


NASA NeMO-Net's Convolutional Neural Network: Mapping Marine Habitats with Spectrally Heterogeneous Remote Sensing Imagery

Alan S. Li , Ved Chirayath, *Member, IEEE*, Michal Segal-Rozenhaimer, Juan L. Torres-Pérez, and Jarrett van den Bergh

Abstract—Recent advances in machine learning and computer vision have enabled increased automation in benthic habitat mapping through airborne and satellite remote sensing. Here, we applied deep learning and neural network architectures in NASA NeMO-Net, a novel neural multimodal observation and training network for global habitat mapping of shallow benthic tropical marine systems. These ecosystems, particularly coral reefs, are undergoing rapid changes as a result of increasing ocean temperatures, acidification, and pollution, among other stressors. Remote sensing from air and space has been the primary method in which changes are assessed within these important, often remote, ecosystems at a global scale. However, such global datasets often suffer from large spectral variances due to the time of observation, atmospheric effects, water column properties, and heterogeneous instruments and calibrations. To address these challenges, we developed an object-based fully convolutional network (FCN) to improve upon the spatial-spectral classification problem inherent in multimodal datasets. We showed that with training upon augmented data in conjunction with classical methods, such as K-nearest neighbors, we were able to achieve better overall classification and segmentation results. This suggests FCNs are able to effectively identify the relative applicable spectral and spatial spaces within an image, whereas pixel-based classical methods excel at classification within those identified spaces. Our spectrally invariant results, based on minimally preprocessed WorldView-2 and Planet satellite imagery, show a total accuracy of approximately 85% and 80%, respectively, over nine classes when trained and tested upon a chain of Fijian islands imaged under highly variable day-to-day spectral inputs.

Index Terms—Convolutional neural network (CNN), deep learning, image segmentation, multispectral imaging.

I. INTRODUCTION

MACHINE learning in the field of computer vision has recently led to dramatic progress in areas of image classification, segmentation, and feature extraction. The increase

Manuscript received May 19, 2020; revised July 27, 2020; accepted August 11, 2020. Date of publication August 24, 2020; date of current version September 16, 2020. This work was supported by the National Aeronautics and Space Administration (NASA) Earth Science Technology Office (ESTO), under the Advanced Information Systems Technology (AIST) and Biodiversity and Ecological Forecasting Programs under Grant AIST-16-0046. (*Corresponding author: Alan S. Li.*)

The authors are with Earth Science Division (Code SG), NASA Ames Research Center, Mountain View, CA 94035 USA (e-mail: alan.s.li@nasa.gov; ved.c@nasa.gov; michal.segalrozenhaimer@nasa.gov; juan.l.torresperez@nasa.gov; jarrett.s.vandenbergh@nasa.gov).

Digital Object Identifier 10.1109/JSTARS.2020.3018719

in abundant data sources in mobile, cloud, and social media technologies is a large contributor to the successes of these methods, which tend to thrive in environments where large quantities of training data are available [1]. In the field of remote sensing, data-rich sources such as daily satellite imagery over the entirety of Earth's surface have steadily become increasingly prevalent and available. The confluence of these factors has led to interest in incorporating deep learning and sensor fusion methods with the traditional remote sensing methodology, particularly, to automate the classification and interpretation of Earth Observation Systems (EOS) data.

Our focus primarily centers upon benthic habitat mapping into biological and nonbiological classes, particularly coral reef ecosystems. These ecologically and economically important marine ecosystems have undergone worldwide degradation due to the increase in frequency and magnitude of extreme events (i.e., bleaching and die-offs) as a consequence of warming oceans, ocean acidification, and anthropogenic factors such as run-off, pollution, and overfishing [2]–[5]. The ocean sciences community as well as NASA has recognized the urgency of these scenarios, labeling these episodes as severe and unexpectedly large in scope with the capability of causing rapid environmental change [6]. Because coral reefs primarily exist in tropical, remote areas, direct human observation is typically lacking or delayed, and thus, the need for space-based observation that is able to cover large geographical extents without the need for costly oceanic or airborne missions. In particular, it was realized that there was a need for new analytic tools with the ability to analyze and exploit large datasets (TB and PB-scale) cross-platform, leveraging information from multiple sources to create a fused data product.

These concerns eventually prompted the Earth Science Technology Office (ESTO) of NASA to support the creation of NeMO-Net, the neural multimodal observation and training network for global coral reef assessment, with the following major objectives.

- 1) Fuse existing datasets from FluidCam and Fluid Lensing technologies [7]–[10] developed at the Laboratory for Advanced Sensing (LAS) at NASA Ames Research Center (ARC) with NASA EOS data as well as commercial imagery where possible. Because targeted aerial surveys are impractical at covering large geographical regions, the goal here is to utilize features that are readily apparent

in local high-resolution imagery to augment extensive, low-resolution data, either through data fusion or super-resolution.

- 2) Utilize deep convolutional neural networks (CNNs) applied to aerial and satellite imagery to spatially determine the percent cover of reef benthic classes, including major living hard and soft coral families as well as other dominant reef organisms (i.e., seagrasses) of the present coral reef ecosystems. Here, we seek to investigate various CNN architectures, training methods, and processes within the context of remote sensing to produce segmented classification maps of benthic cover in an automated manner.
- 3) Develop and deploy an active learning and citizen science module that allows users to label 3-D reconstructed coral reef scenes from structure from motion (SfM) and fluid lensing products, as well as traditional 2-D imagery from satellites. Crowd sourcing in this manner also requires work to aggregate the data in such a way that user classifications of high fidelity are retained, while improper or erroneous classifications are discarded or assigned low reliability when utilized within the training network. More information regarding this aspect of NeMO-Net can be accessed.¹

Within the scope of this article, we address the second objective by constructing and evaluating various CNNs to segment satellite imagery into appropriate benthic classes. One major requirement for these networks is that they possess invariance to input noise and nonlinear spectral shifts that are apparent in satellite imagery due to spatial variation, temporal factors (i.e., time of day, sun angle, and seasonal effects, among others), atmospheric effects, water column physics, and the inherent spectral variability within benthic components. While these disturbances often confound traditional methods of automated classification, humans are often able to bypass these difficulties with relative ease due to our ability to infer context within an image given just a few training examples. We explored deep learning as an alternative and/or extension to the existing methods due to their ability to take advantage of relative spatial and spectral context when classifying images. The rest of this article is organized as follows: Section II provides an overview of the history, traditional methods, and machine learning work as applied to coral reef remote sensing; Section III covers the data sources as utilized within this study; Section IV delves into the CNN methodology; Section V examines the results of our study; Section VI discusses important factors and caveats within our work; and finally, Section VII concludes this article with possible future work.

II. BACKGROUND AND RELATED WORK

A. Coral Reef Remote Sensing

Remote sensing is the primary method in which global-scale observations are made regarding the Earth system barring direct physical contact. In the realm of remote sensing of oceans and marine habitats, airborne platforms such as the airborne

visible/infrared imaging spectrometer (AVIRIS) [11], compact airborne spectrographic imager (CASI) [12], advanced airborne hyperspectral imaging system (AAHIS) [13], and portable remote imaging spectrometer (PRISM) [14] have often provided high resolution, hyperspectral imagery of targeted locations. On larger spatial and temporal scales, on-orbit optical imagers are able to deliver multispectral data of these same locations, albeit at mostly lower resolutions (0.5–30 m). Previous examples include the hyperspectral imager for the coastal ocean (HICO) [15], IKONOS [16], Sentinel-2 [17], Worldview-2 [18], and the Landsat series [19], while newer instruments include HISUI [20] and DESIS [21] aboard the international space station (ISS).

Early space-based remote sensing of coral reefs and near-shore benthic habitats was established with the Landsat thematic mapper (TM) and Satellite pour l'Observation de la Terre (SPOT) high-resolution visible (HRV). These datasets have been available since the 1980s, with resolutions on the order of 30 m [22], [23]. During the period of 1999–2002, the millennium coral reef mapping project was initiated, aimed at understanding, classifying, and mapping coral reef structures worldwide based upon predominantly Landsat 7 images [23]. The launch of the IKONOS and QuickBird satellites in 2000 and 2001 drastically improved the ability to spatially resolve the Earth's surface in the visible spectrum compared to their predecessors, with resolutions of 4 and 2.4 m, respectively [24]–[27]. Within the last decade, the Pleiades and WorldView series of satellites have enabled meter-scale classification and mapping of benthic habitats from space, offering up to eight spectral bands (five of which are water penetrating) at a dynamic range of 11 bits per pixel [28]–[31]. It has become evident that although satellites are disadvantaged in terms of resolution (spatially and spectrally) and signal-to-noise ratios (SNRs) relative to their airborne counterparts, their coverage and scope are unmatched, and thus, necessary in any large reef survey or monitoring effort. To utilize these datasets, however, requires the preprocessing, radiometric calibration, removal of atmospheric and water attenuation effects, and compensation for extraneous factors such as sun glint and cloud cover [32]–[35]. These issues become challenging at large spatial and temporal scales, due to changing weather and environmental conditions.

Regardless of the difficulties, many different methods, supervised and unsupervised, have been applied to the task of mapping reef habitats. Often these methods attempt to segment the imagery into categorical classes that may be noticeably distinct. At other times, separation of spectrally similar classes (i.e., coral versus algae) may depend upon the availability and quality of distinct spectral libraries with hyperspectral or high-resolution texture-detection capabilities [24], [26]. The earliest methods harken to classical machine learning methods, such as principal component analysis (PCA), maximum likelihood estimation (MLE), canonical variate analysis, K-means clustering, and ISODATA unsupervised classification based upon class means and variance, which often analyzed and extrapolated in-field observables to spectral signatures captured by an airborne or spaceborne sensor [24], [36]–[38]. These methods were initially pixel based, but eventually expanded to cluster similar pixels in close proximity as belonging to the same class [30].

¹[Online]. Available: <http://nemonet.info/>

With the availability of high-resolution imagery, the use of object-based image analysis (OBIA) became the standard method to categorically classify a wide range of remote sensing data products [39]. OBIA operates by initially segmenting the imagery in question into small clusters based upon not only spectral reflectance, but also the location, texture, and shape of grouped pixels. These factors, as well as their relation to one another and their nearby environments, then determine how the clusters are eventually assigned classes based upon predefined rules. These methods have been used extensively in the shallow marine remote sensing community for seagrass monitoring, coral mapping, and for delineating geomorphic zones [30], [40]–[43]. However, it has been noted that although OBIA performs well under ideal conditions where physics models can remove and correct for most of the atmospheric, water column, scattering, and sensor effects, true automated classification under less-than-ideal circumstances consistently deliver results that are lower in accuracy than when compared to assignments produced by an expert user. This may be because humans have the innate ability to infer context on multiple scales within an image almost immediately and instinctively, despite disturbances such as color shifts, noisy pixels, and partial obstruction of view.

Recently, the volume of data collected through heterogeneous sensors has grown exponentially, driving the need for full automation of remote sensing classification that can operate across sensors over large temporal and spatial scales. The goal within this article is not to provide highly detailed classifications of particular coral types requiring expert knowledge, but to develop an automated classification baseline to distinguish large geomorphic and benthic cover classes easily identifiable by any human given modest spatial and spectral distortions within the data. To this end, data fusion techniques have already approached this issue through the coupling of disparate types of sensors to provide a multifaceted view of the remote sensing problem, enabling higher classification accuracies where these measurements are available [44]. Multiple state-of-the-art methods have also been proposed to exploit the spatial or spectral nature between datasets. These include dictionary learning, aimed at extracting a sparse spatial-spectral representation of hyperspectral imagery [45], kernel-based methods relying upon canonical correlation analysis (CCA) to calculate the nonlinear transformation between spectral features in Hilbert space [46], and manifold learning, where data from various sources are projected upon a higher order manifold and evaluated via similarity metrics such as proximity graphs [47], [48]. However, to combine spatial and spectral domains within a deep feature context, we turn to CNNs to merge the state-of-the-art techniques in deep learning with remote sensing.

B. CNNs and Remote Sensing

Within the last decade, successful demonstrations of CNNs in applications of image classification [49], natural language processing [50], and self-enabled reinforcement learning [51], have increased the popularity and interest of deep learning in a variety of fields. We will provide here an overview of CNNs and their basic operation, their increasing role within the remote

sensing community, and the challenges they yet face before full adoption alongside existing techniques.

