







## Article

# Quantifying Seagrass Distribution in Coastal Water with Deep Learning Models <sup>†</sup>

Daniel Perez <sup>1</sup>, Kazi Islam <sup>2</sup>, Victoria Hill <sup>3</sup>, Richard Zimmerman <sup>3</sup>, Blake Schaeffer <sup>4</sup>,  
Yuzhong Shen <sup>1</sup> and Jiang Li <sup>2,\*</sup>

<sup>1</sup> Department of Computational Modeling and Simulation Engineering, Old Dominion University, Norfolk, VA 23529, USA; dpere013@odu.edu (D.P.); yshen@odu.edu (Y.S.)

<sup>2</sup> Department of Electrical and Computer Engineering, Old Dominion University, Norfolk, VA 23529, USA; kisl001@odu.edu

<sup>3</sup> Department of Earth & Atmospheric Sciences, Old Dominion University, Norfolk, VA 23529, USA; vhill@odu.edu (V.H.); rzimmerm@odu.edu (R.Z.)

<sup>4</sup> Office of Research and Development, U.S. Environmental Protection Agency, Washington, DC 20004, USA; schaeffer.blake@epa.gov

\* Correspondence: jli@odu.edu

<sup>†</sup> This paper is an extended version of our paper published in PRCV 2018.

Received: 29 April 2020; Accepted: 13 May 2020; Published: 16 May 2020



**Abstract:** Coastal ecosystems are critically affected by seagrass, both economically and ecologically. However, reliable seagrass distribution information is lacking in nearly all parts of the world because of the excessive costs associated with its assessment. In this paper, we develop two deep learning models for automatic seagrass distribution quantification based on 8-band satellite imagery. Specifically, we implemented a deep capsule network (DCN) and a deep convolutional neural network (CNN) to assess seagrass distribution through regression. The DCN model first determines whether seagrass is presented in the image through classification. Second, if seagrass is presented in the image, it quantifies the seagrass through regression. During training, the regression and classification modules are jointly optimized to achieve end-to-end learning. The CNN model is strictly trained for regression in seagrass and non-seagrass patches. In addition, we propose a transfer learning approach to transfer knowledge in the trained deep models at one location to perform seagrass quantification at a different location. We evaluate the proposed methods in three WorldView-2 satellite images taken from the coastal area in Florida. Experimental results show that the proposed deep DCN and CNN models performed similarly and achieved much better results than a linear regression model and a support vector machine. We also demonstrate that using transfer learning techniques for the quantification of seagrass significantly improved the results as compared to directly applying the deep models to new locations.

**Keywords:** capsule networks; convolutional neural networks; seagrass quantification; transfer learning; deep learning

## 1. Introduction

Seagrass constitutes a significantly important economic, ecological, and social well-being component of coastal ecosystems [1,2]. Economically, seagrass ecosystems are valued 33 and 23 times more than oceanic and terrestrial ecosystems, respectively. Ecologically, seagrass provides numerous benefits such as organic fertilization, sediment trapping, or pollution filtering [2]. However, trustworthy information about seagrass distribution is missing in nearly most of the planet due to the excessive costs of its mapping [2].

This paper analyzes different deep learning approaches for quantification of seagrass in satellite images and compares them against traditional machine learning methods. Specifically, our methods quantify the leaf area index (LAI) of each pixel based on multispectral satellite images. LAI is defined as leaf area per square area [3], and it is a critical biophysical component of seagrass [2]. The LAI index is denoted as a floating number ranging from 0 to 10, with '0' as no seagrass and '10' as the largest seagrass density per area.

Our ultimate goal is to automatically quantify LAI index using satellite images with minimum workforce for field observations. To achieve this goal, we need to address the following two questions:

1. Can we train a deep learning model to successfully predict the level of LAI based on multispectral satellite images?
2. Can we generalize a deep model trained with images from one location to predict seagrass LAI levels at a different location?

To address the first question, we develop two deep learning models for seagrass quantification: (1) a convolutional neural network (CNN) for regression of LAI, and a deep capsule network (DCN) that is optimized jointly for simultaneous classification and regression. To answer the second question, we train the deep learning models with one multispectral image and develop a transfer learning approach to generalize the models to other two images collected at different locations. The multispectral images utilized in this study are WorldView-2 satellite images. Each image has a resolution of 1.24 m and a total of 8 visible and near-infrared (VNIR) bands. An experienced operator labeled some pixels in the images as sea, sand, seagrass, or land and applied a physics model [4] to seagrass pixels to obtain LAI index. The physics model has reported an error rate of 10% [4], making it not completely suitable as ground truth for training. To resolve this challenge, an experienced operator identified several regions where the LAI mappings are reliable, and those LAI mappings are used as ground truth. The major contributions of this paper are:

- Two deep learning models for regression of seagrass LAI that outperform traditional methods for regression.
- A transfer learning approach that performs seagrass LAI level mapping at a new location with minimum workforce site observation.

The remainder of the paper is organized as follows. A review of the literature is provided in Section 2. The proposed methods are described in Section 3. Results of the proposed method are presented in Section 4. Finally, discussions and conclusions are given in Sections 5 and 6, respectively.

## 2. Related Work

### 2.1. Deep Learning

Deep learning models achieved superb results in different areas in recent years such as object detection and tracking [5–8], image classification [9–13], remote sensing [12,14–18], speech recognition [19,20], autonomous driving [21,22], cybersecurity [23,24], and medical imaging [25]. Among various structures, CNNs are the most popular models for different applications. A CNN model consists of convolutional layers for feature learning. The convolutional layers are usually followed by fully connected layers to perform classification or regression. Many CNN based image classification methods are able to perform end-to-end learning where feature learning and classification are jointly optimized, and it has been proved that this feature learning capability is a key contributing factor to the success of CNNs.

Though regression problems using CNNs are not as common as classification, some examples can be found in the literature. CNNs have been utilized for regression by adding a regression layer on top of the convolutional layers. Niu et al. [26] proposed a CNN model for age estimation which consists of  $K - 1$  output layers with each performing a sequential classification of whether the input facial

image's age was older than a rank of  $K$ , thus transforming the classification model into an ordinal regression model. Yuan et al. used a generic 2D CNN regression model for simultaneous face detection and segmentation and achieved competitive results [27]. Girshick developed a fast region-based CNN method for object detection [28] where the CNN model produced two outputs: a softmax probability for predicting the presence of the object, and regression outputs defining the position of the detected object. A similar method was used by Gidaris and Komodaris in [29], where they developed a CNN model for bounding box regression, which allowed the authors to refine the location of the detected objects. CNN models can also be used for image synthesis. In [30], Li et al. developed a CNN model to generate Positron Emission Tomography (PET) images from Magnetic Resonance Imaging (MRI) images to improve Alzheimer's disease diagnosis.

