A Query-Specific Representation-Based Sampling Strategy for Imbalanced Classification

Yi Tan¹, Yanpeng Qu¹, Longzhi Yang¹, Hongbo Liu¹, Changjing Shang¹, and Qiang Shen¹

¹Affiliation not available

March 18, 2024
A Query-Specific Representation-Based Sampling Strategy for Imbalanced Classification

Yi Tan, Yanpeng Qu, Longzhi Yang, Hongbo Liu, Changjing Shang and Qiang Shen

Abstract—Data imbalance is a common challenge for many classification tasks where the dataset suffers from disproportionate partitions of samples between classes. Typical techniques for handling class imbalance are either based on algorithmic or data-driven strategies. The algorithmic strategies reduce model bias towards the majority classes by modifying the classification algorithms, and the data-driven strategies revise the datasets in terms of resampling. The current implementations of the data-driven strategy often ignore the correlations between instances but assign a uniform sampled dataset to all queries despite their distinctions. This may eliminate valid instances and retain irrelevant instances during sampling, thereby negatively affecting classification performance. To address this limitation, this paper presents a representation-based sampling (RBS) approach, implemented in a bi-stage hybrid framework, including a training stage to compile the sampling dictionary and a retrieving stage to sample the data with respect to the query. In the first stage, RBS learns the correlations between each object per class via a reconstruction model, and produces a sampling dictionary for each class. In the second stage, for a given query, the sampled data specific to this query in each class are retrieved via the offline sampling dictionary by locating its most related object in each class. Because the number of the sampled data in each class are unified, the classification of the query is guaranteed to be conducted on a class-balance dataset. The systematic experiments using image datasets demonstrate that RBS can effectively solve the data imbalance issue in classification and improve the representation of images with correlated features, leading to a better recognition performance.

Index Terms—Representation-based classification, Imbalanced classification, Data sampling, Image classification.

I. INTRODUCTION

Imbalanced classification refers to a classification task where the quantity of observations per class is not equally distributed. As a result, imbalanced classification poses a challenge for predictive modeling when identifying an object in a minority class. There are many imbalanced scenarios, where categories with fewer objects may play a greater role, such as tumor [1], cancer, and pneumonia [2] detection in the medical field [3], fault detection in mechanical maintenance [4], etc. In general, the negative impacts of imbalanced classification can be comforted via an algorithmic strategy or/and a data-driven strategy.

The algorithmic-based strategy to remedy data imbalance concentrates on the modification of existing models to alleviate their bias towards the majority groups. One solution of this strategy is using ensemble learning algorithms which construct several independent balanced training subsets from the original training dataset to train different models, and aggregate their predictions. This technique keeps the training dataset intact while improving model performance. In practice, many ensemble learning approaches [5], such as RUSBoost [6], AdaBoost [7], logitBoost [8] have been used to defeat the defect of a single classifier on data imbalance. Overall, these decision tree-based ensemble learning approaches perform classification with the use of the optimal cost values for the minority classes in line with the misclassification cost. Then, the decision trees which suffer from inferior accuracies will be reduced from the ensemble learning model. Moreover, the impact of data imbalance on deep learning [9] can be addressed via the loss re-weighting strategy and transfer learning. Specifically, the loss re-weighting strategy aims to assign different losses to different training objects in a class or instance aspect. The class balance loss can vary at a class level in order to match a given data distribution and improve the classification performance of minority classes [10]. A more fine-grained control of the loss, such as focal loss [11], Meta-Weight-Net [12], can also be achieved to comfort data imbalance by increasing the weights of minority classes in the loss function. Also, image classification in a class imbalance situation is performed with category centres of the last convolutional layer or those of the consequent full connection layer [13]. In addition, a hybrid loss function that jointly performs classification and clustering in a single formulation was introduced based on an affinity measure in the Euclidean space [14]. By employing transfer learning, the features are extracted from the majority class and added to the minority class. Recent work includes transfer variance within each categories [15] or deep semantic features within each categories [16].

The data-driven strategy for dealing with imbalanced classes is a process of resampling. The most popular techniques of this strategy are oversampling and undersampling. Specifically, oversampling involves adding copies of samples in the minority class to the training dataset. The extra data instances are usually randomly selected from the minority classes, synthesised through linear interpolation, or synthesised with the support of certain distribution manipulation strategy. Undersampling refers to sampling representative data from the majority class that will only be used during training (i.e. the
removal of the data will be discarded). Data undersampling is often implemented by randomly removing data instances from the minority classes, or eliminating specific data instances from the majority classes which are deemed to be noisy or pose negative impact to classification performance. A hybrid approach uses a combination of two sampling techniques to increase the number of data instances in the minority classes and reduce the number of data instances in the majority classes. Common to all these approaches is that the current sampling method assigns a uniform sampled dataset to all queries, although they are often different. This may mean the retaining of irrelevant instances (by oversampling) or elimination of valid instances (by undersampling), either of which will negatively impact the classification performance.

To solve data imbalance while providing supportive information for each query, this paper proposes a representation-based sampling (RBS) method. This method can be implemented in a hybrid two-stage framework: a training stage to compile the sampling dictionary and a retrieving stage to sample the data with respect to the query. In training stage, each instance in the training dataset is reconstructed by its own class. The resulting representation coefficients can be deemed a linear reconstruction metric (LRM) [17] and used to gauge the correlations between instances. In contrast to one-to-one mappings (e.g. Euclidean distance), LRM reconstructs the query from a one-to-many perspective, whilst aiming simultaneously to minimise the reconstruction error. If the instance counts of a class is lower than the overall average per class, this class Let $K$ denote the overall average amount of instances per class. If the instance counts of a class is larger than $K$, with regard to each object in this class, the $K$ instances associated with the $K$ largest representation coefficients will be undersampled from this class. For other under-represented classes, oversampling will be performed by copying the instances associated with the largest coefficient to the small one, until $K$ instances are presented in each class. For each class, the subscripts of the sampled instances will be recorded in a class-specific sampling dictionary. In retrieving stage, for a given query, the most correlated object in each class is located via LRM again. Then, the $K$ sampled data associated with these chosen objects can be retrieved via the class-specific sampling dictionaries, respectively. As a result, all of these retrieved data will be united into a class-balance dataset and used to conduct the classification of this query.

By applying a pretrained VGG-16 [18] on ImageNet-1K [19] to extract features, the quality of the sampled data by the proposed RBS algorithm is respectively compared against that of those sampled by the CBS [20], SRS [20], ROS [21] and RUS [22] methods, representing a cross section of the most popular sampling approaches. In particular, 12 imbalanced image classification datasets and 5 medical image datasets are employed in the experiment, which are examples of real-world scenarios that suffer from class imbalance. Each of these sampling methods are testified three times, with LRNN [23], CRC [24] and CNRC [25] being used as the classifiers. This experimental evaluation is followed by a comparative study of RBS in reference to other imbalance classification approaches, including focal loss [11], seesaw loss [9], category centres

of the consequent full connection layer (F) [13] and using oversampling for F (OF) [13]. The results of the original VGG-16 are also provided as the baseline. The experimental results demonstrate that, compared to either the alternative sampling methods or the algorithmic solutions for deep learning models, RBS can effectively relieve the challenges of data imbalance by generating the query specific class-balanced information, leading to a better performance in terms of both accuracy and the area under the ROC (AUC).

In general, the novelty and contributions of the proposed RBS algorithm are summarised as follows:

1) The proposed RBS method applies LRM to reveal the internal correlations and generate a sampling dictionary for each class. It is noteworthy that the process of compiling sampling dictionaries is offline. Thus, they can be directly used for a given query without any further training.

