applied sciences

MDPI

Article

A New Imbalanced Encrypted Traffic Classification Model Based on CBAM and Re-Weighted Loss Function

Jiayu Qin, Guangjie Liu * and Kun Duan

School of Electronical and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China
* Correspondence: gjieliu@nuist.edu.cn

Abstract: The accurate classification of traffic data is challenging for network management and security, especially in imbalanced situations. The limitation of the existing convolutional neural networks is that they have problems such as overfitting, instability, and poor generalization when used to classify imbalanced datasets. In this paper, we propose a new imbalanced encrypted traffic classification model. The proposed model is based on the improved convolutional block attention module (CBAM) and re-weighted cross-entropy focal loss (CEFL) function. The model exploits the redefined imbalance degree to construct a weight function, which is used to reassign the weights of the categories. The improved CBAM based on the redefined imbalance degree can make the model pay more attention to the characteristics of the minority samples, and increase the representation ability of these samples. The re-weighted CEFL loss function can be used to expand the effective loss gap between minority and majority samples. The method is validated on the public ISCX Tor 2016 dataset. The experimental results show that the performance of the new classification model is better than the baseline methods, and the proposed method can remarkably push the precision of the minority categories to 93.28% (14.63%↑), recall to 91.71% (16.98%↑), and F1 score to 92.49% (16.23%↑).

Keywords: imbalance ratio; convolutional neural network; attention module; loss function

Citation: Qin, J.; Liu, G.; Duan, K. A New Imbalanced Encrypted Traffic Classification Model Based on CBAM and Re-Weighted Loss Function. *Appl. Sci.* **2022**, *12*, 9631. https://doi.org/10.3390/app12199631

Academic Editor: Christos Bouras

Received: 11 August 2022
Accepted: 23 September 2022
Published: 25 September 2022

1. Introduction

With the rapid development of online applications, the composition of the flow rate is also richer, such as streaming media, instant messaging, online games, etc. Every day, new applications appear and generate network traffic. Meanwhile, newly emerging encrypted techniques are used to protect the privacy of users. However, various cyberattacks pose enormous challenges to network security monitoring. In order to manage the network, it is necessary to classify network traffic through technical methods and obtain relevant information [1].

Traditional traffic classification methods are mainly divided into four categories [2]: port-based, deep packet inspection (DPI), machine learning, and deep learning. With the rapid growth of port camouflage [3], port randomization [4], and tunneling technology [5], the port-based method is quickly invalidated. The DPI-based methods patterns the packets, and then classifies the traffic based on different matching characteristics [6]. However, these methods have a high complexity, thus, it cannot be applied to the encrypted traffic. Machine-learning methods do not require parsing data inside traffic, which can simplify the computational complexity. These methods rely highly on the design of features sets, which directly affect the classification performance, and require a suitable dataset for classification [7]. To tackle the above problems, research in encrypted traffic classification evolved significantly over time. Recently, deep-learning methods automatically learn layer-by-layer from the raw traffic, and classify them based on the generated high-level features. Although promising results were achieved, the classification precision of minority categories is still low, due to the imbalanced distribution of encrypted traffic data [8].

In this paper, we improve the classification accuracies of minority categories by solving the problems caused by an imbalanced dataset. These problems are currently solved mainly by using over-sampling methods such as synthetic minority over-sampling technique (SMOTE). However, the data it generates may not be sufficiently realistic. To handle the classification problems caused by imbalanced data categories, and improve the classification performance of these categories, we propose a new imbalanced encrypted traffic classification model based on CBAM–CEFL. The experiments are conducted on the public dataset ISCX Tor 2016 that contains more than 18 representative applications of Tor and non-Tor encrypted traffic data.

The main contributions of this paper are summarized as follows:

- We propose a redefined imbalance degree for imbalanced encrypted traffic datasets, and take it as the indicator to construct a weight function to reassign the weights of each category;
- According to the performance tendency problems caused by imbalanced datasets, we propose a re-weighted CEFL loss function to expand the inter-class distance and increase the effective loss gap between the majority and minority samples;
- We improved the channel attention module (CAM) in CBAM with the redefined imbalance degree, which can make the model pay more attention to the characteristics of the minority categories, and increase the representation ability of these samples.