## 2.2. Capsule Networks

Capsule networks were introduced in late 2017 by Sabour et al. [31]. In these models, neurons in filter maps are grouped to form a set of capsules, which represent instantiation parameters of an entity in a given image, and information between different capsule layers are communicated through routing. The first implementation of capsule networks achieved a 99.75% accuracy on the MNIST dataset, which still represents the state of the art. Capsule networks have two unique properties as compared to CNNs: being able to identify overlapped objects in images and capable of performing simultaneous classification and regression. The same research team recently proposed a modification of their capsule network and obtained state-of-the-art results on smallNORB [32].

The last capsule layer of a DCN model comprises a set of capsule vectors, where each vector corresponds to one class in the training dataset and the length of the vector is treated as the posterior probability of the class for classification. In addition, the model reconstructs each input image using the corresponding capsule vectors. These reconstructed images are used for regularization during the training process. The errors between input images and their reconstructions are then backpropagated to optimize all the weights in the network. The unique configuration of DCN makes it able to perform classification and regression simultaneously. Sabour et al. demonstrated that the reconstruction stage is also an important contributor to the superb results obtained by the model applied to MNIST [31].

Recently, DCN models have been applied to more complex data. The application of DCN models to the CIFAR-10 dataset was studied in [33], where the authors obtained an accuracy of 77.55%. This performance is significantly worse than the current state-of-the-art results (96.53%). In the medical image analysis field, it has been demonstrated that capsule networks outperform CNNs in different tasks such as classification of brain tumor type [34], diagnosis of thoracic disease [35] and reconstruction of image stimuli from functional MRI [36]. In [37], the authors showed how a capsule network could be successfully implemented in the deep reinforcement learning framework to create intelligence agents in games. Additionally, LaLonde and Bagci applied a capsule network to an object segmentation task [38] and showed that the number of parameters of the capsule network can be reduced by 94.5% as compared to the traditional design, while still improving its accuracy. In our recent study [39], a DCN model was implemented as a generative model to readjust a trained capsule network for classification of seagrass at different locations. To the best of our knowledge, our study is the first attempt to apply capsule networks for seagrass quantification.

## 2.3. Seagrass LAI Mapping

Besides our previous studies on seagrass identification [39] and LAI mapping [40], research in the literature on seagrass mapping is mostly focused on analyzing performances of manual mapping approaches [41,42]. A remote sensing method was developed by Yang et al. [43]. Instead of quantifying the seagrass distribution in satellite images, they manually determined whether seagrass was presented in a given region and achieved an accuracy slightly better than 80%. A few works proposed automatic methods for seagrass quantification. For example, Wicaksono et al. implemented an automatic algorithm for seagrass LAI mapping and achieved a mean square error (MSE) of 0.72 [2]. In [44],

Pu et al. implemented a regression model for LAI quantification that achieved MSEs of 0.78 and 0.59 using data taken by Hyperion (HYP) and Advanced Land Imager (ALI) satellites, respectively. In [45], Dierssen et al. developed a remote sensing strategy to estimate LAI levels of seagrass with MSEs ranging from 0.88 to 0.98. To the best of our knowledge, our team has developed the first approaches for classification [39] and regression [40] of seagrass. In this paper, we focus on the regression part and analyze which deep learning model is more suitable for LAI regression.

#### 2.4. Transfer Learning

Transfer learning aims to solve problems in a target domain where training data samples are limited by using a model trained in a similar source domain where rich training data is available [46]. DeCaf is among the first successful implementations of transfer learning, where Donahouse et al. [47] generalized a CNN model trained on ImageNet [9] to different domains and achieved state-of-the-art performances. In [48], Yosinki et al. froze the first few layers in a model trained in the source domain and re-trained the other layers with labeled data from the target domain. They achieved better accuracies by fine tuning.

While remote sensing images may be widely available, reliable mapping and labeling are often missing due to the associated excessive cost [2,49,50]. For this reason, transfer learning is a common tool used in remote sensing applications. For instance, Hu et al. [51] transferred knowledge from existing pre-trained CNNs on the ImageNet dataset [9] to high-resolution remote sensing images for scene classification, improving accuracy by about 2.5–5% as compared to the state-of-the-art methods that only focus on exploring low-level hand-crafted features. Xie et al. were able to transfer knowledge from a pre-trained CNN model [9] to predict nighttime light intensity from daytime satellite imagery and to estimate poverty with an accuracy of 76.1% [49]. Jun et al. [50] demonstrated how an active learning model for characterizing land cover in satellite images can be trained in one region and then be applied to different regions. Their results showed that the active learning approach produced better results than applying the model directly to the regions with limited training samples.

Transfer learning has been applied to other deep models in the literature. In [52], the authors demonstrated that the performance of a deep belief network (DBN) for post-traumatic stress disorder (PTSD) diagnosis could be significantly improved using transfer learning. Chowdhury et al. [24] developed a few-shot deep learning approach for intrusion detection, in which they extracted features of a small dataset from some trained CNN and DBN models, and trained a simple classifier to improve the intrusion detection performances. Transfer learning with CNN for regression problems has been studied in [49,53]. However, transfer learning with capsule networks has not been investigated for seagrass quantification.

### 3. Methodology

#### 3.1. Datasets

Three multispectral images taken by the WorldView-2 satellite at three different coastal locations in Florida are utilized in this study. The images have eight spectral bands with a resolution of 1.24 m, as shown in Figure 1. The images have spatial sizes of  $12,208 \times 6717$ ,  $8962 \times 7227$  and  $6143 \times 9793$  pixels, respectively. A patch of  $5 \times 5 \times 8$  is extracted for each pixel centered in the patch. The patch then will be classified as sea, land, seagrass, or sand. Additionally, the physics model [4] computed LAI index for each pixel.

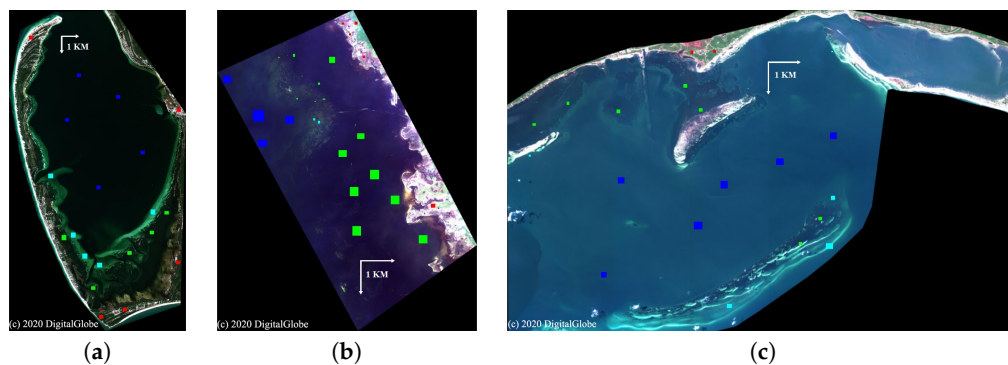
#### 3.2. Data Labeling

The physics model reported a 10% error for LAI mapping [4]. Therefore, we do not use the whole mapped images as ground truth to train our models. However, some regions in the LAI mappings by the physics model are considered to be more reliable than others. In our study, several regions in the images where the LAI mappings are more accurate are selected by an experienced operator

