# Balanced Mini-batch Training for Imbalanced Image Data Classification with Neural Network

Ryota Shimizu, Kosuke Asako, Hiroki Ojima, Shohei Morinaga, Mototsugu Hamada, and Tadahiro Kuroda
Faculty of Science and Technology
Keio University
Yokohama, Japan
r_shimizu@kuroda.elec.keio.ac.jp

***Abstract*—We propose a novel method of training neural networks for industrial image classification that can reduce the effect of imbalanced data in supervised training. We considered visual quality inspection of industrial products as an image-classification task and attempted to solve this with a convolutional neural network; however, a problem of imbalanced data emerged in supervised training in which the neural network cannot optimize parameters. Since most industrial products are not defective, samples of defective products were fewer than those of the non-defective products; this difference in the number of samples causes an imbalance in training data. A neural network trained with imbalanced data often has varied levels of precision in determining each class depending on the difference in the number of class samples in the training data, which is a significant problem in industrial quality inspection. As a solution to this problem, we propose a balanced mini-batch training method that can virtually balance the class ratio of training samples. In an experiment, the neural network trained with the proposed method achieved higher classification ability than that trained with over-sampled or under-sampled data for two types of imbalanced image datasets.**

***Keywords-imbalanced data; visual quality inspection; image classification; convolutional neural network; mini-batch training***

## I. Introduction

To maintain production quality, visual quality inspection is important in the manufacturing process of industrial products. There is a demand to automate quality inspections because the more products to be produced, the more skilled inspectors or time will be required. We suggested visual quality inspection of industrial products as an image-classification task and trained a convolutional neural network with a dataset created from labeled images taken during the manufacturing process. However, since the proportion of non-defective industrial products is usually higher than that of defective ones, the dataset we made was an imbalanced dataset [1] having a large bias on the number of samples for each class; the neural network trained with the imbalanced training dataset tended to show weak generalization ability.

In other cases as well, an imbalanced training dataset adversely affects supervised training of a neural network. For example, in classifying certain input data into two classes. i.e., A or B, the trained neural network tends to classify most inputs as class A when the sample ratio of A to B in the training samples is 99 to 1. A neural network is generally trained with the gradient descent method, which optimizes parameters to minimize gradients of the loss function quantitatively showing the differences between labels and determined classes. However, when the training dataset is imbalanced as in the above example, parameters hardly change because the gradient of the loss function is flat while the neural network classifies all input data as majority class A; in other words, too many samples of majority class A prevent the neural network from understanding the features of inputs of minority class B.

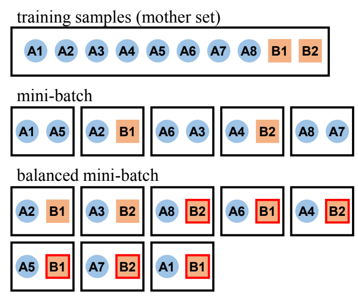


Figure 1. Example conventional and balanced mini-batches generated in one epoch. There are 8 samples of class A and 2 samples of class B in mother set, then mini-batches were generated by batch size of 2.

Unlike datasets for competitions such as the MNIST dataset, an actual target of classification, i.e., industrial products, tends to be imbalanced data, and the breakdown of classes is very imbalanced; when training the neural network with the image data created from real samples, it is necessary to combat the problem of imbalanced data because the classifier that always determines all products as non-defective cannot be used for quality inspection. Conventional data-level methods for rebalancing imbalanced data are, e.g., over-sampling, under-sampling, and a combination of both [1]. Over-sampling is a method of fabricating samples of minority classes by adding random noise to samples, generating average samples from other samples, or just duplicating samples. Under-sampling is a method of randomly discarding samples of majority classes. However, these methods are not without problems. Non-genuine samples fabricated by over-sampling without proper adjustment cause overfitting of the neural network because the features of fabricated samples often differ from those of genuine samples. Repetitive use of the same samples for training also causes overfitting. Under-sampling can also cause overfitting because valuable information is discarded, and remaining samples have low diversity.

