

Learning Instrument Invariant Characteristics for Generating High-resolution Global Coral Reef Maps

Ata Akbari Asanjan*
Universities Space Research
Association
Mountain View, California, USA
aakbariasanjan@usra.edu

Ved Chirayath
NASA Ames Research Center
Moffett Field, California, USA
ved.c@nasa.gov

Kamalika Das†
VMWare Inc.
Palo Alto, California, USA
kamalika.das@gmail.com

Juan Torres-Perez
NASA Ames Research Center
California, USA
juan.l.torresperez@nasa.gov

Alan Li
Bay Area Environmental Research
Institute
Moffett Field, California, USA
alan.s.li@nasa.gov

Soroosh Sorooshian
Center for Hydrometeorology and
Remote Sensing, UC Irvine
Irvine, California, USA
soroosh@uci.edu

ABSTRACT

Coral reefs are one of the most biologically complex and diverse ecosystems within the shallow marine environment. Unfortunately, these underwater ecosystems are threatened by a number of anthropogenic challenges, including ocean acidification and warming, overfishing, and the continued increase of marine debris in oceans. This requires a comprehensive assessment of the world's coastal environments, including a quantitative analysis on the health and extent of coral reefs and other associated marine species, as a vital Earth Science measurement. However, limitations in observational and technological capabilities inhibit global sustained imaging of the marine environment. Harmonizing multimodal data sets acquired using different remote sensing instruments presents additional challenges, thereby limiting the availability of good quality labeled data for analysis. In this work, we develop a deep learning model for extracting domain invariant features from multimodal remote sensing imagery and creating high-resolution global maps of coral reefs by combining various sources of imagery and limited hand-labeled data available for certain regions. This framework allows us to generate, for the first time, coral reef segmentation maps at 2-meter resolution, which is a significant improvement over the kilometer-scale state-of-the-art maps. Additionally, this framework doubles accuracy and IoU metrics over baselines that do not account for domain invariance.

*Part of this work was done when author was affiliated to Center for Hydrometeorology and Remote Sensing, UC Irvine.

†This work was done when author was affiliated to USRA, NASA Ames Research Center.

ACM acknowledges that this contribution was authored or co-authored by an employee, contractor, or affiliate of the United States government. As such, the United States government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for government purposes only.

KDD '20, August 23–27, 2020, Virtual Event, USA

© 2020 Association for Computing Machinery.

ACM ISBN 978-1-4503-7998-4/20/08...\$15.00

<https://doi.org/10.1145/3394486.3403312>

CCS CONCEPTS

• **Computing methodologies** → **Image segmentation**; **Neural networks**; *Feature selection*; • **Applied computing** → **Environmental sciences**.

KEYWORDS

Domain Adaptation; Neural Networks; Super Resolution; Coral Reef Segmentation

ACM Reference Format:

Ata Akbari Asanjan, Kamalika Das, Alan Li, Ved Chirayath, Juan Torres-Perez, and Soroosh Sorooshian. 2020. Learning Instrument Invariant Characteristics for Generating High-resolution Global Coral Reef Maps. In *26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '20)*, August 23–27, 2020, Virtual Event, USA. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3394486.3403312>

1 INTRODUCTION

As one of the most biologically complex and diverse ecosystems within the shallow marine environment, coral reefs are not only of great ecological value [14], but also of economic value [6]. Reef ecosystems support the activities of modern civilization and mitigate changes in our planet's biosphere through a diversity and density of species few other ecosystems possess [16]. At present, however, coral reefs are experiencing one of the most significant changes in their history on Earth, triggered by unprecedented anthropogenic pressures, warming seas, ocean acidification, sea-level rise, habitat destruction, agricultural runoff, and overfishing, among other contributing stressors [1].

Our understanding of the impacts of these rapidly-changing ecosystems is limited by a lack of global baseline habitat mapping data and knowledge of reef makeup over regional areas. This can be attributed to limitations in observational technology that inhibits global sustained imaging of the marine environment. The standardization and normalization of terrestrial remote sensing practices, geo-referencing, and data set processing algorithms are not directly applicable to marine data sets. Additionally, ground truth data is often not representative of the entire population distribution, and corresponds only to reefs within their immediate geographical vicinity, which are known to vary worldwide compositionally and

structurally. Due to the difficulties associated with harmonizing multimodal data sets, performing ecosystem assessment at a global scale is an unsolved problem in this domain.

