PLANKTON CLASSIFICATION ON IMBALANCED LARGE SCALE DATABASE VIA CONVOLUTIONAL NEURAL NETWORKS WITH TRANSFER LEARNING

Hansang Lee, Minseok Park, and Junmo Kim

School of Electrical Engineering, KAIST, South Korea

ABSTRACT

Plankton image classification plays an important role in the ocean ecosystems research. Recently, a large scale database for plankton classification with over 3 million images annotated with over 100 classes was released. However, the database suffers from imbalanced class distribution in which over 90% of images belong to only 5 classes. Due to this class-imbalance problem, the existing classification approaches are limited to label the data only to major classes, ignoring the small-sized classes. In this paper, we propose a fine-grained classification method for large scale plankton database based on convolutional neural networks (CNN). To overcome the class-imbalance problem, we incorporate transfer learning by pre-training CNN with class-normalized data and fine-tuning with original data. The class-normalized data is constructed by reducing the number of data via random sampling, for large-sized classes. In experiments, our method showed superior classification accuracy compared to both CNN without transfer learning and CNN with transfer learning via other data augmentation techniques.

Index Terms— Image classification, plankton images, transfer learning, convolutional neural networks, fine-grained classification

1. INTRODUCTION

Planktons are the most fundamental components of ocean ecosystems. Since the changes of their abundance or distribution, e.g. eutrophication or pollution, are useful indicators for oceanic or climatic events, it is important to observe and monitor their habit and behavior. Earlier scientists have studied the temporal and spatial changes in plankton distribution by manually collecting the samples of planktons using towed nets, pumps, or Niskin bottles, and manually identifying and counting them, which is expensive and time consuming. Recently, imaging-based technologies [1] have been developed to automatically capture plankton images continuously using underwater image sensors. Since these imaging systems often generates a large amount of image data, the presence of automated techniques for pattern recognition and classification of plankton images [2, 3, 4] has also been requested.

Hansang Lee and Minseok Park equally contributed.

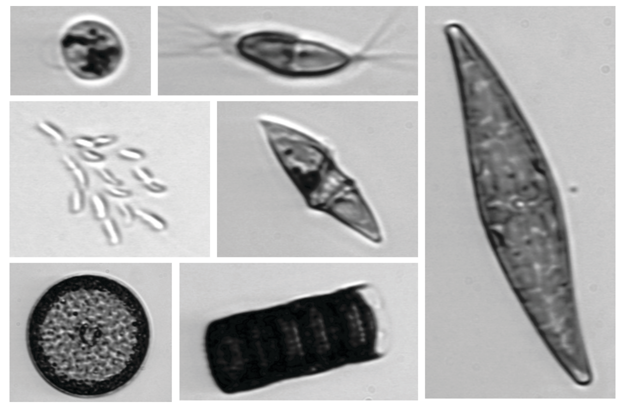
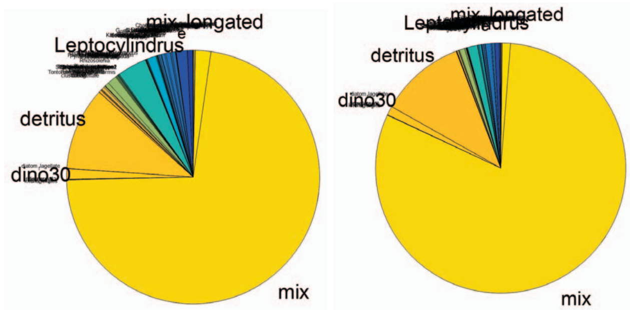


Fig. 1: Sample images of WHOI-Plankton database [5]

Recently, one large scale database for plankton classification, called WHOI-Plankton [5], was publicly released. The WHOI-Plankton consists of over 3.4 million images of planktons, as shown in Fig. 1, captured by the Imaging FlowCytobot (IFCB) system at Woods Hole Oceanographic Institution (WHOI) since 2006. These 8-years-collected images are labelled by experts into 103 classes and distributed publicly as a benchmark for plankton classification. However, the data has a major challenge of class distribution; Over 70% of images belong to one class and over 90% of images belong to only five classes. This *class-imbalance* problem is known to cause a classifier to be biased so that its performance is degraded for small-sized classes [6, 7]. Similar to other image classification tasks, a classifier based on the convolutional neural network (CNN) [8, 9] showed high overall accuracy, but suffered from classifying small-sized classes [5].