978-1-5386-9209-7/18/$31.00 ©2018 IEEE
DOI 10.1109/ai4i.2018.00015

**Algorithm 1** Balanced mini-batch generating.

**Inputs:**

-number of classes: $N$

-Subsets of samples for each class:
$\boldsymbol{S} = \{\boldsymbol{S}_i \mid \boldsymbol{S}_1, \boldsymbol{S}_2, \ldots, \boldsymbol{S}_N\}$

-Numbers of samples for each class:
$\mathbf{C} = \{c_i \mid c_1, c_2, \ldots, c_N\}$

-batch size: $M$

-number of mini-batches in 1 epoch: $K = c_{\max}/(M/N)$

**Algorithm:**

1:Initialize empty mini-batches
$\boldsymbol{B} = \{\boldsymbol{B}_i \mid \boldsymbol{B}_1, \boldsymbol{B}_2, \ldots, \boldsymbol{B}_K\}$
2:**for** $i = 1$ to $K$ **do**:
  3:**for** $j = 1$ to $N$ **do**:
    4:randomly choose $(M/N)$ samples of $\boldsymbol{S}_j$,
    5:add chosen samples to $\boldsymbol{B}_i$
    6:**if** $c_i == c_{\max}$ **do**:
      7:eliminate chosen samples from $\boldsymbol{S}_j$
    8:**end if**
  9:**end for**
10:**end for**

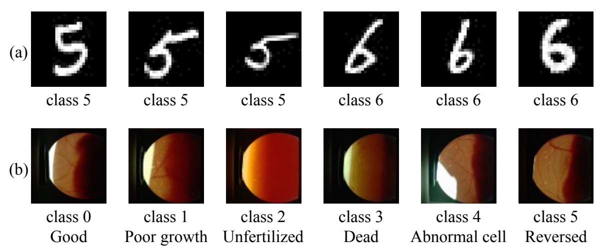


Figure 2. Example of two image datasets; (a): imbalanced MNIST dataset, which has samples of digits "5" and "6". (b): egg image dataset, which has samples of six quality classes.

Table 1. Breakdown of imbalanced MNIST.

| Class | Training samples | | | | Test samples |
|---|---|---|---|---|---|
| | 100% | 10% | 1% | 0.1% | |
| 5 | 5000 | 5000 | 5000 | 5000 | 892 |
| 6 | 5000 | 500 | 50 | 5 | 958 |
| Total | 10000 | 5500 | 5050 | 5005 | 1850 |
| Standard deviation | 0 | 3182 | 3500 | 3532 | 49.669 |

Table 2. Breakdown of egg image dataset.

| Class | Training samples | Test samples |
|---|---|---|
| 0 | 78345 | 100 |
| 1 | 8038 | 100 |
| 2 | 12 | 100 |
| 3 | 1667 | 100 |
| 4 | 2898 | 100 |
| 5 | 152 | 100 |
| Total | 91112 | 600 |
| Standard deviation | 31081 | 0 |

As a solution to this problem, we propose a balanced mini-batch training method that can be used without editing training samples. As shown in Fig. 1, in balanced mini-batch training, the numbers of samples for each class in a mini-batch are restricted to be the same. This method does not modify genuine samples or discard valuable samples, so it can solve the problems with over-sampling and under-sampling. We compared the proposed method with over-sampling and under-sampling methods by using an imbalanced MNIST dataset and new imbalanced egg image dataset created from labeled images taken in the manufacturing process.

The remainder of this paper is organized as follows. In Section 2, we give details of the algorithm of our balanced mini-batch training method. We explain the experimental conditions and present the results in Section 3. Finally, we conclude the paper in Section 4.

## II. Proposed Method

In conventional mini-batch training, instead of feeding all training samples into the neural network, a small number of samples are randomly selected from all training samples then they are fed into the neural network as a mini-batch; therefore, the computation for the backpropagation is smaller than when feeding samples into the neural network one by one. When generating mini-batches, a predetermined number of samples are randomly chosen from the mother set that has all the samples, then the chosen samples are eliminated from the mother set. This is repeated until samples in the mother set run out. This procedure is considered one epoch, and the samples in the mother set are restored every time a new epoch comes. The predetermined number of samples in one mini-batch is called batch size, so the number of mini-batches generated in one epoch is equal to the number of all samples divided by batch size.

However, with conventional mini-batch training, all samples are extracted once without duplication within the same epoch, and mini-batches inherit the sample imbalance for classes in the mother set when the training dataset is imbalanced, as shown in Fig. 1(a); therefore, rebalancing methods, e.g., over-sampling or under-sampling, are necessary when using conventional mini-batch training with an imbalanced dataset. In contrast to conventional mini-batch training, as shown in Fig. 1(b) and Algorithm 1, our balanced mini-batch training method allows overlapping selections of minority samples in the same epoch and limits the number of samples in each class in one mini-batch to be equal to the batch size divided by the number of classes.