In this paper, we present a framework for fusing images from two different remote sensing platforms at multiple spatial scales to arrive at accurate marine habitat classification that can take advantage of labeled data from higher resolution platforms to predict scenes at a lower resolution and then improve the generated labeled scenes to match the high-resolution imagery. High-resolution images (and ground-truth) are rare to obtain, whereas low-resolution imagery is available in abundance, providing better (albeit poor quality) coverage of the global coral reef makeup. Additionally, these low-resolution satellite images are usually not good for manual labeling due to their low visual quality. This necessitates the need for a classifier that can learn from the limited amount of high-resolution ground-truth data and predict classes on previously unseen low-resolution scenes, thereby creating global coral reef maps. Since these reef cover and morphology classification maps allow marine biologists to delve deeper into areas of interest in certain geographical regions, it is essential for these maps to match the resolution of the best remote sensing instrument so that scientists can make use of these images for their assessment and analysis. In order to facilitate that, it is important to synthesize these labeled coral reef maps at a higher resolution. Therefore, we apply super-resolution to our generated reef maps, and show that our approach yields significantly better results than traditional approaches of doing super-resolution on poor quality satellite imagery. State-of-the-art global assessments of coral reef cover and morphology classification are on kilometer-scale satellite data that have a classification error greater than 40% [5]. Using this framework we are able to generate coral reef segmentation maps at 2-meter resolution with double the accuracy of the previous product on an average and having much higher geographical coverage. In details, the main contributions of this study are as follows:

- Introducing an instrument invariant approach for coral reef classification using a deep learning domain adaptation technique.
- Improving the domain adaptation algorithm for image-to-image segmentation rather than image-to-point classification using a U-NET structure [17] to prevent loss of resolution problem.
- Improving the domain invariance capabilities of the model by removing the covariate shifts in the data through batch normalization.
- Increasing the spatial resolution of the instrument invariant coral reef classification by applying an advanced super-resolution deep learning algorithm, downscaling coral reef classification maps from 10 meters to 1.85 meters.

The rest of the paper is organized as follows. In Section 2 we discuss our deep neural net framework that is being used to circumvent existing challenges. Section 3 describes our data sets and the different experimental results obtained for this study. We conclude the paper with discussions on future directions in Section 4.

2 PROPOSED FRAMEWORK

Although deep neural networks have successfully been used for classification and segmentation tasks in remote sensing [3], [4], [19] the issue of amalgamating multi-resolution, multi-spectral, multi-modal remote sensing has remained a pertinent and challenging problem in ecosystem science.

In this paper, we present LAPDANN, our framework for reconciling multimodal remote sensing inputs for improving scientific discovery. The LAPDANN framework consists of two neural networks: (i) an improved version of Domain Adaptation Neural Network (DANN) adapted from [11], and (ii) super-resolution neural network-based on [7]. The DANN network is improved in two main areas: (i) altering the image-to-point classification approach to an image-to-image segmentation structure using a U-NET, and (ii) improving the domain invariance by removing covariate shifts using BatchNormalization layers [12]. We describe each of these components individually, and then show how they fit together in the context of LAPDANN.

2.1 Domain Adaptation

Domain adaptation aims to learn from one or more source domain(s) and successfully transfers the learned task to a target domain through transfer learning. The Domain Adaptation Neural Network (DANN) proposed by Ganin [11] accomplishes this task by implementing a three-part neural network: (1) a generative network to learn and extract domain invariant features, (2) a label classification network to predict class labels of input patches and (3) a domain discriminative network to discriminate the features originated from the source and target domains.

The generative part (G) learns to map the input image of domain A (x_A) to an intermediate feature representation (f) i.e. $G : x_A \rightarrow f$. The proposed generator is a modified version of the generator in the original DANN [11]. The original DANN has three building blocks in the generator consisting of a convolutional layer followed by a Rectified Linear Unit (ReLU) [15]. In this study, we included a Batch Normalization layer [12] in between convolutional and ReLU layers (Figure 1) to remove the internal covariate shifts and eliminate layers inputs' distribution mismatches. Normalizing the layers inputs' distributions allows the model to focus on "real" discrepancies rather than the trivial variance of data batch distributions. A label classification network is defined to classify the input images (x_A) into defined class labels. In alteration of the original DANN architecture, instead of having the classifier predict the center pixel class, we used a U-Net architecture [17] to obtain image segmentation of the input images (Figure 1). Following the logic described in the generator part, the classification network consists of three blocks with a deconvolutional layer followed by Batch Normalization and ReLU activation layers (Figure 2). At each resolution level of the network, a crop and copy connection (blue dotted line) from the generator to the label classifier is established to prevent the loss of resolution in the network architecture (Figure 1). In addition to the three blocks, a convolutional layer with softmax function is applied as the last layer to obtain the class-conditioned probability maps.

The domain discriminator network is designed to determine the likelihood of the features generated by G to be from the source domain or the target domain. To obtain this goal, a network with