Neural networks are loosely modeled after the structure of the human brain, in the sense that they mimic layers of cascading neurons transporting electrical signals. Neurons that activate together tend increasingly to do so over time, which gives rise to complex abilities such as pattern recognition and perception. On a singular level, each “neuron” within a CNN acts according to the function

$$z = f(\mathbf{w} * \mathbf{x} + \mathbf{b}) \# \quad (1)$$

where \mathbf{z} is the output, \mathbf{w} the weights to be learned, \mathbf{x} the input, and \mathbf{b} the bias, with $*$ symbolizing the convolution operator. The nonlinear activation function f relates the importance of the linear output, where currently the rectified linear unit (ReLU) is the most widely adopted for processing and numerical reasons [52]. Note that although the variables in (1) are vectors (\mathbf{R}) in this case, it is not difficult to generalize $\mathbf{W}, \mathbf{X} \in \mathbf{R}^3$ as tensors such that \mathbf{X} covers 2 spatial dimensions and one spectral dimension, thus (1) performs the mapping $\mathbf{R}^3 \rightarrow \mathbf{R}$. In essence, we slide the weight kernel \mathbf{W} over the image both vertically and horizontally in predefined window sizes, repeating the process k times such that we generate k filter maps, each channel describing the strength of the presence of the filter \mathbf{W}_k . Performing these operations again on the filter maps leads to deeper features that build upon previous features, where the importance of sets of combinations of features is determined by the weighting kernels \mathbf{W}_{lk} , with l denoting the layer depth. It is not surprising, therefore, that often the shallower layers convey simple information such as edges and corners, while deeper features will combine these rudimentary patterns to form complex shapes such as buildings or faces. Also note that because we slide similar kernels over the entire image, we learn a type of positional invariance that is independent of the absolute location of where each feature is but rather correlated with its relative position within the context of other features. During the final classification step, a CNN usually deploys fully connected layers (every node connected to every other node) or global average pooling for a full representation of the output, a softmax operation to transform the output values into probabilities, and then, optimize for the cross entropy against the training data.

Two additional important components within a CNN are the pooling and batch normalization layers [53]. The pooling function downscales the feature space spatially such that the CNN focuses on the more beneficial features as well as promoting invariance to small-scale transformations such as translations and rotations. Batch normalization is performed usually between convolutional layers such that we normalize the layer to its batch mean and standard deviation to stabilize the training procedure. Often in addition to these layers, hyperparameters and additional considerations such as the number of layers, kernel size, method of convolution (e.g., stride, atrous, and transpose), weights, regularization, and types of connections (e.g., feed-forward, shortcut, and parallel) have given rise to many types of popular and refined architectures, including AlexNet [49], VGG16 [54], ResNet [55], and GoogLeNet [56].

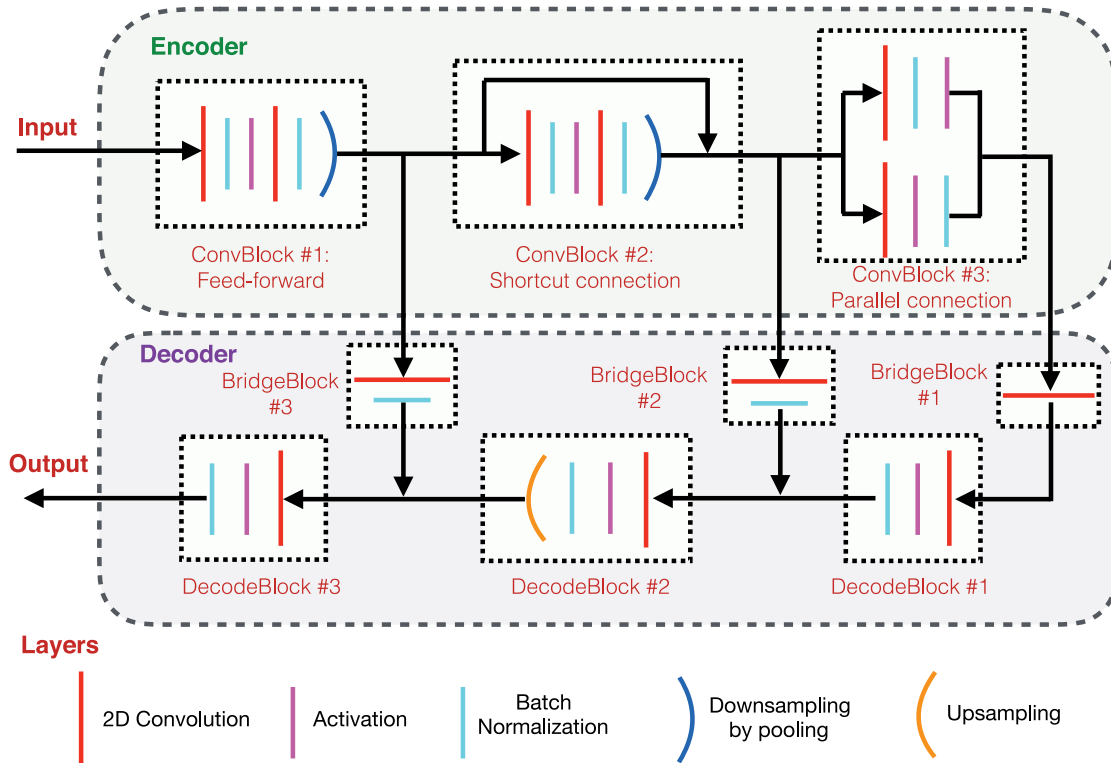


Fig. 1. General structure of an encoder–decoder FCN. Each convolutional block represents convolutional downsampling layers with a pooling operation at the end of each block. Here, feed-forward, shortcut, and parallel connections are shown as examples of possible internal structures. The bridge blocks transfer each encoder block output to the decoder through a convolutional layer, allowing only relevant features to transfer over. The decoder blocks represent the upsampling procedure. Note that convolutional, activation, and batch normalization layers can still exist within decoder blocks as we upsample back to the input size of the original image.

However, for the purposes of remote sensing, we need to introduce another major component to the CNN architecture to output not only a classification for an image, but to reproduce an entire semantically segmented image. Note that although it is possible to predict on a pixel-level scale (i.e., predict upon the center pixel given surrounding pixels), this type of classification often does not take into account the entire surrounding spatial context. Furthermore, the time required scales to the number of pixels classified, and thus, becomes time consuming during run-time for large datasets. Most semantically driven CNNs thus employ a fully convolutional network (FCN) in an encoder–decoder structure, where the encoder functions as a feature detector, while the decoder attempts to reconstruct the image labels at the similar scale of the original by upsampling the high level but spatially compact features, usually through 2-D transpose convolutions (also known as deconvolutional filters). Often this is performed in a hierarchical fashion, upsampling the deepest features first to reconstruct dense interpolations, then combining it with the next lower level features, upsampling again, and so on, as illustrated in Fig. 1. Popular semantic segmentation algorithms include DeepLab [57] (note that DeepLab does not use the encoder–decoder structure, rather it relies upon parallel atrous convolutions), FCN [58], U-Net [59], and SegNet [60]. Often these CNNs will also include a conditional random field (CRF) [61] for postprocessing to

filter the classifications in both value similarity and probability space, such that details often missed by the CNN may be captured.

Given machine learning’s recent successes for the purposes of object identification and scene segmentation, it was only natural for the remote sensing community to quickly realize its potential application. Land use classification became a natural fit for CNN algorithms, as they took direct advantage of semantic segmentation using airborne and satellite multispectral and hyperspectral data, applied to many human-related activities such as urban mapping, agriculture, and forestry mapping [62]–[64]. Widespread adoption, however, has still been met with certain skepticism within the scientific community due to the black-box nature of CNNs and interpretability issues, as well as how to connect CNN predictions to the existing scientific models and principles. Specifically, variable phenomena such as sensor degradation, shadowing, scattering, and atmospheric effects are poorly encapsulated within a CNN. Often, training these structures with high-quality datasets does not yield good results, particularly when contending with the high diurnal heterogeneity that so frequently dominates satellite imagery. Alternatively, engineered metrics and indices [i.e., normalized difference vegetation index (NDVI)] have been used such that the noise is preprocessed or smoothed away, or avoided altogether [65].

TABLE I
CONVERSION OF KSLOF CLASSES TO NEMO-NET CLASSES

KSLOF Classes	NeMO-Net Classes
Shallow fore reef terrace, shallow fore reef slope, deep fore reef slope, coralline algal ridge, back reef rubble dominated, back reef pavement, back reef coral framework, back reef coral bommies, lagoonal pinnacle reefs calcareous red-algal conglomerate, lagoonal pinnacle reefs massive coral dominated, lagoonal pinnacle reefs branching coral dominated, lagoonal <i>Acropora</i> framework, lagoonal patch reefs, lagoonal floor coral bommies, lagoonal fringing reefs, coral rubble, lagoonal floor bommie field, lagoonal coral framework	Coral
Fore reef sand flats, back reef sediment dominated, lagoonal sediment apron sediment dominated, lagoonal sediment apron macroalgae on sediment, lagoonal floor barren, lagoonal floor macroalgae on sediment, dense macroalgae on sediment, lagoonal pavement	Sediment
Dense seagrass meadows	Seagrass
Mangroves, mud flats, terrestrial vegetation	Terrestrial Vegetation
Deep ocean water, deep lagoonal water, inland waters	Deep Water
Beach sand, mud, rock	Beach
Urban, No Data	Other
Clouds	Clouds
N/A	Wave Breaking

Note that many areas that were classified as coral rubble within KSLOF are simply classified as coral for NeMO-Net, since they usually consist of a mix of both live and dead coral colonies. Only one new class was added to NeMO-Net: wave-breaking areas where the ocean meets the coralline algal ridge and reef crest, producing highly visible wave crests.

In regards to the benthic mapping of shallow-water marine ecosystems such as coral reefs, CNNs have been used to annotate diver-scale high-resolution images of coral [66]–[68], although little has been accomplished in applying CNN classifiers to satellite imagery. Currently, many satellite-based reef mapping projects still utilize OBIA methods in software packages such as eCognition to perform much of their classification and segmentation work [69], [70]. This is because the difficulties in mapping marine environments are compounded as compared to the land cover scenario, since for ocean mapping, one has to contend with additional effects such as sun glint, wave action, water column attenuation, and obfuscation of benthic cover due to sediment transport. In addition, classification of coral morphology becomes difficult when the ground sample distance of most visible-range satellite sensors are over 1 m in resolution. Our work with NeMO-Net, however, demonstrates that we can overcome these difficulties on a modest scale. The inspiration for these systems stems from our human experience, since our ability to infer context allows us to easily distinguish simple classification measures despite highly shifted and noisy imagery within the spectral and spatial sphere.

III. DATA SOURCES

To test and apply our algorithms, we focused largely on the fringing reefs surrounding the remote islands of Fiji, namely Cicia, Fulaga, Kobara, Mago, Matuka, Moala, Nayau, Totoya, Tuvuca, Vanua Balavu, and Vanua Vatu, covering an extent of roughly 2300 km². These islands were initially surveyed by the Khaled bin Sultan Living Oceans Foundation (KSLOF) expedition [69] as part of an effort to map the world's remotest coral reef habitats. As part of the project, KSLOF obtained DigitalGlobe's

WorldView-2 (WV-2) multispectral imagery of these islands over multiple days, spanning up to a few months apart, which cover the 400–1050 nm visible and near infrared (NIR) regimes through eight unique spectral bands. The imagery itself has been orthorectified to a per-pixel resolution of 2 m, and was initially delivered as 16-bit digital numbers (DN) before conversion to remote sensing reflectance to just above the water surface.

Our primary source of data was WV-2 multispectral imagery throughout in conjunction with KSLOF data as a starting point for verifying NeMO-Net's results. However, we have reduced the number of spectral bands to the more commonly used four spectral channels: red (624–694 nm), green (506–586 nm), blue (442–515 nm), and NIR (765–901 nm), since we wished to compare the multisensor capabilities of our algorithm against other datasets, such as Sentinel and Planet satellite imagery. Geomorphic classification maps, which classifies according to geomorphology and physical locations, were also provided to us through KSLOF to compare our results against theirs, in which they utilized Definiens eCognition software with the traditional OBIA methods to perform much of their analysis. Note that our classifications are more in line with benthic habitat mapping, as we do not delineate between different geomorphic zones. We have downsampled their 20 or more geomorphic classes down to 9 spectrally distinct classes: coral, sediment, seagrass, waves, deep water, clouds, terrestrial vegetation, beach, and other. An example of this downsampling is shown in Table I. Note that because the geomorphic classes as provided by KSLOF were also largely the result of a classification algorithm, inaccuracies may exist in their segmentation results, and hence, a full truth map is difficult to ascertain. Therefore, we employed additional experts to hand-classify multiple transects as the final test set that we ultimately compare against.

