Transfer Learning and Deep Feature Extraction for Planktonic Image Data Sets

Eric C. Orenstein
Scripps Institution of Oceanography - University of California San Diego
e1orenst@ucsd.edu

Oscar Beijbom
Department of Electrical Engineering and Computer Sciences - University of California Berkeley
obeijbom@eecs.berkeley.edu

Abstract

Studying marine plankton is critical to assessing the health of the world's oceans. To sample these important populations, oceanographers are increasingly using specially engineered in situ *digital imaging systems that produce very large data sets. Most automated annotation efforts have considered data from individual systems in isolation. This is predicated on the assumption that the images from each system are so different that there would be little benefit to considering out-of-domain data. Meanwhile, in the computer vision community, much effort has been dedicated to understanding how using out-of-domain images can improve the performance of machine classifiers. In this paper, we leverage these advances to evaluate how well weights transfer between Convolutional Neural Networks (CNNs) trained on data from two radically different plankton imaging systems. We also examine the utility of CNNs as feature extractors on a third unique plankton data set. Our results indicate that these data sets are perhaps more similar in the eyes of a machine classifier than previously assumed. Further, these tests underscore the value of using the rich feature representations learned by CNNs to classify data in vastly different domains.*

1. Introduction

Marine plankton, the microscopic organisms that populate the worlds oceans, are of critical importance to the eco- and climate systems they inhabit. These tiny creatures form the basis of the food web, link the atmosphere to the deep ocean, and regulate global-scale biogeochemical cycles [1, 14]. Studying and monitoring these important organisms is a challenge due to their microscopic size and the dynamic nature of the ocean [2].

Oceanographers and ecologists are increasingly eschewing established sampling techniques in favor of digital imaging technologies. *In situ* microscopes, in particular, allow scientists to observe undisturbed, natural populations of plankton. Many such systems have been developed and deployed, collectively capturing billions of images [2, 5, 17, 20]. The major impediment to studying these data sets lies in the so-called annotation bottleneck: expert annotation of plankton images is expensive, time consuming, and error prone [7].

Many groups have explored automated methods drawn from the computer vision literature to ameliorate the human cost of exploring their data. Most published techniques utilize a similar workflow: region of interest identification followed by hand-engineered feature extraction. Classification is often done with Support Vector Machines or Random Forests [4, 10, 12, 23].

Deep learning has been slow to appear in the plankton imaging literature. The first documented application was the winning entry to Kaggle's National Data Science Bowl (NDSB) competition to sort data from the In Situ Icthyoplankton Imaging System (ISIIS). The victorious team used a deep architecture to achieve 81.5% accuracy across 121 classes. As per the NDSB contest rules, entrants were not allowed to use data from outside sources. This stipulation ruled out any sort of fine-tuning and led many teams to use extensive data augmentation to mitigate over-fitting.

The benefits of using out-of-domain data to train deep classifiers are well established in the computer vision literature. Improvements have been recognized over baseline accuracies when using out-of-domain data to initialize deep network weights [25]. Likewise, using deep activations as features to train ensemble or margin classifiers in a new domain has improved upon scores achieved with hand-engineered features [9, 21].

In this paper, we build on the insights from the NDSB and investigate how deep nets perform on several data sets of *in situ* plankton images. Given the large number of different imaging systems currently in use, the relative paucity