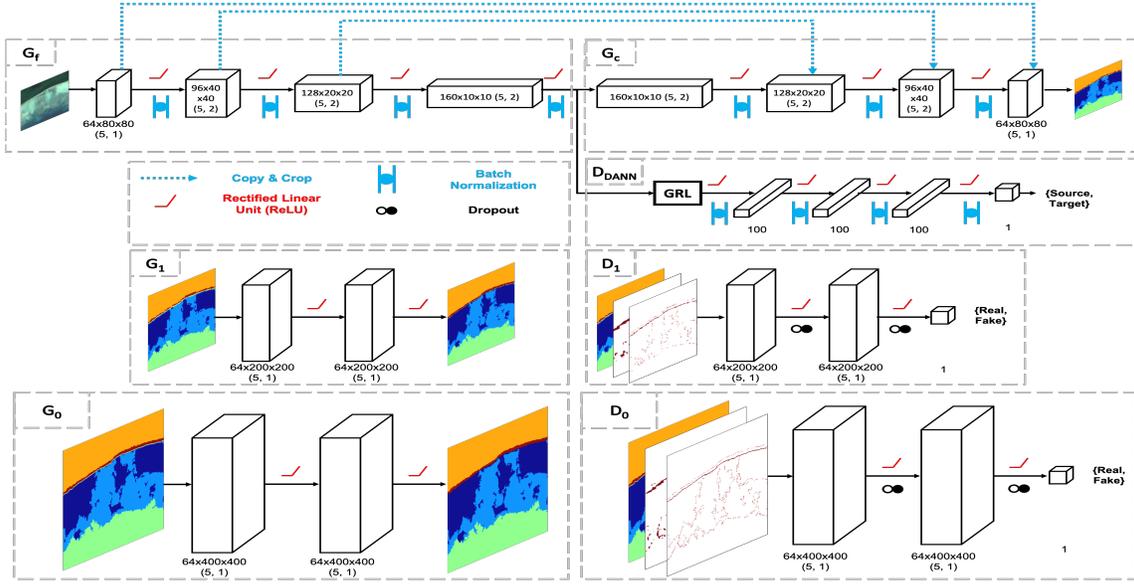


Figure 1: Details of the LAPDANN architecture. Modified DANN structure (top) demonstrates the model specifications. The G_f , G_c and D_{DANN} are the generator, classifier and domain discriminator, respectively. The middle row with G_1 and D_1 is the low-resolution GAN in the LAPGAN architecture and the G_1 and D_1 are the generator and discriminator, respectively. The (bottom) is the high-resolution GAN in the LAPGAN architecture where G_0 and D_0 are the generator and discriminator, respectively. The layer details for convolutional layer are described next to each block in the following format: number of filters \times image height \times width (kernel size, stride size). For fully connected layers, the number of nodes are represented as a single number.

three blocks of fully-connected, Batch Normalization and ReLU layers are used. Following the architecture of the original DANN, the domain discriminator is connected to the generator via a Gradient Reversal Layer (GRL). The GRL does not affect the forward propagation process; it only reverses the gradient coming from the discriminative loss by multiplying the gradient to a negative value of λ .

The DANN framework is trained by optimizing the following equation:

$$E(\theta_f, \theta_y, \theta_d) = \frac{1}{n} \sum_{i=1}^n \ell_y^i(\theta_f, \theta_y) - \lambda \left(\frac{1}{n} \sum_{i=1}^n \ell_d^i(\theta_f, \theta_d) \right) + \frac{1}{n'} \sum_{i=n+1}^N \ell_d^i(\theta_f, \theta_d) \quad (1)$$

In equation 1, θ_f , θ_y , and θ_d are the trainable parameters in the generator, classifier and discriminator parts, respectively. ℓ_y and ℓ_d are the classification and discrimination losses and the λ is the gradient reversal coefficient. n and n' are respectively the number of samples in source and target domains with a total sum equal to N .

Implementing the GRL, as the connection between the generator and the discriminator (Figure 2) units, allows the optimization to find the saddle point of equation 1 and obtain:

$$(\widehat{\theta}_f, \widehat{\theta}_y) = \arg \min_{\theta_f, \theta_y} E(\theta_f, \theta_y, \widehat{\theta}_d) \quad (2)$$

$$\widehat{\theta}_d = \arg \max_{\theta_d} E(\widehat{\theta}_f, \widehat{\theta}_y, \theta_d) \quad (3)$$

$\widehat{\theta}_f$, $\widehat{\theta}_y$, and $\widehat{\theta}_d$ are the optimum set of parameters for the generator, classifier and discriminator parts, respectively.

2.2 Statistical Downscaling

Statistical downscaling refers to the process of generation of high-resolution (HR) images given their low-resolution (LR) counterparts. In this framework, we adapt the Laplacian Generative Adversarial Networks (LAPGAN) model proposed by Denton [7] for downscaling our segmented images to high-resolution coral reefs maps. In this architecture, a Laplacian pyramid [2] representation is used to downscale the given images across each level of the resolution gap. Despite conventional downscaling approaches (e.g. bilinear interpolation, cubic interpolation, etc.) that leverage low-frequency filters, Laplacian pyramid uses high-frequency filters to sharpen the edges of the image resulting in high-quality high-resolution segmentation images. In the LAPGAN approach, the linear invertible operators of the Laplacian pyramid are replaced with Generative Adversarial Networks (GANs) conditioned on images of lower resolution. In the learning process, first the high-resolution images are upsampled to reach the lowest resolution. Then, in the backward process, GANs $\{G_0$ and $G_1\}$ learn the low-frequency residuals conditioned on the lower-resolution images (Figure 2). In this framework, to obtain high-quality high-resolution segmentation maps, we use two GANs $\{G_0$ and $G_1\}$ for scaling by a factor of 2.0 and 2.5 respectively. The generators in both GANs ($\{G_0$ and $G_1\}$) consist of two