2. Related Work

In the literature, deep-learning-based methods dominated encrypted traffic classification tasks in recent years. Due to the advantages of automatically extracting discriminative features rather than relying on manual design, these methods can classify large amounts of encrypted traffic data. Hence, various deep-learning-based techniques were investigated in the literature to classify encrypted traffic. ICLSTM [9] uses long short-term memory networks (LSTMs), and ETCC [10] uses convolutional neural networks (CNNs) to automatically extract representations from raw packet size sequences of the encrypted traffic. CBD [11] is pre-trained with unlabeled data to classify encrypted traffic from the packet level and traffic level. However, these approaches rely on a large amount of balanced data, while the vast majority of datasets cannot meet this requirement. This phenomenon is observed in multiple fields such as banking fraud detection, medical diagnosis, especially in the field of the network security. In the face of these imbalanced datasets, the traditional deep-learning-based models will be biased towards the majority categories, and correspondingly, the minority categories will be ignored or even misclassified. Hence, for datasets with imbalanced categories, the classification performance of traditional deep-learning-based models is always poor.

In imbalanced encrypted traffic classification, the methods used by researchers are divided into data-level methods, algorithm-level methods, and hybrid methods. Data-level methods reduce the imbalance and noise of data through various data sampling methods, such as over-sampling and under-sampling. Bai [12] proposed to generate samples with similar statistical characteristics to other minority categories to build a balanced dataset. However, as the over-sampling method for resolving the imbalance problem involves repeatedly learning the same data, the classification model can overfit the learning data. Meanwhile, the under-sampling methods proposed to address the imbalance problem may cause information loss as they remove data from the original set. Without changing the data distribution, algorithm-level methods increase the emphasis on minority categories by adjusting learning styles or decision-making processes. A. Telikani [13] proposed to use a cost-sensitive learning method to increase the robustness of deep-learning classifiers against the class imbalanced problem in network traffic classification. Feng [14] proposed to use random forest method for imbalanced traffic classification. It is compared with K-nearest neighbor and C4.5 decision tree algorithms, and the experimental results show that it can classify imbalanced encrypted traffic more effectively. Hybrid methods resample the data to reduce noise and imbalance, and then use algorithm-level methods to further reduce

bias towards most classes. COUSS [15] introduces a combined over-sampling and under-sampling method based on the slow-start algorithm. Compared with synthetic minority over-sampling technique (SMOTE) and generative adversarial network over-sampling algorithms, the method improves the F1 score by 8.639% and 4.074%, respectively.

3. Methodology

In this paper, we aim to improve the classification performance of the model in an imbalanced encrypted traffic situation. This section presents the re-weighted CEFL loss function algorithm formulation and the improved CBAM, which are based on redefined imbalanced degree (ID), including a description of other loss functions and the original CBAM, for comprehensive comparisons. The method structure diagram of this paper is shown in Figure 1.

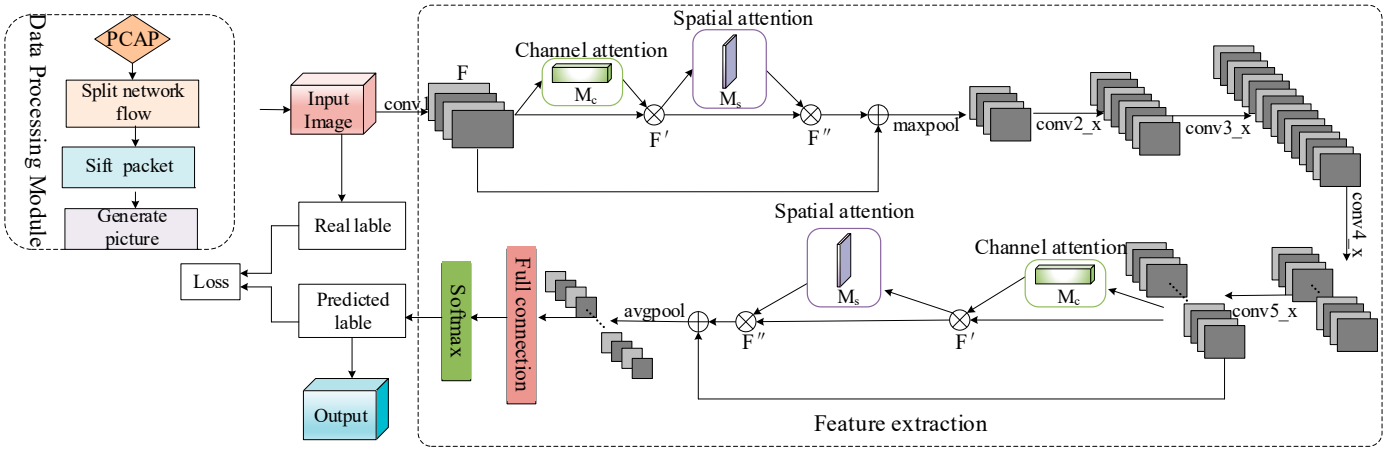

Figure 1. Overview of network framework.