By using our balanced mini-batch training method, samples are rebalanced only when a mini-batch is created; therefore, no over-sampling nor under-sampling is required, and the risk of overfitting is reduced.

## III. Experiment and Result

### A. Image Dataset

We used two imbalanced image datasets; imbalanced MNIST and egg image. Their sample images are shown in Fig. 2.

***Imbalanced MNIST***. To confirm the effectiveness of the proposed method regarding imbalanced training data, we prepared a dataset that is intentionally imbalanced, i.e., imbalanced MNIST. The original MNIST dataset has images of handwritten digits from "0" to "9" and consists of 60,000 samples used for training and 10,000 samples for test, but we intentionally extracted some images of "5" and "6" from the training data to create an imbalanced two-class training dataset. We changed the ratio of class 6 to class 5 in training samples for 100%, 10%, 1%, and 0.1% then evaluated the effect of proposed

Table 3. Contingency table of MNIST test data (0.1%, minibatch).

| Class | | Labels | | |
|---|---|---|---|---|
| | | 5 | 6 | Total |
| Determined | 5 | **892** | 958 | 1850 |
| | 6 | 0 | **0** | 0 |
| | Total | 892 | 958 | 1850 |

Table 4. Contingency table of MNIST test data (0.1%, balanced minibatch).

| Class | | Labels | | |
|---|---|---|---|---|
| | | 5 | 6 | Total |
| Determined | 5 | **891** | 270 | 1161 |
| | 6 | 1 | **688** | 689 |
| | Total | 892 | 958 | 1850 |

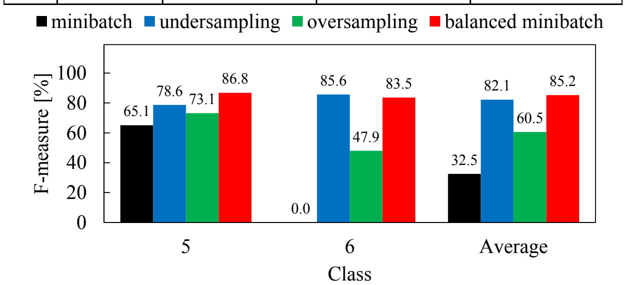


Figure 3. F-measure for each class (imbalanced MNIST, 0.1%).

method in different variance of training samples. For testing, we extracted all images of “5” and “6” from the original test data. The breakdown of samples is shown in Table 1. The original image size of MNIST is 28 x 28 pixels, but we resized it to 224 x 224 pixels to be used with GoogLeNet [3], which is described later.

***Egg image dataset***. We used the labeled images of hens’ eggs as an example of an image dataset of actual industrial products. The eggs were produced in large quantities, and the quality classes were labeled from visual inspection by skilled inspectors. This egg image dataset has six quality classes, and each image is labeled with one class. Since the images were actually taken in the manufacturing process, this dataset is imbalanced; as shown in Table 2, class 0 is a majority class showing non-defective eggs and there are few samples of classes 2 and 5 showing different defective eggs. All images are 224 x 224 pixels.

### *B. Neural Network Configurations and Training Methodology*

The model of a convolutional neural network called GoogLeNet [3] was used as a classifier in our experiment. The output dimensions of the final fully connected layer and auxiliary classifiers were changed from 1000 (original) to 2 (imbalanced MNIST) or 6 (egg image dataset). Other structures were the same as those in the GoogLeNet paper [3]. The hyper parameters were set as follows. Parameters were optimized by Adam [4] with an initial learning rate of 0.001, and the batch size was 60. The initial parameter values were given by truncated normal distribution with a standard deviation of 0.1. Training was stopped after feeding 1000 mini-batches for imbalanced MNIST, or 10000 mini-batches for the egg image dataset, which equals 60000 and 600000 image samples, respectively.