2) By consulting the offline sampling dictionaries, each query can retrieve a specific class-balance dataset for its own classification.

3) Existing sampling methods assign a uniform sampled dataset to all queries, but the proposed RBS method produces a query-specific dataset for each query, which better supports the query and conduces to addressing this query.

The remainder of this paper is structured as follows: Section II reviews the typical data driven class imbalance strategies, the LRM and the concept of locality preserving projection. Section III details the proposed RBS strategy. Section IV discusses the experimentation using imbalanced datasets and analyses the results. The paper is concluded in Section V with a discussion of further work.

II. BACKGROUND

Typical data-driven class imbalance strategies, the classical representation-based classification methods, and a linear reconstruction method are reviewed in this section.

A. Data-driven Class Imbalance Strategy

There are mainly three groups of data-driven strategies, including oversampling, undersampling and hybrid methods.

1) Oversampling:

- Random Oversampling (ROS) [21] solves data imbalance by replicating the instances from the minority class at random. The random oversampling may copy the objects that are not relevant to the query objects and thus degrade the classification performance. To address this, some statistical techniques have been developed to better represent the underlying classes distributions when selecting samples.

- Synthetic Minority Over-Sampling Technique (SMOTE) [26] is an oversampling technique that generates synthetic samples by using linear interpolation between minority class data and its $k$-nearest neighbours.

- Borderline SMOTE [27] is an improved algorithm based on SMOTE, which only uses minority classes samples on
the boundary to synthesise new samples, and improve the class distribution of dataset.

- SMOTE-Edited Nearest Neighbour (ENN) [21] is a two stage sampling method, which employs SMOTE to synthesise new minority instances and then uses ENN to eliminate the noise instances and boundary instances in the majority caused by SMOTE.

- SMOTE-Tomek [28]: The process of SMOTE-Tomek links method consists of two steps. First, new minority class samples are generated via the SMOTE method to obtain an expanded dataset. Secondly, the Tomek links, which are pairs, are removed from the expanded dataset.

- Adaptive Synthetic Sampling [29] mainly consists of 4 steps as follows:
  1. Calculate the number of samples needed to be synthesised;
  2. Calculate the majority class proportion in the k-nearest neighbours of each minority instance;
  3. Calculate the number of instances to be generated for each minority class based on the proportion produced in the second step;
  4. Use SMOTE to oversample each minority class.

2) Undersampling:

- Random Undersampling (RUS) [22] balances data distribution and improves classification performance by eliminating data instances from the majority classes. This approach may suffer from the elimination of samples that improve classification performance, resulting in classification performance degradation.

- One-Sided Selection (OSS) [22] is an under-sampling method resulting from the application of the Condensed Nearest Neighbour (CNN) algorithm to remove noisy and borderline majority class objects. In particular, since borderline objects can be easily misclassified, CNN effectively remove borderline objects from the majority classes.

3) Hybrid Methods:

- Instance-balanced sampling (IBS) [20] samples each instance only once in an epoch with the same probability. Let the probability of an object being sampled from the j-th class be expressed as

\[ p_j = \frac{n_j^2}{\sum_{i=1}^{M} n_i^2} \tag{1} \]

where M is the number of the classes; \( n_i \) is the quantity of the objects in the i-th class. For the IBS sampling algorithm, the value of q in Eq. (1) is 1.

- Class-balanced sampling (CBS) [20] selects each class with the equal probability. That is, the value of q in Eq. (1) is 0. Then, an instance from the selected class will be sampled randomly.

- Square-root sampling (SRS) [20] is a compromise of IBS and CBS strategies. In SRS, the value of q in Eq. (1) is 1/2.

- Progressively-balanced sampling (PBS) [30] is a hybridisation of IBS and CBS algorithms, which performs IBS first, and then CB, according to a preset number of epochs. The sampling probability of a soft version of PBS for class j in the t-th epoch is

\[ p_j^{PBS}(t) = (1 - \frac{t}{T}) p_j^{IBS} + \frac{t}{T} p_j^{CBS}, \tag{2} \]

where \( T \) is the total number of epochs.

B. Linear Reconstruction Measure

Given a data set \( X = [x_1, \ldots, x_N] \in \mathbb{R}^{P \times N} \) consisting of \( N \) objects in a \( P \)-dimensional feature space, for a query \( y \in \mathbb{R}^P \), the result of the following linear reconstruction process

\[ W^* = \arg \min_W \| y - XW \|^2_2 \tag{3} \]

can be employed as the linear reconstruction metric (LRM) [17] to evaluate the similarity between \( X \) and \( y \). Here, \( W \) is the matrix of the reconstruction coefficients. In particular, each element \( w^*_i \) of \( W^* = [w^*_1, \ldots, w^*_N]^T \) indicates the LRM similarity between \( y \) and \( x_i \), \( i = 1, \ldots, N \). In contrast to one-to-one mappings (e.g. Euclidean distance), LRM reconstructs the query from a one-to-many perspective, whilst aiming simultaneously to minimise the reconstruction error. Thus, for the query \( y \), LRM can discover its true nearest-neighbours, which are close to \( y \) in line with their inherent geometric similarities to \( y \) and play a significant role in the neighbourhood estimation due to the full consideration of the neighbourhood relationship of instances. As a result of this positive side-effect, the physical meaning of LRM can be interpreted as an indicator of the probability of the data item being a close neighbour of the query \( y \). In other words, the larger the value of LRM between a data sample and a query \( y \) is, the greater the probability of the data sample being a neighbour of \( y \) is.

In practice, LRM is always implemented with the \( \ell_p \) regularisation item:

\[ W^* = \arg \min_W \| y - XW \|^2_2 + \lambda \| W \|_p, \tag{4} \]

where \( \lambda \) is the regularisation parameter. Specifically, \( p = 1 \) indicates the sparse representation [31] which can prevent overfitting; and \( p = 2 \) indicates the collaborative representation [32], [33] which can overcome singularity.

As an effective correlation measure for locating the neighbourhood of the query object, LRM has been used in representation-based classification, which attempts to classify the query object into a class whose training objects have the maximum representation ability. For instance, by taking account of the nearest neighbour of each sample, the collaborative neighbour representation classification (CNRC) method was proposed in [25], by solving the following optimization problem:

\[ W^* = \arg \min_W \left\{ \frac{1}{2} \| y - XW \|^2_2 + \sigma \sum_{i=1}^{N} w^2_i \| y - x_i \|^2_2 + \lambda \| w \|^2_2 \right\}. \tag{5} \]

From the sense of the least reconstruction error, the optimal solution of Eq. (5) is achieved by:

\[ W^* = \left( X^TX + \lambda I + \left[ \begin{array}{cccc} \| y - x_1 \|^2_2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \| y - x_N \|^2_2 \end{array} \right] \right)^{-1} D^T y, \tag{6} \]
Furthermore, a local within-class collaborative representation based nearest neighbour (LRNN) algorithm was presented in [23]. Assume that the training dataset \( X \) consists of \( M \) classes, and each class \( X_j \) contains \( N_j \) instances, \( j = 1, \ldots, M \). Let \( X_j^{(K)} = [x_j^{(1)}, x_j^{(2)}, \ldots, x_j^{(K)}] \in \mathbb{R}^{P \times K} \) denote the \( K \)-nearest neighbours of the query \( y \) within \( X_j \). The representation of the query \( y \) over \( X_j^{(K)} \) is \( \hat{y}_j = X_j^{(K)} W_j^* \), where

\[
W_j^* = (X_j^{(K)} X_j^{(K)\top} + \lambda^2 \Gamma_{j,y} \Gamma_{j,y})^{-1} X_j^{(K)} y
\]

is the closed-form solution of following optimisation problem:

\[
\min_{W_j} \|y - X_j^{(K)} W_j\|_2^2 + \lambda \|\Gamma_{j,y} W_j\|_2.
\]