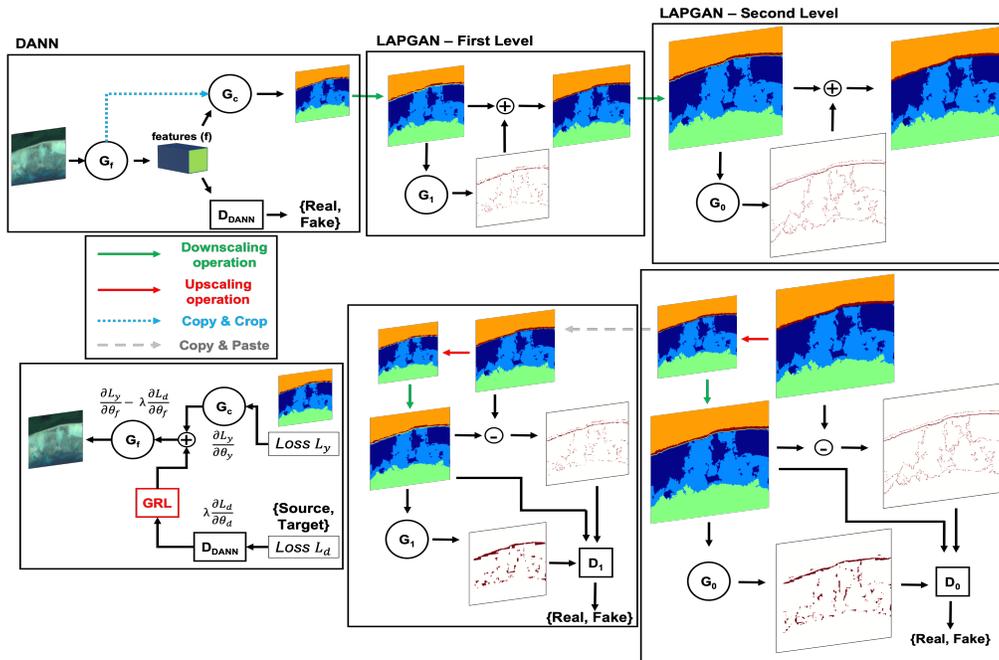


Figure 2: Feedforward (top) and backpropagation (bottom) connections of the LAPDANN framework. From left to right, the DANN, LAPGAN-First Level (for low-resolution) and LAPGAN-Second Level (for high-resolution) are presented.

blocks of convolutional-batch normalization-ReLU. The generators have an additional convolutional layer with a softmax activation function to provide higher resolution segmentation. The discriminator part of the GANs $\{D_0$ and $D_1\}$ is designed using two blocks of convolutional-batch normalization-ReLU, and 50% dropout. As the last layer for the discriminators, a fully connected layer with sigmoid activation is used to predict whether the input is from the source distribution or target distribution.

2.3 LAPDANN

The LAPDANN framework is a three-model architecture, where the first model is DANN and second and third models are the GANs in the LAPGAN architecture (figure 2).

This structure aims to simplify the learning by decoupling the domain adaptation and super-resolution tasks. The input to the DANN framework is high-resolution image patches and corresponding upsampled ground-truth. During the forward pass, the DANN model generates low-resolution segmentations. Then, by downscaling the low-resolution segmentation image using a conventional interpolation method (we used bilinear interpolation in this study) shown by green arrows in figure 2, a mid-resolution segmentation is obtained in the LAPGAN unit. $\{G_1\}$ predicts the high-frequency residuals and the residuals are added to the mid-resolution segmentations. The same process is repeated for $\{G_0\}$ for ultimately obtaining the high-quality high-resolution segmentations. As demonstrated in figure 2, during backpropagation, DANN and the LAPGAN are trained independently. In super-resolution part, $\{G_0\}$ learns the high-frequency residuals consistent with the upsampled high-resolution segmentations, and $\{G_1\}$ is trained to the generate high-frequency

residuals consistent with low-frequency upsampled low-resolution segmentations. We implemented LAPDANN in TensorFlow and used Adam optimizer [13] for tuning LAPDANN parameters. For training DANN model we used the learning rate schedule suggested by [11] and for the LAPGAN parts we used constant learning rate of 0.0008.

3 EXPERIMENTS

Our experiments segment low-resolution images of coral islands in the Pacific Ocean and the Indian Ocean and produce and synthesize high-resolution maps of the reef makeup for these regions. In this section, we describe our data set and the experimental setup. We then present some numerical results to demonstrate the performance of our framework in comparison to some baseline methods.

3.1 Data Description

WorldView-2 (WV-2) satellite is a low-orbit multispectral Earth observation satellite launched by DigitalGlobe in October 2009. The WV-2 satellite has 8 spectral bands ranging from 400nm to 1040nm with 1.85-meter resolution [8], which is well-suited for marine and coral habitat classification.

European Space Agency (ESA)'s *Sentinel-2* (S2) satellite, on the other hand, is a wide-swath multispectral Earth-observing satellite which provides 13 spectral bands with spatial resolutions ranging from 10 to 60 meters [10].