3.1. Imbalance Degree Based on Information Entropy

The symbol ID is generally expressed as the ratio between the maximum and minimum number of samples. Suppose that there are C categories in the dataset N, and the ID calculation formula is shown in Equation (1).

$$\mathrm{ID} = \frac{\max\{N_i\}}{\min\{N_i\}} \tag{1}$$

where N_i represents the number of samples in category i, i = 1, 2, 3, . . . , C. In order to transform the traditional ID into multi-class imbalanced datasets, we redefine the ID′ using information entropy, whose calculation formula is shown in Equation (2).

$$\mathrm{ID}' = -\frac{1}{C}\sum_{i=1}^{C} \log \frac{N_I}{\max\{N_i\}} \tag{2}$$

where C represents the number of categories, N_i represents the number of samples in category i, I = 1, 2, 3, . . . , C, and the log default base is 10.

As shown in Table 1, the redefined ID is more reasonable and accurate. A larger number of ID′ indicates that the distribution of the dataset category is more uneven.

Taking ID′ as the decision indicator, the constructed weight function is defined as:

$$f(\lambda, N_i, N_{max}) = \begin{cases} \left(\frac{N_{max}}{N_i}\right)^{\frac{1}{2}} & \mathrm{ID} \le 1 \\ \left(\frac{N_{max}}{N_i}\right)^{\frac{1}{\lambda}} & \text{other} \end{cases} \tag{3}$$

In Equation (3), $N_{max} = \max_i\{N_i\}$, λ is a hyperparameter.

Table 1. Imbalance of different category distributions.

Category Distribution	ID	ID′
[1,1,1,1,10]	10	0.8000
[2,4,6,8,10]	5	0.2831
[1,10,10,10,10]	10	0.2000
[10,10,10,10,10]	1	0
[1,10,100,1000,10000]	10,000	2

As shown in Table 2, the weight function is used to assign weights to obtain various categories of weight coefficients.

Table 2. Imbalance of different category distributions.

Category Distribution	ID′	$\mathbf{f(\lambda,N_i,N_{max})}$ $(\lambda=4)$
[1,1,1,1,10]	0.8000	[3.1623,3.1623,3.1623,3.1623,1]
[2,4,6,8,10]	0.2831	[2.2361,1.5811,1.2910,1.1180,1]
[1,10,10,10,10]	0.2000	[3.1623,1,1,1,1]
[10,10,10,10,10]	0	[1,1,1,1,1]
[1,10,100,1000,10000]	2	[10,5.6234,3.1623,1.7783,1]

3.2. Channel–Spatial Domain Attention Module

In 2018, ref. [16] proposed to use the CBAM module, which is composed of the channel attention module (CAM) and the spatial attention module (SAM), to optimize adaptive features from the channel dimension and spatial dimension. The CAM can learn more important feature content, and the SAM can learn more important feature positions. However, the CAM simply adds the eigenvectors obtained after the pooling layer without considering the correlation between the imbalanced data and vector [17]. To this end, we improve the CBAM module with the weight function to increase the representation ability of these samples.

As shown in Figure 2, the feature graph obtained after the convolutional layer is used as the input. We compress the spatial dimensions by global max pooling (GMP) and global average pooling (GAP) and obtain two $1 \times 1 \times C$ feature vectors called MaxPool and AvgPool. We input the feature vectors into the same multi-layer perception (MLP) to obtain the GMP-based weight vector W_1 and GAP-based weight vector W_2. We then use the activation function to operate on M, which is obtained by adding the two weight vectors and multiplying the weight function, to obtain the eigen weight vector $M_C(F)$. In addition, we multiply $M_C(F)$ with the input feature map to assign each category the corresponding weight.

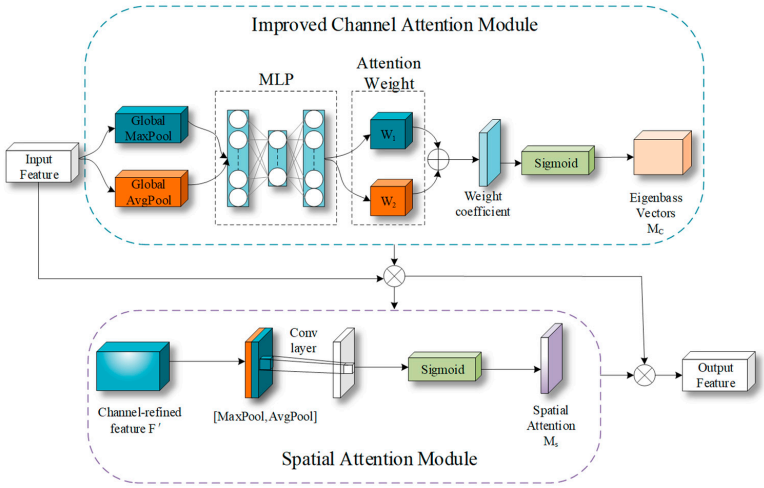

Figure 2. The framework of the proposed method.

