Marine Biodiversity Classification using Dropout Regularization

A.M. Rahimi[1], R.J. Miller[2], D.V. Fedorov[1], S. Sunderrajan[1], B.M. Doheny[2], H.M. Page[2], and B.S. Manjunath[1]

[1] Department of Electrical and Computer Engineering, UCSB
{mohaymen,fedorov,santhosh,manj}@ece.ucsb.edu
[2] Marine Science Institute, UCSB
{miller,doheny,page}@msi.ucsb.edu

Abstract. Coastal marine ecosystems are highly productive and diverse, but biodiversity of underwater habitats is poorly described due to logistical and financial limitations of diving and submersible operations. Imagery is a promising way to address this challenge, but the complexity of diverse organisms thwarts simple automated analysis. We consider the problem of automated annotation of complex communities of sessile marine invertebrates and macroalgae in order to automate percent coverage estimation. We propose an efficient fusion technique amongst diverse classifiers based on the idea of "dropout" in machine learning. We use dropout technique to weight each classifier implicitly and for each specie we optimize the size of the region of interest (ROI) for highest accuracy. The preliminary results are promising and show 20% increase in average accuracy (over 30 species) when compared with the best base performance of Random Forest classifiers. The dataset along with human "ground truth" annotations are available to the public.

Keywords: Image Classification, Ensemble Methods, Underwater Imagery, ROI

1 Introduction

Evidence of the positive relationship of species diversity and ecosystem function in the marine environment is mounting [1]. Ecological mechanisms contributing to this effect include complementarity in resource use among species, positive inter-species interactions, and functional redundancy that provides biological insurance against changes in ecosystem function. Ensuring the continuity of marine ecosystem diversity and functioning requires information on the numbers of species and their abundance in marine habitats over large scales in time and space. Quantifying the diversity and abundance of organisms in sub-tidal marine ecosystems involves long and challenging hours of deep sea diving, counting and identifying plants and animals. Imagery has long been used by sub-tidal ecologists and deep-sea biologists to record biodiversity in marine habitats in an attempt to simplify and formalize the process.

In this study we examine a large image dataset collected in sub-tidal habitats of Santa Barbara, California. The objective of this study is to examine the distribution of an invasive species, the Bryozoan Watersipora Subtorquata, on offshore

978-0-7695-5318-4/14 $31.00 © 2014 IEEE
DOI 10.1109/CVAUI.2014.17

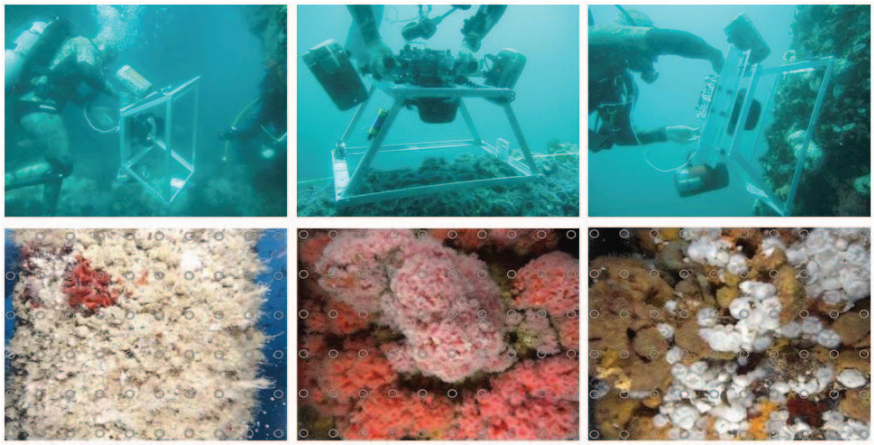

Fig. 1: The complexity of underwater image acquisition and the photographic setup are presented in the first row. The second row shows exemplar images with overlayed percent coverage annotation grid.