Although S2 images are freely available for download through AWS opendata services¹, the high-resolution WV-2 images had to be

¹<https://aws.amazon.com/opendata/>



Figure 3: Worldview2 (left) and Sentinel2 (right) image patches for a South Pacific coral reef island. The resolutions are 1.85m and 10m respectively.

obtained through an agreement with the non-profit Living Oceans Foundation². The ground truth labels accompanying the WV-2 images are generated by the organization through a segmentation software, the output of which is curated manually by marine habitat experts and validated in certain locations through diving missions. This ground truth, though noisy at times, is the best available in this field and is considered gold-standard for all purposes.

In order to have consistency across the two sets of images obtained from WV-2 and Sentinel, we used the Red (630-690nm), Green(510-580nm), Blue(450-510nm), and Near-Infrared 1 (770-895nm) channels from both with WV-2 resolution of 1.85 meters and S2 resolution of 10 meters over the same areas. Figure 3 illustrates, using a patch of a coral island from both WV-2 and S2 instruments, the difference in visual clarity for a 5x change in resolution.

Geographically we focus only on those areas for which we have high-resolution ground-truth data of relatively less noisy and high quality. For this study, we have such data from WV-2 for the Peros Banhos islands in the Indian Ocean, and 4 transects of Fiji islands, namely, Cicia, Fulaga, Mago, and Cobara, in the South Pacific Ocean. The range of years for the different images is 2010 to 2013.

The ground truth data has 10 habitat classes: (i) reef-crest or coralline algae ridge, (ii) fore-reef, (iii) backreef pavement or sediment, (iv) backreef coral framework, (v) lagoon, (vi) terrestrial vegetation, (vii) beach, (viii) seagrass meadows, (ix) deep water, and (x) other. The other class encompasses noise and elements that are not of interest in the context of marine habitats, such as clouds. The first four classes are particularly relevant for coral mapping, but additional classes, when available in a global labeled data set, allow marine biologists to investigate morphology, composition and species makeup of other sea life, depending on which area (class) in the scene they are focusing on. Figure 4 shows how these classes appear in our ground truth data and the imbalance across different classes.

3.2 Data Processing Pipeline

Since these images come from multiple remote sensing instruments there are a variety of discrepancies across the WV-2 and S2 images. Additionally, even images obtained from the same instrument suffer from issues that hinder classification or segmentation performance.

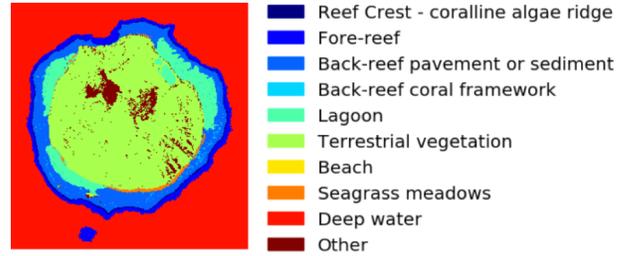


Figure 4: Ground truth color codes for the 10 classes

In this section, we describe the preprocessing steps for preparing our training data.

- (1) Dealing with multimodal data: Preprocessing for harmonizing images from two different sensors include orthorectification and spatial registration on a global reference system for geo-alignment, and image normalization by scaling the pixel values to the same scale of [0-255]
- (2) Cloud removal: The cloud masks supplied with the WV-2 and S2 data sets are of poor quality and cannot be used for creating clean ground truth data for training our network. Therefore, to obtain cloud-free images, we convolve a 40×40 averaging filter over the data and substitute the pixels with RGB values below 40 with the cloud-free pixel values from the same geolocation in a different time step. We apply the same mask to the ground truth data in order to update the labels with appropriate class information.
- (3) Creating training and test sets: We create 400×400 patches from the WV-2 data and decimate the patches down to 80×80 patches using bilinear upsampling. We use 8000 such patches representing uniformly all classes in our training set, that we batch normalize before training. All results are reported on 2000 test patches.

It should be noted here that, other than resolution differences, there are several differences between the images in our data set. These include spectral shifts due to sensing and lighting conditions during imaging. We do not adjust for any of these differences in our preprocessing since we want our framework to learn domain invariant features of these reef classes that go beyond spectral properties.

3.3 Results

In this section, we report the results of several experiments. As mentioned in Section 1, our framework not only doubles the accuracy of state-of-the-art coral reef mapping systems, but also, for the first time, allows for the creation of meter-scale maps. However, in this section, we provide more objective comparisons to baseline methods that could be used instead of this framework for developing these reef maps.

For our comparison with the different baselines, we train on certain patches of the Peros Banhos island, and report our results on unseen patches of the same island.