The $M_C(F)$ calculation formula is shown in Equation (4), where σ represents the sigmoid function.

$$M_C(F) = \sigma[(W_1 + W_2)f(\lambda, N_i, N_{max})] \quad (4)$$

3.3. Heavily Weighted Cross-Entropy Focal Loss Function

The loss function is an operation function that measures the degree of difference between the predicted and the true value. The common loss functions such as CE loss perform well when trained on balanced datasets, such as CIFAR-10/100 datasets and MNIST datasets. However, CE loss processes all samples equally, even if the datasets are imbalanced, which leads to the contribution of the minority categories being ignored [18]. In this paper, we propose an ID′-based re-weighted CEFL loss, which combines the characteristics of CE loss and FL loss. This makes the model pay more attention to the minority categories.

3.3.1. CE Loss Function

The CE loss function is one of the most commonly used loss functions in classification problems. The calculation formula is shown in Equation (5), where C represents the number of sample categories, $y_i \in \{0, 1\}$ represents the true label of the sample, and $\hat{y}_i$ represents the probability that the sample predicts correctly.

$$L_{CE} = -\sum_{i=1}^{C} y_i \log \hat{y}_i \quad (5)$$

During the initial training period of the model, the $\hat{y}_i$ of the same category is almost unchanged, and $\hat{y}_i$ gradually increases to 1 as the training progresses. It means that CE loss is biased towards categories with large sample sizes in imbalanced datasets.

3.3.2. Focus Loss Function

The FL loss function proposed by Huang et al. [19] greatly solves the problems caused by the imbalanced categories. It adjusts the proportion of losses in positive and negative samples by reducing the weight of loss for negative samples, so that the model is not be biased towards the negative samples. The calculation formula is shown in Equation (6), where α and γ are hyperparameters. α is the proportional parameter that controls the importance of positive and negative samples, and the focus parameter γ is used to adjust the rate at which the weight of the sample decreases. When $\gamma = 0$, the FL loss function is equivalent to the weighted CE loss.

$$L_{FL} = -\sum_{i=1}^{C} \alpha y_i (1 - \hat{y}_i)^{\gamma} \log \hat{y}_i \quad (6)$$

3.3.3. CEFL Loss Function

The FL function only reduces the amplitude of losses for the categories with small sample sizes. It does not make a loss allocation of these samples. Therefore, we propose a re-weighted CEFL loss function to control the weights of CE loss and FL loss, which is defined as:

$$L_{CEFL} = \sum_{i=1}^{C} -(1 - \hat{y}_i) \log \hat{y}_i - y_i (1 - \hat{y}_i)^{\gamma} \log \hat{y}_i \quad (7)$$

In Equation (7), the weights of $\hat{y}_i$ and $1 - \hat{y}_i$ are assigned to the CE loss and FL loss, respectively. When the sample prediction probability is small ($\hat{y}_i < 0.5$), the CEFL loss training is closer to the CE loss; conversely, when $\hat{y}_i > 0.5$, the CEFL loss training is closer to the FL loss, which means that for samples of the majority categories, less loss is allocated by FL loss, and for the other categories, more loss is allocated by CE loss [20]. Therefore, the polarization trend of well-classified and poorly classified samples is conducive to reducing the loss in minority samples and increasing the loss in majority samples, making model training pay more attention to minority samples.

To enhance the classification performance of the loss function, we introduce a reconstructed weight function for CEFL loss, and the calculation formula is shown in Equation (8). On this basis, a weight function is added to CEFL loss to enhance the allocation performance of CEFL loss. Considering the performance difference caused by the imbalance in sample number, the re-weighted CEFL loss can widen the distance between classes to a certain extent. It increases the effective loss gap of the majority and minority samples. It is calculated as:

$$L_{CEFL} = \sum_{i=1}^{C}(-(1-\hat{y}_i)\log\hat{y}_i - y_i(1-\hat{y}_i)^{\gamma}\log\hat{y}_i)f(\lambda, N_i, N_{max}) \quad (8)$$

For imbalanced datasets, the re-weighted CEFL loss can widen the inter-class distance, which increases the effective loss gap between the majority and minority samples.