oil platforms and natural reefs, Fig 1. In addition to the abundance of the invader, the diversity and abundance of native species of sessile invertebrates and algae are also quantified to examine whether native communities may provide biotic resistance to the invasion. The proposed solution is to automate the common percent coverage technique using supervised machine learning and computer vision. We use manual percent cover annotations to train the automated classifiers. In order to robustly classify a wide range of species with different visual characteristics we use a set of diverse (14) computer vision feature descriptors. Our goal is to arrive at a consensual decision between predictors quickly and accurately. We introduce a new regularization technique for K-nearest neighbor (K-NN) that identifies a subset of more robust/reliable features for classification. Once the predictions are pruned, final decisions are made with a simple majority vote. The main contributions of this paper include:

- Implicit modeling of mutual dependencies among classifiers with "drop-out" regularization with K-NN.
- Fast and automatic classification with an optimized ROI.
- Introducing a new manually annotated high resolution underwater data set.

2 Related Work

Dropout techniques have recently generated much interest in the machine learning community as an alternative to the regularizers used in neural networks such as [2,3,4]. These are designed to achieve the effect of training a massive number of neural networks and then averaging over their decisions [4]. Dropout achieves this by training a single massive neural network for which sub-networks are used during training. Dropout has also been applied to K-NN classification [5]. The bagging technique [5] uses neural networks and genetic algorithms to train a boosted set of classifiers that together are more accurate than any single classifier. Each of these classifiers is defined over a distance metric. These metrics are

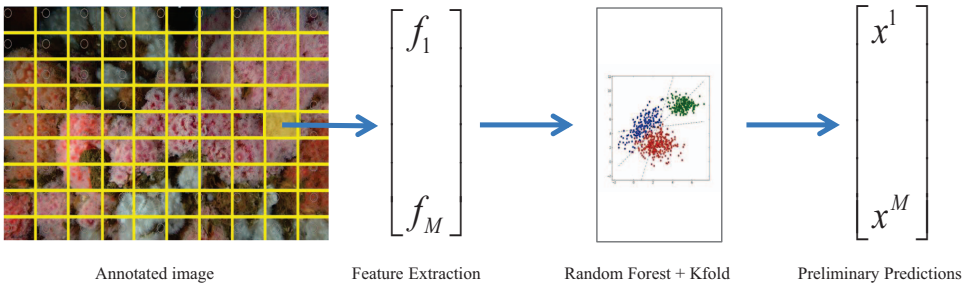

Fig. 2: Each classifier uses one unique feature descriptor independently and individual predictions are obtained with K-fold cross validation.

constructed by ignoring certain entries in the data vectors when selecting the neighbors.

There are multiple ways of parameterizing dropout. The simplest way is to ignore a random portion (say 50%) of the inputs on each metric. Other methods of regularizing K-NN have also been studied [6,7] to be highly robust in removing a portions of the data.

3 Base Level Feature Classifiers

The base level (weak) classifiers work on a diverse set of image derived features. We first partition each image s_i into J blocks so that each segment $s_{i,j}$ corresponds to the j-th patch of the i-th image. Each segment $s_{i,j}$ can only take one label from a label set L where $l_k \in \{l_1, ..., l_N\}$ and N indicates the total number of species including the empty class (water background). The total number of images in the data set is denoted by Q, hence a total of $Q \times J$ annotated segments in our data set. Therefore,

$$\forall s_{i,j} \;\; \exists! \;\; l_k \;\; \text{s.t. } i \in \{1, ..., Q\}, \;\; j \in \{1, ..., J\}, \;\; \text{and } k \in \{1, ..., N\}$$

We choose the random forest classifiers to construct the weak classifiers on each of the M computed features for each segment.

$$\boldsymbol{f}^m_{i,j} \in \{\boldsymbol{f}^1_{i,j}, ..., \boldsymbol{f}^M_{i,j}\} \;\; \text{where} \;\; \boldsymbol{f}^m_{i,j} \in \Re^{d_m} \tag{1}$$

Each weak classifier produces a regression vector $\boldsymbol{c}^m_{i,j}$ of size N indicating the likelihood of a given segment belonging to each class label.

$$\boldsymbol{c}^m_{i,j} = [c^{m,1}_{i,j}, ..., c^{m,k}_{i,j}], \quad \text{where} \quad c^{m,k}_{i,j} = P(l_k \mid m, i, j) \tag{2}$$

Using MAP inference with K-fold cross validation, we generate a vector $\boldsymbol{x}$ for each segment $s_{i,j}$.

$$x^m_{i,j} = \arg\max_k \{c^{m,k}_{i,j}\}, \quad \text{and} \quad \boldsymbol{x}_{i,j} = [x^1_{ij}, ..., x^M_{ij}] \tag{3}$$

These preliminary decision vectors are then aggregated to obtain the final prediction y_{ij}.

