CGAN-PLANKTON: TOWARDS LARGE-SCALE IMBALANCED CLASS GENERATION AND FINE-GRAINED CLASSIFICATION

Chao Wang, Zhibin Yu, Haiyong Zheng, Nan Wang, Bing Zheng*

Department of Electronic Engineering, Ocean University of China
*zhenghaiyong@ouc.edu.cn

ABSTRACT

Plankton classification is becoming critically important as people concentrate more on oceans and global environment changing. Data of plankton species naturally exhibit imbalance in their class distribution. Meanwhile, it arouses fine-grained classification challenge. Although Convolutional Neural Networks (CNNs) have human-level performance on image classification task, they tend to be biased to large classes without considering the imbalance issue. In this paper, we introduce Generative Adversarial Network (GAN) based generative model to overcome these challenges. Our proposed model consists of fully convolutional layers, and includes three parts: a generative model G, a discriminative model D and a classification model C. We train generative and discriminative models on small classes data to learn a discriminative features through D model and reduce mode missing problem to some extent. We implement classification task using shared CNN layers of D model on whole data. Experimental results show that our model significantly improved the F1 score on an imbalanced plankton dataset with well-generated plankton images.

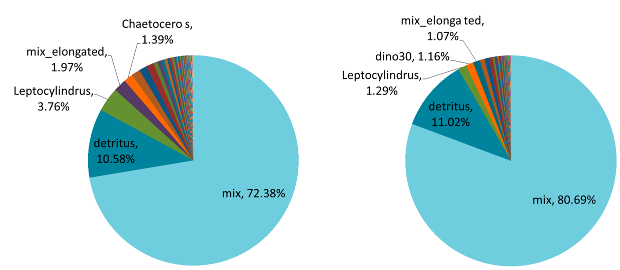
Index Terms— GAN, fine-grained classification, class imbalance, adversarial training, samples generation

1. INTRODUCTION

Study of plankton distribution is vital for global climate and environment protection. Plankton is the main oxygen producer, and accounts for about 90% of all oxygen. With the research of ocean, knowing the amount of plankton and the variation of them becomes more and more critical [1].

Due to the increasingly plankton imaging systems [2], more and more plankton images are acquired *in situ*, which require automatic plankton image classification and analysis urgently [3]. Although plankton image classification has been addressed for more than two decades [4, 5], this issue is still very challenging because: firstly, *in situ* plankton images naturally exhibit severe imbalance problem due to the characteristics of biological community structure and the classes of imbalanced species also change quickly; secondly, plankton image classificaion belongs to fine-grined visual classification problem which makes the classical classifiers with traditional hand-crafted features ineffective [6].

Recently, the Imaging FlowCytobot (IFCB) at WHOI has collected more than 700 million samples with expert labels called WHOI-Plankton [7]. Fig. 1 shows the distribution of the dataset, and it can be seen that the five large classes contribute more than 90% of the whole data. It's extremely imbalanced so that a challenging work for automatic classifier to distinguish each of the classes. Convolutional Neural Networks (CNNs) achieve state-of-the-art performance on image classification task [8, 9, 10]. As expected, a simple CNN model like cifar10 CNN can achieve a notable global accuracy, but a poor average recognition rate [7].



(a) Training dataset (b) Testing dataset

Fig. 1. Data distribution of WHOI-Plankton dataset.

In this paper, we wish to use an advanced generative model called Generative Adversarial Network (GAN) [11] to overcome the class imbalance problem. Our motivation is the idea of constructing a classifier through adversarial training as well as generating new samples from original data. Our proposed model contains a generative model G, a discriminative model D and a classification model C. In adversarial training procedure, when G model has the ability to generate almost real images, it also means that D model has a strong discriminative power of the original data. By learning effective discriminative parameters, D model can learn a hierarchical identification features of these fine-grained classes. The novelty of our model is that we implement classification

This work was supported by the National Natural Science Foundation of China under grant numbers 61301240 and 61271406, and China Postdoctoral Science Foundation under grant number 2016M590658.

shared weights with discriminate model and force D model to concentrate on features of small classes. In one training pipeline, we simultaneously train a generative model and a fine-grained classifier. After adding an extra supervised information, G model can generate more real samples than complete unsupervised training.