4. Experiment and Analysis

4.1. Dataset and Pre-Processing

The encrypted traffic dataset used in this paper is the UNB ISCX Tor 2016 [21], which contains more than 18 representative applications of Tor and non-Tor encrypted traffic, such as Facebook, Skype, Spotify, Gmail, etc. We categorize the dataset by services, forming the dataset with eight categories, namely, browsing, chat, audio-streaming, video-streaming, mail, VoIP, P2P, and file transfer. The sample size and proportion of each type in the dataset are shown in Table 3.

Table 3. Sample size and proportion in the dataset.

Service Category	Sample Size	Proportion (%)
VoIP	684,601	54.20%
File transfer	271,804	21.52%
P2P	228,300	18.07%
Browsing	39,323	3.11%
Video-streaming	16,923	1.34%
Audio-streaming	13,727	1.09%
Mail	5076	0.40%
Chat	3419	0.27%

According to Table 3, the sample size of the eight different categories in the dataset is extremely uneven, such as VoIP accounting for 54.20%, while browsing, chat, audio-streaming, video-streaming, and mail account for less than 5%. The imbalance in the sample sizes severely affects the classification performance of the model. Thus, we propose several optimization methods to improve the classification accuracy of these minority samples in the dataset.

Data pre-processing can reduce the noise in the raw traffic and adjust the traffic to fit the input form of the deep-learning model. The following pre-processing steps contain:

- Data splitting: since the CNN model requires the input data to be of the same size, we split the traffic data at the session level and pick the first 784 bytes;
- Data cleaning: we remove the packets in .pcap format without application layer data to avoid generating bin files with no actual content;
- Image generation: we convert the file with the length of 784 bytes into a grayscale image in binary form, one byte for one grayscale pixel value.

4.2. Evaluate Metrics

Different from the classification of balanced datasets, a single classification accuracy does not fully reflect the classification performance of the model. We evaluate and compare

the performance of our methods by three typical metrics, including precision (PR), recall (RC), and F1, which are designed as:

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

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

$$\text{F1} = \frac{2{*}\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{11}$$

where PR denotes the ratio of correctly predicted samples to total positively predicted samples, RC denotes the ratio of correctly predicted samples to actual positive samples in the dataset, and F1 is the harmonic mean of precision and recall.

4.3. Results and Analysis

4.3.1. Parameter Settings

All experimental models are based on PyTorch framework. The dataset is divided into the training set, the validation set, and the testing set, according to the ratio of 8:1:1. During the model training period, the batch size is 64 and the epoch is 20. The learning rate is set to 1×10^{-3}, the momentum is 0.9, and the weight decay parameter is 1×10^{-4}.

4.3.2. Compare Experiment

In the 2012 ILSVRC Challenge, AlexNet won the championship and was well ahead of second place. This led to extensive research on AlexNet, and led to the belief that "the deeper the network, the higher the accuracy." As VGGNet, Inception v1, Inception v2, and Inception v3 have been continuously verified and strengthened, this belief is increasingly recognized. However, as the number of layers of the network increases, the loss in the training dataset rises, which causes the degradation problem of neural networks. To solve the degradation problem, Kaiming [22] proposed the residual neural network (ResNet), which consists of residual blocks. The structure of the residual block is shown in Figure 3.

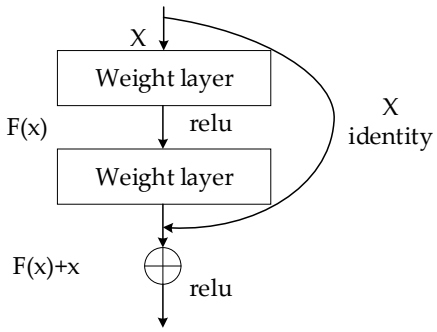

Figure 3. Residual learning: a building block.

In order to evaluate the classification performance of the proposed loss function, we compare the CEFL loss, CE loss, and FL loss used in various ResNet classification models. The experimental results are shown from Tables 4–6.

As shown from Tables 4–6, the proposed CEFL loss without using the weight function obtains better classification performance than CE loss and FL loss, which can be improved by about 1% to 5%. As the number of network layers deepens, the F1 score of CEFL loss increases 1.7% from ResNet-18 to ResNet-34, then decreases by about 1.2% from ResNet-34 to ResNet-50. Finally, the performance of ResNet-50 is increased by 5.2% over ResNet-101. However, the classification performance of models that use the re-weighted CEFL loss function is further improved. In particular, when the parameter $\gamma = 1, \lambda = 4$, the model achieves the best performance on F1 score, with 88.63%.

Table 4. Accuracy rates based on different loss functions of various ResNets.