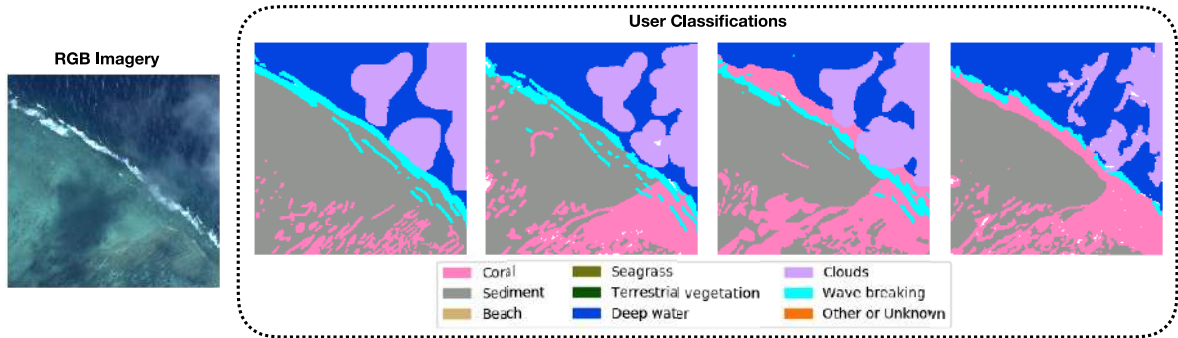


Fig. 2. User classifications as collected from the NeMO-Net citizen science app. Here, four different classifications from separate users for the same area are shown. Note that there exist discrepancies between the input data from each user, although during test time, we found that this difference often enhances the capability of the CNN as it is able to infer areas of high confidence more readily.

We also obtained Planet imagery as part of a NASA effort to gauge the usefulness and applicability of commercial datasets, and it was delivered similarly to that of WV-2 as DN, covering four spectral bands (RGB + NIR), at a per-pixel resolution of 3 m [71]. As will be shown, although the temporal coverage is unparalleled for these datasets, the fact that they suffered from poor calibration posed a challenge to work with.

To provide training data, we collected segmented classifications through our NeMO-Net citizen science active learning app, where users can directly color in areas they believe correspond to specific classes. This is directly shown in Fig. 2, where classifications from four different users are shown for the same area. It is interesting to note that here we see some prominent clouds and cloud shadowing, but from the experiences of each user, they are easily able to filter this out due to their innate knowledge of context, something that is currently still difficult to train an algorithm for. Initially, we were able to collect up to 100 of these classifications internally, mostly covering different regions surrounding Cicia Island. Various methods of cleaning the user data were employed, such as filling in areas that were left unclassified with the most common surrounding classified class, and discarding incomplete or heavily unclassified results.

IV. METHODOLOGY

The overarching goal of NeMO-Net is to facilitate the classification of shallow marine habitats (i.e., coral reefs and seagrass beds) under highly variable spectral and sensor characteristics. To the best of our knowledge, this has not yet been accomplished on a global meter-level scale without significant preprocessing effort and human intervention to align the aforementioned spectral spaces in a consistent manner. We relaxed the problem insofar as to limit the instrument variability (mostly only WV-2 imagery) while allowing for a wide range of spectral variability, with imagery taken often months apart under highly variable conditions mentioned previously (e.g., aerosols, haziness, solar effects, etc). To provide the training data for our algorithm, NeMO-Net utilizes an active learning platform developed in-house on handheld devices to aid in quick classification of coral

reef transects by simply drawing upon the surface of the image [72]. Curation of this data is not within the scope of this article, but the filtered credible results from this application were fed into the CNN.

The NeMO-Net classification algorithm can be summarized into three distinct segments, as shown in Fig. 3. The first segment, preprocessing, attempts to lightly radiometrically calibrate the data by only accounting for factors such as sun angle and sensor characteristics (gain and bias) available from the vendor. Next, we implement the general CNN architecture, which is trained upon the hand-segmented data from the NeMO-Net active learning app. This is done parallel to the land and cloud masking (which is also performed using a similar CNN architecture), the former from available infrared channels and the latter utilizing a CNN cloud and cloud shadow detector developed alongside NeMO-Net [73]. The CNN's priority is to identify the spectral space that the current imagery operates within by focusing on invariant features across all its input data, and to attempt a first-order classification result. The final segment of the algorithm, postprocessing, utilizes the traditional machine learning methods to finalize the classification process utilizing predictions in which the CNN is highly confident in. A conditional random field may be applied after the initial CNN phase and after postprocessing to refine the image boundaries. Note that because of the existence of the postprocessing step, the entire NeMO-Net algorithm cannot be trained end-to-end.

A. Preprocessing

During preprocessing, the DN, are first converted to top of atmosphere (TOA) radiance-based upon vendor calibration values and metadata. Two general corrections are made, one for the spectral channels

$$x_{b,RAD} = \frac{x_{b,DN} * CAL_b}{BW_b} \quad (2)$$

and one for solar angle

$$x_{geo} = \frac{x_{RAD} * d^2}{\cos\left(\frac{\pi}{2} - e\right)} \# \quad (3)$$

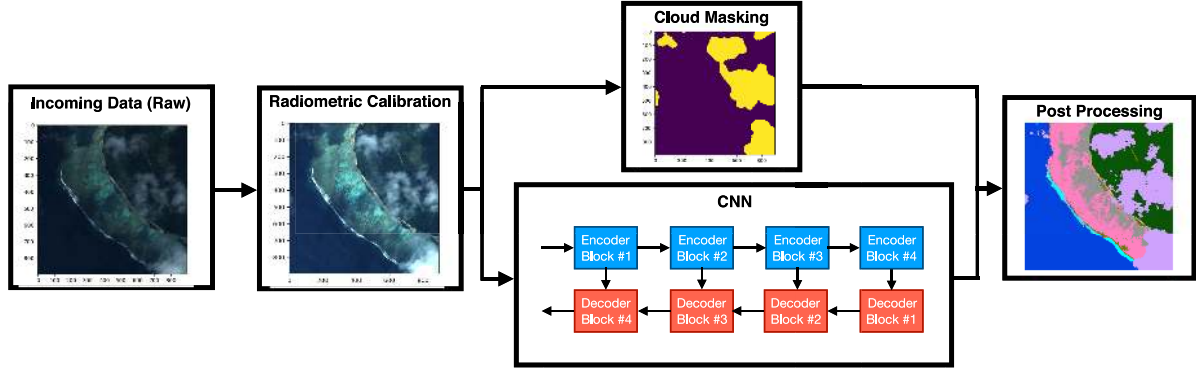


Fig. 3. General flow diagram of the NeMO-Net algorithm. The initial incoming raw data arrive in the format as digital numbers (DN), which is then converted to top of atmosphere (TOA) radiance based upon calibration values and sun angles as given by the vendor. A cloud masking procedure for clouds and cloud shadows is performed in parallel to the CNN section [73], in which the final result is combined from both and passed through to postprocessing.

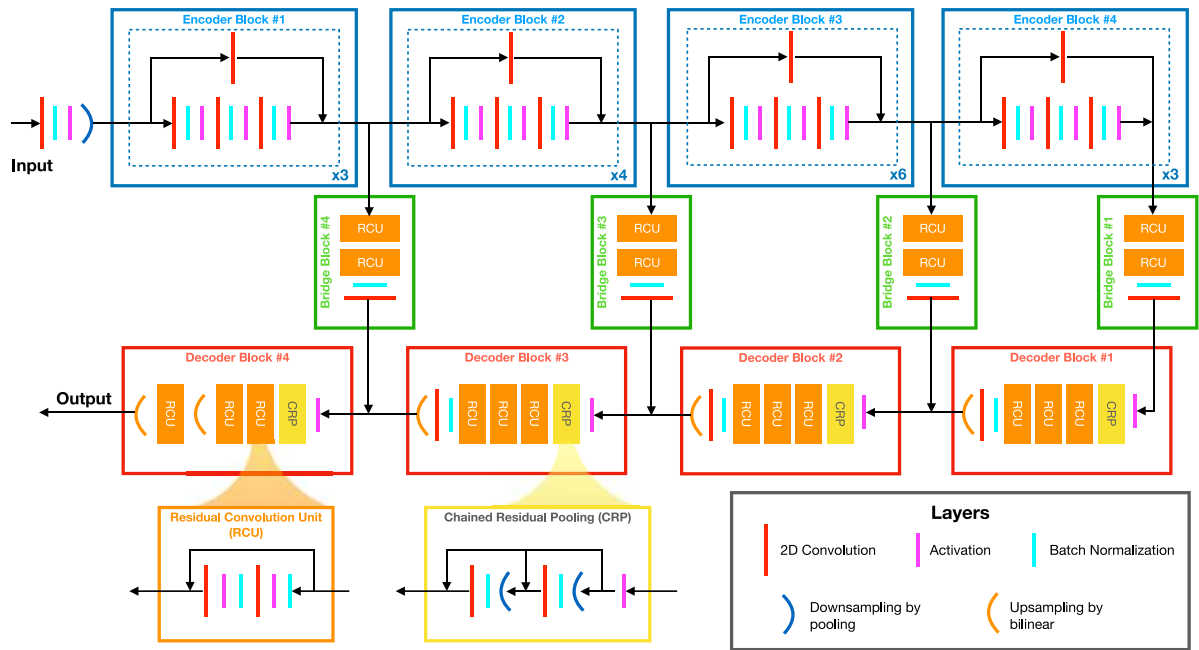


Fig. 4. CNN structure of NeMO-Net. The encoder blocks downsample the data as it extracts deeper and deeper features, the bridge blocks pass the intermediate feature maps to the decoder, and the decoder blocks reconstructs the classification map from the aggregate feature space. Note that the encoder blocks may have slight variations within their inner structure not depicted here, but they follow the general ResNet-50 [55] architecture. The decoder section follows the RefineNet [74] architecture.

where $x_{b,RAD}$ is the TOA radiance value for a specific band, $x_{b,DN}$ is the digital number for a specific band, CAL_b is the band-specific calibration value provided by the satellite vendor to adjust DN to TOA radiance, BW_b is the bandwidth of a specific band, x_{geo} is the sun-angle geometrically calibrated TOA radiance value, d is the sun-earth distance, and e is the sun angle elevation.

Although these corrections were meant to calibrate the data to comparable values across all measurements, there always will remain unaccounted-for phenomena that distort these values, sometimes imperceptibly, which may significantly impact the predictive capabilities of classification algorithms. In other words, these corrections are on the first order, and as such cannot take into account all the variances that exist within datasets taken at different temporal intervals. Yet they provide a

meaningful starting point for utilizing CNNs for remote sensing segmentation and classification, as some semblance of a consistent data product has been attempted.

B. NeMO-Net CNN

NeMO-Net's CNN architecture, shown in Fig. 4, mirrors that of a FCN (see Fig. 1), with an encoder and decoder for achieving deep feature representations and for segmentation reconstruction of the image to its original resolution. Because the input Fiji data are highly variable in size and often huge (on average 5000×5000 pixels in size), as well as the difficulty in obtaining dense segmented truth data, a "patch-based" method to classify the data is utilized through a 256×256 sliding window over the entire transect. The input to the CNN is a

$256 \times 256 \times 4$ image patch with four spectral channels (RGB + NIR), and the output is a classification map of size 256×256 across nine classes. The ResNet-50 encoder section, adapted from [55], is a highly successful architecture within deep feature learning. The name implies a total of 50 convolutional layers, often 3×3 in size, although only 49 of those layers are used within our version (the final layer is often a fully connected layer for classification). Within these layers, we employed ReLU activation and batch normalization layers to improve the training process, with shortcut (or residual) layers that allows for training of deeper features. Each major block is illustrated within Fig. 4, with the output feeding into relevant decoder sections during the upsampling process.

For the decoder, we utilized the general RefineNet [74], [75] architecture, which uses residual convolution units (RCUs) and chained residual pooling (CRP) structures during the upsampling procedure. Each major decoder block takes inputs from the high level but low-resolution features from the previous decoder block combined with the low level but high-resolution features from the convolutional blocks to generate an upsampled feature map to pass to the next decoder block. The final block contains an additional upsampling layer, which segments the image into the relevant classes. Note that although we may use 2-D transpose convolutions for upsampling, simple bilinear upsampling layers were employed here to decrease the amount of required training weights and to speed up the training process.

Particular to this CNN design is the use of RCUs and CRPs, which are constructed to infer greater context from the surrounding areas it seeks to classify. RCUs are similar to residual units used in ResNet, and offer a bypass connection allowing for direct transition between the previous layers and intended outputs without passing through convolutional layers, speeding up training and decreasing the overall dependence on the convolutional weights. CRPs on the other hand pools large areas of the surrounding features successively to form a contextual view of the area, often followed by convolutional layers to determine the importance of said pooling operation. Again, the use of shortcut connections here allows for an alternate independence from CRPs in the case that they are ultimately unnecessary. The combination of residual units, multiresolution fusion, and chained pooling allows for effective segmentation of remote sensing imagery.