978-1-5090-4822-9/17 $31.00 © 2017 IEEE
DOI 10.1109/WACV.2017.125

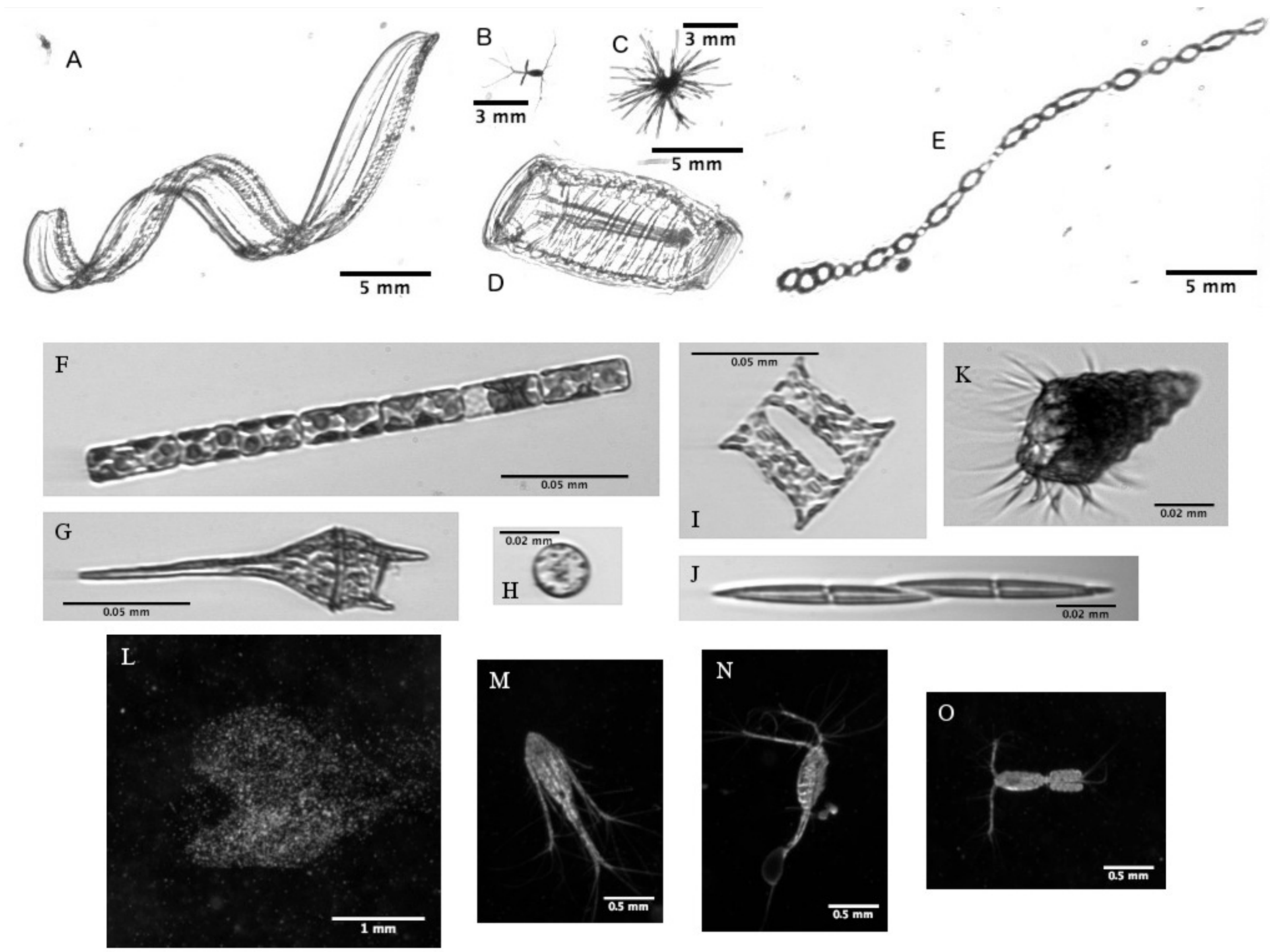

Figure 1. Example Regions of Interest for each system. *A - E* are from ISIIS, *F-J* were captured by the IFCB, and *L-O* are images from the SPC. Note the difference in scale and background color between the three systems.

of labeled data from each system, and the high likelihood of new systems coming online in the future, we focus on evaluating the effects of out-of-domain data on classification accuracy. Successfully implemented, such methods might allow previous annotation efforts to be leveraged towards labeling data from new imaging systems and deployment scenarios.