In this paper, we propose an CNN-based plankton classification method for class-imbalanced large scale database. To overcome the class-imbalance problem, transfer learning-based CNN models with pre-training on class-normalized data is proposed. In pre-training, the class-normalized data is constructed by reducing the number of data via random sampling, for large-sized classes. The classifier pre-trained on class-normalized data is then fine-tuned by re-training on the original data to reflect the class distribution with preventing bias. In experiments, our method showed significantly improved performances on classifying small-sized classes



(a) Training data (b) Testing data

Fig. 2: Class distribution of training and testing data.

Classes	Total	Training	Testing
Mix	73.2%	72.4%	80.1%
Detritus	10.6%	10.6%	11.0%
Leptocylindrus	3.5%	3.8%	1.3%
Mix_elongated	1.9%	2.0%	1.1%
Dino30	1.3%	1.4%	1.2%

Table 1: Class percentages of five largest classes in total, training, and testing data.

while also maintaining its overall classification accuracy.

The rest of the paper is organized as follows. Section 2 reviews the WHOI-Plankton database and its class-imbalance problem. Section 3 presents the proposed transfer learning-based CNN model and Section 4 provides experimental results. Section 5 concludes with discussions.

2. REVIEW ON WHOI-PLANKTON DATABASE

In this section, we briefly review the WHOI-Plankton database and its class-imbalance problem. As mentioned in Sec. 1, WHOI-Plankton consists of 3.4 million expert-labelled plankton images with 103 classes. In experiments, all data from before 2014 was treated as training data and the data in 2014 was treated as testing data. As a result, the training data consists of over 3.2 million images of planktons and the testing data includes about 330,000 images. All images are labelled by experts into 103 classes, including the *mix* class for small, undifferentiated particles.

Fig. 2 shows the class distributions of training and testing data, and Table 1 shows the percentages of five largest classes in total, training, and testing data. The *mix* class dominates over both training and testing data with the ratios of 73.2% and 80.1%, respectively. Furthermore, the total ratios of five largest classes, denoted as *L5*, on training and testing data are 90.2% and 94.7%, respectively. This class-imbalance problem inevitably results in heavy bias of classifier learning. In [5], three baseline classifiers, a random forest (RF), the CIFAR10-based CNN, and the VGG-based CNN models, were applied and they have achieved high average accuracy of 90.8%, 92.8%, and 93.8%, respectively. However, for another measurement called the unweighted average F_1 score, which is computed by averaging the harmonic mean of the precision and recall class-by-class, they have achieved significantly low scores of 0.27, 0.36, and 0.42, for RF, CIFAR10, and VGG, respectively. Since the unweighted average F_1 score is less affected by the distribution of class sizes, it can be observed that the classifiers trained on the data mostly failed to capture the small-sized classes, while they just labeled all testing data into one of the large-sized classes. In this paper, we apply the transfer learning to one of the baseline CNN classifier to overcome the class-imbalance problem.

3. TRANSFER LEARNING WITH CLASS-NORMALIZED DATA

Our proposed model focuses on reducing bias from class-imbalance problem. First, we select one baseline model from CIFAR10 CNN model as a classifier model since CIFAR10 model showed competitive performance compared to that of complex VGG model [8] for this task [5]. The CIFAR10-quick model for Caffe [10] consists of three convolutional layers followed by two fully-connected layers. All convolutional layers have 5×5 filters with padding of 2 pixels and stride of 1 pixel, and have 32, 32, and 64 kernels, respectively. Pooling layer followed by each convolutional layer has a kernel of size 3×3 with stride of 2. We use average pooling layer except for the first max pooling layer. After the third convolutional and pooling layers, the first fully connected layer with 64 neurons is followed and the dropout [11] is applied to the first fully connected layer with the drop-rate of 0.5. Finally, the last fully connected layer has 103 neurons connected to the plankton classes.

