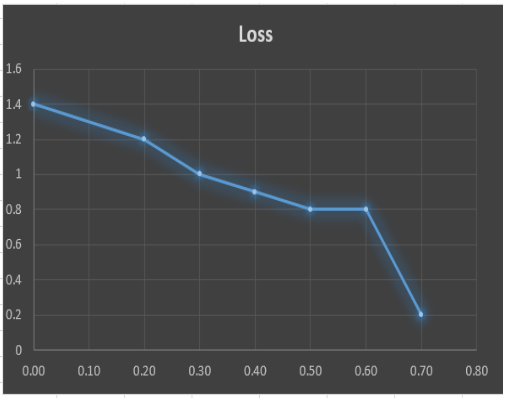

Fig. 3 Convergence of Focal Loss Function with epochs

3.3 Analysis

This section analyses the results presented in the previous section. State of the art primarily focused on increasing overall accuracy, and less strenuous efforts were made for misclassifications and model calibration. The present model achieves better accuracy on the EuroSAT dataset with RGB bands. The proposed technique can be a good choice for LULC scene classification in case of inadequate training samples.

Misclassifications and Focal Loss

Lack of inadequate samples and imbalance in classes in training data has a negative impact on classification performance. The loss function in the Deep learning framework can enhance accuracy and reduce misclassification without altering the model network[12]. Hence, here Focal loss optimization is used instead of the Cross-Entropy Loss function assigning varying weight decay to different samples. Table 4 illustrates accurate predictions in all the classes. There were many misclassifications in the results; many ‘sea lake ‘examples have been misclassified as Forest.

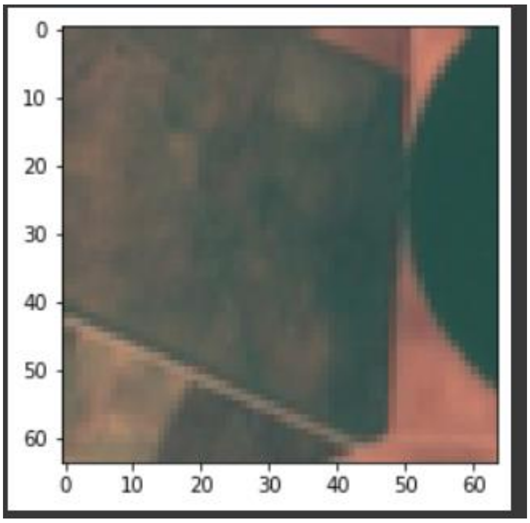

Figure 4. Misclassified Sample Image

This is an 'Annual Crop' example misclassified as 'Permanent Crop.'

Focal Loss reduced misclassifications and increased class accuracy.

4. Conclusions and Future Research Directions

This research proposes a Class Focal Loss optimization with deep networks instead of a Cross-Entropy Loss Function. This function accentuates the concept of assigning differential weight decay to different classes. This idea will be useful in case of imbalanced datasets. Experiments were evaluated on Sentinel – 2 benchmark dataset for Convolutional networks. Results corroborated that this loss function optimization reduces misclassifications and enhances class accuracy.

Further, pre-trained networks mitigate overfitting at the dense layers in the case of supervised classifiers. No prior literature discussed and analyzed the LULC scene multiclass classification with focus on model calibration, misclassification, and class imbalance. Future investigations will compare different class imbalance methods with loss function optimization.

5. References

[1] Jiang, Z. (2018). "A Survey on Spatial Prediction Methods." *IEEE Transactions on Knowledge and Data Engineering* 20(10):1–20. https://doi.org/10.1109/TKDE.2018.2866809.

[2] Ball, J. E., Anderson, D. T., & Chan, C. S. (2017). "A comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community". *Journal of Applied Remote Sensing* 11(04) . https://doi.org/10.1117/1.jrs.11.042609.

[3] Zhang, L., Zhang, L., & Du, B. (2016). Deep learning for remote sensing data: A technical tutorial on state of the art. *IEEE Geoscience and Remote Sensing Magazine* 4(2): 22–40. https://doi.org/10.1109/MGRS.2016.2540798.

[4] Jia, S., Jiang, S., Lin, Z., Li, N., Xu, M., & Yu, S. (2021). "A survey: Deep learning for hyperspectral image classification with few labeled samples." *Neurocomputing* 448: 179–204. https://doi.org/10.1016/j.neucom.2021.03.035.

[5] Marmanis, D., Datcu, M., Esch, T., & Stilla, U. (2016). Deep learning earth observation classification using ImageNet pretrained networks. *IEEE Geoscience and Remote Sensing Letters* 13(1): 105–109. https://doi.org/10.1109/LGRS.2015.2499239.

[6] Naushad, R. & Kaur T(2021). Deep Transfer Learning for Land Use and Land Cover Classification: A comparative study.*Sensors* 21, 8083.https://doi.org/10.3390/s21238083.

[7] Helber, P., Bischke, B., Dengel, A., & Borth, D. (2019). Helber, Eurosat: A novel dataset and deep learning benchmark for land use and land cover Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12(7): 2217–2226. https://doi.org/10.1109/JSTARS.2019.291824.

[8] Helber, P., Bischke, B., Dengel, A., & Borth, D. (2018). Introducing eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *International Geoscience and Remote Sensing Symposium (IGARSS)*, *2018-July*:204–207. https://doi.org/10.1109/IGARSS.2018.8519248.

[9] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. doi:10.48550/ARXIV.1512.03385.

[10] J. Donahue *et al.*,(2013). Decaf: A deep convolutional activation feature for generic visual recognition. Unpublished paper, [Online]. Available: http://arxiv.org/abs/1310.1531.

[11] Wang, M., Zhang, X., Niu, X., Wang, F., & Zhang, X. (2019). Scene Classification of High-Resolution Remotely Sensed Image Based on ResNet. *Journal of Geovisualization and Spatial Analysis 3*(2). https://doi.org/10.1007/s41651-019-0039-9.

[12] Zhu, X., Men, J., Yang, L., & Li, K. (2022). Imbalanced driving scene recognition with class focal loss and data augmentation. *International Journal of Machine Learning and Cybernetics*, *13*(10), 2957–2975. https://doi.org/10.1007/s13042-022-01575-x.

[13] Y. Lin, P. Goyal, and R. Girshick, "Focal loss for dense object detection," Proceedings of the IEEE international conference on computer vision. pp. 2980–2988, 2017.

[14] Bai, H., Cheng, J., Su, Y., Liu, S., Member, G. S., & Liu, X. (2022). Calibrated Focal Loss for Semantic Labeling of High-Resolution Remote Sensing Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *15*, 6531–6547. https://doi.org/10.1109/JSTARS.2022.3197937

[15] Mukhoti, J., Sanyal, A., Golodetz, S., & Torr, P. H. S. (2020). Calibrating Deep Neural Networks using Focal Loss. (NeurIPS).

[16] Wang, Q., Tian, Y., & Liu, D. (2019). Adaptive FH-SVM for Imbalanced Classification. *IEEE Access*, *7*, 130410–130422. https://doi.org/10.1109/ACCESS.2019.2940983.

[17] European Space Agency. (2018) Sentinel online. [Online]. Available: https://sentinel.esa.int/web/sentinel/home.