Although these classifications are informatively dense, it is ultimately not possible to effectively train a CNN that must predict upon imagery over multiple days under varying conditions with such few numbers of training samples. Even so, it is entirely possible to overfit a model given the vast number of parameters within a CNN and only comparatively few training patches. To remedy this, image augmentation was utilized to vary the imagery such that we inject noise and spectral shifts randomly to each sample image patch while maintaining the same respective user classification. Although this was initially successful for a single island, it was not generalizable to multiple islands and days due to the unrealistic augmentations applied. In reality, the spectral variance between scenes is not entirely random, as the relative radiance between spectral channels does possess some consistency. To address this issue, we applied a

simplified spectral shift calibration between scenes taken on multiple days by focusing on the three major classes that are relevant to our classifications of underwater features: coral, sediment, and seagrass. By taking a few random samples of these classes from multiple days, a multivariable polynomial fit can be constructed such that a spectral transfer between datasets can be realized. Mathematically, this is reflected as

$$\begin{aligned} R_a &= F_1 (R_b, G_b, B_b) \\ G_a &= F_2 (R_b, G_b, B_b) \\ B_a &= F_3 (R_b, G_b, B_b) \\ \text{NIR}_a &= F_4 (\text{NIR}_b) \end{aligned} \quad (4)$$

where R , G , and B represents the RGB channels, a and b subscripts represent the datasets, and F represents the spectral transfer function to map one dataset to another. Special treatment is given to the NIR channel since it behaves differently from the RGB channels, in that it is highly apparent over terrestrial vegetation but falls off dramatically over water due to the high absorption of these wavelengths in the first layers of the water column. We applied a simple second-order polynomial for the fitting function F across every channel to avoid overfitting. In practice, this spectral transfer function can also be approximated using radiometric models that can vary atmospheric and capture parameters from scene to scene, and allow further augmentation of the data. However, for our purposes, the simple polynomial fit was sufficient for our training processes as a fast and easy method in which to implement augmentation. Inclusion of additional random spectral shifts and noise following this mapping, an abundance of augmented data were generated to train upon from a small initial set.

The data are initially organized into their respective classes dependent upon the central pixel of the 256×256 scene. As only a small number of training samples were collected, focus was given more toward areas of fringing reefs where an abundance of coral, sediment, and seagrass classes were evident. Often, these scenes contained deep water, wave breaking, terrestrial vegetation, and beach classes as well, particularly toward the corners or edges of the scenes. The presence of clouds and urban classes were fairly random, the latter often suffering from user misclassifications since their appearance were so infrequent. A mean value of 100 and a standard deviation of 100 was chosen to normalize the spectrally calibrated TOA radiance scene to values between roughly -1 and 1 before ingestion into the CNN. The training process was run over minibatch sizes of 8 scenes per batch, 100 steps per epoch, and over 100 epochs. Parameters for the trainings include the usual cross-categorical entropy loss, an ADAM optimizer [76], and an early stopper interrupt after 30 consecutive epochs with no improvement. We have also tested different loss functions such as focal loss [77] and Lovasz loss [78] due to class imbalance within our datasets, although since we use postprocessing after the initial CNN, these metrics have little effect on the end result. A generator function consistently produced randomly spectrally shifted training sets during runtime, and so an epoch was not inherently tied to the number of classified scenes available. This was similarly done

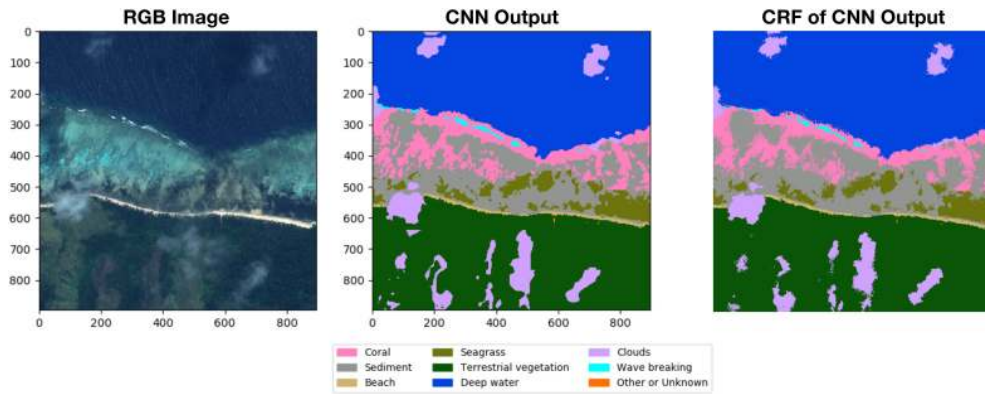


Fig. 5. Sample 896×896 patch showing the RGB image, CNN classification, and conditional random field (CRF) of the CNN output. The classification was generated by taking the central 128×128 classification, while sliding a 256×256 window over a scene of total size 1024×1024 pixels. Note that the CNN output tends to produce classifications with rounded or smooth edges, which can be resolved with a CRF.

for the validation set. The total training time, on average, was 6 hours, utilizing Tensorflow [79] and Keras [80] with an Nvidia 1080 TI GPU.

Parallel to the benthic mapping CNN, we also ran an additional CNN responsible for cloud and cloud shadow masking and detection. Although our main CNN already included a cloud class, we discovered it was more effective to train another CNN unburdened by the benthic classes. We combined through intersection, the results from both CNNs during postprocessing. The details of the cloud masking CNN are detailed in [73].

Although the results from the final trained CNN are more than adequate for segmenting coral habitats (shown in Fig. 5), we noticed that they often generate classifications that are overly smoothed and can still vary over different regions if the contextual information is misleading or not present. The former can be slightly remedied using a conditional random field (CRF) [61], whereas the latter was slightly more difficult to address. The cause is such that although the CNN is spectrally adaptable, large offsets in spectral variation between scenes can still give rise to localized misclassifications due to contextual information often being misinterpreted. For example, slight variations between the TOA radiance can give a lighter appearance to seagrass, resulting in some classifications where they were treated as coral, although their presence close to land should preclude such results. To form a more coherent classification basis across scenes, postprocessing is required such that classifications of high confidence from the CNN is given appropriate treatment, and generalized across the entire dataset.

C. Postprocessing

For the postprocessing algorithm, a K-nearest neighbor (KNN) algorithm was chosen. The reasoning is such that although the CNN is innately adaptable to random spectral shifts, its classification accuracy for boundary cases suffers as a result. In other words, because the CNN is now more tuned to the relative spatial and spectral qualities between classes, it bases its predictions less upon absolute attributes, and so this inherently allows for some uncertainty within data that exhibit qualities that lie close to the boundaries between classes. Furthermore,

if a large number of points could be generated from the CNN output, then the relative nonlinearity between classes could be preserved in regards to the decision boundaries. By combining multiple classifications by sliding the scene window over an area, and by only trusting the central 128×128 patch within the 256×256 scene that is predicts over (essentially discarding the boundaries where spatial contextual information is lacking), it is possible to build up a set of points of high confidence for each class (result shown in Fig. 6).

As we were mainly concerned with coral, sediment, and seagrass, the CNN output is then filtered for these classes where they are represented with high confidence ($>85\%$). Because the number of highly confident points may vary from scene to scene, and because there generally exist more sediments than corals, both of which are more abundant than seagrasses, a relative confidence threshold was set such that the number of coral pixels was roughly three quarters of sediment pixels, and no bounds were set on the number of seagrass pixels. A lower confidence bound of 65% was set for all classes, such that there exists a maximum number of pixels per class possible. From these pixels, we trained a KNN classifier specific to coral, sediment, and seagrass pixels only. The KNN used a distance metric and the closest ten points to determine a class, and thus, all sediment, seagrass, and coral classes were reclassified this way, shown in Fig. 7. A CRF on the KNN classification was applied afterwards so that we decrease the noisiness of the KNN output and refine the boundaries.

D. Application

We tested the capability of the NeMO-Net algorithm upon five scenes of 256×256 patches each. The test area was focused specifically on Cicia Island, but the scenes were taken at differing times under varying conditions, and hence, there exist significant spectral variations across the scenes themselves (shown in Fig. 8). These transects were initially classified on a larger scale of 1024×1024 pixels each such that a mean prediction is generated by stacking smaller 256×256 predictions together. In addition, a larger prediction area was necessary such that the KNN postprocessing step was able to infer the correct

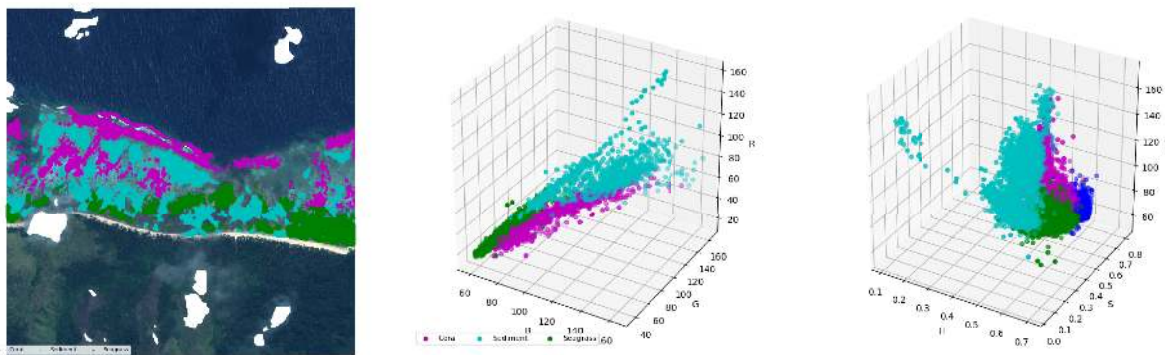


Fig. 6. (Left) Scene where pixels of high confidence (>85%) are illustrated in the following colors: coral (magenta), sediment (cyan), seagrass (green), and clouds (white). (Right) Taking those points of high confidence, a sample representation of the RGB and HSV values of those points are shown (the HSV values also include the deep water pixels shown in blue). There exists a highly nonlinear boundary and significant overlap between classes, which varies from scene to scene.

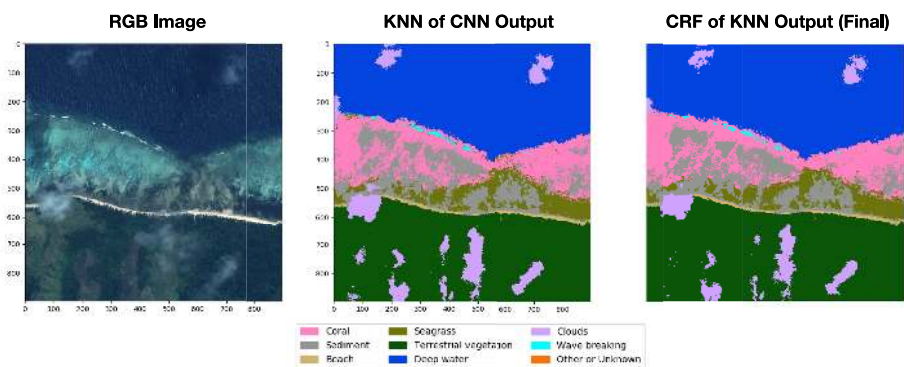


Fig. 7. Sample 896×896 patch showing the RGB image, KNN output and CRF of KNN output derived from the CNN results shown in Fig. 6. The KNN was able to produce a better classification than that of the raw CNN output by utilizing areas of high confidence as predicted by the CNN, although it produced a noisier product. A CRF filter was usually appended after the KNN postprocessing, shown in the rightmost classification.

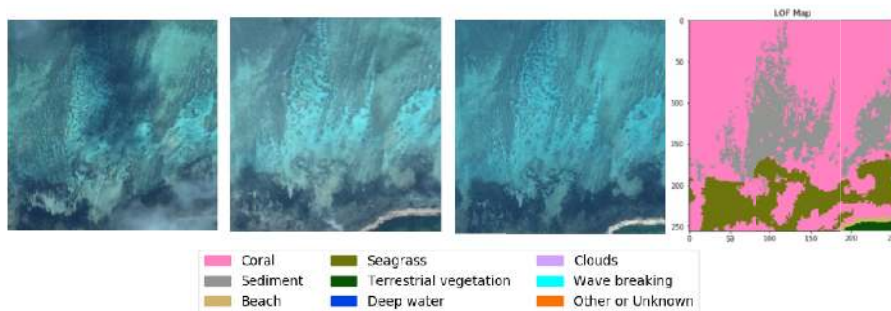


Fig. 8. Scene comparison over multiple days of one specific 256×256 geographic area on the northern side of the Cicia Island. Note that although the location remained the same, due to cloud shadowing, atmospheric effects, time of capture, and sensor calibration, the spectral quality of the transects varied significantly over multiple days. The KSLOF map is included here on the right; note large areas are classified as “coral” (in actuality “back-reef pavement”) and only provides a one-time, static classification of the area.

classification from a larger number of sample points. The central 256×256 patch from the 1024×1024 prediction was taken coinciding with the expert hand-classified segmentation of the same area.

For comparison, we tested multiple traditional CNN architectures utilized for image segmentation against our NeMO-Net algorithm. Specifically, we generated both raw CNN results as well as KNN results derived from the various CNN outputs. The

CNNs attempted in this study include variants of the VGG16-FCN [58], DeepLab v2 [57], SharpMask [75], and NeMO-Net’s RefineNet [74], [75] architectures. We briefly describe each method as follows.