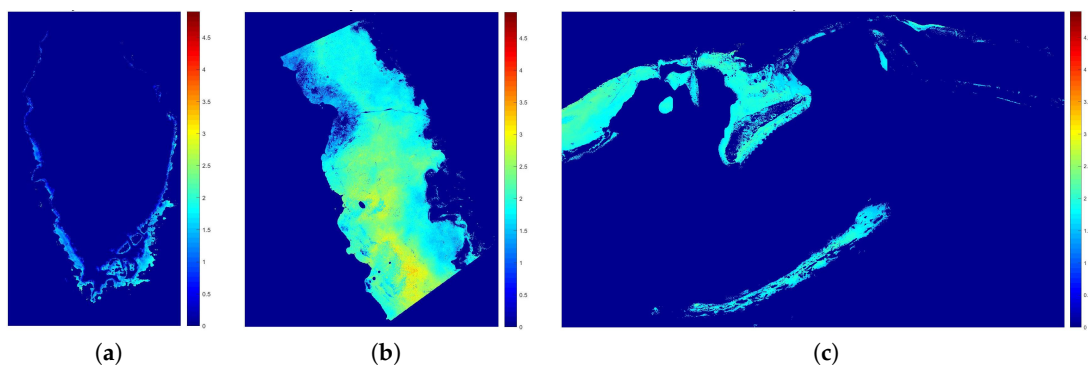
(co-author of the physics model in [4]). These regions are treated as ground truth for training the deep models. Figure 1 shows the selected regions where cyan, blue, red and green boxes represent sand, sea, land, and seagrass, respectively. Additionally, Figure 2 shows the LAI mappings of the whole images. When we train our models, we only use the selected regions highlighted in Figure 1. Table 1 shows the number of pixels in the selected regions per class in each of the satellite images used in the study. Note that the labeled pixels in the selected regions are unbalanced. To address this issue, we balance the training samples by randomly downsampling majority classes and upsampling minority classes to ensure that each class has roughly the same number of training examples.

**Table 1.** Number of patches per class in the selected regions of each satellite image.

Label	St. Joseph Bay	Deckle Beach	St. George Sound
Sea	108,675	240,361	104,094
Land	16,304	7642	23,317
Seagrass	120,375	137,210	26,573
Sand	108,167	34,059	5914



**Figure 1.** Images taken by the WorldView-2 satellite from (a) Saint Joseph Bay, (b) Keeton Beach and (c) Saint George Sound. The images were taken on 14 November 2010, 20 May 2010 and 27 April 2012, respectively. Selected sand, sea, land and seagrass regions are represented by cyan, blue, red and green boxes, respectively.



**Figure 2.** Mappings of seagrass LAI level obtained by the physics model [4] at (a) Saint Joseph Bay, (b) Keeton Beach and (c) Saint George Sound.

### 3.3. Joint Optimization of Classification and Regression in Capsule Networks for Seagrass Mapping

We design a DCN model for simultaneous classification (sea, sand, seagrass, land) and regression (LAI mapping) in multispectral satellite images as shown in Figure 3. Inputs of the model are image patches of size  $5 \times 5 \times 8$ . The first layer is a convolutional layer with  $32 \ 2 \times 2 \times 8$  filters and a stride of 1. This layer is followed by a second convolution layer with  $64 \ 2 \times 2 \times 32$  kernels and a stride of 1. The output of this layer is organized as 8 blocks of capsules of size  $3 \times 3 \times 8$  in the *PrimaryCaps* layer. The reconstruction part of the original DCN model [31] is replaced by a linear regression layer

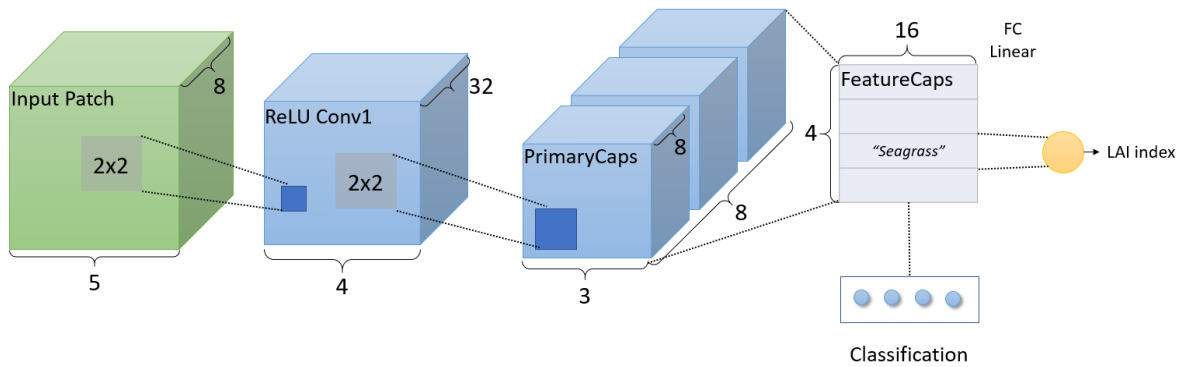


for seagrass mapping. This layer quantifies the LAI level of seagrass based on the seagrass capsule vector from *FeatureCaps*. We define LAI of an image patch as the LAI of its center pixel. This structure allows us to jointly optimize LAI regression and seagrass classification. The *FeatureCaps* layer performs classification for the four classes (sea, land, seagrass, and sand) with a separate margin loss for the  $k$ th class [31]:

$$L_k = T_k \max(0, m^+ - \|\mathbf{v}_k\|) + \lambda(1 - T_k) \max(0, \|\mathbf{v}_k\| - m^-)^2$$

where  $T_k = 1$  if the class of  $k$  is present,  $m^+ = 0.9$ ,  $m^- = 0.1$  and  $\mathbf{v}_k$  is the magnitude of the  $k$ th vector in *FeatureCaps* representing the posterior probability for the  $k$ th class.  $\lambda$  is set as the default value of 0.5 and the total loss for the four classes is the sum of each individual loss.

We set the number of routings from the *PrimaryCaps* layer to the *FeatureCaps* layer in the DCN model to 3. During training, if a seagrass image patch is fed as input, the seagrass vector in the *FeatureCaps* layer is used to train the regression model for LAI quantification. Then, the error of the regression is used during back-propagation to update the weights of the DCN model, jointly optimizing classification and regression. For other types of image patches (sea, sand, land), the regression step is skipped, and only the classification model is optimized.