Here, \( \Gamma_{j,y} = \begin{bmatrix} ||y - x_j^{(1)}||_2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & ||y - x_j^{(K)}||_2 \end{bmatrix} \) is a biasing Tikhonov matrix specific to \( X_j \) and \( y \). Then, for CNRC and LRNN, the class label of \( y \) can be determined by the minimum residual error between \( \hat{y}_j \) and \( y \), i.e.,

\[
l(y) = \arg \min_j \|\hat{y}_j - y\|_2^2.
\]

### C. Locality Preserving Projection

Locality preserving projection (LPP) is a linear version of the Laplacian eigenmap algorithm [34], which can preserve the local neighbour structure of high dimension data in its low-dimensional embedding. Given a training dataset \( X = [x_1, \ldots, x_N] \in \mathbb{R}^{P \times N} \), LPP aims to seek for a transformation matrix \( A \) such that the projection \( Y = AX \) fulfils the following optimisation problem:

\[
\min_A \sum_{i,j} ||y_i - y_j||_2^2 w_{ij}, i, j = 1, \ldots, N.
\]

The weight \( w_{ij} \) can be a Gaussian kernel which is defined as follows:

\[
w_{ij} = \begin{cases} \exp\left(-||x_i - x_j||^2\right) & x_i \in N_k(x_j) \text{ or } x_i \in N_k(x_j) \\ 0 & \text{otherwise} \end{cases},
\]

where \( N_k(x_i) \) denotes the set of the \( k \) nearest neighbours of \( x_i \). Given weight matrix \( W \), consisting of \( w_{ij} \), \( i, j = 1, \ldots, N \), the minimisation of the objective function Eq. (10) of LPP will ensure \( x_i \) and \( x_j \) are similar, and the projected values \( y_i \) and \( y_j \) are close to each other.

Because the optimisation problem as expressed in Eq. (10) has multiple solutions, a constraint \( YDY^\top = I \) is imposed to this optimisation problem to generate a unique solution. Here, \( \text{tr} (\cdot) \) represents the matrix trace. In so doing, the optimisation problem of Eq. (10) is formulated as follows:

\[
\min_{Y} \frac{1}{2} \sum_{i,j} ||y_i - y_j||_2^2 w_{ij}
\]

\[
= \min_{YDY^\top = I} \text{tr} (Y (D - W) Y^\top)
\]

\[
= \min_{YDY^\top = I} \text{tr} (YL^\top),
\]

where \( D \) denotes the diagonal matrix whose elements are column sum of matrix \( W \), and \( L = D - W \) is a Laplacian matrix.

### III. REPRESENTATION-BASED SAMPLING ALGORITHM FOR IMBALANCE CLASSIFICATION

Current data-driven strategies towards imbalance classification, or the sampling methods, sample a uniform dataset for all queries, regardless of their requirement of distinct information for each queries. This may unnecessarily keep irrelevant instances (by oversampling) or incorrectly remove valid instances (by undersampling) for queries, either of which will negatively impact the classification performance.

To sample a query-specific dataset in an effort to solve data imbalance, a representation-based sampling (RBS) method is proposed in this work. This method is implemented in a hybrid bi-stage framework. The first stage is a training stage, which compiles sampling dictionaries; and the second stage is a retrieving stage, which locates the data with respect to the query. The flowchart of the proposed model is illustrated in Fig. 1.

#### A. Training Stage: Sampling Dictionary Compilation

Assume that the training dataset \( X \) consists of \( N \) objects belonging to \( M \) classes, and each class \( X_j \) contains \( N_j \) instances, \( j = 1, \ldots, M \). Different from the sparse representation used in [31] and the collaborative representations used in [32], [33], LRM in this paper is implemented via LPP and \( \ell_2 \) regularisation terms to reveal the local representation correlations between instances in each class. For \( X_j, j = 1, \ldots, M \), the optimisation problem is:

\[
\min_{W_j} \{||X_j - X_j W_j||_2 + \rho_1 ||W_j||_2 + \rho_2 \text{tr} (W_j^T X_j^T L_j X_j W_j)\},
\]

where \( L_j \) denotes the Laplacian matrix representing the relation information between pairs of training samples in \( X_j \). In Eq. (13), the regularisation term associated with \( \rho_1 \) is responsible for the collaborative relation of data in each class; and the regularisation term associated with \( \rho_2 \) facilitates LPP by optimally preserving the local neighbourhood information in a certain sense [35].

The closed-form solution of Eq. (13) is:

\[
W_j^* = (X_j^T X_j + \rho_1 I + \rho_2 X_j^T L_j X_j)^{-1} X_j^T X_j.
\]

Assuming that \( X_j \) consists of \( N_j \) objects, each element \( w_{jj}^* (s, t) \) of \( W_j^* \), \( s, t = 1, \ldots, N_j \) is an LRM that can gauge the correlation between the \( s \)-th and the \( t \)-th objects in \( X_j \). In this work, to overcome data imbalance, each class are sampled into \( K \) instances uniformly. The value of \( K \) is defined as the average number of the instances per class:

\[
K = \left\lceil \frac{N}{M} \right\rceil.
\]

Specifically, if \( N_j < K \), \( X_j \) is oversampled by copying the instances associated with the \( (K - N_j) \) largest components in \( w_{jj}^* (\cdot, t) \) with respect to the \( t \)-th object in \( X_j, t = 1, \ldots, N_j \). Otherwise, if \( N_j \geq K \), the \( K \) sampled instances with regard
to the $t$-th object in $X_j$ are those which enjoy the $K$ largest components in $w_j^*(\cdot, t)$. Here,

$$w_j^*(\cdot, t) = [w_j^*(1, t), w_j^*(2, t), \ldots, w_j^*(N_j, t)]^T$$

denotes the representation coefficients of the $t$-th object in $X_j$, i.e., the $t$-th column in $W_j^*$, $t = 1, \ldots, N_j$.

Let $\mathcal{N}_j \in \mathbb{R}^{K \times N_j}$ summarise the sampled instances with regard to each object in $X_j$. Induced by $\mathcal{N}_j$, the sampling dictionary $D_j \in \mathbb{R}^{K \times N_j}$ is used to record the subscript of each instance in $\mathcal{N}_j$. The process of building $D_j$ for $X_j$ is outlined in Alg. 1.

**Algorithm 1: Dictionary of Sampled Instances of $X_j$**

**Input:**
- $\rho_1$, $\rho_2$: regularisation coefficients in (13);
- $N$, the number of the objects in $X$;
- $M$, the number of the classes in $X$;
- $X_j$, the objects of the $j$-th class;
- $N_j$, the number of the objects in $X_j$.

**Output:** $d_j$.

1. $W_j^* \leftarrow (14)$ over $X_j$;
2. $K = \frac{N_j}{M}$;
3. **if** $N_j < K$ **then**
   4. $\mathcal{N}_j \leftarrow$ Oversampling $X_j$ by copying the instances associated with the $(K - N_j)$ largest components in $w_j^*(\cdot, t)$, $t = 1, \ldots, N_j$;
5. **else**
6. $\mathcal{N}_j \leftarrow$ Selecting the instances associated with the $K$ largest components in $w_j^*(\cdot, t)$, $t = 1, \ldots, N_j$;
7. **end**
8. $D_j \leftarrow$ The subscript of each instance in $\mathcal{N}_j$;
9. **return** $D_j$.