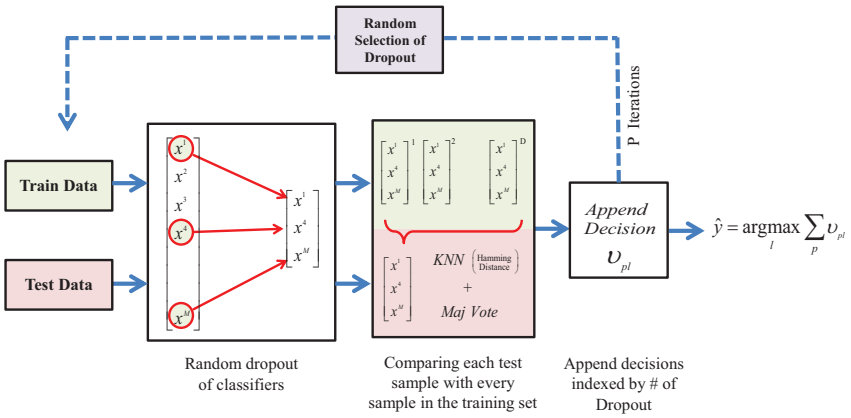

Fig. 3: Overview of the aggregation technique using dropouts. Classifiers are randomly dropped from both training and test sets. Selected classifiers are then aggregated with KNN and majority vote. This process is repeated and the final decision is achieved with majority vote.

4 Implicit Aggregation Technique with Dropout

4.1 K-NN + Dropout

The classical K-NN [8] is defined by i) a training data set, ii) a parameter K that acts as a regularizer, and iii) a distance metric. The regularizer K is a small positive integer for which larger values refer to greater regularization. Choosing a distance metric used for K-NN can be posed in various learning frameworks [9]. Selecting K can be done efficiently given a distance metric by computing the classification error for various settings of K.

We iteratively dropout number of classifiers and keep a fraction D of the classifiers to compute the K-NN. Each iteration is essentially a new predictor, hence it implicitly creates a new metric $p \in 1, ..., P$ with a random subset of classifiers. At test time, for a given test image $\boldsymbol{x}_{i,j}$, we search the entire training set for the K most similar images. We collect each predictor's decision and sort them by their similarity scores. As shown in Fig. (3) we reach the final consensus by taking the majority vote over the labels of top K predictors. Here D is exhaustively searched, while P is set to a sufficiently large value. We show later in the paper why selecting as large a P as possible is ideal. Ideally, we would use all possible metrics P but that is often prohibitively expensive.

The decisions of the predictors for a given test point is denoted as $\mathbf{y}$. We denote a new similarity measure v indexed by p and l based on the distance between the test sample and all training samples.

$$y = \arg\max_l \sum_p v_{pl} \tag{4}$$

Here $v_{pl} = 1$ if label l has the plurality of the K nearest neighbors according to a given predictor p and $v_{pl} = 0$ otherwise. Given v a decision $y_{i,j}$ is produced by determining which label is most common amongst the predictors as indicated in (4).