- Baseline 1: In this setting, we test the performance without domain adaptation and GAN-based super-resolution. That is, we switch off the domain classifier branch of the DANN

²<https://www.livingoceansfoundation.org/>

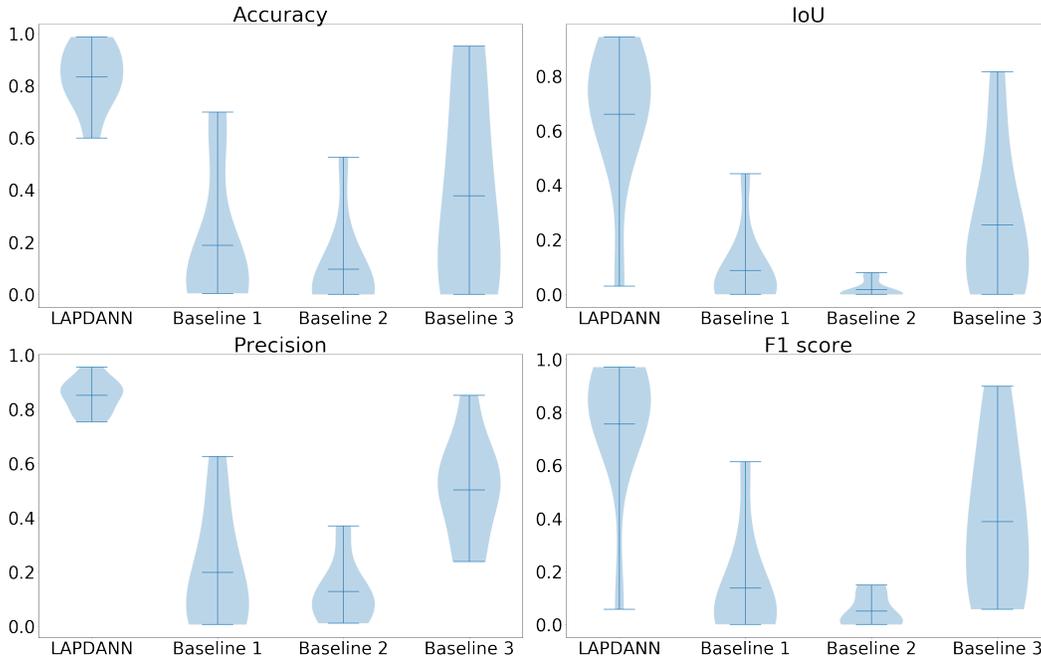


Figure 5: Violin plots for demonstrating the performance of accuracy (top left), IoU (top right), precision (bottom left), and F1 measure (bottom right) for the LAPDANN framework in comparison to the three baselines described in Section 3.3. The center lines indicate the average value of the performance metrics and the shapes indicate the distribution of scores over all classes. A narrow tail of the plot indicates very few observations in that metric value range. For example although LAPDANN has a relatively high range for both IOU and F1 scores, its narrow tail towards the bottom indicates higher concentration of high metric value (desired) for most of the data.

network and evaluate what would have happened, had we trained only on WV-2 data and then tested the classifier on S2 data. To make sure all metrics are reported at the same frame of reference, we scale the output of the classifier using bilinear interpolation.

- Baseline 2: This is a slight variant of Baseline 1, where instead of using a portion of the DANN network as the shallow classifier learned on WV-2, we use a deep CNN which is a variant of the VGG16 architecture [18] that was trained with optimal hyperparameter settings for WV-2 data to have 86% accuracy on an average overall classes on WV-2. In addition, an SRCNN network [9] is trained to synthesize high-resolution versions of the S2 data and the classifier trained on WV-2 is tested on these synthesized data sets.
- Baseline 3: Baseline 3 is a variant of Baseline 2 where the VGG16 architecture is trained on decimated WV-2 data, rather than on the original resolution and then tested on Sentinel. The output segmented image is then bilinearly interpolated to WV-2 resolution.

Figure 5 shows the performance of our LAPDANN framework in comparison to the above baseline cases using violin plots. Violin plots show the distribution of the error metrics over the classes, in addition to the mean and variance of the metric. For reporting performance, we chose the following metrics: accuracy, precision, F-1 score, and intersection over union (IoU).

As seen in Figure 5, the mean performance of LAPDANN is better compared to the baselines for all metrics. In addition to that, the variance of the errors for all classes is low for the accuracy and precision metrics. The narrow tail of the violin plot for the IoU and F1-score indicates that for a specific class our framework performed poorly compared to the average performance across other classes. The mean performance metrics are significantly lower for all the baseline methods, indicating that irrespective of how good a classifier is, it does not scale to other domains if not trained for domain adaptation. Additionally, baseline 2 performance exceptionally poorly indicating that super-resolution attempts on the original satellite imagery are a futile exercise for classification purposes, for a 5x resolution difference. When trained over a single island, the results show a minimum of 120% improvement in accuracy and 160% improvement in IoU.