To illustrate the working process of Alg. 1, an exemplar dataset, used as a running example throughout the paper, is given in Table I, consisting of 8 objects belonging to 2 categories: $X_1 = [x_1, x_2, x_3]$ and $X_2 = [x_4, x_5, x_6, x_7, x_8]$. For this toy example, the $\ell_2$ normalisation approach [24] is utilised leading to a normalised data set as shown in Table II.

**TABLE I: Exemplar dataset**

<table>
<thead>
<tr>
<th>Object</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>$c_1$</td>
</tr>
<tr>
<td>$x_2$</td>
<td>$c_1$</td>
</tr>
<tr>
<td>$x_3$</td>
<td>$c_1$</td>
</tr>
<tr>
<td>$x_4$</td>
<td>$c_1$</td>
</tr>
<tr>
<td>$x_5$</td>
<td>$c_2$</td>
</tr>
<tr>
<td>$x_6$</td>
<td>$c_2$</td>
</tr>
<tr>
<td>$x_7$</td>
<td>$c_2$</td>
</tr>
<tr>
<td>$x_8$</td>
<td>$c_2$</td>
</tr>
</tbody>
</table>

**TABLE II: Normalised dataset**

<table>
<thead>
<tr>
<th>Object</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>$c_1$</td>
</tr>
<tr>
<td>$x_2$</td>
<td>$c_1$</td>
</tr>
<tr>
<td>$x_3$</td>
<td>$c_1$</td>
</tr>
<tr>
<td>$x_4$</td>
<td>$c_1$</td>
</tr>
<tr>
<td>$x_5$</td>
<td>$c_2$</td>
</tr>
<tr>
<td>$x_6$</td>
<td>$c_2$</td>
</tr>
<tr>
<td>$x_7$</td>
<td>$c_2$</td>
</tr>
<tr>
<td>$x_8$</td>
<td>$c_2$</td>
</tr>
</tbody>
</table>

By applying Eq. (14), the respective representation coeffi-
eTable 1: \( W_1^* \) and \( W_2^* \) for the reconstruction coefficients of \( X_1 \) and \( X_2 \) respectively.

\[
W_1^* = \begin{bmatrix}
0.2423 & 0.1397 & 0.1382 \\
0.1736 & 0.2905 & 0.1887 \\
0.2157 & 0.2346 & 0.3512
\end{bmatrix},
\]
\[
W_2^* = \begin{bmatrix}
0.1901 & 0.0806 & 0.1063 & 0.0954 & 0.1474 \\
0.1213 & 0.2442 & 0.1267 & 0.1290 & 0.1388 \\
0.1348 & 0.1154 & 0.2373 & 0.1169 & 0.1293 \\
0.0838 & 0.0869 & 0.0884 & 0.1907 & 0.1463 \\
0.1746 & 0.1211 & 0.1266 & 0.1747 & 0.1916
\end{bmatrix}.
\]

In this example, \( N = 8 \) and \( M = 2 \). Thus, according to Eq. (15), \( K = 4 \). Since \( N_1 = 3 \) is smaller than \( K \) and \( (K - N_1) = 1 \), according to Alg. 1, \( X_1 \) will be oversampled by the instance associated with the top 1 component in \( w_j^* \). Induced by \( \tilde{V}_1 \) in \( [0.2423, 0.1736, 0.2157]^T \), \( w_1^*(1, 1) = 0.2423 \) is the largest component. Thus, for \( x_1 \), \( X_1 \) is supposed to be oversampled with \( x_1 \). As a result, the first column of \( \mathcal{N}_1 \) is \([x_1, x_2, x_3]^T \). Similarly, for \( w_j^* \) \((j = 2, 3)\), \( w_2^*(2, 2) = 0.2905 \) and \( w_3^*(3, 3) = 0.3512 \) are the largest components, respectively. Thus, the second and third column of \( \mathcal{N}_1 \) are \([x_1, x_2, x_3]^T \) and \([x_1, x_2, x_3]^T \), respectively.

As for \( X_2 \), since \( N_2 = 5 \) > \( K = 4 \), the \( K \) nearest neighbours of each object in \( X_2 \) will be deemed the associated sampled instances. Specifically, in the light of \( W_2 \), the respective \( 4 \) nearest neighbours of \( x_4, x_5, x_6, x_7 \) and \( x_8 \) in \( X_2 \) are \([x_4, x_5, x_6, x_8]^T \), \([x_4, x_5, x_7, x_8]^T \), \([x_5, x_6, x_7, x_8]^T \) and \([x_4, x_5, x_7, x_8]^T \). In summary, \( \mathcal{N}_1 \) and \( \mathcal{N}_2 \) are derived as follows.

\[
\mathcal{N}_1 = \begin{bmatrix}
x_1 & x_1 & x_1 \\
x_1 & x_2 & x_2 \\
x_2 & x_2 & x_3 \\
x_3 & x_3 & x_3
\end{bmatrix}, \quad \mathcal{N}_2 = \begin{bmatrix}
x_4 & x_5 & x_4 & x_5 & x_4 \\
x_5 & x_6 & x_5 & x_6 & x_5 \\
x_6 & x_7 & x_6 & x_7 & x_7 \\
x_8 & x_8 & x_8 & x_8 & x_8
\end{bmatrix}.
\]

Induced by \( \mathcal{N}_1 \) and \( \mathcal{N}_2 \), the respective dictionary of sampled instances, \( D_1 \) and \( D_2 \), in \( X_1 \) and \( X_2 \) are as follows.

\[
D_1 = \begin{bmatrix}
1 & 1 & 1 \\
1 & 2 & 2 \\
2 & 2 & 3 \\
3 & 3 & 3
\end{bmatrix}, \quad D_2 = \begin{bmatrix}
4 & 5 & 4 & 5 & 4 \\
5 & 6 & 5 & 6 & 5 \\
6 & 7 & 6 & 7 & 7 \\
8 & 8 & 8 & 8 & 8
\end{bmatrix}.
\]

B. Retrieving Stage: Query-Specific Data Identification

Assume that the representation coefficients \( V_j^* = [v_{j1}, \ldots, v_{jN_j}]]^T \) to code the query \( y \) over \( X_j \), \( j = 1, \ldots, M \) has been obtained by solving following optimisation problem:

\[
V^* = \arg \min_V \| y - XV \|_2^2 + \rho_3 \| V \|_2.
\]

Amongst the reconstruction coefficients \( v_{j1}^*, \ldots, v_{jN_j}^* \), the maximum element

\[
\tilde{v}_j^* = \max_t \{ v_{jt}^* \}, \quad t = 1, \ldots, N_j
\]

is used to locate its associated instance \( \tilde{x}_j^* \in X_j \), which can be regarded as the nearest neighbour of \( y \) in \( X_j \).

By consulting \( D_j \), the \( K \) sampled instances with regard to \( \tilde{x}_j^* \) in \( X_j \) are registered in \( \hat{X}_j \), \( j = 1, \ldots, M \). Apparently, the objects in \( \hat{X}_j \) are highly related to the query \( y \). As a result, the union of \( \hat{X}_j \), i.e.,

\[
\hat{X} = \bigcup_{j=1}^M \hat{X}_j
\]

consists of \( M \) classes, and each class has \( K \) instances. This class-balance subset of \( X \) will be used to further identify \( y \).