Four different classifiers (same models) were trained with four different methods, i.e., conventional mini-batch training with imbalanced data (minibatch), conventional mini-batch training with over-sampled data (oversampling), conventional mini-batch training with under-sampled data (undersampling), and our balanced mini-batch training with imbalanced data (balanced minibatch), to compare the generalization abilities of the classifiers. The mixup-based data-augmentation method [2] was used as the over-sampling method, which fabricates pseudo images from the weighted sum of two images of the same class; one image is chosen as the base image, its brightness values are weighted by 0.8, and the other image is chosen as a noise image, and its brightness values are weighted by 0.2. The weighted sum of the base and noise images is used as a new pseudo sample of the class. With over-sampling, the procedure was repeated until the sample number of each class reached the sample number of majority class. With under-sampling, samples were randomly discarded until the sample number of each class reached the sample number of minority class.

### *C. Evaluation*

After training, the generalization abilities of the classifiers for the test data were evaluated based on F-measure calculated from precision and recall, using Equations 1, 2, and 3;

$$(Precision)_x = \frac{y_{x,\,x}}{\sum_{i=1}^{6}(y_{i,\,x})}, \tag{1}$$

$$(Recall)_x = \frac{y_{x,\,x}}{\sum_{i=1}^{6}(y_{x,\,i})}, \tag{2}$$

$$(F\text{-}measure)_x = \frac{2\cdot(Precision)_x\cdot(Recall)_x}{(Precision)_x+(Recall)_x}, \tag{3}$$

where variable $y_{a,\,b}$ indicates the number of samples whose labeled quality class is $a$ and the class determined by the classifier is $b$. $(Precision)_x$ is the percentage of samples whose label is $x$ among all samples determined as class $x$. High precision to the class means the classification for the class by the classifier is reliable. $(Recall)_x$ is the percentage of samples determined as class $x$ among all samples whose label is $x$. High recall to the class indicates high detection ability of the classifier for that class. $(F\text{-}measure)_x$ is the harmonic mean of precision and recall to the class and shows the comprehensive ability of the classifier for the class $x$.

### *D. Results*

***Imbalanced MNIST***. Tables 3 and 4 are contingency tables for the test data provided using conventional and our balanced mini-batch with imbalanced MNIST (0.1%). F-measures for each class calculated from the contingency tables are shown in Fig. 3. The classifier trained with minibatch failed to classify class 6 samples; therefore, F-measure of class 6 was quite low. This problem was mitigated with the other three methods, and average F-measures of class 5 and class average were the highest when the case used our balanced mini-batch training method.

***Egg image dataset***. Tables 5 and 6 are contingency tables and Fig. 4 is the graph of F-measure for each class. As with imbalanced MNIST, the conventional mini-batch training

Table 5. Contingency table of egg test data (minibatch).

| | Class | Labels | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | 4 | 5 | Total |
| Judged class | 0 | **99** | 17 | 0 | 0 | 17 | 0 | 133 |
| | 1 | 0 | **79** | 0 | 2 | 0 | 0 | 81 |
| | 2 | 0 | 0 | **0** | 0 | 0 | 0 | 0 |
| | 3 | 0 | 3 | 100 | **98** | 0 | 0 | 201 |
| | 4 | 1 | 1 | 0 | 0 | **82** | 64 | 148 |
| | 5 | 0 | 0 | 0 | 0 | 1 | **36** | 37 |
| | Total | 100 | 100 | 100 | 100 | 100 | 100 | 600 |

Table 6. Contingency table of egg test data (balanced minibatch).

| | Class | Labels | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | 4 | 5 | Total |
| Judged class | 0 | **93** | 12 | 0 | 0 | 6 | 0 | 111 |
| | 1 | 1 | **80** | 0 | 2 | 0 | 0 | 83 |
| | 2 | 0 | 0 | **83** | 0 | 0 | 0 | 83 |
| | 3 | 0 | 5 | 17 | **98** | 0 | 0 | 120 |
| | 4 | 6 | 3 | 0 | 0 | **93** | 7 | 109 |
| | 5 | 0 | 0 | 0 | 0 | 1 | **93** | 94 |
| | Total | 100 | 100 | 100 | 100 | 100 | 100 | 600 |

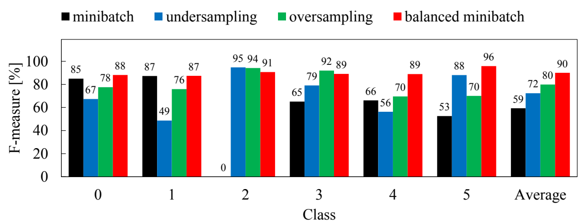


Figure 4. F-measure for each class (egg image dataset).

method with imbalanced data made the classifier have weak recall for minority class 2, then F-measure for the class was low. With over-sampling or down-sizing, F-measure for class 2 improved over 90% but that for class 0 worsened. Only our balanced mini-batch succeeded in improving F-measures for all classes, and the average F-measure was 10% or more higher than that of over-sampling and under-sampling.