The following experiments consider labeled data from three different domains: the ISIIS data set provided for the NDSB[1], the extensively annotated Imaging Flow Cytobot (IFCB) image library[2], and the Scripps Plankton Camera System (SPC)[3]. The ISIIS and IFCB data sets are used to train a variety of deep networks, both by fine-tuning and from scratch. The classifiers are then used to classify within domain images directly and as feature extractors for out-of-domain data. For comparison, a suite of hand-engineered features are computed from the images and used to train Random Forests.

This work addresses three core issues: (1) Training deep architectures to two very different plankton data sets, IFCB and ISIIS. (2) Performing fine-tuning experiments to evaluate the similarity of these data sets. (3) Testing the utility of deep networks, trained on both natural and plankton images, as feature extractors on all three plankton data sets.

1.1. Related Work

The computer vision community's work on domain adaptation and transfer learning guided the design of the following experiments. Domain adaptation is the problem of training and testing a machine classifier on data drawn from different underlying class distributions [8]. Transfer learning addresses scenarios where the source and target domains are completely different [19]. Deep learning methods

[1] data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.nodc:0127422
[2] darchive.mblwhoilibrary.org/handle/1912/7341
[3] spc.ucsd.edu

have substantially outperformed other learning frameworks in both types of situations [3, 9, 11].

Automated classification of plankton images has not, to our knowledge, been treated with attention to these issues. Both are of concern in this realm since plankton populations are known to change dramatically on short time- and spatial-scales [13]. The experiments presented here focus on transfer learning, rather than domain adaptation, to highlight the utility of out-of-domain data in classifying plankton.

Deep learning methods also provide a way of estimating the distance between data sets. Yosinski et al. [25] demonstrated that when a source and target transfer learning task are very different, the accuracy of a fine-tuned network declines. Based on this interpretation of fine-tuning performance, we attempt to establish how similar the plankton data sets are to one another.

2. Data sets

Unlike cameras used to capture natural images, plankton-imaging systems are specifically engineered to sample a particular population. The three instruments considered in this paper were all designed to observe different segments of the planktonic kingdom in disparate regions of the ocean (Figure 1). They all use different optical schemata and are deployed under different protocols. Numbers of classes and labeled samples for each data set considered are summarized in Table 1.

2.1. In Situ Icthyoplankton Imaging System (ISIIS)

ISIIS was developed to study organisms in the mm to cm range. The system uses a shadowgraph illumination set-up with a line scan sensor. ISIIS is towed by a research vessel along a transect line while imaging continuously, effective sampling $\sim$70 L/sec. A single 8 hour deployment can generate $\sim$500,000 images.

The labeled data used in the NDSB is a subset drawn from $\sim$50 million images captured during an 18 day cruise in the Straits of Florida. The training set provided for the NDSB contains 30,336 expert annotated ROIs distributed among 121 classes (see Figure 1 *A-E* for examples). The test set has 30,400 samples taken from images captured on the same cruise. The labels are generally taxonomic, but also include semantic descriptors. There are also several noise categories that contain images of detritus that do not have a taxonomic identification.

2.2. Imaging Flow Cytobot (IFCB)

The IFCB is an *in situ* flow cytometer designed for high temporal resolution, long duration deployments. The data considered here is drawn from an instrument has been continuously operated at the Martha's Vineyard Coastal Observatory (MVCO) since 2006 [17, 23]. To date, the IFCB at MVCO has logged nearly one billion ROIs of organisms in the 10 - 100 μm size range. The current labeled image set contains $\sim$3.5 million samples, separated into 102 classes (examples in Figure 1 *F-J*). Again, the labels are generally taxonomic with several semantic descriptors and noise categories. There are 100s to 100,000s of samples per class ranging in size from 100s to 10,000s of pixels per ROI.

For the purposes of these experiments, the labeled data set was split into training and test sets by year as suggested in [18]. All annotations from 2006-2013 were treated as training data and all labeled samples from 2014 were used as test data. The split was chosen to simulate a deployment scenario: training a classifier on historic data and applying it to new data. It is important to note that there is a high degree of variability in the class distribution from year-to-year.