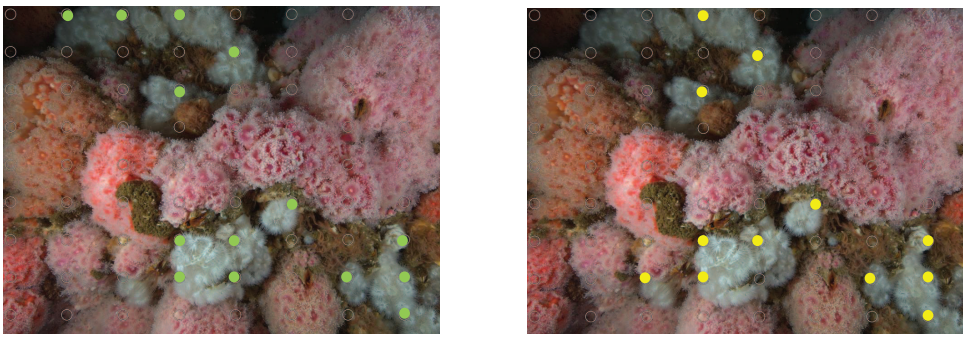

Fig. 4: Comparison of manual (left) and automated (right) annotation of Anemone (*Metridium senile*) on a photograph taken at approximately 10 meters below the surface. This figure is best viewed in color.

4.2 Justification for the K-NN+Dropout Approach

As in random forest methods, the decision of each predictor is of equal importance. We should understand the effect of using different dropout fractions. If we use predictors that have low dropout then we are likely to have nearest neighbors that are very close (in the feature space) to the test point and likely to be uninformative. In the extreme case of dropping out all values but one is akin to Naive Bayes. The number of predictors with a given amount of dropout M is $\binom{N}{M}$. This quantity is maximized by setting $M = \frac{N}{2}$. Having very low dropout may not provide the diversity in the space of predictors needed to make an accurate classification.

We now describe why having more predictors helps. Consider that the average predictor is only slightly better than chance. The output of a single random predictor follows a multinomial over N possible classes. The multinomial distribution is described by vector $\boldsymbol{c}_{i,j}^{m}$ where $c_{i,j}^{m,k}$ is the probability that a random predictor produces a label l_k. Given P random predictors, the empirical average value has a gaussian distribution with mean $c_{i,j}^{m,k}$ and standard deviation $\frac{\sqrt{PF_l(1-c_{i,j}^{m,k})}}{P}$. Notice that as P goes to ∞ the difference between $c_{i,j}^{m,k}$ and the empirical mean goes to zero. Thus, in the case where the largest element of $c_{i,j}^{m,k}$ for a given example tends to be the correct one, and in which the predictors make different types of mistakes, then having large numbers of predictors is beneficial.

5 Results

5.1 Dataset and annotations

We analyzed underwater photographic images of surfaces covered with marine invertebrates and algae on oil platform support structures and natural reefs at depths of 5-20m off the coast of Santa Barbara, California. Thus far our analysis has focused on the oil platform images. Images are taken by SCUBA divers (Figure 1) using a housed SLR camera (Canon 6D) fitted with a 14mm lens and two strobes (Nikonos SB-104) mounted to a rigid quadrapod designed to capture an image of a fixed area of $0.25m^2$. We use Bisque [10,11] to organize and annotate the datasets used in the experiments. Bisque is a distributed, web

Species	Anemone	Echinodermata	Bryozoa	Barnacle	Sponges
Extracted ROI (pixels)	128×128	64×64	128×128	256×256	128×128
Pixel resolution (cm)	1.4×1.4	0.7×0.7	1.4×1.4	2.8×2.8	1.4×1.4
F-Score	0.75	0.36	0.21	0.38	0.17

Table 1: Estimated optimum image resolutions for five different species. We experimented with image resolutions by extracting different block sizes: 64, 128 and 256 pixels. All feature descriptors were computed using 64x64 pixel blocks thus downsizing larger blocks. Although, the original pixel resolutions were the same, features computed from larger blocks were effectively using lower pixel resolutions.

based platform for scientific image management and analysis, offering web-based annotation tools for multi-dimensional (2D-5D) imagery.

The quadrapod eliminates variation in camera-to-subject distance as well as camera movement. At each of 3 depths ($\approx 6m, \approx 12m, \approx 18m$), divers take at least 16 photos distributed around the platform to capture spatial variability at that depth, and an additional 16 photos on horizontal beams at the shallowest depth where the bryozoan is typically most abundant. This growing dataset currently consists of >1500 images from 14 platforms on which we identify 30 different species or categories of data. Images are stored in RGB Canon RAW format, 5496×3670 pixels at 14 bits per channel. We overlaid 100 small circles on each image in a 10x10 grid. Each of these grid elements are then annotated by marine biologists (by naming the species). Dataset and annotations are both available on Bisque[3].