2. RELATED WORK

People start paying great attention to class imbalance problem from AAAI2000 Workshop [12]. Previous work can be mainly divided into two groups: sampling based method [13, 14, 15] and cost-sensitive learning [16, 17, 18]. On first released WHOI-Plankton dataset with 73 classes, [7] provides three baseline classifiers: random forest, CIFAR10 CNN model, and VGG16 model which achieve 90.8%, 92.8% and 93.8% weighted accuracy respectively. However, the classifier can be fooled by global accuracy, so they only get 0.27, 0.36 and 0.42 F1 scores respectively. Then [19] implemented transfer learning method to overcome imbalance challenge on WHOI-Plankton dataset, they used a CIFAR10 CNN model and transfer learned from sampled data, which have improved F1 score and keep a higher weighted accuracy.

Representation learning is commonly a main area of unsupervised learning. Generative model have been studied for many years. In 2014, Ian Goodfellow *et al.* proposed a new framework for learning a generative model via adversarial training [11]. GAN gives a novel way to represent the data and meanwhile provides a powerful discriminator of training data. Several recent papers focused on using GAN to generate new samples and solving the related problems, [20] proposed a recurrent generative model that can be trained using adversarial strategy, and [21, 22] focused on improving the stability of training and getting perceptual quality of GAN samples. In this paper, we introduce a GAN based method trying to solve the imbalance problem using G model to produce missing data and fine-grained classification problem implementing C model for supervised classification.

3. METHOD

GAN is a remarkable approach to train a generative model. The basic idea is that it contains two models, a generative model G and a discriminative model D. The task of the generator is to estimate the real distribution of raw dataset. It tries to make the generative images similar to the original one. Corresponding to the generator, discriminative model is trying to determine whether the given image looks like a real image from the dataset or an image created by generator. Both G and D could be a non-linear mapping function, and it is a multi-layer perception in the original GAN model. Suppose p_g is generated data distribution, it learns from mapping a prior noise distribution $p_z(z)$ to data $G(z;\theta_g)$. And $D(x;\theta_d)$ gives a probability to represent that x is from origin data or noise p_g. Then G and D will be simultaneously trained: we train G to adjust net weights for minimizing $log(1-D(G(z)))$ and D for minimizing $logD(x)$. It seems that they are playing a minimax game with value function $V(G,D)$:

$$\begin{aligned} \underset{G}{min}\,\underset{D}{max}\, V(D,G) &= \mathbb{E}_{x\sim p_{data}(x)}[logD(x)] \\ &+ \mathbb{E}_{z\sim p_z(z)}[log(1-D(G(z)))] \end{aligned} \tag{1}$$

Our proposed model contains three parts: generative model, discriminative model and classification model (share weights with discriminator), as shown in Fig. 2. Input of generative model is a 100-dimensional vector which is randomly sampled from a uniform distribution with $a=1$ and $b=-1$. The vector is inputted to a fully connected layer and then reshaped to $1024\times4\times4$ to be stored into the fractional-strided convolutional layer. All convolutional layers have the same kernel size 5×5 and 512, 256, 128 kernels. At the end of G model, use a convolutional layer with 1 kernel to generate a 64×64 gray image. Discriminative model is a mirror of G model, it contains four convolutional layers without any pooling or fc layer. Then the classification model which shares weights with discriminative model, has an extra fc layer to learn a deep features of the data and improve classification accuracy. Specially, it is a softmax classifier to take x as input and output a class possibility through computing $p(y^{(i)}=j|x^{(i)};\theta)=\frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^{k}e^{\theta_l^T x^{(i)}}}$. After implementing classification task, G model can generate realistic-looking samples. In each convolutional layer we use Batch Normalization [23] and scale. All convolutional layers use ReLU activation function except the last layer in G model who stores $tanh$ function.