2.3. Scripps Plankton Camera System (SPC)

The SPC consists of a pair of *in situ*, dark field microscopes designed to image objects that range in size from 10's of μm to cm [22]. The system employs no nets, filters, or pumps; the cameras see only what floats through its field of view. Annotation of the SPC data is in its early stages. The current work makes use of a 4-class labeled image set subsampled from data collected by the large volume instrument moored at Scripps Pier. The ROIs were annotated in an effort to study a specific type of zooplankton called copepods. *Oithona similis* is found in all ocean basins and has begun showing up in the SPC image data with a parasite.

The ROIs were labeled from randomly selected days from between March and September of 2015. Each ROI in each day of data was then annotated as other, Oithona, Oithona with parasite, or Oithona with eggs (Figure 1 *L-0*). Examples of Oithona, including those with eggs and parasites, make up only $\sim 2\%$ of the ROIs examined in each day.

3. Methods

Images from the ISIIS and IFCB data sets were used to train several CNNs using the AlexNet architecture in Caffe [15]. Data from the SPC were only used in ensemble learning tests due to the small number of samples. The CNNs were trained either from scratch on only plankton data or

	# unaug. train	# aug. train	# test	# classes
ISIIS	30336	37000	30400	37
IFCB	53239	95000	65913	95
SPC	3200	-	800	4

Table 1. Data set information. Note that SPC data was not augmented since it was not used to train deep models.

		Training Method						
		ifcb	*ImageNet⇒ifcb*	*ImageNet⇒ isiis⇒ifcb*	*isiis*	*ImageNet⇒isiis*	*ImageNet⇒ ifcb⇒isiis*	*ImageNet*
CNNs	IFCB	0.78	**0.86**	**0.86**	-	-	-	-
	ISIIS	-	-	-	0.71	**0.83**	**0.83**	-
Deep Feat	IFCB	-	-	-	0.65	0.77	-	**0.81**
	ISIIS	0.56	**0.65**	-	-	-	-	0.63
	SPC	0.57	0.69	0.67	0.52	0.65	0.65	**0.71**
Deep + Hand	IFCB	-	-	-	0.68	0.78	-	**0.81**
	ISIIS	0.65	**0.70**	-	-	-	-	0.66
	SPC	0.67	0.76	0.74	0.69	0.74	0.74	**0.77**

Table 2. Top-1 classification accuracies on plankton data sets using CNNs, based on the AlexNet architecture, applied directly and as feature extractors. Columns display the training method and rows indicate the target data set.

fine-tuned from AlexNet, trained on ImageNet, weights provided by the BLVC. AlexNet was chosen for these experiments in the interest of extensibility and training expediency.

Each network was then applied directly to the independent test set from the training domain and used as a feature extractor for the two image sets not seen during training. The deep features from each network were used to train Random Forest (RF) classifiers. For reference, RFs were also trained with a suite of hand-engineered features.

3.1. Data preprocessing

Several classes from both the ISIIS and IFCB data sets were combined. The ISIIS label set contained many orientational subclasses. These labels were combined to focus on the taxonomic designations as indicated in [6]. Likewise, several IFCB classes were merged to avoid semantic subgroups of the same taxonomic designations. Three IFCB classes were ignored all together as recommended in [24]. The resulting ISIIS and IFCB label sets had 37 and 95 classes respectively.

All images from ISIIS and IFCB were resized to 256 x 256 using a perspective preserving transform along the longest axis of the ROIs. Rescaling is necessary because the ROIs vary greatly in dimension and simple resizing significantly distorts the samples. The background of each image was filled with the median pixel value of the entire image set in each respective domain.

For the purposes of this work, the class distribution of the training sets were made even by requiring each class to contain 1000 samples. Classes with many ROIs were randomly subsampled. For categories with fewer than 1000 samples, data augmentation was performed using a random assortment of image transformations (rotation, perspective, translation, and shearing).