Following the running example in section III-A, with its normalised dataset listed in Table II, given a query:

\[
y = [0.365, 0.4756, 0.4009, 0.3624, 0.2817, 0.4295, 0.291]^T,
\]

the respective representation coefficients of \( X_1 \) and \( X_2 \) via Eq. (4) with \( p = 2 \) are:

\[
V_1^* = [0.0859, 0.0845, 0.0707]^T,
\]
\[
V_2^* = [0.066, 0.0729, 0.0704, 0.0813, 0.0809]^T.
\]

According to Eq. (17), for \( X_1 \),

\[
\tilde{v}_1^* = \max \{ 0.0859, 0.0845, 0.0707 \} = 0.0859.
\]

Thus, in \( X_1 \), the nearest neighbour of \( y \) is the first object, i.e.,

\[
\tilde{x}_1^* = x_1.
\]

In line with the first column of \( D_1 \), the subscripts of the instances sampled from \( X_1 \) are 1, 1, 2 and 3. Thus, \( \hat{X}_1 \) is \([x_1, x_1, x_2, x_3]^T \).

In the same manner, in \( X_2 \), since

\[
\tilde{v}_2^* = \max \{ 0.066, 0.0729, 0.0704, 0.0813, 0.0809 \} = 0.0813,
\]

the fourth object is the nearest neighbour of \( y \), i.e., \( \tilde{x}_2^* = x_7 \). By consulting the fourth column of \( D_2 \), the subscripts of the instances sampled from \( X_2 \) are 5, 6, 7 and 8. Thus, \( \hat{X}_2 \) is \([x_5, x_6, x_7, x_8]^T \).

According to Eq. (18), the sampled training dataset for \( y \) is:

\[
\hat{X} = \hat{X}_1 \cup \hat{X}_2 = \{x_1, x_1, x_2, x_3, x_5, x_6, x_7, x_8\}.
\]

Apparently, each category in \( \hat{X} \) consists of \( K \) instances. In this case, the classification of \( y \) is performed over a class-balance dataset. The overall process of the proposed RBS algorithm is summarised in Alg. 2. It is worth noting that for each class in \( X \), the procedure shown in lines 4 to 8 can be operated in parallel, so as to expedite the training stage separately by class to better obtain the local information of the data.

IV. EXPERIMENTAL EVALUATION

This section presents a systematic experimental evaluation of the proposed RBS algorithm. The results and discussion are divided into two parts, following a description of the data and experimental setting. In Section IV-B, RBS is compared against a cross-section of the most popular sampling approaches, including CBS, SRS, ROS and RUS. Section IV-C performs a comparative study of RBS in reference to 4 loss functions of deep learning for imbalanced classification, including focal loss [11], seesaw loss [9], category centres of the consequent full connection layer (F) [13] and using ROS for F (OF) [13]. All of the comparative studies are
Algorithm 2: Representation-based Sampling Algorithm

Input:
\( X \): training dataset;
\( y \): the query;
\( \rho \): regularisation coefficient in (16);

Output:
\( \hat{X} \): the sampled training dataset for \( y \).

1. For the \( j \)-th class \( X_j \), build \( D_j \) via Alg. 1, \( j = 1, \ldots, M \);
2. \( \hat{X} \leftarrow \emptyset \);
3. for \( j = 1 \) to \( M \) do
4. \( V_j^* \leftarrow \) Eq. (16);
5. \( \bar{v}_j \leftarrow \) Eq. (17);
6. Locate \( \tilde{x}_j^* \) via \( \bar{v}_j \);
7. Find \( \hat{X}_j \) of \( \tilde{x}_j^* \) by looking up \( D_j \);
8. \( \hat{X} = \hat{X} \cup \hat{X}_j \);
9. end
10. return \( \hat{X} \).

conducted on 6 imbalanced image classification datasets, 6 artificially imbalanced image datasets and 5 medical image datasets, which are examples of real-world scenarios that suffer from class imbalance. The experiments are carried out on a Dell EMC PowerEdge T640 server running an Intel Xeon Gold 5218 CPU at 2.3 GHz, 320GB of RAM and two GeForce RTX 3090 GPUs.

A. Experimental Settings

The information of the image datasets used in this experimental study is summarised in Table III, where the imbalance rate (IR) is defined as follows:

\[
IR = \frac{\max_i \{N_i\}}{\min_i \{N_i\}},
\]

where \( \max_i \{N_i\} \) and \( \min_i \{N_i\} \) denote the maximum and minimum numbers of data in a single class, respectively. Moreover, for all datasets, 80% of the data are taken as the training set and the remaining 20% as the test set, except the PneumoniaMNIST dataset, which is divided according to the references in [3].

For completeness, a brief summary of the used datasets is provided below.

- PneumoniaMNIST [2], [3] is an image classification dataset for pneumonia detection. It has two categories representing the normal chest X-ray images and the chest X-ray image collected from subjects who were infected by pneumonia.
- BreastMNIST [3], [36] is a breast tumour classification data set, which contains three categories: benign, malignant and normal. This dataset was labelled using two strategies. BreastMNIST_2class is used in the original data set [36], which distinguishes benign from malignant and leads to a three categories classification tasks. BreastMNIST_3class [3] combines benign and malignant into one label, so this problem becomes a binary classification.
- Brain tumour [1] includes a total of 4,600 images which consists of two categories: health and brain tumour.
- Tuberculosis (TB) [37] has two classes: the chest X-ray images of normal (3,500) and patients with TB (700). This dataset is publicly accessible.
- Caltech-101 [38] consists of 101 categories. There are 40 to 800 images per class. Most categories have around 50 images. The size of each image is roughly \( 300 \times 200 \) pixels.
- Caltech-256 [39] is collected similarly with several improvements, compared to Caltech101. The number of categories in Caltech256 is more than doubled, the minimum number of images in any category is increased from 40 to 80.
- ECUSTFD [40] contains 20 kinds of food photos taken under the same background, different lighting conditions, and different angles.
- Animal images [41] contains 11,959 images representing 6 categories of animals.
- Office-31 [42] includes 31 types of office items in three domains. Each class involves 164 images at most and 68 images at least. In this work, the impact of the domain on image classification is ignored.
- Modern office-31 [43] is a revamped version of the commonly used Office-31 dataset. Like office-31, the impact of the domain is ignored in this work.
- Cifar-10 [44] and Cifar-100 [44] are labelled subsets of the 80 million tiny images datasets [45]. Cifar-10 has 10 classes and Cifar-100 has 100 classes. In order to better support the evaluation of the performance of the RBS method in a class imbalance situation, 3 imbalanced datasets with different imbalance rates, i.e. IR=m, 10 and 20, are intercepted from each of Cifar-10 and Cifar-100. In particular, IR=m denotes a linear changing of the ratio. Each imbalance type is produced through RUS in the original balance training data and validation data. The detailed interception method [13] of the resulting 6 datasets is shown in Table V and VI. The labels of Cifar-10 and Cifar-100 are translated into numbers, ranging

