ZooplanktoNet: Deep Convolutional Network for Zooplankton Classification

Jialun Dai, Ruchen Wang, Haiyong Zheng*, Guangrong Ji
College of Information Science and Engineering
Ocean University of China
Qingdao 266100, China
*Corresponding author: zhenghaiyong@ouc.edu.cn

Xiaoyan Qiao
College of Mathematic and Information Science
Shandong Institute of Business and Technology
Yantai 264005, China

***Abstract*—Zooplankton are quite significant to the ocean ecosystem for stabilizing balance of the ecosystem and keeping the earth running normally. Considering the significance of zooplantkon, research about zooplankton has caught more and more attentions. And zooplankton recognition has shown great potential for science studies and mearsuring applications. However, manual recognition on zooplankton is labour-intensive and time-consuming, and requires professional knowledge and experiences, which can not scale to large-scale studies. Deep learning approach has achieved remarkable performance in a number of object recognition benchmarks, often achieveing the current best performance on detection or classification tasks and the method demonstrates very promising and plausible results in many applications. In this paper, we explore a deep learning architecture: ZooplanktoNet to classify zoolankton automatically and effectively. The deep network is characterized by capturing more general and representative features than previous pre-defined feature extraction algorithms in challenging classification. Also, we incorporate some data augmentation to aim at reducing the overfitting for lacking of zooplankton images. And we decide the zooplankton class according to the highest score in the final predictions of ZooplanktoNet. Experimental results demonstrate that ZooplanktoNet can solve the problem effectively with accuracy of 93.7% in zooplankton classification.**

I. Introduction

Zooplankton are essential to the ocean ecosystem in which zooplankton function at many levels in oceans food web, as consumer, producer and prey to stabilize food chain. And they are also major contributors to elemental cycling and vertical fluxes. Zooplankton are sensitive and reactive to external perturbations and work as indicators of environmental change, such as global warming or a rapid increase in carbon dioxide in the atmosphere.

Loss of zooplankton population can have devastating impacts in our ecosystem, but the boom of zooplankton also can result in a huge disaster for the whole ecosystem. Thus, measuring and monitoring zooplankton are fundamental to the wellbeing of our society, which indicates the urgent need for the development of zooplankton recognition. The wide usage and great potential of zooplankton also have promoted more science studies and environmental researches aboout zooplankton recognition. However, traditional recognition of zooplankton always depend on some experts or some researchers with professional knowledge and experiences. Therefore, it is high-time to find an automatic method in zooplankton recognition with effectivity and efficiency.

Howerer, there are still many problems about zooplankton recognition to hamper its development. The challengings about zooplankton recognition mainly include the following parts: 1) the images of zooplankton are obscure for the low resolution and the object in image just looks like a dark spot, which makes it hard to identify. It will be mistaken as the wrong class easily by human, even worse by machine. 2) compared to dataset in other fields, the size of dataset is small. For such small dataset, it is difficult to train a deep model to work the problem out effectively. Even the numbers of each zooplankton class are quite different from each other in the dataset, where the unbalance may lead to some variances when training, and the accuracy may degrade consequently. 3) the simple information extracted may be not discriminative enough in zooplankton recognition. And some traditional feature exatracted by methods in computer vision, such as SIFT or LBP, may underperform in such difficult problem.

Neural network has experienced a resurgence thanks to breakthroughs in deep learning that has achieved remarkable performance in a number of object recognition benchmarks [1]. Such object recognition tasks where deep learning has achieved the best results include the MNIST hand-written digit dataset, CIFAR10 and the ImageNet Large-Scale Visual Recognition Challenge.

CNN (Convolutional Neural Network), one of the classical deep learning methods, has been inspired by proposed models of the human visual cortex and the concept of local receptive fields that take advantage of the topology. And CNN enforces a particular geometric knowledge by constructing an architecture that learns features based upon local region. Through local receptive fields, shared weights and subsampleing, CNN enables backward propagation (superviased learning) to train deep architectures [2].