- 1) The VGG16-FCN structure follows closely of the original VGG16, encompassing 16 convolutional layers on the encoder side and a simple bilinear upsampling with convolutional layers on the decoder end. Encoder outputs are

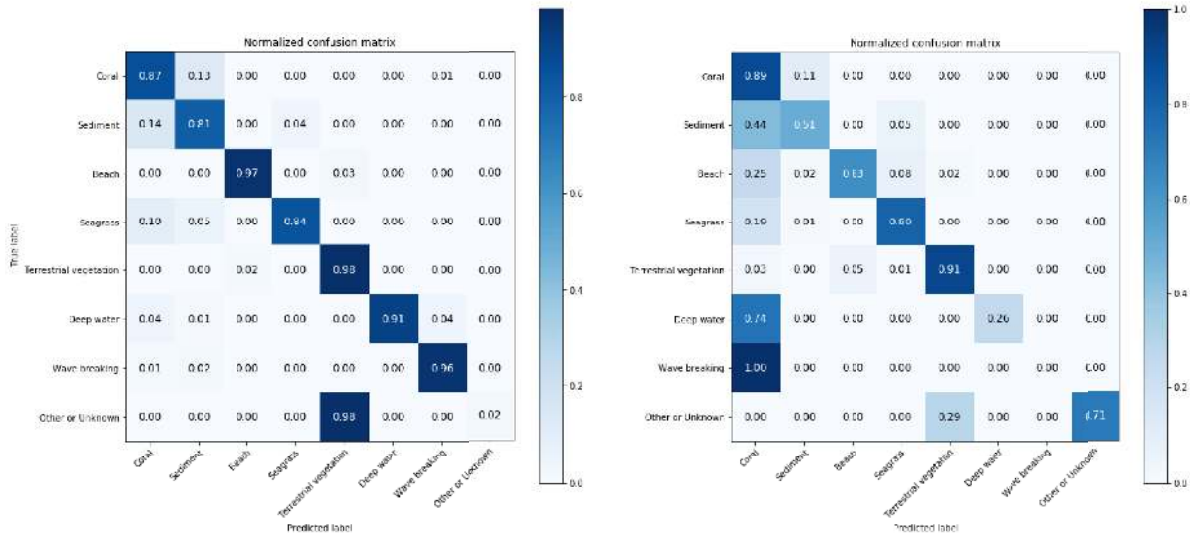


Fig. 9. Normalized confusion matrices for (left) NeMO-Net prediction and (right) KSLOF eCognition results against expert hand-classified segmentation over five 256×256 patches focused on the Cicia Island where little to no cloud or cloud shadowing were present. The apparent higher classification errors of the KSLOF object-based methodology might be due to the intrinsic heterogeneity of reef benthos and classifying large areas as “coral” although these areas might in reality be a mixture of living and dead coral framework.

passed directly to the decoder with little to no operations in between. This structure was one of the first to explore multiresolution combinations for the purposes of image segmentation.

- 2) DeepLab does not explicitly use the FCN encoder–decoder structure. Rather, following the standard VGG16 convolutional layers, the resulting output is split into four parallel channels in which atrous convolution (otherwise known as dilated convolution) is administered with differing stride values. The output of these parallel branches are combined afterwards and a final convolutional layer produces the output. We explicitly utilize an architecture based upon Deeplab v2 within our comparison.
- 3) SharpMask was developed by Facebook for the purposes of object detection and semantic segmentation, and is a stepping stone between the older FCN structures and that of RefineNet. The usual encoder–decoder structure uses a ResNet-50 architecture on the encoder side, with more involved convolutional layers with upsampling procedures on the decoder end.
- 4) NeMO-Net’s CNN is based upon RefineNet, which was covered in detail in Section IV. The most important contribution of RefineNet is the use of RCUs and CRPs within the decoder section to infer additional context during multiresolution fusion and upsampling.

We also evaluated the KSLOF predictions generated from eCognition using OBIA methods, although we note that since large portions of imagery were classified as “back-reef pavement,” which we translate roughly to “coral,” KSLOF predictions can often be misleading particularly for islands that were poorly surveyed.

To fully test the adaptability of the NeMO-Net algorithm, we applied it directly to the 4-band (RGB + NIR), 3-m resolution Planet imagery gathered over Cicia Island over recent years,

focusing on days where little to no cloud cover were present. These datasets were taken with the PlanetScope constellation of satellites, and as such their spectral qualities vary wildly from one platform to the next, often with incomplete calibration values. Because of the lower spatial resolution encountered here, we initially performed a simple sharpening convolution over the image, bringing out high frequency information that the CNN would otherwise overlook. In addition, we performed a slight mean and standard deviation fix to the data (targeted over a specific 512×512 scene), such that the normalization mirrors that of the WV-2 input (recall that WV-2 input used a 100 mean value, 100 standard deviation value to normalize the input data). Beyond these simple modifications, the CNN algorithm with KNN-CRF post-processing was applied as before, with no additional retraining.

V. RESULTS

The resulting confusion matrix for both the NeMO-Net classification as well as KSLOF eCognition results are shown in Fig. 9. Note that we generalized large areas of “back-reef pavement” as “coral” for the purposes of downscaling the number of classes. It has also been in our experience that the KSLOF labels often classified large swathes as “back-reef pavement” even though there exist clear delineations between pavement and sediment areas, possibly due to the fact that a mixture of live and dead coral colonies make up such regions. The truth data utilized here was rigorously classified using the NeMO-Net citizen science app by expert users.

We report for this study the common metrics of mean accuracy, mean precision, mean recall, and frequency-weighted intersection over union (IoU, or otherwise known as the Jaccard Index). The use of the weighted IoU was due to the large class imbalances per scene. We also report these metrics for a

TABLE II
ACCURACY, PRECISION, RECALL, AND FREQUENCY-WEIGHTED IOU METRICS FOR THREE POPULAR CNN-BASED SEGMENTATION ALGORITHMS COMPARED AGAINST NEMO-NET'S REFINE NET-BASED ALGORITHM

Method	Accuracy	Mean Precision	Mean Recall	Frequency-weighted IOU
CNN-CRF only (all classes)				
VGG16-FCN	81.5%	74.0%	77.5%	69.2%
DeepLab	61.0%	62.0%	55.9%	43.6%
SharpMask	79.9%	68.8%	63.0%	66.6%
NeMO-Net (RefineNet) (Focal)	75.9%	73.4%	70.2%	61.5%
NeMO-Net (RefineNet) (Lovasz)	81.2%	70.4%	61.6%	69.6%
NeMO-Net (RefineNet) (CE)	82.1%	68.3%	67.8%	69.9%
Post-processing with KNN (all classes)				
VGG16-FCN	81.0%	73.7%	78.3%	68.8%
DeepLab	78.5%	67.4%	74.0%	65.4%
SharpMask	80.3%	68.3%	64.3%	67.5%
NeMO-Net (RefineNet) (Focal)	80.6%	73.8%	72.8%	67.7%
NeMO-Net (RefineNet) (Lovasz)	83.6%	69.3%	65.4%	72.4%
NeMO-Net (RefineNet) (CE) No CRF	83.5%	67.3%	69.6%	71.8%
NeMO-Net (RefineNet) (CE) with CRF	84.3%	77.6%	79.5%	72.9%
Post-processing with KNN-CRF (Coral, sediment, and seagrass only)				
VGG16-FCN	79.6%	76.2%	82.5%	66.7%
DeepLab	80.7%	77.8%	82.9%	67.9%
SharpMask	79.8%	78.2%	70.9%	66.2%
NeMO-Net (RefineNet)	83.6%	82.6%	84.4%	72.0%
KSLOF Ecognition Prediction	65.1%	68.9%	59.0%	48.6%

Results from three types of loss metrics are displayed: focal loss, Lovasz loss, and categorical cross entropy (CE). CNN-CRF shows the results when only predicting using RefineNet with a CRF filter afterwards, while postprocessing with KNN without CRF, and KNN with CRF results is shown afterwards. The KSLOF eCognition OBIA predictions are given in comparison. The test areas consist of five 4-band, 2-m resolution, 256×256 WV-2 patches where little to no cloud cover and cloud shadowing were present over the Cicia Island. Highest metrics are bolded.

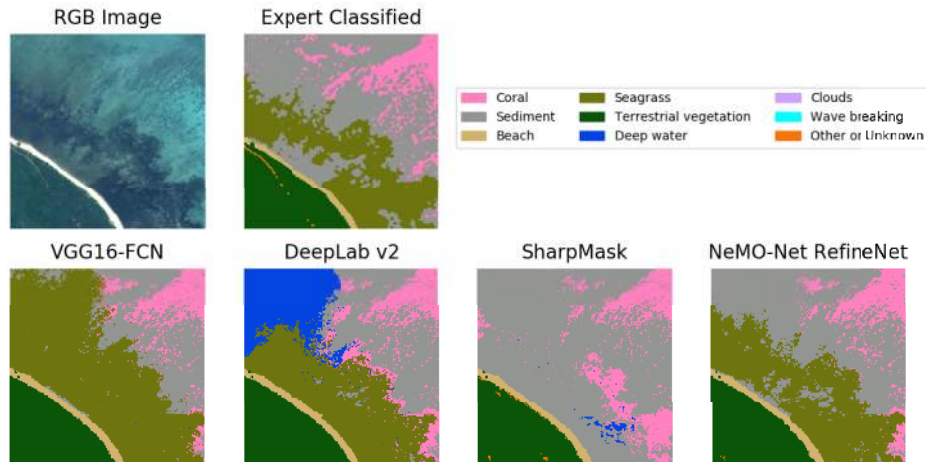


Fig. 10. Segmentation and classification results from all methods across a WV-2 256×256 patch, compared against an expert hand-classified result. In this particular example, VGG16 overestimated seagrass, DeepLab misclassified deep water, and SharpMask underestimated seagrass, while NeMO-Net's RefineNet very closely matches the expert level classification. Results taken from the KNN-CRF method for all classes with cross-entropy loss.

subsample of our three more important classes: coral, sediment, and seagrass, with regards to KNN-CRF postprocessed results across all CNN architectures. All CNNs were trained using the same training data, with identical image augmentation criteria.

The resulting tables are organized as follows.

- 1) Table II shows the results for five 256×256 patches over Cicia Island where little to no cloud cover and cloud shadowing were present. Also displayed are the three types of losses used: focal loss, Lovasz loss, and categorical cross entropy. To show the effectiveness of the CRF after

KNN postprocessing, its results are compared as well to the KNN-CRF method. Fig. 10 shows classification results across all CNNs over one particular 256×256 patch.

- 2) Table III repeats the aforementioned process over eight different 256×256 patches, each taken from a different island (Fulaga, Kobara, Mago, Matuka, Moala, Nayau, Totoya, Tuvuca, and Vanua Vatu) on different days to test the generalizability of the algorithm. Fig. 12 in the Appendix shows sample classified transects from these test sites.

TABLE III
FINAL ACCURACY, PRECISION, RECALL, AND FREQUENCY-WEIGHTED IOU METRICS FOR THREE POPULAR CNN-BASED SEGMENTATION ALGORITHMS COMPARED AGAINST NEMO-NET'S REFINE-NET-BASED ALGORITHM

Method	Accuracy	Mean Precision	Mean Recall	Frequency-weighted IOU
Post-processing with KNN with CRF (all classes)				
VGG16-FCN	79.0%	67.4%	70.0%	66.0%
DeepLab	73.4%	55.2%	54.7%	59.4%
SharpMask	73.2%	57.4%	54.1%	60.2%
NeMO-Net (RefineNet)	83.3%	64.9%	65.8%	71.5%
Post-processing with KNN with CRF (Coral, sediment, and seagrass only)				
VGG16-FCN	82.8%	80.1%	85.4%	70.7%
DeepLab	77.4%	68.5%	82.2%	64.1%
SharpMask	81.6%	77.9%	81.9%	69.1%
NeMO-Net (RefineNet)	85.1%	86.8%	87.9%	74.0%
KSLOF Ecognition Prediction	48.9%	56.9%	44.5%	26.2%

The test areas consist of nine 4-band, 2-m resolution, 256×256 WV-2 patches where little to no cloud cover and cloud shadowing were present over nine geographically diverse Fiji Islands: Fulaga, Kobara, Mago, Matuka, Moala, Nayau, Totoya, Tuvuca, Vanua Balavu, and Vanua Vatu. Highest metrics are bolded. Example transects are shown in Fig. 12 within the Appendix.