| Table III: Imbalanced datasets used in the experiments |
|-----------------------------------------------|--|--|--|--|
| **Datasets** | **Objects** | **Training** | **Test** | **Classes** | **IR** |
| PneumoniaMNIST | 5836 | 5232 | 624 | 2 | 2.88 |
| BreastMNIST_2class | 780 | 624 | 153 | 2 | 2.71 |
| BreastMNIST_3class | 780 | 624 | 153 | 3 | 3.23 |
| BrainTumor | 4600 | 3680 | 1747 | 2 | 1.20 |
| Tuberculosis | 4200 | 3360 | 840 | 2 | 2 |
| Caltech101 | 8734 | 6987 | 1747 | 1 | 27.8 |
| Caltech256 | 30185 | 24148 | 6037 | 256 | 13.51 |
| ECUSTFD | 8734 | 6987 | 1747 | 1 | 27.8 |
| Animal images | 11959 | 9567 | 2392 | 6 | 19.9 |
| Office-31 | 4110 | 3288 | 822 | 31 | 2.41 |
| Modern office-31 | 4626 | 3700 | 926 | 31 | 2.41 |
| Cifar-10 (IR=m) | 30300 | 20300 | 10000 | 10 | m |
| Cifar-10 (IR=10) | 32000 | 22000 | 10000 | 10 | 10 |
| Cifar-10 (IR=20) | 21800 | 11800 | 10000 | 10 | 20 |
| Cifar-100 (IR=m) | 30300 | 20300 | 10000 | 100 | m |
| Cifar-100 (IR=10) | 32000 | 22000 | 10000 | 100 | 10 |
| Cifar-100 (IR=20) | 21800 | 11800 | 10000 | 100 | 20 |
from 0 to 9 and from 0 to 99, respectively. For instance, the first column of Table VI indicates that each of the first 50 classes in Cifar-100 has 40 objects and each of the last 50 classes in Cifar-100 has 400 objects. Each class in the test set of Cifar-10 and Cifar-100 contains 1000 and 100 instances, respectively.

**TABLE V: Imbalance subsets sampled from Cifar-10**

<table>
<thead>
<tr>
<th>Label</th>
<th>IR=10</th>
<th>IR=20</th>
<th>IR=m</th>
<th>Test_set</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>400</td>
<td>200</td>
<td>1000</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>400</td>
<td>200</td>
<td>600</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>400</td>
<td>200</td>
<td>1000</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>400</td>
<td>400</td>
<td>1400</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>400</td>
<td>400</td>
<td>1800</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>400</td>
<td>400</td>
<td>2200</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>400</td>
<td>2000</td>
<td>2600</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>400</td>
<td>2000</td>
<td>3000</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>400</td>
<td>2000</td>
<td>3500</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>400</td>
<td>400</td>
<td>4000</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE VI: Imbalance subsets sampled from Cifar-100**

<table>
<thead>
<tr>
<th>Label</th>
<th>IR=10</th>
<th>IR=20</th>
<th>IR=m</th>
<th>Test_set</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0-9</td>
<td>40</td>
<td>20</td>
<td>20</td>
<td>100</td>
</tr>
<tr>
<td>10-19</td>
<td>40</td>
<td>20</td>
<td>60</td>
<td>100</td>
</tr>
<tr>
<td>20-29</td>
<td>40</td>
<td>20</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>30-39</td>
<td>40</td>
<td>40</td>
<td>140</td>
<td>100</td>
</tr>
<tr>
<td>40-49</td>
<td>40</td>
<td>40</td>
<td>180</td>
<td>100</td>
</tr>
<tr>
<td>50-59</td>
<td>40</td>
<td>40</td>
<td>220</td>
<td>100</td>
</tr>
<tr>
<td>60-69</td>
<td>40</td>
<td>200</td>
<td>260</td>
<td>100</td>
</tr>
<tr>
<td>70-79</td>
<td>40</td>
<td>200</td>
<td>300</td>
<td>100</td>
</tr>
<tr>
<td>80-89</td>
<td>40</td>
<td>200</td>
<td>350</td>
<td>100</td>
</tr>
<tr>
<td>90-99</td>
<td>40</td>
<td>400</td>
<td>400</td>
<td>100</td>
</tr>
</tbody>
</table>

The VGG-16 model pre-trained on ImageNet-1K is used in this experimentation to extract features from the above listed datasets. The settings of the parameters of the tested algorithms used in the experimental studies are summarised in Table VII.

**TABLE VII: Parameters settings**

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Parameters</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>RBS</td>
<td>$\rho_1=1$, $\rho_2=1$, $\rho_3=1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRC [24]</td>
<td>$\lambda=0.1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CNRC [25]</td>
<td>$\lambda=0.1$, $\sigma=5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LRNN [23]</td>
<td>$K$ is set as Eq. (15), $\lambda=4$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VGG16</td>
<td>Learning_rate=0.01, num_epochs=20, batch_size=12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Focal loss [11]</td>
<td>Learning_rate=0.01, num_epochs=20, batch_size=32, $\gamma=2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seesaw loss [9]</td>
<td>Learning_rate=0.01, num_epochs=20, batch_size=32, $\rho=0.8$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**B. Comparison against Sampling Strategies**

In this experiment, the quality of the sampled instances resulted from the proposed RBS is compared against that from CBS, SRS, ROS and RUS, which jointly represent a cross-section of the most popular sampling approaches. Each of these sampling methods are testified with three classifiers, including LRNN, CRC and CNRC. The popular metrics, including the classification accuracy (ACC) and the area under the ROC curve (AUC), are also employed in this experiment, with the results shown in Tables IV and VIII, respectively.

It can be observed that the results in terms of both ACC and AUC achieved by RBS are superior to the others in general. Occasionally, for the Office-31 dataset, the highest values of ACC and AUC are gained by the SRS sampling approach working with the CNRC classifier. And the combination of CBS sampling approach and the CRC classifier performs better than RBS and CRC on the Caltech-101 dataset in terms of ACC, and the Animal images dataset in terms of AUC, respectively. But for the rest cases, RBS consistently returns the best ACC and AUC scores, as well as the average performance. Moreover, as the IR increases, the advantage of RBS against the other sampling strategies is becoming increasingly significant. This demonstrates that RBS is not only valid to address class imbalance, but also contributing positively for the sampling of relevant objects for image recognition.

The ACC results of RBS for different subsets of Cifar-10 and Cifar-100 are illustrated in Fig. 2. Specifically, these values are collected for each class on Cifar-10, and the average
### TABLE VIII: Comparison against alternative sampling strategies on AUC