In this paper, we explore a deep learning architecture based on Convolutional Neural Network: ZooplanktoNet to classify zooplankton effectively and automatically, and the framework is shown in Fig. 1. With the capacity of deep network, the framework can capture more general and representative features than previous pre-defined feature extraction algorithms to improve the accuracy. In order to achieve high

978-1-4673-9724-7/16/$31.00 ©2016 IEEE

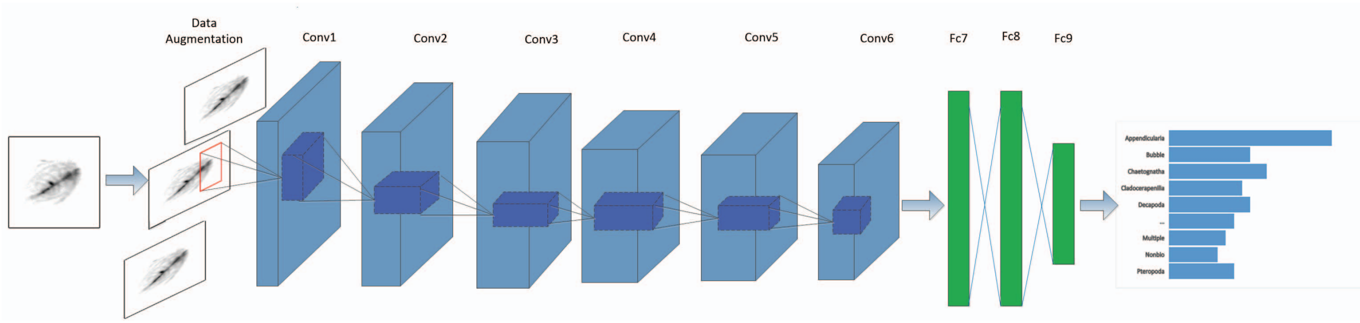


Fig. 1. The framework of our deep convolutional network for zooplankton classification: ZooplanktoNet.

accuracy in zooplankton classification, our work involves the following aspects. First, the deficiency of zooplankton images may cause heavy overfitting in such classification. To reduce the overfitting, we use several forms of data augmentation to aim at inflatting the dataset by different transformations, which is a common and effective to overcome the overfitting. Second, we also design some experiments to investigate some important factors in performances of different architectures for zooplankton classification. Thrid, compared with some deep models with time and space comsuption, a new deep learning architecture with less training time and paramters, ZooplanktoNet, is designed for zooplankton classification. It is strongly inspired by AlexNet [3] and VGGNet [4]. Experimental results demonstrate that the proposed architecture can work well on the zooplankton classification with accuracy over 93%.

The structure of our paper is organized as follows. Some related researches and technologies about zooplankton are introduced in Section 2. The proposed approach is illustrated in Section 3, where we give a detail description about processing before training. In Section 4, we talk about the dataset and experimental setting. And experimental evalutaion of the proposed approach and experimental results are discussed in this part. The final conclusion is given in Section 5.

II. Related Work

There are some efforts to devote to the study of zooplankton classification, whose most methods are about feature extraction and design of classifier. In the early years, Xiaoou Tang *etc.* [5] have done some works about automatic plankton classification, whose properties and problems are similar to the zooplankton classification. Their approach combines traditional invariant moment features and fourier boundary descriptors with grayscale morphological granulometries to form a feature vector, capturing both shape and texture information of plankton images.

The work of Philippe Grosjean [6] mostly focuses on image-processing on subsampling and they try several methods of machine learning in zooplankton classification, such as Support Vector Machine and K-nearest neighbours. In their experiments, Random Forest and Support Vector Machine work better than other algorithms in zooplankton classification. With a small number of simple features used in machine learning algorithms, the result of zooplankton classification may be not so satisfying.