Classification Model / Loss Function	ResNet-18	ResNet-34	ResNet-50	ResNet-101
CE Loss	69.95%	73.71%	79.82%	81.34%
FL ($\alpha = 0.25\ \gamma = 2$)	78.42%	82.52%	84.65%	85.35%
FL ($\alpha = 0.75\ \gamma = 2$)	76.04%	77.23%	81.53%	83.79%
FL ($\alpha = 1\ \gamma = 2$)	75.18%	74.26%	81.81%	84.92%
CEFL ($\gamma = 1$)	82.17%	78.82%	85.35%	88.50%
CEFL ($\gamma = 2$)	81.86%	77.74%	85.73%	88.63%
CEFL ($\gamma = 5$)	82.39%	79.91%	86.49%	89.14%
CEFL ($\gamma = 1\ \lambda = 4$)	87.42%	89.91%	90.25%	92.70%
CEFL ($\gamma = 2\ \lambda = 4$)	87.65%	87.56%	89.12%	93.77%
CEFL ($\gamma = 5\ \lambda = 4$)	88.07%	87.77%	91.05%	92.35%

Table 5. Recall rates based on different loss functions of various ResNets.

Classification Model / Loss Function	ResNet-18	ResNet-34	ResNet-50	ResNet-101
CE Loss	72.33%	71.62%	74.07%	71.19%
FL ($\alpha = 0.25\ \gamma = 2$)	72.99%	67.75%	72.20%	71.16%
FL ($\alpha = 0.75\ \gamma = 2$)	69.94%	69.14%	66.78%	71.66%
FL ($\alpha = 1\ \gamma = 2$)	71.40%	72.83%	69.05%	68.55%
CEFL ($\gamma = 1$)	68.83%	79.62%	63.29%	76.13%
CEFL ($\gamma = 2$)	73.10%	83.36%	67.50%	78.50%
CEFL ($\gamma = 5$)	71.94%	77.25%	69.84%	76.78%
CEFL ($\gamma = 1\ \lambda = 4$)	89.87%	85.25%	82.66%	84.61%
CEFL ($\gamma = 2\ \lambda = 4$)	87.83%	88.36%	80.21%	85.02%
CEFL ($\gamma = 5\ \lambda = 4$)	87.12%	88.77%	81.79%	85.42%

Table 6. F1-score based on different loss functions of various ResNets.

Classification Model / Loss Function	ResNet-18	ResNet-34	ResNet-50	ResNet-101
CE Loss	71.12%	72.65%	76.84%	75.93%
FL ($\alpha = 0.25\ \gamma = 2$)	75.61%	74.41%	77.93%	77.61%
FL ($\alpha = 0.75\ \gamma = 2$)	72.86%	72.96%	73.42%	77.25%
FL ($\alpha = 1\ \gamma = 2$)	73.24%	73.54%	74.89%	75.86%
CEFL ($\gamma = 1$)	74.91%	79.22%	72.68%	81.85%
CEFL ($\gamma = 2$)	77.23%	80.45%	75.53%	83.26%
CEFL ($\gamma = 5$)	76.81%	78.56%	77.28%	82.50%
CEFL ($\gamma = 1\ \lambda = 4$)	88.63%	87.52%	86.29%	88.47%
CEFL ($\gamma = 2\ \lambda = 4$)	87.74%	87.96%	84.43%	89.18%
CEFL ($\gamma = 5\ \lambda = 4$)	87.59%	88.27%	86.17%	88.75%

In order to verify the effectiveness of the improved CBAM module, this section introduces the improved CBAM module, traditional CBAM module, and the single-domain attention module to the CNN model for comparative experiments. The experimental results are shown from Tables 7–9.

From the results presented above, it can be observed that the various ResNet models introduced to the improved CBAM module increase significantly in their PR, RC, and F1 score. This is because the CBAM module improves with the redefined imbalance degree, and can extract features that are more conducive to sample classification. It makes the model pay more comprehensive attention to the characteristics of minority samples, enhance classification accuracy, and thus, improves the classification effect of the model.

Table 7. Accuracy rates of classification models under each attention mechanism.

Classification Model / Added Modules	ResNet-18	ResNet-34	ResNet-50	ResNet-101
No attention module	75.21%	74.17%	78.47%	77.24%
Spatial domain attention module	77.85%	75.88%	78.89%	77.17%
Channel domain attention module	75.59%	77.29%	79.27%	79.55%
CBAM module	79.83%	79.25%	80.88%	82.23%
Improved CBAM module	83.52%	85.81%	87.22%	85.64%

Table 8. Recall rates of classification models under each attention mechanism.