3.2. Convolutional Neural Networks (CNN)

CNNs were trained using the AlexNet architecture: 5 convolutional and 3 fully connected layers [16]. The Caffe reference AlexNet, trained on ImageNet, was used in lieu of larger networks to ensure portability and to expedite training and analysis [15]. 20% of each training set was split off for validation. Training was done in 40000 iterations with a learning rate of 0.0002 on an NVIDIA Tesla K40.

A total of six CNNs were trained using the IFCB and ISIIS data sets (Table 2). Each network is referred to according to the data it has seen. For example, the *isiis* and *ifcb* were trained from scratch in their eponymous domains. Fine-tuned versions of AlexNet are called *ImageNet⇒isiis* and *ImageNet⇒ifcb*. The two fine-tuned networks were then fine-tuned again using the training data of the other domain, denoted *ImageNet⇒isiis⇒ifcb* and *ImageNet⇒ifcb⇒isiis*.

All fine-tuning was done by randomly initializing the final convolutional layer and allowing the whole network to adapt. *ImageNet⇒isiis* and *ImageNet⇒ifcb* were trained starting from weights drawn from AlexNet trained on natural images. The double fine-tuned networks, *ImageNet⇒isiis⇒ifcb* and *ImageNet⇒ifcb⇒isiis*, retained the weights for the first seven layers of the networks previously trained on plankton images. The final layer was again initialized randomly and the network fine-tuned.

Each plankton network was applied directly to an independent test set drawn from the final image set used to train it. For example, *ImageNet⇒isiis⇒ifcb* was used to classify IFCB data. Each network was further used a feature extractor in domains it had not seen. *ImageNet⇒isiis*, for instance, drew features from IFCB and SPC test data.

An unfine-tuned version of the BVLC reference AlexNet was also used as a feature extractor in all three plankton domains. This network is referred to as *ImageNet* in keeping with the labeling convention of the fine-tuned classifiers.

This experiment was done to assess how well a generic classifier, trained on natural images, fares when applied to images of microscopic plankton.

3.3. Random Forests

Random Forests (RF) were trained for each domain with hand-engineered features, deep features, and a combination of hand and deep features. No data augmentation was done for the RF tests; only the original, unaugmented images from each domain were used. Each RF was trained with 500 trees. No further parameter optimization was performed, though this could have improved accuracies.

Unresized ROIs were put through a feature extraction routine that pulled 72 standard metrics including morphological indicators, texture descriptors, and gray level co-occurrence matrices. These features have been used for classification in the oceanographic literature and have been shown to be discriminative for smaller spaces of classes [4, 23]. The resulting feature vectors were used to train and test an RF in each domain.

RFs were also trained and evaluated using features extracted from the final layer of CNNs trained with out-of-domain data. The deep features were combined with hand-engineered features in a separate experiment.

4. Results

Top-1 accuracies from all experiments with CNNs are displayed in Table 2.

4.1. Deep networks

A total of six CNNs were trained using the ISIIS and IFCB data sets individually and in combinations. The networks trained from scratch on plankton data, *isiis* and *ifcb*, were outperformed by the fine-tuned networks. This performance improvement is consistent with previous studies of fine-tuning of CNNs trained on natural image data sets.

Extra fine-tuning with out-of-domain plankton data did not make a remarkable difference in classifier accuracy. The two double fine-tuned networks, *ImageNet⇒ifcb⇒isiis* and *ImageNet⇒isiis⇒ifcb*, were marginally more successful than the other tested CNNs. The accuracy gains, however, were less than a percent. Results are summarized in Table 2 under the **CNNs** heading.

4.2. Random Forests

4.2.1 Hand-engineered features

The 72, hand-engineered features were used to train RF classifiers for all three domains. The classifiers were then applied to the within domain test set. Performance scaled with the number of classes. The IFCB RF, with 95 classes, had the lowest average accuracy of 0.58. The RF trained to classify ISIIS had an accuracy of 0.62. The SPC RF, with 4 classes, had the highest accuracy of 0.69.