TABLE IV
FINAL ACCURACY, PRECISION, RECALL, AND FREQUENCY-WEIGHTED IOU METRICS FOR THREE POPULAR CNN-BASED SEGMENTATION ALGORITHMS COMPARED AGAINST NEMO-NET'S REFINE-NET-BASED ALGORITHM

Method	Accuracy	Mean Precision	Mean Recall	Frequency-weighted IOU
Post-processing with KNN with CRF (all classes)				
VGG16-FCN	78.1%	59.2%	59.2%	67.3%
DeepLab	71.2%	50.7%	52.0%	56.5%
SharpMask	77.2%	56.1%	50.1%	66.4%
NeMO-Net (RefineNet)	79.7%	60.2%	60.9%	68.6%
Post-processing with KNN with CRF (Coral, sediment, and seagrass only)				
VGG16-FCN	72.5%	67.1%	78.4%	58.8%
DeepLab	72.5%	67.9%	75.5%	57.7%
SharpMask	71.0%	65.7%	60.0%	56.4%
NeMO-Net (RefineNet)	73.4%	70.1%	75.2%	58.9%

Only the KNN-CRF postprocessing results are shown for each CNN. The test areas consist of seven 4-band, 3-m resolution, 256×256 PlanetScope patches where little to no cloud cover and cloud shadowing were present over Cicia Island. Highest metrics are bolded.

3) Table IV shows the results from 4-band Planet data for seven 256×256 image patches, where we initially predicted the surrounding 512×512 transect for context, and then, took the center prediction against our expert hand-classified dataset. A sample classification over one specific area is shown in Fig. 11.

Analyzing the results from Table II, we concluded that in general, NeMO-Net's predictions utilizing a RefineNet structure in combination with KNN postprocessing produced the best results across all metrics. Furthermore, applying a CRF filter after the KNN postprocessing yielded even better accuracy, precision, and recall metrics. However, it is interesting to note that although NeMO-Net's raw CNN (i.e., no postprocessing) accuracy fared better than the other architectures, its mean precision and recall was much lower than that of VGG16-FCN's raw CNN results. This was remedied through the postprocessing step, implying that RefineNet may initially misclassify more of the less apparent classes. KSLOF's eCognition prediction using OBIA methods, shown alongside, did not perform as well due to large generalizations and inability to contextualize spatial information.

One major aspect to note is the relative lack of improvement in regards to utilizing the KNN-CRF postprocessing step for both

VGG16 and SharpMask CNNs. This may indicate that these architectures have already saturated their maximum predictive capability at the CNN phase, and relatively little can be done to improve their classification results. Qualitatively, we observed that both VGG16 and SharpMask in general produced less predictions that are of high confidence, resulting in a sparser KNN that did not improve upon the original result. The complete opposite is true for DeepLab, where the raw CNN results were exceedingly poor but the KNN-CRF vastly improved the results during postprocessing. This can be explained by DeepLab's lack of an encoder-decoder structure, rather relying upon atrous convolutions on parallel branches that eventually sum together, after which a simpler bilinear upsampling layer is applied. The lack of multiresolution fusion produces results that are overly smoothed and relatively inaccurate, but the general contextual information is still preserved and can be exploited during postprocessing. NeMO-Net's RefineNet tangentially exploits this capability with CRPs in its decoder sections, allowing for more contextual information to be retained and exploited later.

In regards to coral, sediment, and seagrass predictions, NeMO-Net's algorithm exceeded the other predictions in every aspect. In general, accuracy across these three specific classes

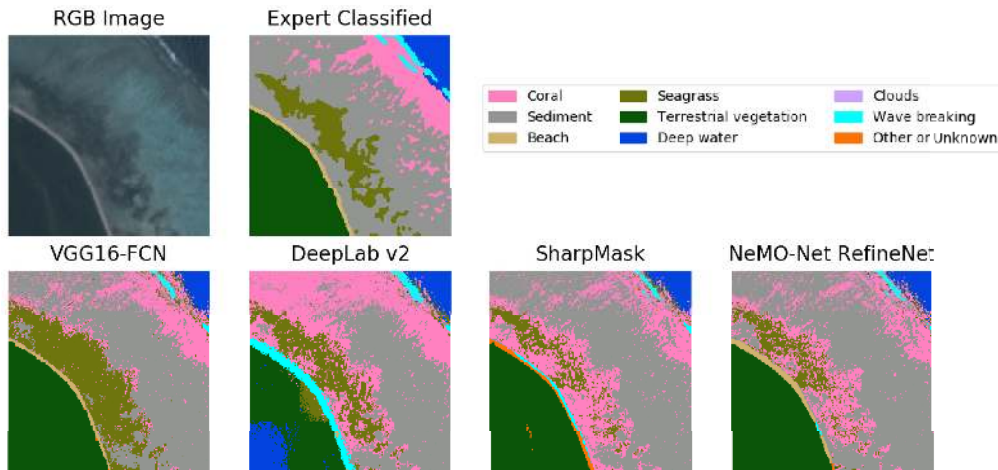


Fig. 11. Segmentation and classification results from all methods across a PlanetScope 256×256 patch, compared against an expert hand-classified result. In this particular example, VGG16 overestimated seagrass, DeepLab misclassified deep water, while both SharpMask and NeMO-Net's RefineNet matched the expert classification more closely. Results taken from the KNN-CRF method for all classes with cross-entropy loss.

decreased compared to the overall accuracy for all CNN architectures with the exception of DeepLab, indicating that each CNN in general was able to make better predictions across classes that are easily distinguishable (e.g., terrestrial vegetation and deep water).

Given results from the various Fijian Islands in Table III, we observed that NeMO-Net's RefineNet outperformed the other CNN architectures in terms of accuracy, as well as in all parameters when only predicting upon coral, sediment, and seagrass. Mean precision and mean recall appeared higher for VGG16-FCN, since it performs better for classes that appear infrequently. It is postulated that due to the complexity of RefineNet, less training data in these less apparent classes lead to incorrect classifications due to little training examples, whereas VGG16-FCN due to its simplicity was able to better tune its weights faster and more accordingly.

For Planet data shown in Table IV, NeMO-Net's algorithm was once again able to outperform the other models, although the results here are closer since all CNNs encounter a certain level of difficulty when attempting to generalize. We observed that the results were also noisier, since the different spectral space forced the boundaries between classes to be less distinct, which caused the KNN to misclassify more often. We observed that all CNNs except for DeepLab were able to generalize fairly adequately across all classes, despite being originally trained upon a dataset from a different instrument and a different resolution altogether.

VI. DISCUSSION

Our attempts at employing CNNs for shallow water benthic classification went through multiple evolutionary phases. The first implementation was pixel-based and took the popular AlexNet [49] and VGG16 [54] architectures, predicting upon a center pixel given a small surrounding context (approximately 32×32). However, it was soon realized that this small spatial context was insufficient to produce accurate classifications, and

predicting at large scales over millions of pixels became increasingly time intensive. Further experimentation led to the use of FCNs and segmentation specific CNNs, with the use of atrous convolutions and encoder-decoder structures. Atrous or dilated convolutions were able to capture larger context within an image, but they were prohibitively expensive as one approached deeper layers with large numbers of features, both in compute time and in memory requirements. The use of augmented data derived from spectral shifts as observed from satellite data was motivated by the fact that the original CNN would perform superbly on images taken on one day, but would fail when attempting to do so for another day. At this point, domain adaptive neural networks (DANNs) [81] were also introduced to handle the domain invariance across image sets, but while they performed well for small-scale CNNs, they were unable to converge for larger scale CNNs that consisted of up to 100 layers. Finally the postprocessing step was introduced as a final method to clean up the predictions as made by the CNN, which would generally achieve overall correct classifications but was lacking in the specific details within an image.

For NeMO-Net's implementation, we specifically avoided much of the preprocessing procedures that traditional classification and mapping algorithms employ, such as detailed corrections for atmosphere, water surface, water attenuation, dark pixel subtraction, and NIR-based corrections. Our purpose was to create an algorithm that did not require these adjustments and could identify the classes based upon relative spatial and spectral context alone. Although successful, we postulate that inclusion of the aforementioned preprocessing steps before ingestion into the CNN would further improve the CNN's capability and accuracy. In addition, while we augmented our data through a simplified polynomial fit to mimic day-to-day spectral variations within the sensor and environment, this component can be replaced with a high fidelity radiometric model that is better able to capture the physics of the environment, as well as allow for further variation beyond the spectral differences observed. However, we caution that all preprocessing inherently does contain some level of noise

and unknown bias, and the purpose of preprocessing is to map all imagery datasets onto one specific feature space where less generalization of the algorithm is required. If this cannot be maintained, then it is better to allow the CNN to infer the context rather than force an incorrect calibration onto the dataset itself.

In regards to the types of loss metrics utilized to train the CNN, there was relatively little difference between the three metrics used: focal loss [77], Lovasz loss [78], and categorical cross entropy (CE), in the final results after postprocessing. Mainly this is because NeMO-Net is not trained end-to-end, and the KNN filter does much of the final fine-tuning work, using only the CNN's relevant probability estimates. We observe that Lovasz loss due to its complexity trains at a much slower pace than the other two, but the results are comparable to CE.

During postprocessing, we observed that the CNN was very accurate for predictions that it is highly confident in, since it is able to perceive the "big picture." However, the CNN might suffer at classifying highly localized areas that are only a few pixels across. As such, we experimented with a number of machine learning techniques to sample from high confidence predictions, such as support vector machines (SVMs) and tree-based methods. SVMs were generally unable to preserve the high nonlinearity between classes, performing at least 5% lower in all metrics, while random forest regression trees (tested with 1000 trees) performed close to that of KNNs, with accuracies up to 83.2% and frequency-weighted IoUs of 71.4%. We settled on the KNN classifier, and although it worked well within our context, it did suffer from noise, as it was unable to infer context unlike a CNN. Eventually, the KNN classifier can be replaced with more traditional OBIA methods, although this is much a topic for future work.

Within NeMO-Net, we observed that although the CNN is able to generalize for modest shifts in spectral space, it cannot compensate for overly dramatic changes that manifest themselves quite frequently in a number of scenarios. First and foremost is the existence of clouds and cloud shadows. While the former is less of an issue due to robust filtering algorithms and the CNNs innate ability to identify clouds, the latter poses a significant obstacle to accurate classification. For our purposes, we employed a separate CNN to identify cloud shadowing, but we realize that a human is often able to easily identify the class being shadowed with relative ease due to their innate understanding of relative context. Another area in which dramatic shifts manifest themselves is the application of NeMO-Net's algorithm to an entirely different satellite, as evidenced through our application to Planet data. Here, the shift was entirely unpredictable, where again significant preprocessing may be required for traditional algorithms to function properly. Although NeMO-Net's CNN was able to generalize, it was able to do so only after slight adjustments were made to the mean and standard deviation of the new dataset. This is not entirely unexpected, as this correction is one of the most basic adjustments necessary for any algorithm to work on entirely new datasets, but it hints at the necessity of obtaining the correct normalization parameters beforehand. This may eventually be derived from specific satellite statistics in its imaging operation, sensitivity, and calibration in a more robust manner.

Finally, our analysis as presented in this article was focused mostly on cloud-free days centered upon fringing corals within Fiji, and more broadly, the Pacific. It should be noted that an abundance of both spectral information and dense contextual spatial data is required to train any CNN properly, and if certain areas are underrepresented, incorrect predictions may result. This was evident in our analysis particularly for deep lagoonal regions, where our training data were insufficient in representing these areas. As a result, deep lagoonal regions would frequently be overclassified as coral since they appeared darker than their surroundings but not as impenetrable as deep ocean water. To remedy this situation, one needs to ensure that the training data encompasses every region that may appear, which may be difficult globally (i.e., Caribbean versus Pacific). In this case, it is recommended that different CNNs be trained to be region-specific, such that geographical context is inherently ingrained during training.

As to the topic of future work, there exists an abundance of topics in which NeMO-Net's algorithm may be improved. As referenced earlier, the KNN postprocessing step can be replaced with the traditional OBIA methods that segment an image into smaller partitions, in which the CNN can aid in classifying. This would decrease the level of noise that the KNN classifier is prone to, while promoting more homogeneous predictions. The addition of different types of data (preprocessing or otherwise) can be incorporated into the CNN as additional channels, such as bathymetry estimates, NIR corrections, or distance from land. With these parameters, it is possible to identify geomorphic zones, either directly through the CNN or during postprocessing. Finally, with the appearance of high temporal satellite data, NeMO-Net may be able to adapt its CNN to incorporate time-series data (such as in long-short term memory, or LSTM) [82] that can correlate between previous predictions to build an improved classification product.

VII. CONCLUSION

The NeMO-Net project was created for the purpose of classifying and mapping coral reef habitats worldwide. To that end, we have developed a CNN-KNN hybrid algorithm that is able to infer contextual spectral and spatial information within a scene to produce predictions on shallow marine benthic cover. The method was based upon the RefineNet encoder-decoder FCN architecture, trained upon 100 original images but augmented to encompass varying spectral effects and day-to-day variances. During postprocessing, a KNN classifier and CRF filter utilized the CNN predictions of high confidence to further enhance the CNN results. We compared our results against expert human-level segmentation over 14 different 256×256 patches taken over different islands on different days under a wide variety of spectral variances, resulting in highly consistent overall accuracies of 83%–85% over nine different classes. We demonstrably showed that a combination of CNNs with traditional machine learning methods was able to yield results that were not only highly accurate, but generalizable across entirely different regimes and even across different satellite instruments entirely.