<table>
<thead>
<tr>
<th>Datasets</th>
<th>LRNN CNRC</th>
<th>ROS CNRC</th>
<th>CBS CNRC</th>
<th>SRS CNRC</th>
<th>RBS CNRC</th>
</tr>
</thead>
<tbody>
<tr>
<td>PneumoniaMNIST</td>
<td>0.849</td>
<td>0.838</td>
<td>0.835</td>
<td>0.841</td>
<td>0.833</td>
</tr>
<tr>
<td>BreastMNIST</td>
<td>0.788</td>
<td>0.793</td>
<td>0.742</td>
<td>0.748</td>
<td>0.739</td>
</tr>
<tr>
<td>BreastMNIST2class</td>
<td>0.862</td>
<td>0.859</td>
<td>0.873</td>
<td>0.826</td>
<td>0.850</td>
</tr>
<tr>
<td>Brain Tumor</td>
<td>0.982</td>
<td>0.977</td>
<td>0.979</td>
<td>0.980</td>
<td>0.975</td>
</tr>
<tr>
<td>Tuberculosis</td>
<td>0.979</td>
<td>0.982</td>
<td>0.942</td>
<td>0.985</td>
<td>0.963</td>
</tr>
<tr>
<td>Caltech-101</td>
<td>0.955</td>
<td>0.952</td>
<td>0.932</td>
<td>0.954</td>
<td>0.936</td>
</tr>
<tr>
<td>Caltech-256</td>
<td>0.883</td>
<td>0.872</td>
<td>0.858</td>
<td>0.875</td>
<td>0.885</td>
</tr>
<tr>
<td>ECUSTFID</td>
<td>0.989</td>
<td>0.996</td>
<td>0.996</td>
<td>0.973</td>
<td>0.966</td>
</tr>
<tr>
<td>Animal images</td>
<td>0.918</td>
<td>0.937</td>
<td>0.939</td>
<td>0.929</td>
<td>0.920</td>
</tr>
<tr>
<td>Office31</td>
<td>0.944</td>
<td>0.946</td>
<td>0.956</td>
<td>0.943</td>
<td>0.953</td>
</tr>
<tr>
<td>Modern-office31</td>
<td>0.973</td>
<td>0.969</td>
<td>0.958</td>
<td>0.965</td>
<td>0.960</td>
</tr>
<tr>
<td>Cifar-10 (IR=m)</td>
<td>0.901</td>
<td>0.878</td>
<td>0.877</td>
<td>0.873</td>
<td>0.881</td>
</tr>
<tr>
<td>Cifar-10 (IR=10)</td>
<td>0.891</td>
<td>0.825</td>
<td>0.803</td>
<td>0.886</td>
<td>0.889</td>
</tr>
<tr>
<td>Cifar-10 (IR=20)</td>
<td>0.881</td>
<td>0.791</td>
<td>0.687</td>
<td>0.871</td>
<td>0.876</td>
</tr>
<tr>
<td>Cifar-100 (IR=m)</td>
<td>0.775</td>
<td>0.767</td>
<td>0.759</td>
<td>0.714</td>
<td>0.718</td>
</tr>
<tr>
<td>Cifar-100 (IR=10)</td>
<td>0.765</td>
<td>0.747</td>
<td>0.735</td>
<td>0.712</td>
<td>0.733</td>
</tr>
<tr>
<td>Cifar-100 (IR=20)</td>
<td>0.741</td>
<td>0.704</td>
<td>0.686</td>
<td>0.705</td>
<td>0.714</td>
</tr>
<tr>
<td>Average</td>
<td>0.887</td>
<td>0.873</td>
<td>0.856</td>
<td>0.869</td>
<td>0.870</td>
</tr>
</tbody>
</table>

![Fig. 2: Classification accuracies for the different categories of Cifar-10 and Cifar-100](image)

C. Comparison against Loss Functions

To comprehensively evaluate the performance of the proposed RBS algorithm, this section presents a comparative study of RBS in reference to typical loss function-based solutions for imbalanced classification, including focal loss, seesaw loss, category centres of the consequent full connection layer (F) and using oversampling for F (OF). The results of the original VGG-16 are also provided as the baseline.

The ACC and AUC results are collected in Tables IX and X, respectively. It can be seen that RBS, especially working in partnership with CNRC, outperforms the referenced loss function-based strategies for imbalance classification on most datasets in terms of the highest ACC and AUC scores. In addition, the combination of RBS and CNRC also show the best average outcomes. These observations reconfirm that RBS...
TABLE IX: Comparison against loss functions on ACC

<table>
<thead>
<tr>
<th>Datasets</th>
<th>VGG16</th>
<th>Focal_loss</th>
<th>Seesaw_loss</th>
<th>F</th>
<th>OF</th>
<th>RBS+LRNN</th>
<th>RBS+CNRC</th>
<th>RBS+CRC</th>
</tr>
</thead>
<tbody>
<tr>
<td>PneumoniaMNIST</td>
<td>0.762</td>
<td>0.873</td>
<td>0.865</td>
<td>0.833</td>
<td>0.835</td>
<td>0.886</td>
<td>0.879</td>
<td>0.865</td>
</tr>
<tr>
<td>BreastMNIST_2class</td>
<td>0.825</td>
<td>0.833</td>
<td>0.827</td>
<td>0.762</td>
<td>0.756</td>
<td>0.868</td>
<td>0.874</td>
<td>0.851</td>
</tr>
<tr>
<td>BreastMNIST_3class</td>
<td>0.775</td>
<td>0.785</td>
<td>0.787</td>
<td>0.641</td>
<td>0.647</td>
<td>0.846</td>
<td>0.856</td>
<td>0.807</td>
</tr>
<tr>
<td>Brain Tumor</td>
<td>0.880</td>
<td>0.893</td>
<td>0.870</td>
<td>0.874</td>
<td>0.810</td>
<td>0.987</td>
<td>0.995</td>
<td>0.979</td>
</tr>
<tr>
<td>Tuberculosis</td>
<td>0.988</td>
<td>0.987</td>
<td>0.991</td>
<td>0.960</td>
<td>0.958</td>
<td>0.989</td>
<td>1.000</td>
<td>0.988</td>
</tr>
<tr>
<td>Caltech-101</td>
<td>0.874</td>
<td>0.894</td>
<td>0.891</td>
<td>0.872</td>
<td>0.871</td>
<td>0.908</td>
<td>0.908</td>
<td>0.893</td>
</tr>
<tr>
<td>Caltech-256</td>
<td>0.797</td>
<td>0.798</td>
<td>0.793</td>
<td>0.729</td>
<td>0.728</td>
<td>0.810</td>
<td>0.812</td>
<td>0.781</td>
</tr>
<tr>
<td>ECUSTFD</td>
<td>0.944</td>
<td>0.956</td>
<td>0.976</td>
<td>0.832</td>
<td>0.825</td>
<td>0.998</td>
<td>0.998</td>
<td></td>
</tr>
<tr>
<td>Animal images</td>
<td>0.897</td>
<td>0.908</td>
<td>0.905</td>
<td>0.820</td>
<td>0.821</td>
<td>0.910</td>
<td>0.915</td>
<td>0.905</td>
</tr>
<tr>
<td>Office31</td>
<td>0.877</td>
<td>0.865</td>
<td>0.871</td>
<td>0.789</td>
<td>0.787</td>
<td>0.914</td>
<td>0.913</td>
<td>0.918</td>
</tr>
<tr>
<td>Modern-office31</td>
<td>0.907</td>
<td>0.913</td>
<td>0.915</td>
<td>0.827</td>
<td>0.824</td>
<td>0.952</td>
<td>0.952</td>
<td>0.945</td>
</tr>
<tr>
<td>Cifar-10 (IR=m)</td>
<td>0.850</td>
<td>0.862</td>
<td>0.841</td>
<td>0.865</td>
<td>0.862</td>
<td>0.835</td>
<td>0.825</td>
<td>0.805</td>
</tr>
<tr>
<td>Cifar-10 (IR=10)</td>
<td>0.787</td>
<td>0.793</td>
<td>0.794</td>
<td>0.826</td>
<td>0.830</td>
<td>0.817</td>
<td>0.815</td>
<td>0.708</td>
</tr>
<tr>
<td>Cifar-10 (IR=20)</td>
<td>0.770</td>
<td>0.771</td>
<td>0.786</td>
<td>0.772</td>
<td>0.790</td>
<td>0.794</td>
<td>0.798</td>
<td>0.668</td>
</tr>
<tr>
<td>Cifar-100 (IR=m)</td>
<td>0.573</td>
<td>0.588</td>
<td>0.564</td>
<td>0.577</td>
<td>0.565</td>
<td>0.541</td>
<td>0.562</td>
<td>0.553</td>
</tr>
<tr>
<td>Cifar-100 (IR=10)</td>
<td>0.524</td>
<td>0.532</td>
<td>0.523</td>
<td>0.550</td>
<td>0.545</td>
<td>0.544</td>
<td>0.555</td>
<td>0.494</td>
</tr>
<tr>
<td>Cifar-100 (IR=20)</td>
<td>0.434</td>
<td>0.473</td>
<td>0.453</td>
<td>0.476</td>
<td>0.468</td>
<td>0.493</td>
<td>0.511</td>
<td>0.428</td>
</tr>
<tr>
<td>Average</td>
<td>0.769</td>
<td>0.779</td>
<td>0.803</td>
<td>0.761</td>
<td>0.743</td>
<td>0.828</td>
<td>0.829</td>
<td>0.799</td>
</tr>
</tbody>
</table>