4.2.2 Deep features

Weights were pulled from the final fully connected layer of networks during a forward pass of out-of-domain images. There were then three feature sets each for IFCB and ISIIS; *ImageNet* and two plankton CNNs. All networks were used as feature extractors for SPC since none of the CNNs were trained with data from that domain. The results are displayed in Table 2 under **Deep Feat**.

For IFCB, features from *ImageNet* achieved the highest accuracy at 81%. The RFs trained on weights from *ImageNet⇒isiis* had better performance than the network trained on ISIIS data alone. RFs tuned on weights pulled from any network outperformed the RF trained only on IFCB hand-engineered features.

Weights pulled from the *ImageNet⇒isiis* yielded the highest scoring RF for ISIIS. Like the experiments done on IFCB, features from the *ifcb* network, which only saw plankton data, were the least discriminative. The RF trained on *ImageNet* weights underperformed *ImageNet⇒ifcb* by $\sim$2%

ImageNet features were the most discriminative for sorting SPC data. The RF trained on *ImageNet* weights beat the next highest score by 2%. All fine-tuned networks provided more salient features for classifying SPC data than either network trained exclusively on plankton data. *ImageNet* weights were also more discriminative than the RF trained exclusively on hand-engineered features pulled from SPC images.

4.2.3 Deep with hand-engineered features

Combining deep and hand-engineered features boosted the accuracies of all the ensemble classifiers. The amount varied by data set and the deep method used to extract features (Table 2 under **Deep+Hand**).

Classification accuracy on IFCB data was boosted an average of $\sim$2% when the 72 image descriptors were added to weights drawn from networks trained on ISIIS data, *isiis* and *ImageNet⇒isiis*. The RF trained on *ImageNet* features improved by less then 1% with the addition of the image metrics but remained the most accurate.

Each classifier applied to ISIIS became more accurate when hand-engineered measures were included in the feature vector. The scores increased by $\sim$5.5% for each test. The most discriminative classifier used weights from *ImageNet⇒ifcb*.

All classifiers used on SPC improved with the addition of hand-engineered features. The RF trained on ISIIS features improved by 17% when image metrics were included. All the SPC classifiers with deep and hand-engineered features were more accurate than the RF trained on hand-engineered features alone.

5. Discussion

In this paper, we have applied a variety of learning schemes to three very different plankton image data sets. The bigger labeled image data sets, IFCB and ISIIS, were used to train CNNs both from scratch and by fine-tuning. The resulting networks were applied to within domain images and as feature extractors on out-of-domain data.

In keeping with current computer vision literature, CNNs were more effective classifiers than RFs on both the IFCB and ISIIS data sets. The networks fine-tuned from *ImageNet* were the most accurate, beating CNNs trained exclusively on plankton data by about 10%. This implies that general to medium scale features from networks trained on natural images are salient for plankton classification.

Double fine-tuning with IFCB and ISIIS had a slight positive effect (less than 1%) on classifier accuracy in both target domains. Following the logic of Yosinki *et al*. [25], this result indicates that the CNN views the IFCB and ISIIS data sets as quite similar. In [25], the researchers demonstrated that the effectiveness of feature transfer is related to the similarity between the source and target data. When the two domains are very similar, feature transfer provides a strong accuracy boost. When the domains are extremely different, the network accuracy can actually decline as more layers are retained for fine-tuning.

Since the performance gains were modest for double fine-tuning with plankton data, we conclude that the IFCB and ISIIS sets occupy a middle ground. They are close enough that double fine-tuning does not hinder performance, nor are they so similar that it has a significant positive effect. This suggests that weights from a highly tuned network for one planktonic image set could be used effectively in another plankton domain. The proximity of the data sets from a machine learning stand point makes sense biologically: the organisms sampled by the two systems are from distinct, but similar, regions of the taxonomic tree.