Classification Model / Added Modules	ResNet-18	ResNet-34	ResNet-50	ResNet-101
No attention module	72.39%	73.28%	72.98%	75.42%
Spatial domain attention module	71.84%	75.14%	74.51%	77.09%
Channel domain attention module	74.51%	75.53%	75.26%	76.15%
CBAM module	77.37%	77.88%	78.16%	79.27%
Improved CBAM module	84.21%	85.42%	86.29%	85.51%

Table 9. F1 score of classification models under each attention mechanism.

Classification Model / Added Modules	ResNet-18	ResNet-34	ResNet-50	ResNet-101
No attention module	73.77%	73.72%	75.61%	77.74%
Spatial domain attention module	74.71%	75.51%	76.63%	78.99%
Channel domain attention module	74.05%	76.40%	77.21%	78.65%
CBAM module	79.99%	78.56%	81.81%	80.25%
Improved CBAM module	83.86%	85.61%	86.75%	86.07%

We combine the re-weighted CEFL loss function and the improved CBAM module for a comprehensive comparative experiment. The results are shown in Tables 10–12.

Table 10. Accuracy rates for each classification model.

Classification Model	ResNet-18	ResNet-34	ResNet-50	ResNet-101
Not CEFL loss	75.21%	74.17%	78.47%	77.24%
CEFL loss ($\gamma = 1$ $\lambda = 4$)	87.42%	89.91%	90.25%	92.70%
Improved CBAM module	83.52%	85.81%	87.22%	85.64%
Improved CBAM module + CEFL loss ($\gamma = 1$ $\lambda = 4$)	89.04%	90.53%	90.75%	93.28%

Table 11. Recall rates for each classification model.

Classification Model	ResNet-18	ResNet-34	ResNet-50	ResNet-101
Not CEFL loss	72.39%	73.28%	72.98%	75.42%
CEFL loss ($\gamma = 1$ $\lambda = 4$)	89.87%	85.25%	82.66%	84.61%
Improved CBAM module	84.21%	85.42%	86.29%	85.51%
Improved CBAM module + CEFL loss ($\gamma = 1$ $\lambda = 4$)	90.26%	89.84%	90.17%	91.71%

Table 12. F1-score for each classification model.

Classification Model	ResNet-18	ResNet-34	ResNet-50	ResNet-101
Not CEFL loss	73.77%	73.72%	75.61%	77.74%
CEFL loss ($\gamma = 1$ $\lambda = 4$	88.63%	87.52%	86.29%	88.47%
Improved CBAM module	83.86%	85.61%	86.75%	86.07%
Improved CBAM module + CEFL loss ($\gamma = 1$ $\lambda = 4$)	91.14%	90.68%	91.44%	92.49%

From the results presented from Tables 10–12, it can be observed that after using the CEFL loss function and the improved CBAM module, the performance of each classification model is significantly enhanced, which is better than only introducing one improved scheme. Further, validation accuracy is also compared with other models named as AAE (87.3%) and GAN (82.8%). After comparison, we can find that the classification accuracies of similar architectures are not as good as the architecture mentioned in this paper. Therefore, the effectiveness of the encrypted traffic classification model proposed in this paper is confirmed.

5. Conclusions, Discussion, and Future Research

In this paper, we propose a new imbalanced encrypted traffic classification model, which is based on the improved CBAM and CEFL loss function, to solve the problems caused by imbalanced datasets. We construct a weight function with a redefined imbalance degree to reassign the weights of each category. To expand the inter-class distance, we propose a reweighted CEFL loss, which increases the effective loss gap between the majority and minority samples. In addition, we take the redefined imbalance degree as an indicator to improve the CBAM. It makes the model pay more attention to the characteristics of the minority categories, and increases the representation ability of these samples. The results confirm the superior performance of the proposed classification model by pushing the precision, recall, and F1 to 93.28% (14.63%↑), 91.71% (16.98% ↑), and 92.49% (16.23%↑), respectively. In the future, we would like to investigate the ability of the model to predict new classes of samples and to resist sample attacks.