For zooplankton classification, one representative work is done by Gaby Gorsk [7]. In their work, they present a semi-automatic approach that entails automated classification of images followed by manual validation, which allows rapid and accurate classification for zooplankton images. In the process, more than sixty features are extracted from zooplankton images and six classification algorithms with these features are compared to obtain one algorithm that performs best in zooplankton classification, the classifier of Random Forest. But their method requires too many simple features, such as the position of zooplankton in image and grayscale of zooplankton. However, these features are not discriminative as other human-designed features. Therefore, it still requires a more accurate method to work zooplankton classification out.

A similar work is done by Jeffrey Ellen [8]. They have experimented several different methods on the zooplankton images and find out a method with best performance among the results. The GBC and SVM_RBF are two best single performing algorithms on zooplankton dataset respectively, and the method that combines GBC with SVM is proved to achieve better performance for zooplankton classification. And it obtains a accuracy of 88.6% on 8 zooplankton classes, where per class includes 4000 training images and 1000 test images. But the dataset size in their experiment is quite larger than ours, which may improve accuracy consequently.

Another recent work about plankton classification is about deep learning based on Convolutional Neural Network[1]. The baisc idea is to investigate code vector (output of second last fully-connected layer) of CNN models to do model assembly. They introduce PCA, Ridge and Ramdonized Ridge model assembly methods and show that the randomized Ridge model outperforms the simple probability averaging method in the model assembly. But training is a multi-stage pipeline, not an end-to-end manner, and it is also expensive in time and space for computation storage.

III. Methods

At first, we give a brief introduction of CNN (Convolutional Neural Network) about how the network is effective in cap-

[1]http://vision.stanford.edu/teaching/cs231n/reports/ymkuang_project.pdf

turing representative features from low dimensionality to high dimensionality. Then we talk about some data augmentation and pre-processing in the dataset before training in order to overcome the deficiency of zooplankton images and reduce the overfitting during training.

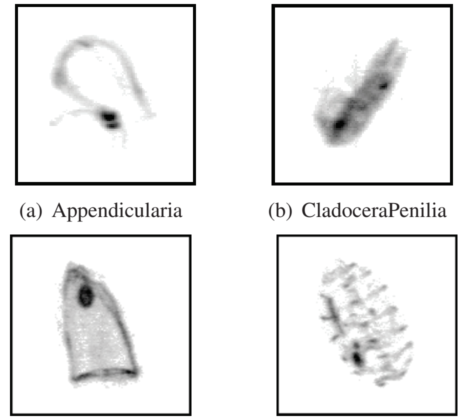
A. Convolutional Neural Network

To learn discriminative patterns from various and obscure zooplankton images, we need a model with powerful learning capacity. Recently, deep learning models have been used to extract multiple layers of features in a hierarchical structure. The constructed structure can capture abstract and translation invariant features, which demonstrates very promising and plausible results in many applications [3], [9]–[12]. Deep learning methods are good at characterizing the high-level abstractions of visual data by using a deep architecture, which is composed of multiple non-linear transformations. Among them, CNN (Convolutional Neural Network) has demonstrated an excellent ability in capturing image characteristics, which can be classified as a class of biologically-inspired, multi-layered neural network. It is inspired by biological processes and multi-layer perceptrons. The learned individual neurons in the feature map respond to overlapping regions in the visual field [13]. One major advantage of convolutional networks is the use of shared weight in convolutional layers, which both reduces required memory size and improves performance. And compared to other image classification algorithms, CNN uses relatively little pre-processing. The lack of dependence on prior-knowledge is another major advantage for CNN.

B. Data Augmentation and Pre-processing

1) Dataset: The dataset consists of microscopic and grayscale zooplankton images captured by ZooScan system [7], in which ZooProcess and Plankton Identifier software is an integrated analysis for acquisition and classification of digital zooplankton images from preserved zooplankton samples. The zooplankton dataset involves 13 classes with 9460 images. Some zooplankton images are shown in Fig. 2. And the following pre-processing and data augmentation are used to overcome this problem caused by small size of dataset and poor quality of images. After data augmentation, we divide these images into two part: training set and validation set. The proportion of images in training and validation is set as 4:1 in the experiments. The unbalance of image number in each class in dataset may produce some errors during training. However, the unbalance in dataset meets the zooplanktonic class distribution in realistic environment so that the meansured accuracy is meaningful.