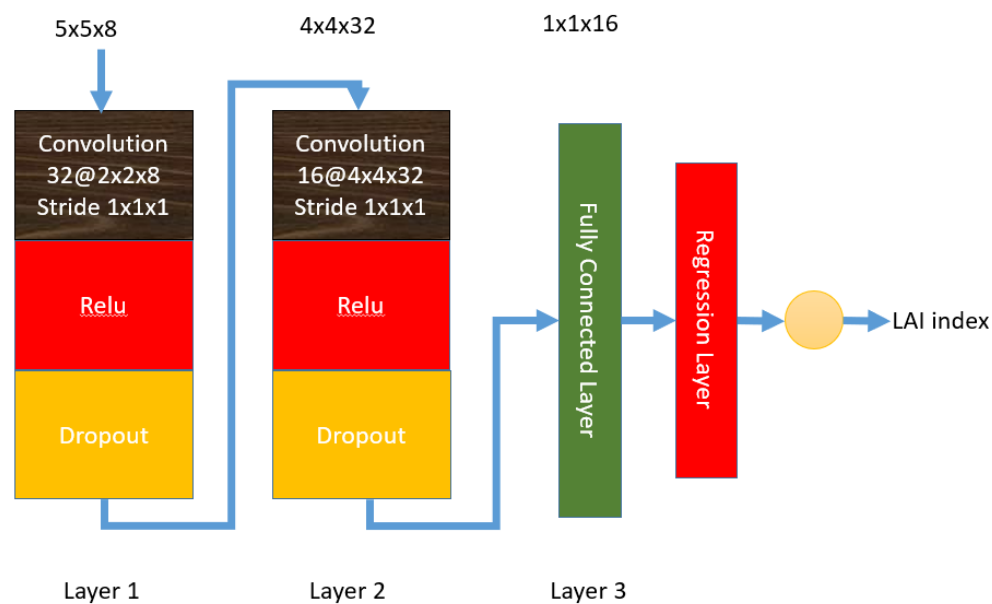


**Figure 3.** Diagram of the proposed deep capsule network for end-to-end seagrass identification and LAI mapping.

### 3.4. Convolutional Neural Network for Seagrass Mapping

Figure 4 shows the CNN model implemented for regression of LAI. The CNN Model has 2 convolutional layers for representation learning. The first convolutional layer has 32 kernels with a size of  $2 \times 2 \times 8$ , and the second layer has 16 filters of size  $4 \times 4 \times 32$ . The fully connected layer has a total of 16 hidden units, which matches the size of the vectors in the *FeatureCaps* layer in the DCN model. The last layer uses this representation to compute LAI through linear regression.

Additionally, we develop an SVM model and a linear regression model to quantify LAI based on the image patches directly. These models offer baseline performances for comparison.



**Figure 4.** CNN structure for LAI regression using 8-channel pan-sharpened multispectral images.

### 3.5. Transfer Learning for Seagrass Mapping at Different Locations

Seagrass distribution differs significantly among different locations, making it challenging to generalize a regression model trained at one location to new locations. Our transfer learning approach uses the features from *FeatureCaps* and generalizes a trained DCN model to a different location with minimum information from the new location. Specifically, the transfer learning approach for DCN model consists of the following steps:

1. Train a DCN model with all labeled samples from the selected regions in the satellite image taken at St. Joseph Bay (Figure 1a).
2. Select a small portion of the training samples from the satellite image taken at Keeton Beach.
3. Classification Step:
  - (a) Pass the labeled samples through the trained DCN model as shown in Figure 3 and output the 64 features from the *FeatureCaps* layer as new representations for the labeled samples.
  - (b) Use the labeled new representations to classify the rest of the unlabeled samples from Keeton Beach using 1-nearest neighbor (1-NN) rule.
4. Regression Step:
  - (a) Use the seagrass vector (16 features) in the labeled new representations from Keeton Beach to train a linear regression model to quantify LAI levels of seagrass.
  - (b) For every unlabeled patch that is classified as seagrass by the 1-NN rule, predict its LAI value using the linear regression model trained in the previous step. LAI for every non-seagrass patch is set to '0.'
5. These procedures are repeated for the image taken at St. George Sound for LAI prediction.

The transfer learning approach is also applied to the CNN model. When performing transfer learning with CNN, we extract the features from the last fully connected layer (16 features) in both the classification and regression step. The other parts of the transfer learning approach are identical to the method using the DCN model.

## 4. Experiments and Results

### 4.1. Model Structure Determination

We determine the hyper-parameters of the proposed models through 3-fold cross-validation (CV) in the patches from the selected regions in the image taken at St. Joseph Bay (Figure 1a). After several experiments, we conclude that the best patch size is  $5 \times 5 \times 8$  pixels and the DCN model has two convolutional layers with 32 and 64 filters of size  $2 \times 2$ , respectively. To ensure a fair comparison, we choose that both the DCN and CNN models have roughly the same number of parameters for representation learning. Specifically, each model has around 9000 parameters for feature learning and 17 parameters for LAI regression (including bias).

It is worthy to note that though the numbers of parameters for representation learning are the same in CNN and DCN, there are around 38,000 parameters in DCN's capsule layers for routing. In total, there are approximately 47K parameters in the DCN model, which are nearly four times more parameters than that in the CNN model.

### 4.2. Cross-Validation in the Selected Regions

In our first experiment, we determine if the proposed models are able to quantify LAI for seagrass in the regions selected by the experienced operator. We perform 3-fold CV in the selected regions in each satellite image as shown in Figure 1. We train both deep learning models until their learning losses converge, which generally happens before 100 training epochs. We use root mean squared error (RMSE) as our metric to assess the performance of each model. Table 2 shows the results for each model. It can be seen that the deep learning models (CNN and DCN) outperform linear regression and SVM. The performances of CNN and DCN are similar, but generally CNN produces the best results, achieving an average RMSE of 0.19.

**Table 2.** RMSEs obtained by *three-fold* CV in the selected regions. Best results are highlighted in bold.

Image	Linear Regression	SVM	CNN	DCN
St. Joseph Bay	0.58	0.57	<b>0.45</b>	0.46
Keeton Beach	0.16	0.16	<b>0.04</b>	0.07
St. George Sound	0.12	0.10	<b>0.08</b>	0.12
Mean	0.29	0.28	<b>0.19</b>	0.21