Numerous approaches have been studied for handling with class-imbalanced data [6, 7, 12] by sampling the large-sized data [13, 14] or duplicating the small-size data [15, 16]. We construct the *class-normalized* data based on the original data, denoted as `full` data, and data thresholding with random sampling. For the data threshold N, the classes which size is larger than N are first identified as *large* classes. For each large class, N images are randomly sampled and other images are rejected to form the class-normalized data, denoted as `thresh` data. Throughout this data thresholding, the class-normalized data has thresholded class distribution in which no class is larger than N. The behavior of the `thresh` data depends on the choice of the threshold N; If N is too small, classification bias will be significantly reduced but the specificity will be lost. If N is too large, classification bias will not be reduced. In this paper, N is experimentally determined as 5000.

The classifier trained with the class-noramlized data `thresh` is less biased. However, since the class-imbalance in WHOI-Plankton comes from natural population of planktons,

Classifiers	All classes	*L5*	Rest
`full`	92.79	95.01	48.68
`noise`	70.40	71.52	48.31
`aug`	80.31	81.97	47.21
`thresh`	82.71	84.04	56.17
`noise+full`	92.00	94.29	46.57
`aug+full`	92.22	94.43	48.17
`thresh+full`	**92.80**	**94.77**	**53.61**

Table 2: Average accuracy rates for the proposed and comparative classifiers on testing data.

Classifiers	All classes	*L5*	Rest
`full`	0.1773	0.7773	0.1548
`noise`	0.2465	0.5409	0.3599
`aug`	0.2726	0.5776	0.3700
`thresh`	0.3086	0.6510	0.4044
`noise+full`	0.3038	0.7531	0.2971
`aug+full`	0.3212	0.7668	0.3156
`thresh+full`	**0.3339**	**0.7791**	**0.3262**

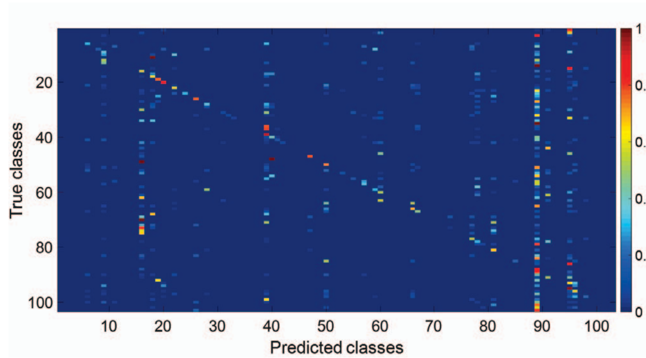
Table 3: Unweighted average F_1 scores for the proposed and comparative classifiers on testing data.

normalizing data brings another side effect of losing population information for classification. Thus, to restore the population information with reducing the bias, we apply transfer learning [17, 18, 19] for our classifier trained with `thresh` data, by re-training the classifier with the original `full` data. This re-trained classifier, denoted as `thresh+full`, not only reduces class bias but also reflects class population information to enhance its classification performance.