As G model produces many "fake" samples of small classes, D model is trained like a binary classification task, so it performs more robust on samples of small classes. Based on this insight, D model could be an efficient classifier when implementing classification task on the whole imbalanced training data. By adding softmax loss with GAN loss, our proposed model has more power to represent the true distribution and become more stability to produce real images. After training on the whole data, we implement the trained model to generate samples and supply as true data for classification task. There still misses efficient quality evaluation of generated samples and all most papers use cross entropy as well as human eyes, here our model uses generated samples for real classification task to ensure its efficiency and sample diversity.

4. EXPERIMENTS

We perform unsupervised generative adversarial training on WHOI-Plankton small classes and supervised classification task on the whole WHOI data. The GAN approach not only provides a good generative model, but also trains a strong discriminator. Our GAN model architecture is mainly based on

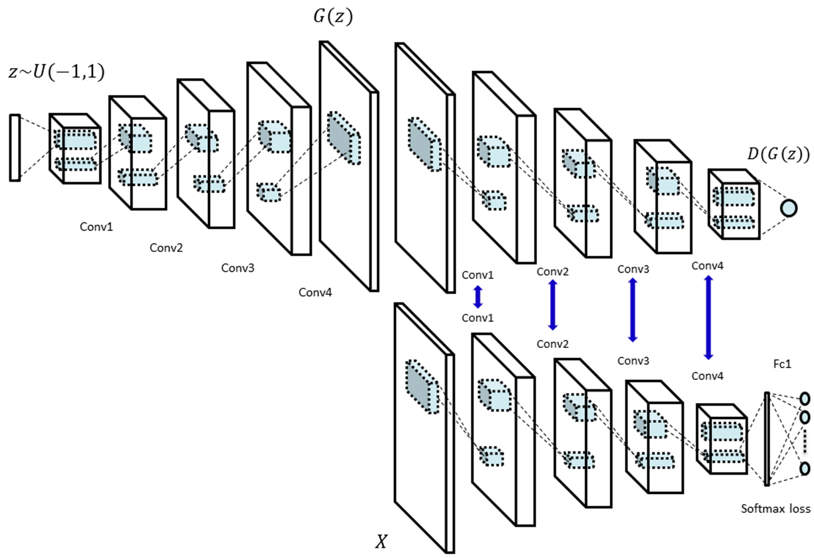


Fig. 2. Our CGAN-Plankton architecture. A 100 dimensional vector sampled from uniform distribution is the input of G model. D model has a series of stride convolution layers and maps $G(z)$ to a possibility score. C model is a parallel model sharing weights with D model and softmax loss. An additional full connection layer is added to enhance the classification performance.

DCGAN [22]. All experiments are deployed on caffe deep learning framework [24] and implemented with 4 NVIDIA TITAN X GPUs.

In experiments, we take the whole dataset into two parts: small classes and large classes. We define large class if the class has more than 5,000 samples and take the rest as small classes. Simply, we want G model to estimate the distribution of small classes and generate images like real ones, so we deploy unsupervised learning of small classes on the dataset. Since D and G model play a minimax game, if G can generate realistic images, it means that D model learns the fine-grained discriminate pattern of training samples.

Table 1 shows the comparison of the results on our proposed GAN model and other popular CNN models, and Fig. 3 and 4 are corresponding confusion matrices respectively. Compared with the classification-based models, our CGAN-Plankton outperforms by F1 score because of learning a well-represented dataset.

Table 1. Comparison of classification results between our CGAN-Plankton and other CNN models.

dataset	model	accuracy	F1 score
full	CIFAR10 CNN model	0.9279	0.1773
full	AlexNet	0.9395	0.3837
full	VGG16	0.9475	0.4461
full + sample	transfer learning [19]	0.9280	0.3339
full	**CGAN-Plankton**	**0.9425**	**0.4777**

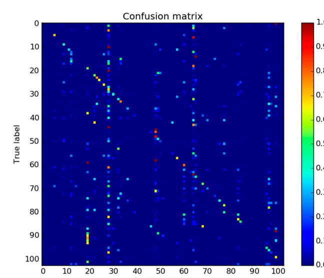


Fig. 3. Confusion matrix of CIFAR10 CNN model.