The results from the ensemble classifier experiments are also in keeping with current literature: features drawn from CNNs are more salient than hand-engineered features [9, 21]. Perhaps the most interesting outcome of these tests is how well weights from *ImageNet* compared to the fine-tuned networks. On the IFCB and SPC data sets, *ImageNet* features yielded the best classifier both with and without standard image descriptors (Table 2).

The success of *ImageNet* is particularly curious for the SPC test domain. The ISIIS set had two classes, 'copepod calanoid' and 'copepod cycloid,' that are taxonomically and morphologically very similar to the organisms of interest in the SPC data (compare, for example, Fig. 1 *B* with 1 *M-O*). Yet features pulled from any CNN trained with ISIIS data were not as discriminative as those taken from *ImageNet*.

This behavior is likely due to the relatively small set of classes and number of images in the ISIIS and SPC data sets. Previous work has shown that weights from generic visual classifiers, such as *ImageNet*, perform well when used as features for sparsely labeled image sets [9]. The success of RFs trained on features drawn from *ImageNet⇒ifcb* on the ISIIS data set also demonstrates this behavior – the larger data set and label space of the IFCB imparts the CNN with a richer feature representation. Likewise, RFs trained on features from *ImageNet⇒ifcb* for SPC classification were within 1-2% of those trained from *ImageNet* features (See Appendix A for further discussion of these results using a probe into the average node activation of each network).

Future work will attempt to include more image data from plankton domains to improve classification accuracy. We believe an interesting avenue could be combining data from many plankton imaging systems to build a more generic classifier of microscopic images. Feature visualization techniques such as those described in [26, 27] could aid this effort by further illuminating the behavior observed in these experiments. Hopefully, these tools will guide the training of better, more efficient classifiers that make optimal use of all available labeled plankton data

It is important to note that the CNNs, feature extraction methods, and RFs used in this work could all be further optimized for any particular application. The goal with these experiments was to evaluate the utility of out-of-domain data for improving classification scores on plankton data sets. We believe that the results show that the use of out-of-domain data is effective and that future development of plankton classifiers should make use of pretrained CNNs.

Acknowledgments

The authors gratefully acknowledge the time and effort of the WACV reviewers for supplying valuable feedback on this work. EO would also like to thank Paul Roberts, Devin Ratelle, and Jules Jaffe for critical readings of the manuscript.

EO was funded under the National Science Foundation's BIGDATA Initiative (NSF IIS 15-46351). The NVIDIA Tesla K40 GPU used in these experiments was generously supplied by NVIDIA under their Academic Hardware Donation Program.

References

[1] K. R. Arrigo. Marine microorganisms and global nutrient cycles. *Nature*, 437(7057):349–355, 2005.

[2] M. C. Benfield, P. Grosjean, P. F. Culverhouse, X. Irigoien, M. E. Sieracki, A. Lopez-Urrutia, H. G. Dam, Q. Hu, C. S. Davis, and A. Hansen. RAPID: research on auto-

mated plankton identification. *Oceanography*, 20(2):172–187, 2007.

[3] Y. Bengio. Deep Learning of Representations for Unsupervised and Transfer Learning. *ICML Unsupervised and Transfer Learning*, 27:17–36, 2012.

[4] M. B. Blaschko, G. Holness, M. A. Mattar, D. Lisin, P. E. Utgoff, A. R. Hanson, H. Schultz, and E. M. Riseman. Automatic in situ identification of plankton. In *Application of Computer Vision, 2005. WACV/MOTIONS'05. Seventh IEEE Workshops on*, volume 1, pages 79–86. IEEE, 2005.

[5] R. K. Cowen and C. M. Guigand. In situ ichthyoplankton imaging system (ISIIS): system design and preliminary results. *Limnol. Oceanogr. Methods*, 6(2):126–132, Feb. 2008.