TABLE X: Comparison against loss functions on AUC

<table>
<thead>
<tr>
<th>Datasets</th>
<th>VGG16</th>
<th>Focal_loss</th>
<th>Seesaw_loss</th>
<th>F</th>
<th>OF</th>
<th>RBS+LRNN</th>
<th>RBS+CNRC</th>
<th>RBS+CRC</th>
</tr>
</thead>
<tbody>
<tr>
<td>PneumoniaMNIST</td>
<td>0.856</td>
<td>0.864</td>
<td>0.853</td>
<td>0.813</td>
<td>0.823</td>
<td>0.851</td>
<td>0.877</td>
<td>0.856</td>
</tr>
<tr>
<td>BreastMNIST_2class</td>
<td>0.798</td>
<td>0.807</td>
<td>0.740</td>
<td>0.702</td>
<td>0.697</td>
<td>0.799</td>
<td>0.815</td>
<td>0.796</td>
</tr>
<tr>
<td>BreastMNIST_3class</td>
<td>0.856</td>
<td>0.836</td>
<td>0.813</td>
<td>0.730</td>
<td>0.735</td>
<td>0.874</td>
<td>0.893</td>
<td>0.858</td>
</tr>
<tr>
<td>Brain Tumor</td>
<td>0.871</td>
<td>0.894</td>
<td>0.885</td>
<td>0.881</td>
<td>0.876</td>
<td>0.981</td>
<td>0.987</td>
<td>0.978</td>
</tr>
<tr>
<td>Tuberculosis</td>
<td>0.932</td>
<td>0.935</td>
<td>0.955</td>
<td>0.898</td>
<td>0.891</td>
<td>0.989</td>
<td>1.000</td>
<td>0.986</td>
</tr>
<tr>
<td>Caltech-101</td>
<td>0.936</td>
<td>0.942</td>
<td>0.945</td>
<td>0.935</td>
<td>0.934</td>
<td>0.967</td>
<td>0.954</td>
<td>0.951</td>
</tr>
<tr>
<td>Caltech-256</td>
<td>0.894</td>
<td>0.899</td>
<td>0.896</td>
<td>0.864</td>
<td>0.873</td>
<td>0.904</td>
<td>0.905</td>
<td>0.895</td>
</tr>
<tr>
<td>ECUSTFD</td>
<td>0.970</td>
<td>0.977</td>
<td>0.982</td>
<td>0.911</td>
<td>0.908</td>
<td>0.993</td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td>Animal images</td>
<td>0.945</td>
<td>0.934</td>
<td>0.941</td>
<td>0.892</td>
<td>0.897</td>
<td>0.946</td>
<td>0.949</td>
<td>0.941</td>
</tr>
<tr>
<td>Office31</td>
<td>0.936</td>
<td>0.930</td>
<td>0.938</td>
<td>0.891</td>
<td>0.890</td>
<td>0.956</td>
<td>0.955</td>
<td>0.957</td>
</tr>
<tr>
<td>Modern-office31</td>
<td>0.952</td>
<td>0.955</td>
<td>0.960</td>
<td>0.910</td>
<td>0.909</td>
<td>0.975</td>
<td>0.975</td>
<td>0.971</td>
</tr>
<tr>
<td>Cifar-10 (IR=m)</td>
<td>0.891</td>
<td>0.903</td>
<td>0.896</td>
<td>0.901</td>
<td>0.901</td>
<td>0.908</td>
<td>0.905</td>
<td>0.895</td>
</tr>
<tr>
<td>Cifar-10 (IR=10)</td>
<td>0.862</td>
<td>0.872</td>
<td>0.873</td>
<td>0.876</td>
<td>0.880</td>
<td>0.898</td>
<td>0.898</td>
<td>0.833</td>
</tr>
<tr>
<td>Cifar-10 (IR=20)</td>
<td>0.834</td>
<td>0.854</td>
<td>0.858</td>
<td>0.842</td>
<td>0.851</td>
<td>0.885</td>
<td>0.896</td>
<td>0.785</td>
</tr>
<tr>
<td>Cifar-100 (IR=m)</td>
<td>0.783</td>
<td>0.792</td>
<td>0.764</td>
<td>0.792</td>
<td>0.795</td>
<td>0.788</td>
<td>0.797</td>
<td>0.774</td>
</tr>
<tr>
<td>Cifar-100 (IR=10)</td>
<td>0.742</td>
<td>0.775</td>
<td>0.759</td>
<td>0.775</td>
<td>0.781</td>
<td>0.774</td>
<td>0.785</td>
<td>0.744</td>
</tr>
<tr>
<td>Cifar-100 (IR=20)</td>
<td>0.695</td>
<td>0.728</td>
<td>0.713</td>
<td>0.721</td>
<td>0.719</td>
<td>0.733</td>
<td>0.762</td>
<td>0.735</td>
</tr>
<tr>
<td>Average</td>
<td>0.868</td>
<td>0.880</td>
<td>0.870</td>
<td>0.859</td>
<td>0.862</td>
<td>0.895</td>
<td>0.903</td>
<td>0.879</td>
</tr>
</tbody>
</table>

Fig. 3: Classification accuracy on the Cifar dataset compared with combination of different algorithms.
can locate high-quality instances for a specific query and thus effectively address data imbalance.

The comparison between RBS+CNRC and other loss functions on ACC for different classes in revised Cifar-10 and Cifar-100 datasets are visually illustrated in Fig. 3. Similar to the results demonstrated in Fig. 2, RBS outperforms its competitors on the minority classes in most cases. Occasionally, for Cifar-10 (IR=20), RBS is not sensitive to the minority classes 1 and 2. This further verifies the ability of RBS in detecting minority instances.

In summary, by examining all of the obtained results, there are clear evidence that RBS can effectively address data imbalance and improve the representation of images with correlated features, jointly leading to a better recognition performance. This outperformance of RBS is mainly due to the use of the sampling dictionaries, which imply the intra-class correlations of each class and the correlated instances, to form a balanced dataset specific to a given query, so as to improve the quality of the dataset.

V. CONCLUSION

To address the data imbalance issue for classification, this paper presented a representation-based sampling (RBS) method which is able to generate a balanced query-specific dataset, leading to better classification performance. The method includes two stages. In the first stage, each instance in the training dataset is reconstructed by its own class, leading to a class-specific sampling dictionary which records the sampled instances. In the second stage, for a given query, the sampled data in each class are retrieved in line with the most correlated object to the query in this class. The proposed approach was evaluated in reference to alternative sampling methods or the algorithmic solutions for deep learning models, and the experimental results demonstrate the superiority of RBS in addressing data imbalance for classification.

Topics for further research include a more comprehensive development of the RBS method through collaboration with alternative regularisation strategies, such as \( \ell_1 \) regularisation term, and metric learning [46], [47], to target more complicated applications. Also, the use of the sampling dictionary in RBS is an effective means of revealing intra-class correlations of each class. Via these offline sampling dictionaries, an investigation on alternative strategies to retrieve the specific information for a given query remains active. Since data imbalance is a common issue in many scenarios, such as anomaly detection and fault detection, further probes into the innovative applications of RBS would construct the foundation for a broader spectrum of future research.

REFERENCES