2) Pre-processing: We performed very little pre-processing, other than two simple steps: rescaling the images and subtracting the mean value over the training set from each pixel. Rescaling the images is necessary because these images vary in size a lot: the smallest ones are less than 40×40 pixels, whereas the largest ones are up to 400×400 pixels. We experiment with various rescaling ways and decide to work with 256×256 pixels image with maintaining aspect ratio. We also do some experiments about rescaling with forgoing aspect radio, whose result is not as good as the ones with maintaining aspect ratio. The method of subtracting the mean value in dataset is widely used for normalization to guarantee training set and test set to be comparable.



(a) Appendicularia (b) CladoceraPenilia (c) Copepoda (d) Pteropoda

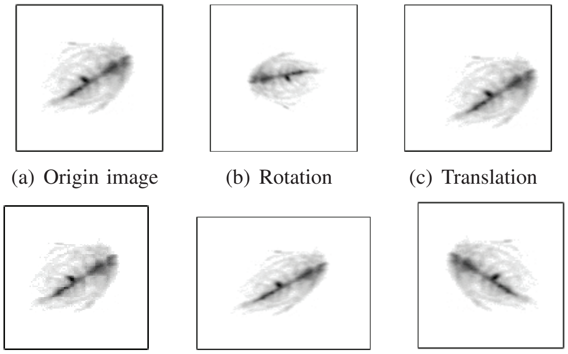
Fig. 2. The zooplankton images in the dataset. Labels under images are the classes that the zooplankton belong to.

3) Data Augmentation: The size of our zooplankton dataset is smaller than other big dataset used for deep learning, and it is possible to train a deep model that does not generalise well to result in overfitting. We augment the data artificially to increase the size of dataset, which is an effective way to overcome the problem of overfitting in training process. For zooplankton images, rotation invariance and translation invariance are very important properties. So transforms of rotation and translation are used for data augmentation and some other transforms are also added to make ZooplanktoNet generalise better. We end up with five augmentation methods: 1) rotation; 2) translation; 3) rescaling; 4) shearing; 5) flipping. The origin image and some images after augmentation are shown in Fig. 3. And data augmentation is employed before training and this improves the accuracy significantly.

- **rotation**: random with angle from $0°$ to $360°$ with $90°$ increments
- **translation**: random with shift -40 or 40 piexls in horizontal and vertical direction
- **rescaling**: random with scale factor between $1/1.2$ and 1.2
- **shearing**: random with angle between $-20°$ and $20°$
- **flipping**: random in horizontal direction or vertical direction

IV. Experiments

We experiment several popular classification models in zooplankton dataset at first and then our experiments have investigated the performance of several important factors in network to observe respective result. And accuracy, training loss, validation loss and complexity (depth and width of the network architecture) are all considered to evaluate performance in our experiments.



(a) Origin image (b) Rotation (c) Translation

(d) Rescaling (e) Shearing (f) Flipping

Fig. 3. The origin image and the images after augmentation.

A. performances of state-of-art architectures on zooplankton dataset without/with data augmentation

Some famous architectures that have achieved remarkable performances in ImageNet, such as AlexNet [3], CaffeNet [14], VGGNet [4] and GoogleNet [15], are implemented in the zooplankton dataset. TABLE I shows the performances of these architectures on zooplankton dataset without data augmentation and the results are similar. It is possible that though some architectures are more capable in common classification, it can not generalise and achieve better performance due to the small size of dataset and poor quality of images. To obtain higher accuracy and overcome overfitting, data augmentation is incorporated before training and we train these architectures again based on data augmentation. Details of the performances are shown in TABLE II. The loss of training and validation degrade obviously, which proves the effectivity of data augmentation.