[6] R. K. Cowen, S. Sponaugle, K. L. Robinson, and J. Luo. PlanktonSet 1.0: Plankton imagery data collected from F.G. Walton Smith in Straits of Florida from 2014-06-03 to 2014-06-06 and used in the 2015 National Data Science Bowl (NCEI accession 0127422). *NOAA National Centers for Environmental Information*, 2015.

[7] P. F. Culverhouse, N. Macleod, R. Williams, M. C. Benfield, R. M. Lopes, and M. Picheral. An empirical assessment of the consistency of taxonomic identifications. *Marine Biology Research*, 10(1):73–84, Jan. 2014.

[8] H. Daume III and D. Marcu. Domain adaptation for statistical classifiers. *Journal of Artificial Intelligence Research*, 26:101–126, 2006.

[9] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell. DeCAF: A deep convolutional activation feature for generic visual recognition. In *Proceedings of the 31st International Conference on Machine Learning*, pages 647–665, 2014.

[10] J. Ellen, H. Li, and M. D. Ohman. Quantifying California current plankton samples with efficient machine learning techniques. In *OCEANS 2015 - MTS/IEEE Washington*, pages 1–9, Oct. 2015.

[11] X. Glorot, A. Bordes, and Y. Bengio. Domain adaptation for large-scale sentiment classification: A deep learning approach. In *Proceedings of the 28th International Conference on Machine Learning*, pages 513–520, 2011.

[12] 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: Journal du Conseil*, 61(4):518–525, 2004.

[13] L. Haury, J. A. McGowan, and P. H. Wiebe. Patterns and processes in the time-space scales of plankton distributions. In *Spatial pattern in plankton communities*, pages 277–327. 1978.

[14] G. C. Hays, A. J. Richardson, and C. Robinson. Climate change and marine plankton. *Trends in Ecology & Evolution*, 20(6):337–344, June 2005.

[15] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, 2014.

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

[17] R. J. Olson and H. M. Sosik. A submersible imaging-in-flow instrument to analyze nano-and microplankton: Imaging FlowCytobot. *Limnology and Oceanography-Methods*, 5:195–203, 2007.

[18] E. C. Orenstein, O. Beijbom, E. E. Peacock, and H. M. Sosik. WHOI-plankton-a large scale fine grained visual recognition benchmark dataset for plankton classification. *arXiv preprint arXiv:1510.00745*, 2015.

[19] S. J. Pan and Q. Yang. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, Oct. 2010.

[20] M. Picheral, L. Guidi, L. Stemmann, D. M. Karl, G. Iddaoud, and G. Gorsky. The Underwater Vision Profiler 5: An advanced instrument for high spatial resolution studies of particle size spectra and zooplankton. *Limnology and Oceanography-Methods*, 8:462–473, 2010.

[21] A. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson. CNN Features Off-the-Shelf: An Astounding Baseline for Recognition. In *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 512–519, June 2014.

[22] P. L. Roberts, J. S. Jaffe, E. C. Orenstein, B. Laxton, P. J. S. Franks, C. Briseno, M. L. Carter, and M. Hilbern. Pier Recognition: An In Situ Plankton Web Camera. In *Ocean Optics XXII*, Portland, ME, USA, Oct. 2014.

[23] H. M. Sosik and R. J. Olson. Automated taxonomic classification of phytoplankton sampled with imaging-in-flow cytometry. *Limnology and Oceanography-Methods*, 5:204–216, 2007.

[24] H. M. Sosik, E. E. Peacock, and E. F. Brownlee. Annotated plankton images - data set for developing and evaluating classification methods. *Imaging FlowCytobot at Woods Hole Oceanographic Institute*, 2016.

[25] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson. How transferable are features in deep neural networks? In *Advances in Neural Information Processing Systems 27*, pages 3320–3328. Curran Associates, Inc., 2014.

[26] J. Yosinski, J. Clune, A. Nguyen, T. Fuchs, and H. Lipson. Understanding neural networks through deep visualization. *arXiv preprint arXiv:1506.06579*, 2015.

[27] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. In *European Conference on Computer Vision*, pages 818–833. Springer, 2014.