The difference in generalization ability in the egg image dataset was more significant than that in imbalanced MNIST. This is because of the differences in the strength of imbalance in the datasets, which is represented by the standard deviation in Tables 1 and 2. Fig. 5 shows F-measure for majority classes of imbalanced MNIST (class 5), and egg image dataset (class 0) in different ratio of minority samples to majority samples. Fig. 6 shows that of minority classes: class 6 for imbalanced MNIST, and sum of classes 1, 2, 3, 4, and 5 for egg image dataset. Fig. 7 shows average F-measures of majority and minority classes. From these figures, it can be noticed that proposed method succeeded to improve generalization ability for majority class, i.e., non-defective products, but that for minority class, i.e., defective products, were not so changed. From these results, our balanced mini-batch method can improve generalization ability for majority class of non-defective products without worsening that for minority class of defective products. In addition, proposed method is more effective than over-sampling or under-sampling for creating a reliable classifier when training data are highly imbalanced.

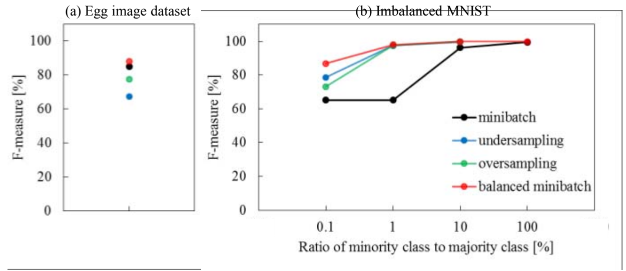


Figure 5. F-measure for majority class.

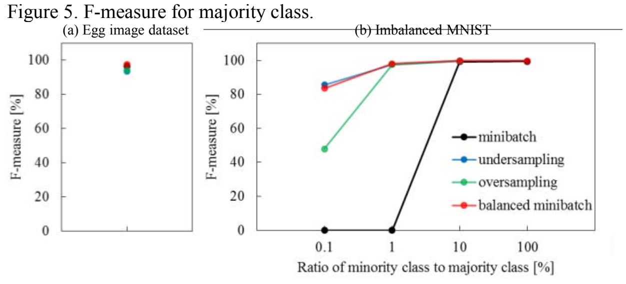


Figure 6. F-measure for minority class.

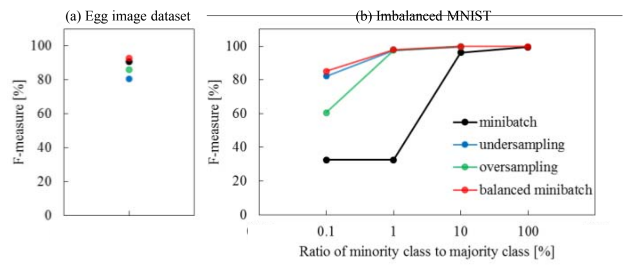


Figure 7. Average F-measure of minority and majority classes.

## IV. Conclusion

We proposed a method of training a neural network for classification, which is effective for problems of imbalanced training data. In the experiment, the classifier trained with our method with imbalanced image datasets showed higher generalization ability than those trained with over-sampling or under-sampling methods. Therefore, out balanced mini-batch training is effective for training a classifier for quality inspection of industrial products.


## Acknowledgment

This work is supported by JST ACCEL Grant Number JPMJAC1502, Japan. The egg images we used were kindly provided by Shikoku Instrumentation Co. Ltd.


## References


[1] N. Japkowicz, "Learning from Imbalanced Data Sets: A Comparison of Various Strategies," in AAAI Workshop, 2000.

[2] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "Mixup: beyond empirical risk minimization," arXiv:1710.09412, 2017.

[3] C. Szegedy. *et al.*, "Going Deeper with Convolutions," IEEE Conference on Computer Vision and Pattern recognition, 2015.

[4] D. P. Kingma and J. L. Ba, "Adam: a method for stocastic optimization," Proceedings of International Conference of Learning Representations, 2015.