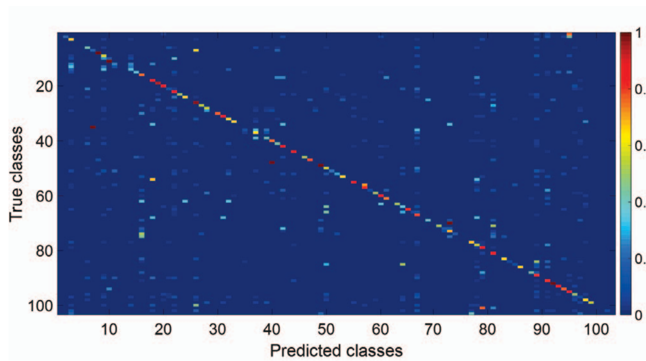
4. EXPERIMENTAL RESULTS

To validate our model, we evaluated the performances of our classifier and comparative methods. We first pre-processed the WHOI-Plankton data by resizing them to 64×64 with mean-value-padding, as in [5]. We constructed comparative methods based on two class-normalized data, `noise` [16] and `aug` [15], in which `noise` is formed with duplicating small-sized classes by adding white noise, and `aug` is formed with augmenting data in small-sized classes by rotation, scaling, translation, and flipping. The data threshold for `noise` and `aug` is experimentally determined as 1000. We also applied transfer learning to the classifiers trained with these two normalized data, to generate two classifiers denoted as `noise+full` and `aug+full`. Two measurements of average accuracy rate and unweighted average F_1 score are calculated for evaluation, as in [5], excluding the fact that day-by-day average is not incorporated for F_1 score.

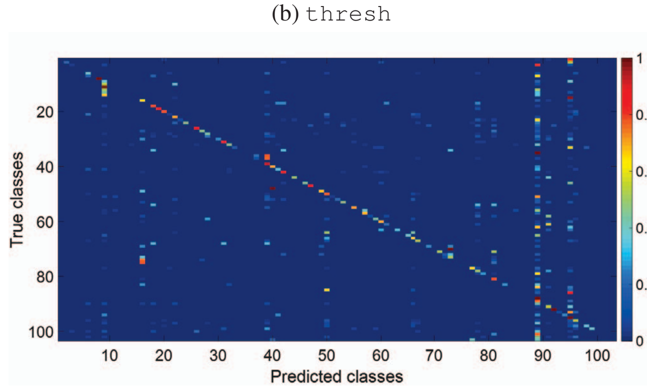
Table 2 and 3 show the average accuracy rates and unweighted average F_1 scores of our classifier and comparative methods, respectively. In Table 3, since we computed the score on all testing data, the scores of `full` classifier is different from those in [5], which have averaged the mean daily scores. As shown in the tables, the average accuracy of all classes is heavily affected by that of larger *L5* classes, while the unweighted average F_1 score of all classes is rather affected by that of rest classes. In `full` classifier, the classifier achieved high accuracy for *L5* classes, but failed to classify the smaller classes so that it scored F_1 of only 0.1773. In `thresh` classifier, the classification accuracy of the smaller classes was improved to 56.17%, but the overall accuracy was



(a) `full`



(b) `thresh`



(c) `thresh+full`

Fig. 3: Classification matrices of the true versus predicted classes for (a) `full`, (b) `thresh`, and (c) `thresh+full` classifiers.