In order to find a well-performance architecture on zooplankton dataset, we start with AlexNet to experiment with different factors in network, including layers of network, size or number of convolution, Local Response Normalization and ReLU. Among these factors, layers of network and size or number of convolution are regarded as the most significant factors.

B. layers of network

Recent researches reveal that depth, layers of network, is of crucial importance, and the leading results on the ImageNet dataset all exploit very deep models in recent years. Therefore, we regard the depth as the most important factor in our network and change from 8 layers to 16 layers to observe the final predictions. According to the conclusion in VGGNet, a stack of small convolutional layers that replace a single big convolutional layer can reduce memory comsuption and computation time, but also make the depth of network increased. This strategy is also used here. But the accruacy decreases and it consumes more time in training when depth keeps increasing. The reason that causes this situation is that the notorious problem of vanishing or exploding gradients [2], [16], which hamper convergence during training. TABLE III describes 5 configurations to investigate effects of depth in network. The results are shown in TABLE IV. And LRN is incorporated in the two convolutional layers from the beginning in all network.

TABLE VI
COMPARISON ON PERFORMANCES WITH AND WITHOUT LRN

Architecture	Accuracy	Time
C (LRN)	93.6%	39 min
C (no LRN)	93.2%	41 min

C. size and numbers of convolution

Except for the depth of network, the width of network which means the size and number of convolution in network is the other important factor to influence on the predictions. We have tried with different sizes of convolution, 13×13, 11×11, 7×7 or 5×5 in the first three convolutional layers to search size of convolution to work on the zoopankton datatset best. And different numbers of convolution, 256, 384 and 512, are also taken into account in the last several layers. TABLE V shows the results of different sizes and numbers of convolution. Normally, 11×11 or 7×7 can work well on such classification task. However, in zooplankton classification, 13×13 convolution in the first convolutional layer and 7×7 convolution in the second convolutional layer can achieve better performance, which may cause some computation comsuption and increase complexity of network. It is possible that zooplankton object occupies in almost whole image. Similarly, more numbers of convolution may capture features from more dimensions so that 384 and 512 convolution can achieve high performance on zooplankton dataset.

D. Local Response Normalisation

The research [4] shows Local Response Normalisation does not improve the performance, but leads to increased memory consumption and computation time. To prove whether it works on zooplankton dataset, we remove LRN in C configuration of 11 layers in TABLE V and the result is shown in TABLE VI. Under the condition that LRN is not provided in network, the accuacy decreases and training time increases a little. Thus, we still employ LRN in ZooplanktoNet.

E. ReLU

ReLU (Rectified Linear Units) is significant for recent success of deep networks. It expedites convergence of the training procedure and leads to better solutions than conventional sigmoid-like units. Some researches have proposed Learky PeLU [17] or Parametric ReLU [18] with focusing on the properties of the rectifiers to obtain better result. PReLU, which learns the parameters of the rectifiers adaptively and improves accuracy at negligible extra computational cost, is utilized in zooplankton dataset to compare with normal ReLU. The result is based on the above C configuration of 11 layers with LRN in TABLE V and shown in TABLE VII. Though PReLU just has little impact on accuracy and time, it still works in ZooplanktoNet.

TABLE I

COMPARISON OF POPULAR MODELS ON ZOOPLANKTON IMAGES WITHOUT DATA AUGMENTATION.

Models	Accuracy	Loss(train)	Loss(val)	Layers	Complexity	Time
AlexNet	82.1%	0.4273	0.5385	5 Conv + 3 FC	1.0	8 min
CaffeNet	81.6%	0.3422	0.5496	5 Conv + 3 FC	1.0	8 min
VGGNet	81.9%	0.3921	0.5175	13 Conv + 3 FC	1.5	14 min
GoogleNet	82.1%	0.4618	0.5356	-	1.7	15 min

TABLE II

COMPARISON OF POPULAR MODELS ON ZOOPLANKTON IMAGES WITH DATA AUGMENTATION.