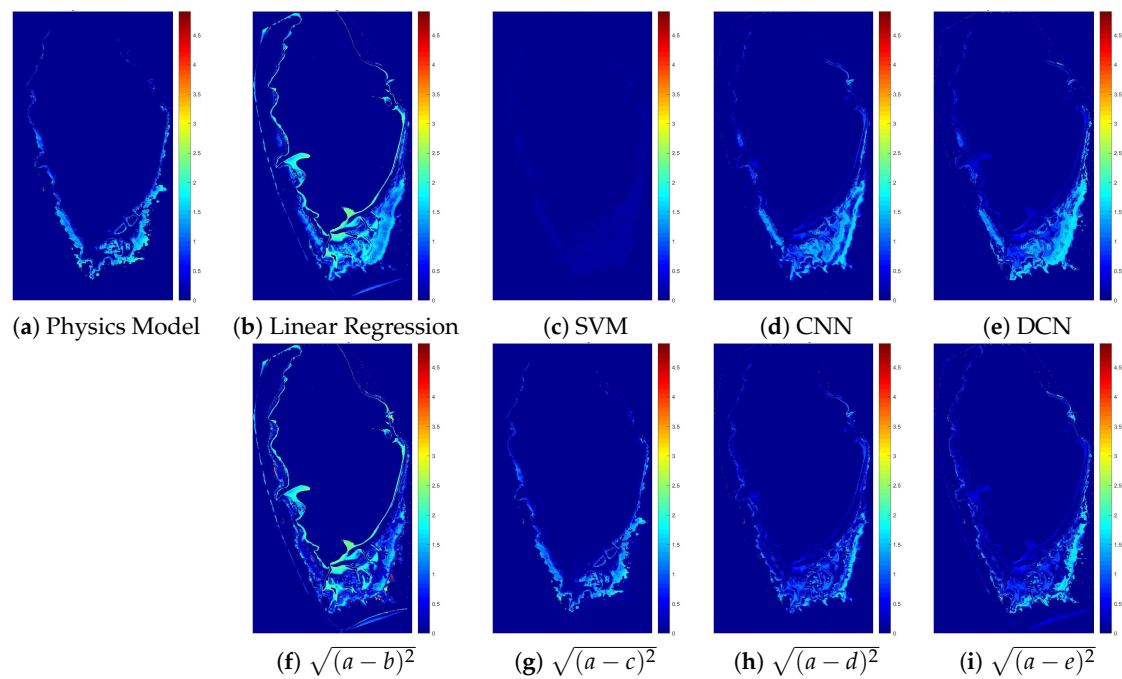
### 4.3. End-to-End LAI Mapping

To perform end-to-end mapping, we train the deep learning models using image patches from the selected regions and predict LAI for the whole image. During training, the DCN model first classifies a given patch as sea, sand, seagrass, or land. Then, it sets LAI to '0' if the patch is classified as non-seagrass. If the patch is classified as seagrass, it is then mapped to the predicted LAI index. Using this method, the model is able to perform end-to-end mapping by jointly optimizing classification and regression.

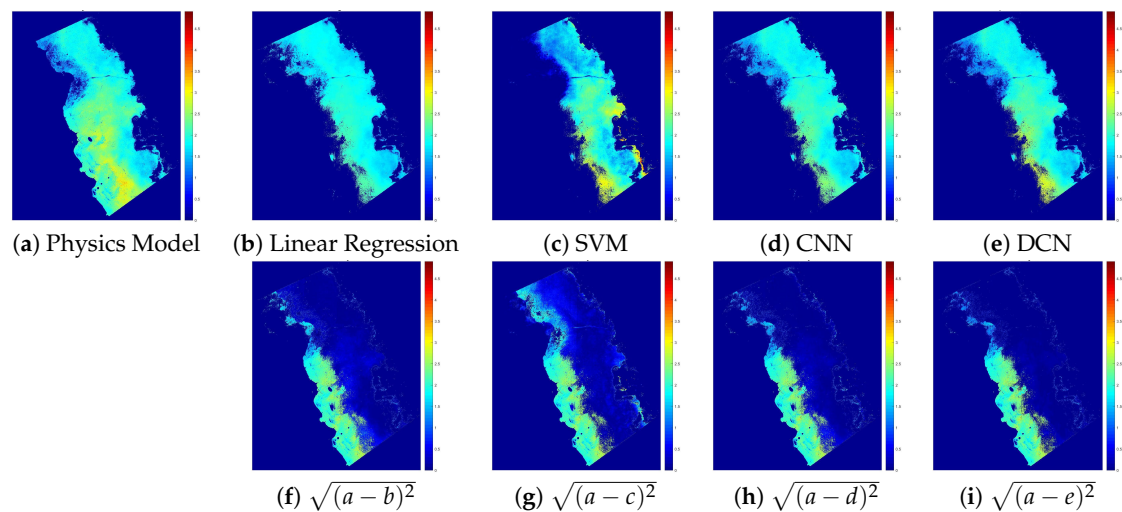
We perform two experiments to illustrate the effect of end-to-end learning. In the first experiment, the direct linear regression model, SVM, and CNN are trained with seagrass patches only as pure regression models, but the produced LAI mappings are masked out by the classification results of the capsule network (Figures 5–7) as they would not be able to correctly quantify LAI for non-seagrass patches. In the second experiment, we use seagrass and non-seagrass (LAI = 0) patches from the selected regions to train our models and obtain 'raw' mappings of LAI with no mask applied to them (Figures 8–10). It is important to note that, since the LAI quantification obtained by the physics model is not considered as ground truth, these figures are not reliable indicators of the performance of our



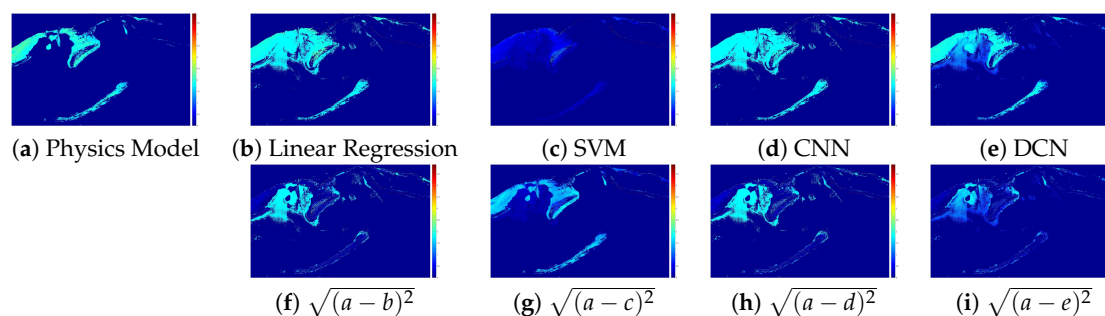
model and are shown in this paper only for visualization purposes. RMSEs obtained in the selected regions (Table 2) are true performance metrics for comparison.