Feature Extraction: We compute 14 visual descriptors: Haralick-Edge, Scalable Color Descriptor (SCD), Color Structure Descriptor (CSD), Color Layout Descriptor (CLD), Homogeneous Texture Descriptor (HTD), Scale-Invariant Feature Transform (SIFT), Speeded Up Robust Features (SURF), Pixel-Intensity-Statistics-Hue, Edge Histogram Descriptor (EHD), Threshold Adjacency Statistics (TAS), Local binary Patterns (LBP), GIST, Region-based Shape Descriptor (RSD), and Brief (ORB).

Optimizing image resolution: We extracted ROIs as square regions centered on an annotated point with sizes ranging over 64×64, 128×128 and 256×256 pixels. These ROIs are then down-scaled to 64x64 pixels and followed by feature extraction and classification. We thus obtain a sequence of M predictions for each ROI (distinct feature descriptors) and apply our aggregation technique to make the final prediction. Table 1 shows estimated optimum ROI sizes for five species based on the classification performance.

5.2 K-NN with Dropout Regularization

Given the predictors we select a single value of K that produces the best results on the data sets. We set $K = 9$ for this experiment and observed that as we increase P we receive an increase in peak results but a broadening out of the range of near optimal settings of dropout. We plot the performance as a function

[3] `http://bisque.ece.ucsb.edu/client_service/view?resource=http://bisque.ece.ucsb.edu/data_service/dataset/6395104`

Classifier	Over 30 Species			Top 5 Species		
	AVG	MV	Proposed	AVG	MV	Proposed
Random Forests	0.23	0.32	0.52	0.44	0.62	0.79
AdaBoost	0.2	0.29	0.51	0.49	0.54	0.68
SVM	0.23	0.27	0.48	0.52	0.61	0.73
Naive Bayes	0.11	0.14	0.23	0.37	0.57	0.61

Table 2: Classification performance (F-Score) comparison of the aggregation techniques: proposed vs. majority vote (MV) and averaging (AVG). The results are averaged over 30 species. We observed a significant variation in the overall classification performance between the top 5 performing classes and the rest, though in each case the proposed method outperforms traditional classifiers by a significant margin.

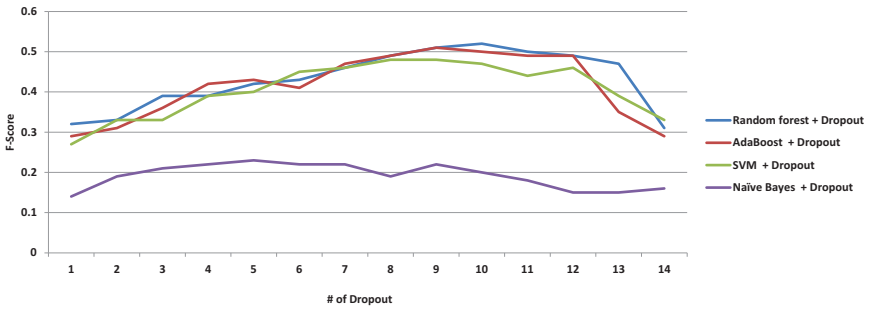

Fig. 5: Comparison of the classification performance as a function of dropout. The peak performance around 10 dropouts indicates that most classifiers are in fact making too many mistakes. By selecting a large number of predictors over a small set of classifiers we are able to select only the top K neighbors and ignore the rest implicitly.

of M (Figure 5). Notice that the optimal quantity of dropout is towards the middle of the set of possibilities. This is a very promising result showing K-NN plus dropout improves the result over simply using the mode decision of the classifiers.

We also used a validation set to find the optimum dropout per species. The final result of our classifier is shown in Figure 6. We compare our result with two other cases. First, with the average performance of individual classifiers (where, for every given species we compute the performance of each classifier first and then average them to get the overall performance.) Second, with the performance of majority vote aggregation. In this case we take the majority vote classifier's output and then evaluate the prediction with ground truth for each species. Figure 4 compares manual annotation of Anemone vs. automated one produced using our method.