Models	Accuracy	Loss(train)	Loss(val)	Layers	Complexity	Time
AlexNet	91.3%	0.0122	0.2263	5 Conv + 3 FC	1.0	28 min
CaffeNet	91.0%	0.0775	0.2435	5 Conv + 3 FC	1.0	28 min
VGGNet	92.2%	0.0144	0.2011	13 Conv + 3 FC	1.5	54 min
GoogleNet	91.9%	0.0041	0.2454	-	1.7	57 min

TABLE III

FIVE CONFIGURATIONS OF ARCHITECTURE DEPTH ON ZOOPLANKTON IMAGES.

<table>
<tr><th>8 layers(AlexNet)</th><th>9 layers</th><th>11 layers</th><th>13 laeyrs</th><th>16 layers</th></tr>
<tr><td colspan="5">input (227 $\times$ 227 image crop)</td></tr>
<tr><td>conv11-96</td><td>conv11-96</td><td>conv11-96</td><td>conv11-96</td><td>conv7-96
conv7-96</td></tr>
<tr><td colspan="5">3×3 maxpool</td></tr>
<tr><td>conv5-256</td><td>conv5-256</td><td>conv5-256</td><td>conv5-256</td><td>conv3-256
conv3-256</td></tr>
<tr><td colspan="5">3×3 maxpool</td></tr>
<tr><td>conv3-384
conv3-384
conv3-256</td><td>conv3-384
conv3-384</td><td>conv3-256
conv3-256
conv3-256</td><td>conv3-256
conv3-256</td><td>conv3-256
conv3-256
conv1-256</td></tr>
<tr><td colspan="5">3×3 maxpool</td></tr>
<tr><td rowspan="2"></td><td>conv3-384
conv3-384</td><td>conv3-384
conv3-384
conv3-384</td><td>conv3-384
conv3-384
conv1-384</td><td>conv3-384
conv3-384
conv1-384</td></tr>
<tr><td colspan="4">3×3 maxpool</td></tr>
<tr><td rowspan="2"></td><td rowspan="2"></td><td rowspan="2"></td><td>conv3-384
conv3-384
conv1-384</td><td>conv3-384
conv3-384
conv1-384</td></tr>
<tr><td colspan="2">3×3 maxpool</td></tr>
<tr><td colspan="5">FC-4096</td></tr>
<tr><td colspan="5">FC-4096</td></tr>
<tr><td colspan="5">FC-13</td></tr>
<tr><td colspan="5">softmax layer</td></tr>
</table>

TABLE IV

RESULTS ON ARCHITECUTURES OF DIFFERENT LAYERS.

Depth	Accuracy	Loss(train)	Loss(val)	Complexity	Time
8 layers	91.3%	0.0122	0.2263	1.0	28 min
9 layers	92.1%	0.0201	0.2255	1.1	29 min
11 layers	92.8%	0.0401	0.2722	1.3	34 min
13 layers	92.5%	0.0415	0.2612	1.5	44 min
16 layers	92.3%	0.0455	0.2627	1.7	53 min

F. ZooplanktoNet

To conclude all above experiments, we propose ZooplanktoNet with 11 layers to achieve the best performance, whose final accuracy is about 93.7%, compared to some popular architectures or other configurations on zooplankton dataset. Accuracy, loss value, traininig time and complexity of architecture are all considered for experimental evaluation. And the details of proposed framework are shown in Fig. 1.

V. CONCLUSION

In this paper, we propose a new network architecture, ZooplanktoNet, for zooplankton classification. Data augmentation is employed to overcome the overfitting and increase size

TABLE V
RESULTS ON ARCHITECUTURES OF DIFFERENT SIZES AND NUMBERS OF CONVOLUTION.