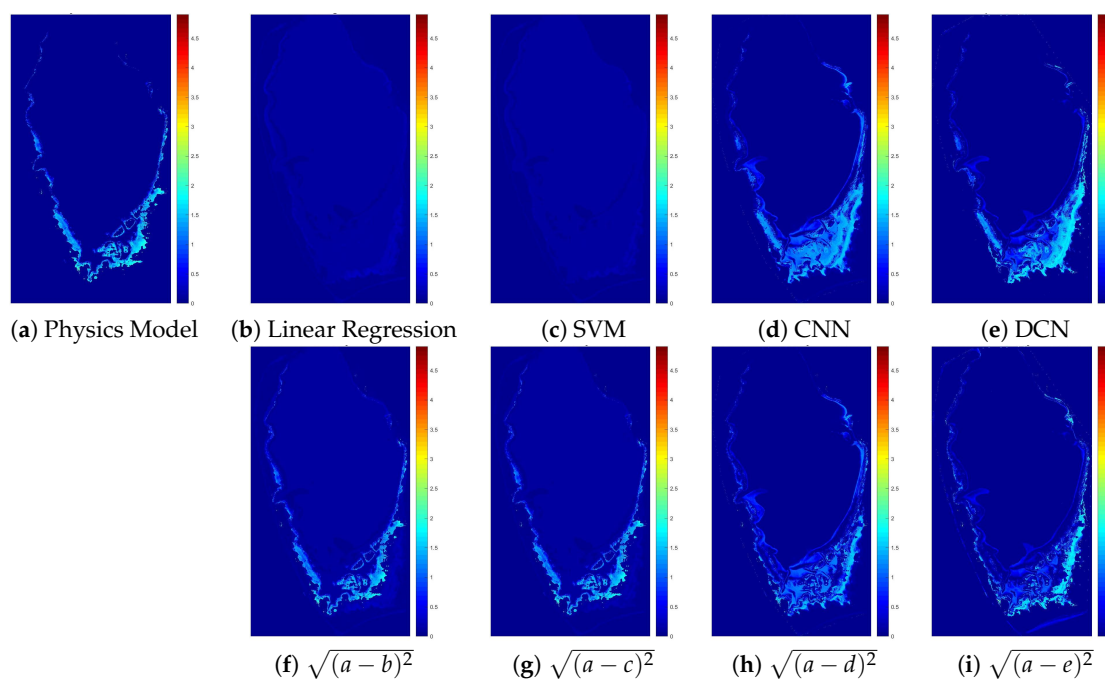
**Figure 5.** LAI mapping at Saint Joseph Bay produced by models trained with patches from the selected regions. The LAI maps are masked to show only those pixels that are classified as seagrass by the capsule network.



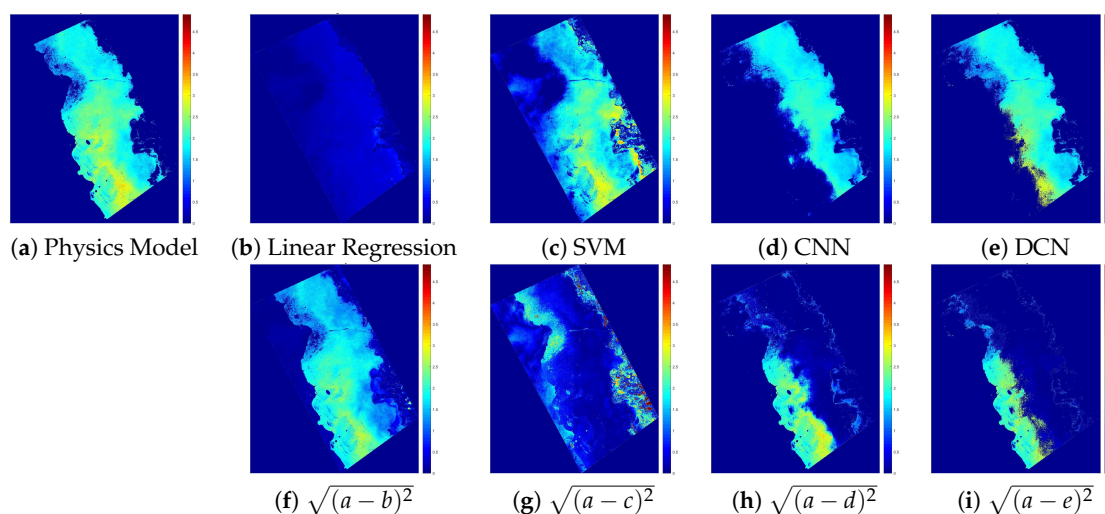
**Figure 6.** LAI mapping at Keeton Beach produced by models trained with patches from the selected regions. The LAI maps are masked to show only those pixels that are classified as seagrass by the capsule network.



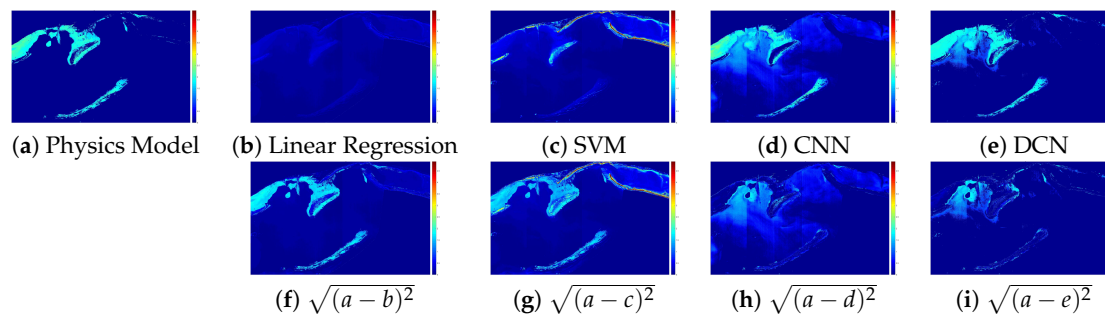
**Figure 7.** LAI mapping at Saint George Sound produced by models trained with patches from the selected regions. The LAI maps are masked to show only those pixels that are classified as seagrass by the capsule network.



**Figure 8.** LAI mapping at St. Joseph Bay produced by models trained with patches from the selected regions. No mask was applied in this case.



**Figure 9.** LAI mapping at Keeton Beach produced by models trained with patches from the selected regions. No mask was applied in this case.



**Figure 10.** LAI mapping at Saint George Sound produced by models trained with patches from the selected regions. No mask was applied in this case.

#### 4.4. Transfer Learning with Deep Models

We first train our deep learning models for LAI quantification with all the selected patches from the satellite image taken at St. Joseph Bay, and then we use the trained models as a feature extractor to transfer their knowledge to the other two locations (Keeton Beach and St. George Sound). Finally, we randomly select 50, 100, 500, and 1000 patches from the two new locations to train a linear regression model for LAI quantification at each new location. These selected patches are balanced among the four classes. To train a linear regression model for a new location, the selected labeled image patches are passed through the trained DCN/CNN model, and outputs from the *FeatureCaps*/FC layer are stored as training data. The outputs belonging to seagrass patches are then used to train a linear regression model for LAI quantification. For the remaining unlabeled image patches, we extract new representations at the *FeatureCaps*/FC layer and classify them to one of the four classes using the stored training data samples based on the 1-NN rule. If an image patch is classified as seagrass, we predict its LAI using the trained linear regression model. Otherwise, we set its LAI to 0.