6 Conclusion

We proposed an efficient aggregation method for correlated classifiers and showed that we can remove outliers in prediction i.e. by estimating the ratio of good classifiers vs. bad ones. Once this ratio is known (dropout) we can use a combination of K-NN and the dropout technique to optimize for the final decision. We showed that the mode decision of the predictors did not perform nearly as well as the K-NN+dropout. Finally, as noted in Table 2, the overall classification performance varies significantly among the species. The top 5 species and their corresponding

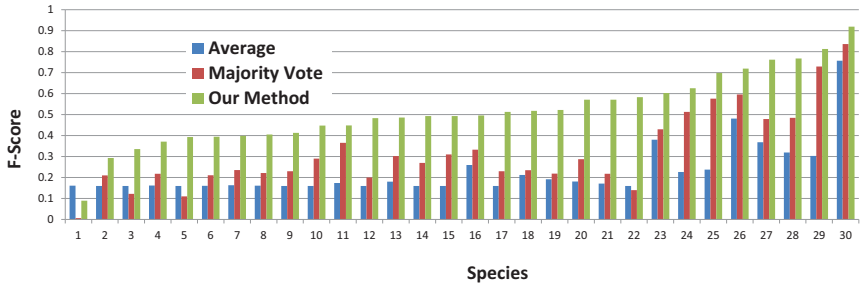

Fig. 6: Classification performance comparison of the aggregation techniques per specie. Where averaging is in blue, majority vote in red and our technique in green.

classifiers perform much better than the rest of the classes, and we speculate that this is possibly due to the limitations of the visual features used and/or the number of available samples in the training set. However, the proposed method outperforms other classifiers that we have compared with in each of these cases. The top 5 species also happen to be the more common ones in the database, and even the current classification accuracy of 80% greatly facilitates expert assisted annotation. **Acknowledgments:** This work is supported in part by the grants NSF-III #0808772, ONR #N000141210503, and the Bureau of Ocean Energy Management (Department of the Interior) Co-Op # M13AC00007.

References

1. Worm, B., Barbier, E.B., Beaumont, N., Duffy, J.E., Folke, C., Halpern, B.S., Jackson, J.B., Lotze, H.K., Micheli, F., Palumbi, S.R., et al.: Impacts of biodiversity loss on ocean ecosystem services. science **314**(5800) (2006) 787–790
2. Shao, Y., Taff, G.N., Walsh, S.J.: Comparison of early stopping criteria for neural-network-based subpixel classification. Geoscience and Remote Sensing Letters, IEEE **8**(1) (2011) 113–117
3. Geurts, P.: Dual perturb and combine algorithm (January 2001)
4. Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. Science **313**(5786) (2006) 504–507
5. García-Pedrajas, N., Ortiz-Boyer, D.: Boosting k-nearest neighbor classifier by means of input space projection. Expert Systems with Applications **36**(7) (2009) 10570–10582
6. Breiman, L.: Bagging predictors. Machine learning **24**(2) (1996) 123–140
7. Grabowski, S.: Voting over multiple k-nn classifiers. In: Modern Problems of Radio Engineering, Telecommunications and Computer Science, 2002. Proceedings of the International Conference, IEEE (2002) 223–225
8. Dhanabal, S., Chandramathi, D.S.: Article:a review of various k-nearest neighbor query processing techniques. International Journal of Computer Applications **31**(7) (October 2011) 14–22 Foundation of Computer Science, New York, USA.
9. Weinberger, K., Blitzer, J., Saul, L.: Distance metric learning for large margin nearest neighbor classification. Advances in neural information processing systems **18** (2006) 1473
10. Kvilekval, K., Fedorov, D., Obara, B., Singh, A., Manjunath, B.: Bisque: A platform for bioimage analysis and management. Bioinformatics **26**(4) (2010) 544–552
11. Miller, R.J., Hocevar, J., Stone, R.P., Fedorov, D.V.: Structure-forming corals and sponges and their use as fish habitat in bering sea submarine canyons. PLoS ONE **7**(3) (03 2012) e33885