	11 layers	A	B	C	D
Configuration	conv11-96	conv13-96	conv13-96	conv13-96	conv13-96
	conv5-256	conv5-256	conv7-256	conv7-256	conv7-256
	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256	conv3-384 conv3-384 conv3-384	conv3-512 conv3-512 conv3-512
	conv3-384 conv3-384 conv3-384	conv3-384 conv3-384 conv3-384	conv3-384 conv3-384 conv3-384	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512
Accuracy	92.8%	92.9%	93.1%	93.6%	93.6%
Time	34 min	35 min	36 min	39 min	43 min

TABLE VII
COMPARISON ON PERFORMANCES OF PRELU AND RELU IN NETWORK

Architecture	Accuracy	Time
C (ReLU)	93.6%	39 min
C (PReLU)	93.7%	38 min

of dataset to improve accuracy. Some typical and popular deep learning architectures are introduced to zooplankton classification, and we investigate some important factors in network, such as depth and width, to prove that ZooplanktoNet can perform better when accuracy, loss value, training time and model complexity are taken into cosideration. And Experimental results demonstrate that ZooplanktoNet is capable to achieve high performance for zooplankton classification.

ACKNOWLEDGMENT

The authors wish to thank Laboratoire d'Océanologie de Villefranche-sur-Mer for providing zooplankton images collected by ZooScan system. This work was supported by the National Natural Science Foundation of China under Grant Nos. 61271406, 61301240, 61401255, and the Fundamental Research Funds for the Central Universities under Grant No. 201562023.

REFERENCES

[1] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[2] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[4] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[5] X. Tang, W. K. Stewart, L. Vincent, H. Huang, M. Marra, S. M. Gallager, and C. S. Davis, "Automatic plankton image recognition," *Artificial Intelligence Review*, vol. 12, pp. 177–199, 1998.

[6] P. Grosjean, M. Picheral, C. Warembourg, and G. Gorsky, "Enumeration, measurement, and identification of net zooplankton samples using the zooscan digital imaging system," *ICES Journal of Marine Science*, vol. 61, no. 4, pp. 518–525, 2004.

[7] G. Gorsky, M. D. Ohman, M. Picheral, S. Gasparini, L. Stemmann, J.-B. Romagnan, A. Cawood, S. Pesant, C. García-Comas, and F. Prejger, "Digital zooplankton image analysis using the zooscan integrated system," *Journal of Plankton Research*, vol. 32, no. 3, pp. 285–303, 2010.

[8] J. Ellen, H. Li, and M. D. Ohman, "Quantifying california current plankton samples with efficient machine learning techniques," in *Proceedings of OCEANS'15 MTS/IEEE Washington*, 2015.

[9] Y. Bengio, P. Lamblin, D. Popovici, H. Larochelle *et al.*, "Greedy layer-wise training of deep networks," in *Advances in Neural Information Processing Systems*, 2007, pp. 153–160.

[10] K. Jarrett, K. Kavukcuoglu, M. Ranzato, and Y. LeCun, "What is the best multi-stage architecture for object recognition?" in *Proceedings of the IEEE International Conference on Computer Vision*, 2009, pp. 2146–2153.

[11] Y. Zhou, H. Chang, K. Barner, P. Spellman, and B. Parvin, "Classification of histology sections via multispectral convolutional sparse coding," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3081–3088.

[12] H. Chang, Y. Zhou, P. Spellman, and B. Parvin, "Stacked predictive sparse coding for classification of distinct regions in tumor histopathology," in *Proceedings of the International Conference on Computer Vision*, 2013, pp. 169–176.

[13] X. Wang, R. Guo, and C. Kambhamettu, "Deeply-learned feature for age estimation," in *Proceedings of the IEEE Conference on Applications of Computer Vision*, 2015, pp. 534–541.

[14] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the ACM International Conference on Multimedia*, 2014, pp. 675–678.

[15] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.

[16] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Conference on Artificial Intelligence and Statistics*, 2010, pp. 249–256.

[17] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the International Conference on Computer Vision*, 2015, pp. 1026–1034.