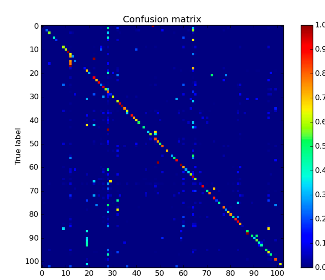
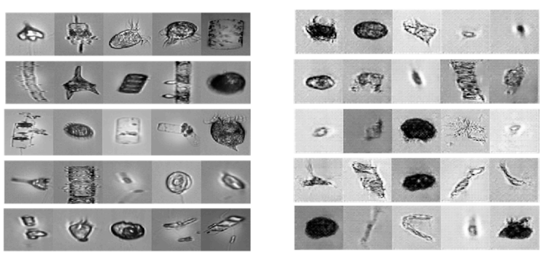


Fig. 4. Confusion matrix of CGAN-Plankton.

In order to check the details of the average accuracy especially the images from small classes, we want to investigate the contribution of generated samples. For specific classes, we train with each class and generate 200 samples of each. We add generated samples to original data and get an enhanced dataset. Training on the same model, we can get an improvement with these additional generated data respectively. Fig. 5 shows the generated images of several small classes. Samples generated by G model are all from a 100 dimensional vector randomly sampled from uniform distribution stored into 4 stride convolutional layers. We use batch normalization in both G model and D model. Experimental results show that our model works well. We can hardly distinguish the "fake" samples from all samples (Fig. 5 and 6). So our proposed model not only promises generating realistic samples but learns a strong discrimination of each class from mini-batch. Our presented samples vary in shape, size and rotation transformation. From Table 2, it can be seen that obvious improvement of F1 score on whole dataset after adding generated samples, and the corresponding confusion matrix is shown in Fig. 7. It is critically important that we get improvement through generated samples so it can become a promising and reliable method to solve general data lack problem. To evaluate our method performance, we use the generated samples on transfer learning method [19] and get a higher F1 score as well as a higher global accuracy than original data.



(a) WHOI-Plankton original data.
(b) WHOI-Plankton generated data.

Fig. 5. Comparison of WHOI-Plankton generated samples and original samples.

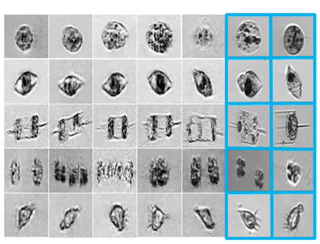


Fig. 6. Samples from G model. The rightmost 2 columns show the original data.

5. DISCUSSION AND CONCLUSION

There are many applications of GAN to generate new samples, however, it is still difficult for GAN to generate reasonable samples while the real images become complex. We

Table 2. Classification results after adding generated samples.

dataset	model	accuracy	F1 score
full	CIFAR10 CNN	0.9332	0.2583
full	transfer learning [19]	0.9344	0.3838
full	CGAN-Plankton	0.9443	0.4992

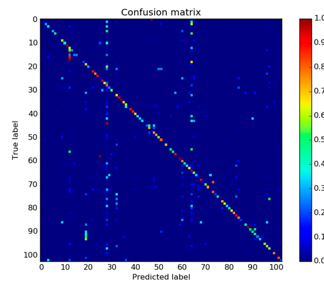


Fig. 7. Confusion matrix of CGAN after adding generated samples to whole data.

find that the generated results can be improved sincerely via a shared weights of D model as classification model. Until now there is no efficient way to evaluate the quality of generated samples from G model. Most of papers use cross entropy as well as human eyes. We show the generated samples in Fig. 5 and 6. To evaluate the efficiency of generated samples, we assemble all the 10 small classes generative images as training set and trained on CIFAR10 CNN model while training the original data as comparison. From Table 3, generated data get a comparative accuracy which means that samples are not only visually appealing to original data but have general features for training a classifier. In WHOI-Plankton 10 extremely small classes, training on the real-like generated data is able to distinguish samples in 60.17% of cases. And after add generated samples to origin dataset, in general, about 50% real and 50% artificial data CIFAR10 CNN classifier improve the whole accuracy and more significantly 7% improvement of F1 score.