APPENDIX

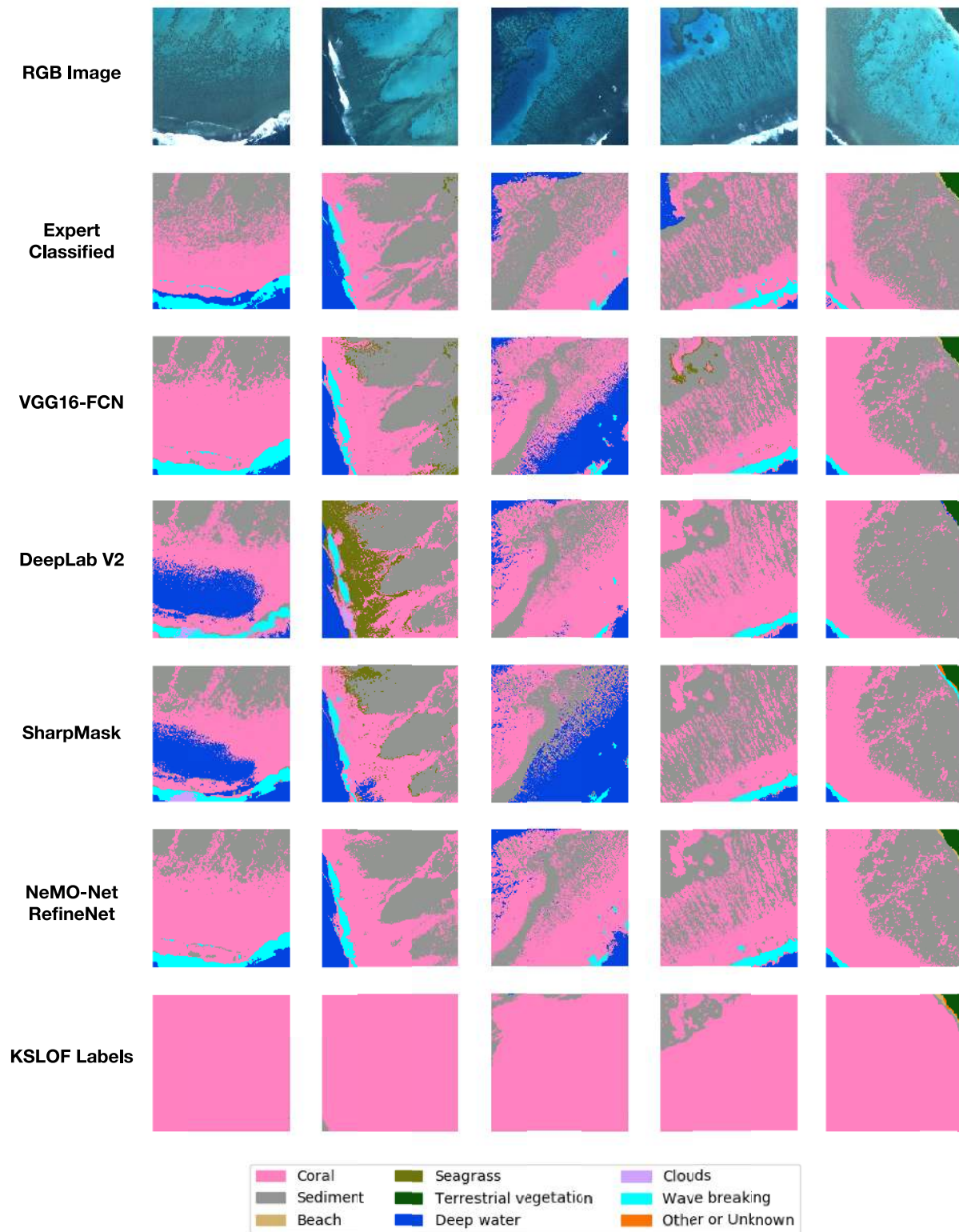


Fig. 12. Comparison between different classifications over a number of 256×256 WV-2 patches taken over a variety of Fiji Islands. The RGB image, expert classified transects, and KSLOF classifications are shown as comparison to the four CNN methods tested within this article.

ACKNOWLEDGMENT

Analysis of DigitalGlobe and Planet data was supported through NASA's 2019 Commercial Data Buy Assessment Program. NeMO-Net's codebase will be made opensource and publicly available upon project completion in 2020. The authors would like to thank the Khaled bin Sultan Living Oceans Foundation (KSLOF), A. Dempsey, and S. Purkis and A. Gleason from the Rosenstiel School of Marine and Atmospheric Science at the University of Miami for providing WV-2 Fiji imagery and habitat classes. They would also like to thank G. Gutman for his aid in arranging NASA's Commercial Data Buy Assessment Program, the app development team for their work on NeMO-Net's citizen science active learning and classification app, and Planet for providing their satellite imagery for the authors' analysis.

REFERENCES

- [1] J. Dill, "Big data," in *Proc. Adv. Inf. Knowl. Process.*, 2019, pp. 11–31.
- [2] J. A. Kleypas and C. M. Eakin, "Scientists' perceptions of threats to coral reefs: Results of a survey of coral reef researchers," *Bull. Mar. Sci.*, vol. 80, pp. 419–436, 2007.
- [3] S. F. Heron, J. A. Maynard, R. Van Hooidonk, and C. M. Eakin, "Warming trends and bleaching stress of the world's coral reefs 1985–2012," *Sci. Rep.*, vol. 6, 2016, Art. no. 38402.
- [4] T. D. Ainsworth *et al.*, "Climate change disables coral bleaching protection on the Great Barrier Reef," *Science*, vol. 352, pp. 338–342, 2016.
- [5] T. P. Hughes *et al.*, "Global warming transforms coral reef assemblages," *Nature*, vol. 556, pp. 492–496, 2018.
- [6] National Academies of Sciences Engineering and Medicine, *Thriving on Our Changing Planet: A Decadal Strategy for Earth Observation from Space*. Washington, DC, USA: The National Academies Press, 2018.
- [7] V. Chirayath and S. A. Earle, "Drones that see through waves—Preliminary results from airborne fluid lensing for centimetre-scale aquatic conservation," *Aquatic Conserv. Mar. Freshwater Ecosyst.*, vol. 26, pp. 237–250, 2016.
- [8] V. Chirayath and A. Li, "Next-generation optical sensing technologies for exploring ocean worlds—NASA FluidCam, MiDAR, and NeMO-Net," *Front. Mar. Sci.*, vol. 6, 2019, Art. no. 521.
- [9] V. Chirayath and R. Instrella, "Fluid lensing and machine learning for centimeter-resolution airborne assessment of coral reefs in American Samoa," *Remote Sens. Environ.*, vol. 235, 2019, Art. no. 111475.
- [10] A. Silver, "Drone takes to the skies to image offshore reefs," *Nature*, vol. 570, 2019, Art. no. 545.
- [11] L. Guild *et al.*, "NASA airborne AVIRIS and DCS remote sensing of coral reefs," in *Proc. 32nd Int. Symp. Remote Sens. Environ., Sustain. Develop. Global Earth Observ.*, 2007, pp. 616–620.
- [12] C. Theriault, R. Scheibling, B. Hatcher, and W. Jones, "Mapping the distribution of an invasive marine alga (*Codium fragile* spp. *tomentosoides*) in optically shallow coastal waters using the compact airborne spectrographic imager (CASI)," *Can. J. Remote Sens.*, vol. 32, pp. 315–329, 2006.
- [13] M. Q. Topping, J. E. Pfeiffer, A. W. Sparks, K. T. C. Jim, and D. Yoon, "Advanced airborne hyperspectral imaging system (AAHIS)," *Proc. SPIE*, vol. 4816, pp. 1–11, 2002.
- [14] P. Mouroulis *et al.*, "Portable remote imaging spectrometer coastal ocean sensor: Design, characteristics, and first flight results," *Appl. Opt.*, vol. 53, pp. 1363–1380, 2014.
- [15] M. R. Corson, D. R. Korwan, R. L. Lucke, W. A. Snyder, and C. O. Davis, "The hyperspectral imager for the coastal ocean (HICO) on the international space station," in *Proc. Int. Geosci. Remote Sens. Symp.*, 2008, pp. IV-101–IV-104.
- [16] G. Dial, H. Bowen, F. Gerlach, J. Grodecki, and R. Oleszczuk, "IKONOS satellite, imagery, and products," *Remote Sens. Environ.*, vol. 88, pp. 23–26, 2003.
- [17] *Sentinel-2 User Handbook*, Noordwijk, the Netherlands, European Space Agency (ESA), Noordwijk, the Netherlands, 2015.
- [18] WorldView-2, DigitalGlobe, Westminster, CO, USA, 2009.
- [19] D. P. Roy *et al.*, "Landsat-8: Science and product vision for terrestrial global change research," *Remote Sens. Environ.*, vol. 145, pp. 154–172, 2014.
- [20] T. Matsunaga *et al.*, "HISUI Status Toward 2020 Launch," in *Proc. Int. Geosci. Remote Sens. Symp.*, 2019, pp. 4495–4498.
- [21] K. Alonso *et al.*, "Data products, quality and validation of the DLR earth sensing imaging spectrometer (DESI)," *Sensors*, vol. 19, 2019, Art. no. 4471.
- [22] S. Andréfouët, M. Claereboudt, P. Matsakis, J. Pagès, and P. Dufour, "Typology of atoll rims in Tuamotu Archipelago (French Polynesia) at landscape scale using SPOT HRV images," *Int. J. Remote Sens.*, vol. 22, pp. 987–1004, 2001.
- [23] S. Andréfouët, F. E. Muller-Karger, E. J. Hochberg, C. Hu, and K. L. Carder, "Change detection in shallow coral reef environments using Landsat 7 ETM+ data," *Remote Sens. Environ.*, vol. 78, pp. 150–162, 2001.
- [24] P. J. Mumby and A. J. Edwards, "Mapping marine environments with IKONOS imagery: Enhanced spatial resolution can deliver greater thematic accuracy," *Remote Sens. Environ.*, vol. 82, pp. 248–257, 2002.
- [25] S. Andréfouët *et al.*, "Multi-site evaluation of IKONOS data for classification of tropical coral reef environments," *Remote Sens. Environ.*, vol. 88, pp. 128–143, 2003.
- [26] E. J. Hochberg and M. J. Atkinson, "Capabilities of remote sensors to classify coral, algae, and sand as pure and mixed spectra," *Remote Sens. Environ.*, vol. 85, pp. 174–189, 2003.
- [27] S. J. Purkis, "A 'reef-up' approach to classifying coral habitats from IKONOS imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1375–1390, Jun. 2005.
- [28] A. Collin and J. L. Hench, "Towards deeper measurements of tropical reefscape structure using the WorldView-2 spaceborne sensor," *Remote Sens.*, vol. 4, pp. 1425–1447, 2012.
- [29] J. D. Hedley *et al.*, "Remote sensing of coral reefs for monitoring and management: A review," *Remote Sens.*, vol. 8, 2016, Art. no. 118.
- [30] C. Roelfsema *et al.*, "Coral reef habitat mapping: A combination of object-based image analysis and ecological modelling," *Remote Sens. Environ.*, vol. 208, pp. 27–41, 2018.
- [31] S. J. Purkis, "Remote sensing tropical Coral Reefs: The view from above," *Ann. Rev. Mar. Sci.*, vol. 3, pp. 149–168, 2018.
- [32] D. R. Lyzenga, "Passive remote sensing techniques for mapping water depth and bottom features," *Appl. Opt.*, vol. 17, pp. 379–383, 1978.
- [33] D. A. Siegel, M. Wang, S. Maritorena, and W. Robinson, "Atmospheric correction of satellite ocean color imagery: The black pixel assumption," *Appl. Opt.*, vol. 39, pp. 3582–3591, 2000.
- [34] T. Cooley *et al.*, "FLAASH, a MODTRAN4-based atmospheric correction algorithm, its applications and validation," in *Proc. Int. Geosci. Remote Sens. Symp.*, vol. 3, 2002, pp. 1414–1418.
- [35] M. L. Zoffoli, R. Frouin, and M. Kampel, "Water column correction for coral reef studies by remote sensing," *Sensors*, vol. 14, pp. 16881–16931, 2014.
- [36] S. Purkis, J. A. M. Kenter, E. K. Oikonomou, and I. S. Robinson, "High-resolution ground verification, cluster analysis and optical model of reef substrate coverage on Landsat TM imagery (Red Sea, Egypt)," *Int. J. Remote Sens.*, vol. 23, pp. 1677–1698, 2002.
- [37] K. A. Call, J. T. Hardy, and D. O. Wallin, "Coral reef habitat discrimination using multivariate spectral analysis and satellite remote sensing," *Int. J. Remote Sens.*, vol. 24, pp. 2627–2639, 2003.
- [38] E. J. Hochberg, M. J. Atkinson, and S. Andréfouët, "Spectral reflectance of coral reef bottom-types worldwide and implications for coral reef remote sensing," *Remote Sens. Environ.*, vol. 85, pp. 159–173, 2003.
- [39] T. Blaschke, "Object based image analysis for remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 65, pp. 2–16, 2010.
- [40] S. L. Benfield, H. M. Guzman, J. M. Mair, and J. A. T. Young, "Mapping the distribution of coral reefs and associated sublittoral habitats in Pacific Panama: A comparison of optical satellite sensors and classification methodologies," *Int. J. Remote Sens.*, vol. 28, pp. 5047–5070, 2007.
- [41] S. R. Phinn, C. M. Roelfsema, and P. J. Mumby, "Multi-scale, object-based image analysis for mapping geomorphic and ecological zones on coral reefs," *Int. J. Remote Sens.*, vol. 33, pp. 3768–3797, 2012.
- [42] C. Roelfsema, S. Phinn, S. Jupiter, J. Comley, and S. Albert, "Mapping coral reefs at reef-to-reef-system scales, 10s-1000s km², using object-based image analysis," *Int. J. Remote Sens.*, vol. 34, pp. 6367–6388, 2013.