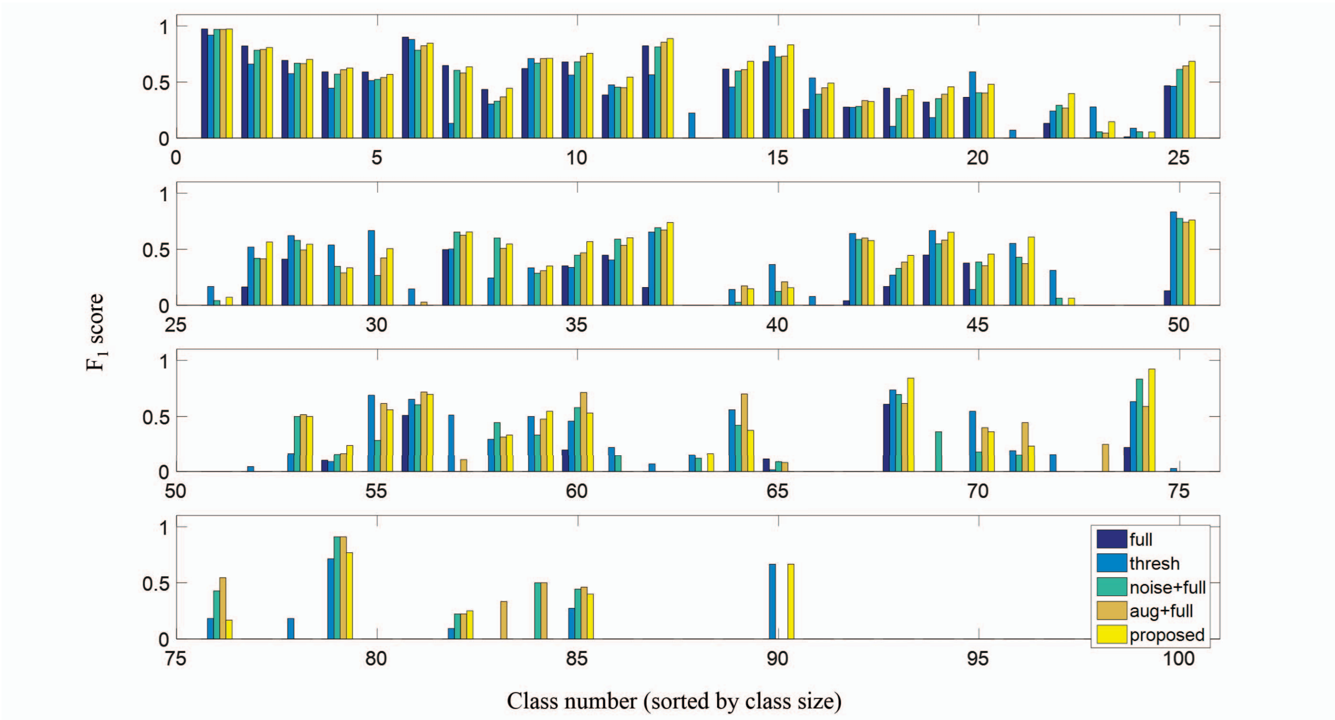


Fig. 4: F_1 scores per class for five methods. Class number is sorted by the class size in which the class 1 is the largest *mix* class in the testing data. Since 9 classes out of 103 have no images in the testing data, there are actually 94 valid classes.

degenerated due to its misclassification in *L5* classes. In the proposed `thresh+full` classifier, the classifier not only maintained the classification power of the smaller rest classes, but also improved its accuracy in the larger *L5* classes, so that it achieves high overall accuracy and the highest F_1 score of 0.3339. In data normalization perspective, it can be observed that the proposed data thresholding better reduced the classification bias than noise addition or data augmentation.

The classification behaviors of our model and comparative methods are visualized in Figs. 3 and 4. Fig. 3 shows the correlation matrices between true and predicted classes for three classifiers. Since `full` model is *L5*-biased, the correlation values in the diagonal are low and several vertical lines are shown. The `thresh` model is less biased so that the diagonal is enhanced, but the values in *L5* is decreased. Finally, the `thresh+full` model is shown more biased than the `thresh` classifier, but the diagonal values in *L5* is enhanced. Fig. 4 shows the class-by-class F_1 scores for our classifier and the comparative methods, in which the class number is sorted by the class size. Due to the *L5*-bias, the scores of the `full` classifier are distributed at the large-sized classes. The `thresh` classifier achieves high scores in small-sized classes, but gets low scores in large-sized classes. Finally the three transfer learning-based classifiers improves their scores in large-sized classes while maintaining the scores in small-sized classes. Furthermore, the proposed model shows the best results compared to those of other two models.

5. DISCUSSIONS AND CONCLUSIONS

In this paper, we investigated the recently released large scale data for plankton classification, and proposed the transfer learning-based CNN classifier model to overcome the class-imbalance problem. To reduce the class bias on small-sized classes, we constructed the class-normalized data by data thresholding for large-sized classes, and trained the CIFAR10 CNN model with this normalized data. To complement the pre-trained model by reflecting the population information, we apply transfer learning by re-training our model with the original full data. As a result, our model not only better performed on small-sized classes compared to the classifier trained with full data only, but also preserved the overall accuracy with the help of transfer learning.