Author Contributions: Conceptualization, J.Q.; methodology, J.Q.; software, J.Q.; validation, J.Q., G.L. and K.D.; formal analysis, J.Q.; investigation, J.Q.; resources, J.Q.; data curation, J.Q.; writing—original draft preparation, J.Q.; writing—review and editing, G.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China and Nanjing University of Information Science and Technology Talent Start-up Fund Project, grant number 61931004, 62072250, 2020r061 and U21B2003.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dong, C.; Zhang, C.; Lu, Z.; Liu, B.; Jiang, B. CETAnalytics: Comprehensive effective traffic information analytics for encrypted traffic classification. *Comput. Netw.* **2020**, *176*, 107258. [CrossRef]
2. Wu, H.; Zhang, X.; Yang, J. Deep Learning-Based Encrypted Network Traffic Classification and Resource Allocation in SDN. *J. Web Eng.* **2021**, *20*, 2319–2334. [CrossRef]
3. Mills, G.A.; Pomary, P.; Togo, E.; Sowah, R.A. Detection and Management of P2P Traffic in Networks using Artificial Neural Networks. *J. Netw. Syst. Manag.* **2022**, *30*, 26. [CrossRef]
4. Islam, F.U.; Liu, G.; Zhai, J.; Liu, W. VoIP Traffic Detection in Tunneled and Anonymous Networks Using Deep Learning. *IEEE Access* **2021**, *9*, 59783–59799. [CrossRef]
5. Huang, Y.F.; Lin, C.B.; Chung, C.M.; Chen, C.M. Research on QoS Classification of Network Encrypted Traffic Behavior Based on Machine Learning. *Electronics* **2021**, *10*, 1376. [CrossRef]
6. Lin, X.; Xiong, G.; Gou, G.; Li, Z.; Shi, J.; Yu, J. ET-BERT: A Contextualized Datagram Representation with Pre-training Transformers for Encrypted Traffic Classification. In Proceedings of the 2022 ACM Web Conference (WWW), Lyon, France, 25–29 April 2022; pp. 633–642.
7. Zhang, X.Q.; Zhao, M.; Wang, J.Y.; Li, S.; Zhou, Y.; Zhu, S.N. Deep-Forest-Based Encrypted Malicious Traffic Detection. *Electronics* **2022**, *11*, 977. [CrossRef]
8. Yao, H.P.; Liu, C.; Zhang, P.Y.; Wu, S.; Jiang, C.X.; Yu, S. Identification of Encrypted Traffic Through Attention Mechanism Based Long Short Term Memory. *IEEE Trans. Big Data* **2022**, *8*, 241–252. [CrossRef]

9. Lu, B.; Luktarhan, N.; Ding, C.; Zhang, W. ICLSTM: Encrypted Traffic Service Identification Based on Inception-LSTM Neural Network. *Symmetry* **2021**, *13*, 1080. [CrossRef]
10. Li, Y.; Lu, Y. ETCC: Encrypted Two-Label Classification Using CNN. *Secur. Commun. Netw.* **2021**, *2021*, 6633250. [CrossRef]
11. Hu, X.; Gu, C.; Chen, Y.; Wei, F. CBD: A Deep-Learning-Based Scheme for Encrypted Traffic Classification with a General Pre-Training Method. *Sensors* **2021**, *21*, 8231. [CrossRef] [PubMed]
12. Bai, L.; Lu, H.; Liu, Y. High-Efficiency Observations:Compressive Sensing and Recovery of Seismic Waveform Data. *Pure Appl. Geophys.* **2020**, *177*, 469–485. [CrossRef]
13. Telikani, A.; Gandomi, A.H.; Choo, K.R.; Shen, J. A Cost-Sensitive Deep Learning-Based Approach for Network Traffic Classification. *IEEE Trans. Netw. Serv. Manag.* **2022**, *19*, 661–670. [CrossRef]
14. Zhang, F.; Shang, T.; Liu, J. Imbalanced Encrypted Traffic Classification Scheme Using Random Forest. In Proceedings of the 2020 International Conferences on Internet of Things (iThings), Rhodes, Island, 2–6 November 2020; pp. 837–842.
15. Park, S.; Park, H. Combined oversampling and undersampling method based on slow-start algorithm for imbalanced network traffic. *Computing* **2021**, *103*, 401–424. [CrossRef]
16. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the 2018 European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
17. Wu, S. Research on Smoke Detection Algorithm Based on Convolutional Neural Network. Ph.D. Thesis, Southwest Jiaotong University, Chengdu, China, 2020.
18. Peng, X.T. Method Study on Classification of Unbalanced Data Sets Based on Deep Learning. Master's Thesis, Beijing University of Chemical Technology, Beijing, China, 2021.
19. Huang, G.; Liu, Z.; Maaten, L.M.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269.
20. Mahdy, A.M.S. A numerical method for solving the nonlinear equations of Emden-Fowler models. *J. Ocean. Eng. Sci.* **2022**. [CrossRef]
21. Choorod, P.; Weir, G. Tor Traffic Classification Based on Encrypted Payload Characteristics. In Proceedings of the 2021 National Computing Colleges Conference (NCCC), Taif, Saudi Arabia, 27–28 March 2021; pp. 1–6.
22. He, K.M.; Zhang, X.Y.; Ren, S.Q. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.