- [43] C. M. Roelfsema *et al.*, "Multi-temporal mapping of seagrass cover, species and biomass: A semi-automated object based image analysis approach," *Remote Sens. Environ.*, vol. 150, pp. 172–187, 2014.
- [44] C. Zhang, "Applying data fusion techniques for benthic habitat mapping and monitoring in a coral reef ecosystem," *ISPRS J. Photogramm. Remote Sens.*, vol. 104, pp. 213–223, 2015.
- [45] Z. Wang, N. M. Nasrabadi, and T. S. Huang, "Spatial-spectral classification of hyperspectral images using discriminative dictionary designed by learning vector quantization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4808–4822, Aug. 2014.
- [46] M. Volpi, G. Camps-Valls, and D. Tuia, "Spectral alignment of multi-temporal cross-sensor images with automated kernel canonical correlation analysis," *ISPRS J. Photogramm. Remote Sens.*, vol. 107, pp. 50–63, 2015.
- [47] C. M. Bachmann, T. L. Ainsworth, and R. A. Fusina, "Improved manifold coordinate representations of large-scale hyperspectral scenes," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2786–2803, Oct. 2006.
- [48] D. Tuia, R. Flamary, and N. Courty, "Multiclass feature learning for hyperspectral image classification: Sparse and hierarchical solutions," *ISPRS J. Photogramm. Remote Sens.*, vol. 105, pp. 272–285, 2015.
- [49] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. 25th Int. Conf. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
- [50] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics*, 2014, pp. 655–665.
- [51] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, pp. 529–533, 2015.
- [52] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 807–814.
- [53] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
- [54] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Representation*, 2015, *arXiv:1409.1556*.
- [55] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [56] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9.
- [57] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.* vol. 40, no. 4, pp. 834–848, Apr. 1, 2018.
- [58] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.
- [59] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput-Assisted Intervention*, 2015, pp. 234–241.
- [60] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
- [61] J. Lafferty, A. McCallum, and F. C. N. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proc. 18th Int. Conf. Mach. Learn.*, 2001, pp. 282–289.
- [62] K. Nogueira, O. A. B. Penatti, and J. A. dos Santos, "Towards better exploiting convolutional neural networks for remote sensing scene classification," *Pattern Recognit.*, vol. 61, pp. 539–556, 2017.
- [63] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Oct. 2016.
- [64] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Fully convolutional neural networks for remote sensing image classification," in *Proc. Int. Geosci. Remote Sens. Symp.*, 2016, pp. 5071–5074.
- [65] M. L. Clark and N. E. Kilham, "Mapping of land cover in northern California with simulated hyperspectral satellite imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 119, pp. 228–245, 2016.
- [66] A. King, S. M. Bhandarkar, and B. M. Hopkinson, "A comparison of deep learning methods for semantic segmentation of coral reef survey images," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, 2018, pp. 1507–1515.
- [67] I. Alonso, A. Cambra, A. Muñoz, T. Treibitz, and A. C. Murillo, "Coral-segmentation: Training dense labeling models with sparse ground truth," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2018, pp. 2874–2882.
- [68] A. Mahmood *et al.*, "Automatic annotation of coral reefs using deep learning," in *Proc. MTS/IEEE Monterey OCEANS*, 2016, pp. 1–5.
- [69] S. J. Purkis *et al.*, "High-resolution habitat and bathymetry maps for 65,000 sq. km of Earth's remotest coral reefs," *Coral Reefs*, vol. 38, pp. 467–488, 2019.
- [70] J. Li, S. R. Schill, D. E. Knapp, and G. P. Asner, "Object-based mapping of coral reef habitats using planet dove satellites," *Remote Sens.*, vol. 11, 2019, Art. no. 1445.
- [71] P. Team, *Planet Application Program Interface: In Space for Life on Earth*. San Francisco, CA, USA: Planet, 2017.
- [72] J. van den Bergh, "NeMO-Net citizen science app for coral reef image segmentation," *PLoS One*, to be published.
- [73] M. Segal-Rozenhaimer, A. Li, K. Das, and V. Chirayath, "Cloud detection algorithm for multi-modal satellite imagery using convolutional neural networks (CNN)," *Remote Sens. Environ.*, vol. 237, Feb. 2020, Art. no. 111446.
- [74] G. Lin, A. Milan, C. Shen, and I. Reid, "RefineNet: Multi-path refinement networks for high-resolution semantic segmentation," in *Proc. 30th IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1925–1934.
- [75] R. Kemker, C. Salvaggio, and C. Kanan, "Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning," *ISPRS J. Photogramm. Remote Sens.*, vol. 145, pp. 60–77, 2018.
- [76] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic gradient descent," in *Proc. Int. Conf. Learn. Representation*, 2015. [Online]. Available: <http://arxiv.org/abs/1412.6980>.
- [77] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal Loss for Dense Object Detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2999–3007.
- [78] M. Berman, A. R. Triki, and M. B. Blaschko, "The Lovász-Softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4413–4421.
- [79] M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning," in *Proc. 12th USENIX Symp. Operating Syst. Design Implementation*, 2016, pp. 265–283.
- [80] F. Chollet, "Keras," GitHub, 2015. [Online]. Available: <https://keras.io>
- [81] Y. Ganin *et al.*, "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 1–35, 2016.
- [82] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, pp. 1735–1780, 1997.



Alan S. Li was born in Shanghai, China, in 1987. He received the B.S. degree in mechatronics engineering from the University of Waterloo, Waterloo, ON, Canada, in 2009, and the M.S. and Ph.D. degrees in aeronautics and astronautics engineering from Stanford University, Stanford, CA, USA, in 2011 and 2017, respectively.

In the summer of 2011, he interned with Planet, San Francisco, CA, with a focus on dynamic modeling of satellites. Since 2017, he has been a Research Engineer with the Laboratory for Advanced Sensing, Earth Sciences Division, NASA Ames Research Center, Mountain View, CA. His research interests include next-generation remote sensing technologies and machine learning as applied to remote sensing datasets.

Dr. Li is a member of the American Geophysical Union and was the recipient of the Outstanding Paper Award for Young Scientists at the 41st COSPAR Scientific Assembly, in 2016.



Ved Chirayath (Member, IEEE) received the B.Sc. degree in physics and astrophysics, and the M.Sc. and Ph.D. degrees in aeronautics and astronautics from Stanford University, Stanford, CA, USA, in 2011, 2014, and 2016, respectively.

He is the Director with the NASA Ames Laboratory for Advanced Sensing in Silicon Valley, Ames Research Center, Mountain View, CA. His research is directed at inventing next-generation advanced sensing technologies for NASA's Earth Science Program to better understand the natural world around us,

extending our capabilities for studying life in extreme environments on Earth, and searching for life elsewhere in the universe. He leads a multidisciplinary team developing new instrumentation for airborne and spaceborne remote sensing, validates instrumentation through scientific field campaigns around the world, and develops machine learning algorithms to process big data on NASA's supercomputing facility. He invented the fluid lensing algorithm and is the PI of the NASA FluidCam instrument. He is also the inventor of the MiDAR system for active multispectral remote sensing and the principal investigator (PI) of the MiDAR instrument. As the PI of the NASA NeMO-Net project, he is currently helping create the world's largest neural network for global coral reef assessment using data fusion of fluid lensing data to augment other airborne and spaceborne remote sensing instruments. His work was recently featured in a 2019 special issue publication for *Frontiers in Marine Science*, Next-Generation Optical Sensing Technologies for Exploring Ocean Worlds—NASA FluidCam, MiDAR, and NeMO-Net.

Dr. Chirayath is a member of the American Geophysical Union, Oceanographic Society, Optical Society of America, American Institute of Aeronautics and Astronautics, American Astronomical Society, American Association for the Advancement of Science, and the International Union for the Conservation of Nature.



Michal Segal-Rozenhaimer was born in Tel-Aviv, Israel, in 1975. She received the B.Sc. degree in chemical engineering, and the M.Sc. and Ph.D. degrees in environmental engineering, focusing on the investigation of pesticides fate in the atmosphere using nondestructive methods such as FTIR, and standoff detection of hazardous aerosols by open-path FTIR, from the Technion —Israel Institute of Technology, Haifa, Israel, in 1999, 2002, and 2011, respectively.

After Ph.D., she received the NASA Postdoctoral Fellowship with the NASA Ames Research Center to

work on a new hyperspectral airborne remote sensing instrument and to develop retrieval algorithms for clouds, gases, and aerosols. She continued her work with the NASA Ames Research Center, under the Bay-Area Environmental Research Institute (BAERI) cooperative, working on remote-sensing and machine learning approaches for retrievals of various atmospheric constituents. She is currently holding a dual-affiliation as an Associate Researcher with the NASA Ames research center (with BAERI) and as an Assistant Professor of atmospheric sciences with the Department of Geophysics, Tel-Aviv University, Tel Aviv, Israel. She has authored and coauthored several papers on retrievals of clouds and cloud properties using neural-network approaches.

Dr. Segal-Rozenhaimer is a member of the American Geophysical Union and American Meteorological Society (AMS).



Juan L. Torres-Pérez was born in Puerto Rico, USA, in 1970. He received the B.S. degree in biology, the M.S. degree in geological oceanography, and the Ph.D. degree in biological oceanography, all from the University of Puerto Rico (UPR), San Juan, PR, USA, in 1993, 1998, and 2005, respectively. His major field of study is bio-optical oceanography as it applies to shallow-water tropical marine ecosystems.

He worked as a Coral Reef Specialist with the NOAA National Marine Fisheries Service in Puerto Rico during 2004–2006. Then, he worked as an Aux-

iliary Professor with the Department of Biology, UPR, Rio Piedras Campus. In 2011, he began working with the NASA Ames Research Center as a Postdoctoral Researcher working on the spectrometry of Caribbean coastal and marine organisms. In 2013, he continued working with Ames under a Cooperative Agreement with the Bay Area Environmental Research Institute. He was the Science principal investigator (PI) with a NASA-funded interdisciplinary project on Human Impacts to Coastal Ecosystems in Puerto Rico (HICE-PR), and the PI of a citizen science project in Puerto Rico (CoralBASICS) aimed at training local members of the Recreational Diving Community on the collection of scientific data for coral reef assessment and remotely sensed data validation. He is the Science Advisor with the NASA DEVELOP and a Trainer with the NASA ARSET capacity building programs, and a co-investigator (Co-I) of the NeMO-Net project. He has a number of peer-review publications on coral reef issues, and recently published, along with two other editors, a book on the invertebrate fauna of Puerto Rico.

Dr. Torres-Pérez has been an active member of the US Coral Reef Task Force since 2013 and a former Co-Chair of the Education and Outreach Working Group. He is also a member of the American Geophysical Union and the American Society for Limnology and Oceanography.



Jarrett van den Bergh was born in Atlanta, GA, USA, in 1996. He received the B.Sc. degree in computer science: video game design from the University of California Santa Cruz, Santa Cruz, CA, USA, in 2018.

He became a member with the NASA Ames Laboratory for Advanced Sensing, in 2016. He currently works as a Research Associate with the Bay Area Environmental Research Institute, Moffett Field, CA. Prior to this, he worked as an App Developer with Clarity Products and a Camp Instructor with iD Tech.

He is the Creator of the NeMO-Net classification app and its associated products. He continues to bring his expertise in game design and programming to explore new possibilities within the realm of remote sensing.