Table 3. CIFAR10 CNN model classification results on WHOI-Plankton 10 extremely small classes.

dataset	model	accuracy	F1 score
original	CIFAR10 CNN	0.7109	0.6744
generated	CIFAR10 CNN	0.6017	0.4877
generated + original	CIFAR10 CNN	0.7374	0.7259

Considering the experiments above, we have validated the usefulness of CGAN-Plankton model's generated samples. For future work, we will combine GAN with RNN to generate more real images to overcome class imbalance and fine-grained classification problem.

6. REFERENCES

[1] Ursula Hofer, "Marine microbiology: Changing climate, changing plankton communities," *Nature Reviews Microbiology*, vol. 14, no. 1, pp. 2, 2016.

[2] Jules S. Jaffe, "Underwater optical imaging: the past, the present, and the prospects," *IEEE Journal of Oceanic Engineering*, vol. 40, no. 3, pp. 683–700, 2015.

[3] Norman MacLeod, Mark Benfield, and Phil Culverhouse, "Time to automate identification," *Nature*, vol. 467, no. 7312, pp. 154–155, 2010.

[4] Mark C. Benfield, Philippe Grosjean, Phil F. Culverhouse, Xabier Irigoien, Michael E. Sieracki, Angel Lopez-Urrutia, Hans G. Dam, Qiao Hu, Cabell S. Davis, Allen Hansen, Cynthia H. Pilskaln, Edward M. Riseman, Howard Schultz, Paul E. Utgoff, and Gabriel Gorsky, "RAPID: research on automated plankton identification," *Oceanography*, vol. 20, no. 2, pp. 172–187, 2007.

[5] Jeffrey S. Erickson, Nastaran Hashemi, James M. Sullivan, Alan D. Weidemann, and Frances S. Ligler, "*In situ* phytoplankton analysis: theres plenty of room at the bottom," *Analytical Chemistry*, vol. 84, no. 2, pp. 839–850, 2011.

[6] Bo Zhao, Jiashi Feng, Xiao Wu, and Shuicheng Yan, "A survey on deep learning-based fine-grained object classification and semantic segmentation," *International Journal of Automation and Computing*, pp. 1–17, 2017.

[7] 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," *arXiv preprint arXiv:1510.00745*, 2015.

[8] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "ImageNet classification with deep convolutional neural networks," in *NIPS*, 2012.

[9] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich, "Going deeper with convolutions," in *CVPR*, 2015.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *CVPR*, 2016.

[11] Ian Goodfellow, Jean Pougetabadie, Mehdi Mirza, Bing Xu, David Wardefarley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets," in *NIPS*, 2014.

[12] Foster Provost, "Machine learning from imbalanced data sets 101," in *AAAIW*, 2000, pp. 1–3.

[13] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[14] Chris Drummond and Robert C. Holte, "C4.5, class imbalance, and cost sensitivity: Why under-sampling beats oversampling," in *ICMLW*, 2003, pp. 1–8.

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

[16] Kai Ming Ting, "A comparative study of cost-sensitive boosting algorithms," in *ICML*, 2000.

[17] Zhi-Hua Zhou and Xu-Ying Liu, "Training cost-sensitive neural networks with methods addressing the class imbalance problem," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 1, pp. 63–77, 2006.

[18] Y. Tang, Y. Q. Zhang, N. V. Chawla, and S. Krasser, "SVMs modeling for highly imbalanced classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, no. 1, pp. 281–288, 2009.

[19] Hansang Lee, Minseok Park, and Junmo Kim, "Plankton classification on imbalanced large scale database via convolutional neural networks with transfer learning," in *ICIP*, 2016.

[20] Daniel Jiwoong Im, Chris Dongjoo Kim, Hui Jiang, and Roland Memisevic, "Generating images with recurrent adversarial networks," *arXiv preprint arXiv:1602.05110*, 2016.

[21] Emily Denton, Soumith Chintala, Arthur Szlam, and Rob Fergus, "Deep generative image models using a laplacian pyramid of adversarial networks," in *NIPS*, 2015.

[22] Alec Radford, Luke Metz, and Soumith Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

[23] Sergey Ioffe and Christian Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *ICML*, 2015.

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