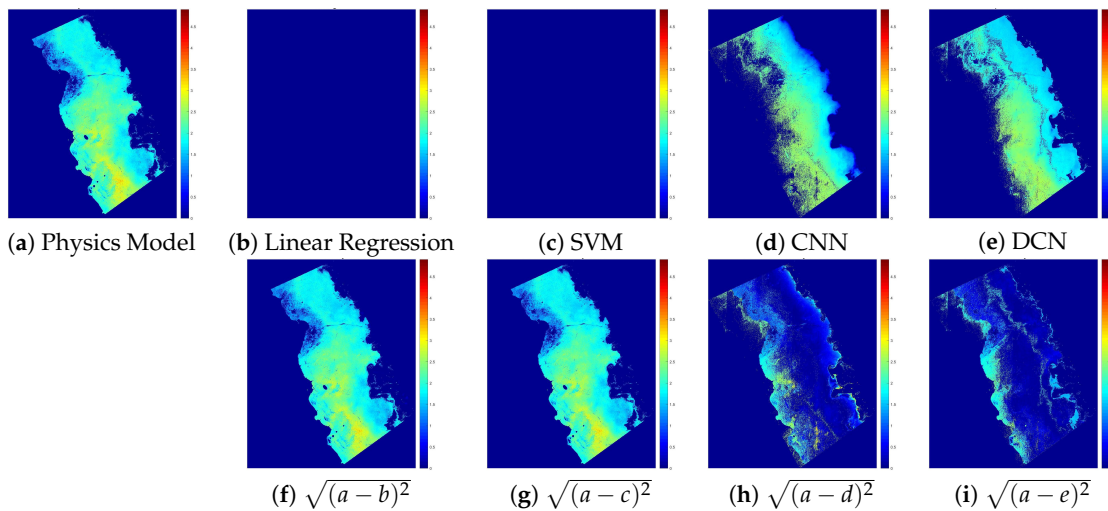
We perform each experiment five times and show results of the 1-NN classification accuracies for the labeled patches at Keeton Beach and St. Joseph Bay in Table 3. It can be seen that while the classification results are very similar between both models, the DCN generally performs slightly better than the CNN. The RMSE results of LAI quantification by transfer learning are shown in Table 4. When the transfer learning approach is applied to the image taken at Keeton beach, the DCN outperforms the CNN in cases with a small number of training samples (50, 100). In cases with a larger amount of training samples, there is no significant difference between CNN and DCN. Fine tuning always makes both the DCN and CNN worse indicating that over-fitting may happen. At St. George Sound, the DCN outperforms the CNN in transfer learning regardless of the number of training samples from St. George Sound. However, the best results at this location are always obtained when performing fine tuning with the CNN. In all cases, our transfer learning approach significantly outperforms direct mapping using linear regression and SVM. Additionally, it can be seen that the errors when using the networks without transfer learning (0 samples) are significantly larger. For visualization purposes, we show the transfer learning results after fine tuning with 500 labeled patches from the new locations in Figures 11 and 12. Note again that these two Figures are for visualization only, the models' comparison should be based on the accuracies computed in the selected regions (Tables 3 and 4) where LAI mappings by the physics model are more reliable.

**Table 3.** Accuracies obtained by our transfer learning approach utilizing different number of samples from new locations. Five experiments are performed for each sample size. The table shows the results as mean  $\pm$  std. Best results are highlighted in bold.

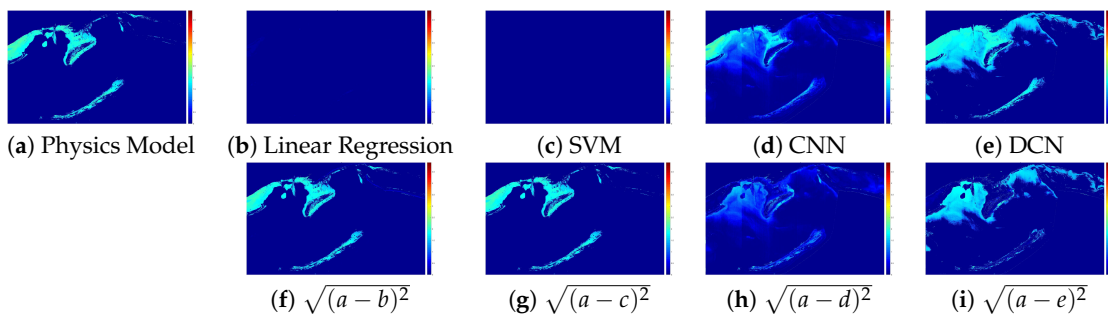
Image	Model	50 Patches	100 Patches	500 Patches	1000 Patches
Keaton Beach	CNN	0.9145 $\pm$ 0.04	0.9514 $\pm$ 0.01	0.9853 $\pm$ 0.003	0.9902 $\pm$ 0.0007
	DCN	<b>0.9311 <math>\pm</math> 0.03</b>	<b>0.9676 <math>\pm</math> 0.01</b>	<b>0.9867 <math>\pm</math> 0.002</b>	<b>0.9908 <math>\pm</math> 0.001</b>
St. George Sound	CNN	<b>0.9615 <math>\pm</math> 0.007</b>	0.9635 $\pm$ 0.007	0.9761 $\pm$ 0.008	0.9868 $\pm$ 0.005
	DCN	0.9529 $\pm$ 0.008	<b>0.9721 <math>\pm</math> 0.01</b>	<b>0.9839 <math>\pm</math> 0.002</b>	<b>0.9896 <math>\pm</math> 0.001</b>

**Table 4.** RMSEs obtained by our transfer learning approach and fine tuning utilizing different number of samples from new locations. Five experiments are performed for each sample size. The table shows the results as mean  $\pm$  std. Best results are highlighted in bold.

Image	Method	0 Samples	50 Samples		100 Samples		500 Samples		1000 Samples	
			Transfer Learning	Fine Tuning	Transfer Learning	Fine Tuning	Transfer Learning	Fine Tuning	Transfer Learning	Fine Tuning
Keaton Beach	CNN	2.76	0.69 $\pm$ 0.19	1.35 $\pm$ 0.14	0.52 $\pm$ 0.08	1.17 $\pm$ 0.297	<b>0.28 <math>\pm</math> 0.03</b>	0.66 $\pm$ 0.33	<b>0.24 <math>\pm</math> 0.007</b>	0.91 $\pm$ 0.47
	DCN	1.72	<b>0.63 <math>\pm</math> 0.12</b>	1.30 $\pm$ 0.14	<b>0.46 <math>\pm</math> 0.06</b>	1.16 $\pm$ 0.02	0.29 $\pm$ 0.02	0.73 $\pm$ 0.31	0.25 $\pm$ 0.02	0.69 $\pm$ 0.03
	LR	–	1.57 $\pm$ 0.003	–	1.61 $\pm$ 0.01	–	1.62 $\pm$ 0.01	–	1.60 $\pm$ 0.01	–
	SVM	–	1.57 $\pm$ 0.003	–	1.62 $\pm$ 0.003	–	1.52 $\pm$ 0.0007	–	1.62 $\pm$ 0.003	–
St. George Sound	CNN	0.61	0.35 $\pm$ 0.03	<b>0.14 <math>\pm</math> 0.01</b>	0.31 $\pm$ 0.04	<b>0.14 <math>\pm</math> 0.005</b>	0.23 $\pm$ 0.05	<b>0.09 <math>\pm</math> 0.006</b>	0.18 $\pm$ 0.03	<b>0.09 <math>\pm</math> 0.01</b>
	DCN	0.56	0.34 $\pm$ 0.04	0.24 $\pm$ 0.03	0.25 $\pm$ 0.07	0.20 $\pm$ 0.04	0.19 $\pm$ 0.008	0.11 $\pm$ 0.01	0.15 $\pm$ 0.005	0.13 $\pm$ 0.03
	LR	–	0.71 $\pm$ 0.01	–	0.73 $\pm$ 0.002	–	0.72 $\pm$ 0.01	–	0.71 $\pm$ 0.01	–
	SVM	–	0.71 $\pm$ 0.0003	–	0.72 $\pm$ 0.0002	–	0.73 $\pm$ 0.0007	–	0.73 $\pm$ 0.0005	–