In addition to training our framework to data from only the Indian Ocean, we experiment with training a more generalizable framework that can learn the domain invariant features from multiple geographical regions. This poses an additional challenge in terms of learning, given that coral reefs in different regions have different morphological characteristics. The metrics for this scenario are reported in Table 1. Although we observe here that our algorithm significantly outperforms baselines for all metrics, it should be noted that the overall performance drops in this case compared

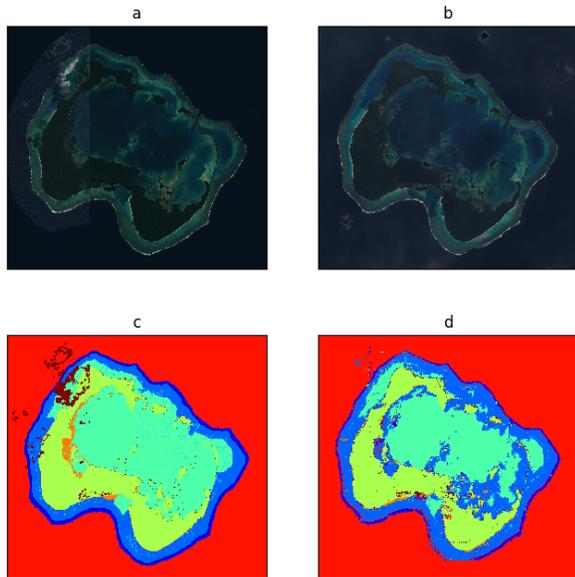


Figure 6: Segmentation results for the island of Fulaga, that is unknown during training. Images (a) and (b) are observations of Fulaga from WV-2 and S2 respectively. Image (c) is WV-2 based groundtruth. Image (d) is the output of the LAPDANN framework.

to the previous set of results based on a single region. This is expected given the additional complexity of the domain adaptation problem. However, we anticipate that more training data, over multiple islands, capturing the variations in sensing conditions will mitigate this performance issue.

	LAPDANN	Baseline 1
Accuracy	0.466	0.106
IoU	0.303	0.069
Precision	0.440	0.148
F1 score	0.411	0.121

Table 1: Table showing performance metrics for LAPDANN trained on Pacific Ocean and Indian Ocean islands simultaneously.

Lastly, we also experiment with the scenario where we train on a subset of islands and test on an island that is new. This scenario is particularly important for creating global maps, given the fact that labeled data is not available for all regions on the map. Figure 6 shows the segmentation of the S2 image of the island of Fulaga that has been left out from the training data. While we can see classification error on certain classes such as seagrass meadows, but in general the quality of the coral reef maps, especially for the coral classes is of comparable quality to the ground truth.

A summary of the results for the three scenarios is shown in Figure 7. This figure allows a visual comparison of the quality of results for the different scenarios we experimented with. The first two rows illustrate results for training on Peros Banhos island alone.

The next two rows show results when training is done on different geographical locations simultaneously, while the last two rows are for testing on an unknown island (Fulaga). Although the quality of the results progressively goes down with increasing problem difficulty, the overall results still significantly improve state-of-the-art performance.

4 CONCLUSION

Aquatic ecosystems, particularly coral reefs, remain quantitatively misrepresented due to low-resolution remote sensing as a result of refractive distortion from ocean waves, optical attenuation, and remoteness. Harmonizing multimodal data sets acquired using different remote sensing instruments presents additional challenges such as spectral shifts across domains that lead to poor generalization performance of trained classifiers. This significantly hinders our understanding of patterns and processes in marine biodiversity at a time when these ecosystems are experiencing unprecedented anthropogenic stress. In this work, we develop LAPDANN, a deep learning-based framework for extracting domain invariant features from two different remote sensing imagery and creating high-resolution global maps of coral reefs. Although the DANN framework has been used for multi-domain classification in the past, it has never been used for images with varying resolutions. Additionally, synthesizing high-resolution remote sensing scenes, given their low-resolution counterparts, is a non-trivial problem even with the most sophisticated deep neural nets and a significant amount of training data. Therefore, our idea of training the domain adaptation part of LAPDANN on decimated high-resolution data and then using the LAPGANN network to synthesize the high-resolution segmented map significantly boosted the performance of our framework, as illustrated in Section 3.3. This framework allows us to generate, for the first time, coral reef segmentation maps at 2-meter resolution, which is a significant improvement over the kilometer-scale state-of-the-art maps. Additionally, this framework improves performance metrics over baselines by significant margins across all test scenarios.

To improve the performance of LAPDANN, we plan to extend our work to incorporate multiple options. We are working with marine habitat experts to gather feedback on scenes that are segmented using LAPDANN, in order to collect additional and refine existing training data. We are also designing a game using high-resolution scenes of coral islands for collecting labels through citizen science. We envision using active learning for incorporating good quality labels collected in this way to improve the performance of our framework. Additional domain knowledge incorporation through smoothing functions might also help with eliminating noisy segmentation results.

ACKNOWLEDGMENTS

This work was supported by the NASA ESTO Advanced Information Systems Technology Program through grant AIST-00019 and NASA’s Biodiversity Ecological Forecasting Program. We also want to acknowledge the supports from the University of California, Irvine, Henry Samueli School of Engineering, and Center for Hydrometeorology and Remote Sensing.