Our future works will focus on further investigation of data statistics. Though we separated the data into two categories according to their sizes, the sizes of the classes in the same category also differ. For instance, in *L5*, the size of the largest class is 2.3 million while that of the smallest one is 45,000. We expect that the generation of normalized data will be more effective with the consideration of finding appropriate threshold. Furthermore, sequential transfer learning with multiple data thresholding in which the classifier first learns to classify large-sized data and sequentially classifies small-sized data, can be considered. Applying our model on other class-imbalanced data also remains as future applications.

6. REFERENCES

[1] Robert J. Olson and Heidi M. Sosik, "A submersible imaging-in-flow instrument to analyze nano-and microplankton: Imaging flowcytobot," *Limnology and Oceanography: Methods*, vol. 5, no. 6, pp. 195–203, 2007.

[2] Feng Zhao, Feng Lin, and Hock Soon Seah, "Binary {SIPPER} plankton image classification using random subspace," *Neurocomputing*, vol. 73, no. 10??2, pp. 1853 – 1860, 2010, Subspace Learning / Selected papers from the European Symposium on Time Series Prediction.

[3] Xiaoou Tang, Feng Lin, S. Samson, and A. Remsen, "Binary plankton image classification," *Oceanic Engineering, IEEE Journal of*, vol. 31, no. 3, pp. 728–735, July 2006.

[4] Heidi M. Sosik and Robert J. Olson, "Automated taxonomic classification of phytoplankton sampled with imaging-in-flow cytometry," *Limnology and Oceanography: Methods*, vol. 5, no. 6, pp. 204–216, 2007.

[5] Eric C. Orenstein, Oscar Beijbom, Emily E. Peacock, and Heidi M. Sosik, "Whoi-plankton- A large scale fine grained visual recognition benchmark dataset for plankton classification," *CoRR*, vol. abs/1510.00745, 2015.

[6] Haibo He, Edwardo Garcia, et al., "Learning from imbalanced data," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 21, no. 9, pp. 1263–1284, 2009.

[7] Giovanna Menardi and Nicola Torelli, "Training and assessing classification rules with imbalanced data," *Data Mining and Knowledge Discovery*, vol. 28, no. 1, pp. 92–122, 2014.

[8] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.

[9] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, Eds., pp. 1097–1105. Curran Associates, Inc., 2012.

[10] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell, "Caffe: Convolutional architecture for fast feature embedding," *arXiv preprint arXiv:1408.5093*, 2014.

[11] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[12] Miroslav Kubat, Stan Matwin, et al., "Addressing the curse of imbalanced training sets: one-sided selection," in *ICML*. Nashville, USA, 1997, vol. 97, pp. 179–186.

[13] Kate McCarthy, Bibi Zabar, and Gary Weiss, "Does cost-sensitive learning beat sampling for classifying rare classes?," in *Proceedings of the 1st international workshop on Utility-based data mining*. ACM, 2005, pp. 69–77.

[14] Nathalie Japkowicz and Shaju Stephen, "The class imbalance problem: A systematic study," *Intelligent data analysis*, vol. 6, no. 5, pp. 429–449, 2002.

[15] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, pp. 321–357, 2002.

[16] Laurens Maaten, Minmin Chen, Stephen Tyree, and Kilian Q Weinberger, "Learning with marginalized corrupted features," in *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, 2013, pp. 410–418.

[17] Sinno Jialin Pan and Qiang Yang, "A survey on transfer learning," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 22, no. 10, pp. 1345–1359, 2010.

[18] Maxime Oquab, Leon Bottou, Ivan Laptev, and Josef Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. IEEE, 2014, pp. 1717–1724.

[19] Dikan Xing, Wenyuan Dai, Gui-Rong Xue, and Yong Yu, "Bridged refinement for transfer learning," in *Knowledge Discovery in Databases: PKDD 2007*, pp. 324–335. Springer, 2007.