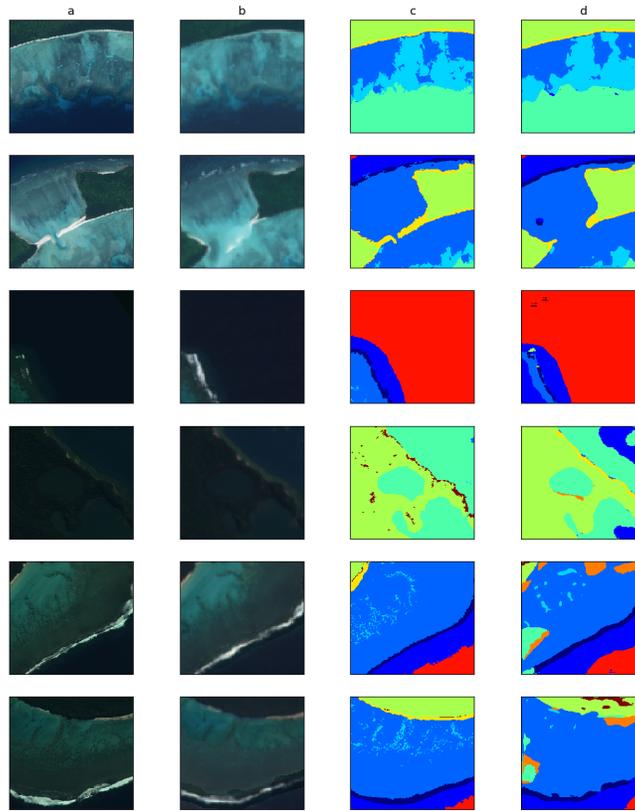


Figure 7: Segmentation results on image patches from Sentinel. Columns (a) and (b) are WV-2 and S2 image patches from the test set respectively. Column (c) shows high-resolution ground truth on WV-2 images. Column (d) presents the results of the LAPDANN framework. The first two rows are results of training and testing based on a single island. Rows 3 and 4 show results when the framework is trained and tested on all available islands. The last two rows correspond to results obtained on patches of the Fulaga island, not included during training.

REFERENCES

- [1] David Bellwood, Terence Hughes, Carl Folke, and M Nyström. 2004. Confronting the Coral Reef Crisis. *Nature* 429 (07 2004), 827–33.
- [2] Peter Burt and Edward Adelson. 1983. The Laplacian pyramid as a compact image code. *IEEE Transactions on communications* 31, 4 (1983), 532–540.
- [3] Marco Castelluccio, Giovanni Poggi, Carlo Sansone, and Luisa Verdoliva. 2015. Land Use Classification in Remote Sensing Images by Convolutional Neural Networks. *CoRR* abs/1508.00092 (2015).
- [4] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi. 2016. Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks. *IEEE Transactions on Geoscience and Remote Sensing* 54, 10 (Oct 2016), 6232–6251.
- [5] V. Chirayath and R. Instrella. 2016. Airborne Fluid Lensing and Machine Learning for Automated Coral Classification. *AGU Fall Meeting 2016* (2016).
- [6] Robert Costanza, Ralph d’Arge, Rudolf S. de Groot, Stephen Farber, Monica Grasso, Bruce Hannon, Karin E. Limburg, Shahid Naeem, Robert V. O’Neill, José M. Paruelo, Robert G. Raskin, Paul Sutton, and Marjan van den Belt. 1997. The value of the world’s ecosystem services and natural capital. *Nature* 387 (1997), 253–260.
- [7] Emily L Denton, Soumith Chintala, Rob Fergus, et al. 2015. Deep generative image models using a laplacian pyramid of adversarial networks. In *Advances in neural information processing systems*. 1486–1494.
- [8] Digital Globe 2010. *The Benefits of the 8 Spectral Bands of WorldView-2*. Digital Globe.
- [9] C. Dong, C. C. Loy, K. He, and X. Tang. 2016. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38, 2 (Feb 2016), 295–307.
- [10] European Space Agency 2015. *Sentinel-2 User Handbook*. European Space Agency. Issue 1.
- [11] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. 2016. Domain-adversarial training of neural networks. *The Journal of Machine Learning Research* 17, 1 (2016), 2096–2030.
- [12] Sergey Ioffe and Christian Szegedy. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167* (2015).
- [13] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- [14] Fredrik Moberg and Carl Folke. 1999. Ecological Goods and Services of Coral Reef Ecosystems. *Ecological Economics* 29 (05 1999), 215–233.
- [15] Vinod Nair and Geoffrey E Hinton. 2010. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*. 807–814.
- [16] Andy Ridgwell and Richard E. Zeebe. 2005. The role of the global carbonate cycle in the regulation and evolution of the Earth system. *Earth and Planetary Science Letters* 234, 3 (2005), 299–315.
- [17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*. Springer, 234–241.
- [18] Karen Simonyan and Andrew Zisserman. 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR* (2014).
- [19] Yanfei Zhong, Feng Fei, Yanfei Liu, Bei Zhao, Hongzan Jiao, and Liangpei Zhang. 2017. SatCNN: satellite image dataset classification using agile convolutional neural networks. *Remote Sensing Letters* 8, 2 (2017), 136–145.