**Figure 11.** LAI mapping at Keaton Beach produced by our transfer learning approach using 500 patches.



**Figure 12.** LAI mapping at Saint George Sound produced by our transfer learning approach using 500 patches.

#### 4.5. Computational Complexity

We carry out the experiments using a computer with 64 GB of RAM and an Intel Xeon E5-2687W v3 @ 3.10 GHz (10 cores). On average, one epoch of training requires 85.39 s and 13.17 s by the DCN and CNN models, respectively. Testing the DCN model takes 0.13 milliseconds/patch, while testing on the CNN model takes 0.023 milliseconds/patch. In total, testing on one entire image takes about 1.5 h with DCN and 0.42 h with CNN. Table 5 includes the training and testing time by each model.

**Table 5.** Average training and testing CPU times by DCN and CNN.

Model	Training Time (s/Epoch)	Testing Time (ms/Patch)
DCN	85.39	0.13
CNN	13.17	0.023

### 5. Discussions

Our experimental results show that seagrass quantification using deep learning models is feasible and are better options than traditional machine learning methods. While all models use linear regression to predict LAI, DCN, and CNN perform LAI regression using new representations learned from raw image patches. In contrast, the linear regression and SVM models use raw pixel values in image patches of size  $5 \times 5 \times 8$  to quantify LAI. Our results show that both DCN and CNN outperform linear regression and SVM, which demonstrates that information in the learned representations is more suitable than raw image patches for LAI quantification.

The CNN model performs better than DCN when quantifying seagrass within the selected regions in the same image (Section 4.2). Specifically, RMSEs obtained by CNN/DCN are 0.04/0.07, 0.08/0.12, and 0.45/0.46 at Keeton Beach, St. George Sound, and St. Joseph Bay respectively. While these performance differences are not significant, training the CNN model takes approximately 6.5 times less than training DCN, which makes CNN the best option in this case.

Figures 5–12 show that our models can effectively perform end-to-end seagrass LAI mapping. The results reported in these figures may be significantly different with respect to the physics model. However, the physics model has a reported error of 10% and should not be treated as ground truth. To assess the performance of each model, we should only compare RMSEs achieved by the models within the selected regions, as reported in Tables 2–4. The end-to-end mappings shown in this paper are for visualization purposes only. New on-site validation is planned to generate more ground truth data for model evaluation.

Our results show that the new representations learned by the DCN and CNN models are much better than the raw image patches for seagrass identification and LAI quantification in transfer learning. The DCN model achieves slightly better classification accuracies (Table 3) than the CNN model at the two new locations. For LAI mapping, the DCN model generally can achieve better results than CNN (without fine tuning) as shown in Table 4. If fine tuning is applied, performances of both the CNN and DCN models drop at Keeton Beach, indicating that over-fitting may happen and degrades the models' performances. At St. George Sound, DCN always performs better than CNN and fine tuning improves the performances of both models. Overall, transfer learning with DCN and CNN significantly improves seagrass quantification at different locations using as less as 50 samples from the new locations for training, as compared to the direct regression models including linear regression and SVM. With these experiments, we demonstrate that generalization of deep models using a limited number of samples from different locations for seagrass quantification is feasible.

Our study has limitations. First, the LAI mapping produced by the physics model has a regression error of about 10% [4]. While the mapping in the regions selected as ground truth for training by the experienced operator is more reliable, the exact mapping error in the selected regions has not been investigated. Results shown in Tables 2–4 are based on the selected regions where the accuracies



are more reliable, while the end-to-end mappings shown in Figures 5–12 are for the whole images, and should be viewed as for visualization purpose only. Second, all regions considered in this study are from the coast of Florida. While the distribution of seagrass differs in each of the images, it is likely that the distribution difference will be even larger if the images are from different parts of the world. The ultimate goal of this work is to map seagrass distribution globally, and the model developed in this work needs to be further validated with more images from different locations, ideally from different hemispheres.

## 6. Conclusions

In this paper, we studied the quantification of seagrass distribution using two deep learning models: a convolutional neural network and a deep capsule network. We evaluated the proposed model for seagrass quantification at three different locations in Florida. The proposed models achieved significantly better results than a linear regression model and a support vector machine. We demonstrated that using transfer learning techniques for the quantification of seagrass significantly improved the results as compared to directly applying the deep models to new locations. Using our technique, seagrass can be accurately quantified with minimum workforce site observation.

**Author Contributions:** D.P. carried out the experiments and wrote the manuscript with support from K.I., J.L. and Y.S. V.H., R.Z. and B.S. provided the satellite images and the physics model to predict LAI levels of seagrass. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by NASA under Grant NNX17AH01G. Subjected to review by the National Exposure Research Laboratory and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use by the U.S. Government. The views expressed in this article are those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Hemminga, M.A.; Duarte, C.M. *Seagrass Ecology*; Cambridge University Press: Cambridge, UK, 2000.
2. Wicaksono, P.; Hafizt, M. Mapping seagrass from space: Addressing the complexity of seagrass LAI mapping. *Eur. J. Remote Sens.* **2013**, *46*, 18–39. [CrossRef]
3. Breuer, L.; Freede, H. Leaf Area Index—LAI. 2003. Available online: <https://www.staff.uni-giessen.de/~gh1461/plapada/lai/lai.html> (accessed on 14 May 2020).
4. Hill, V.; Zimmerman, R.; Bissett, W.; Dierssen, H.; Kohler, D. Evaluating light availability, seagrass biomass and productivity using hyperspectral airborne remote sensing in Saint Joseph’s Bay, Florida. *Estuaries Coasts* **2014**, *37*, 1467–1489. [CrossRef]
5. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. *arXiv* **2017**, arXiv:1612.08242.
6. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015; pp. 91–99.
7. Lee, H.; Grosse, R.; Ranganath, R.; Ng, A.Y. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In Proceedings of the 26th Annual International Conference on Machine Learning, Montreal, QC, Canada, 14–18 June 2009; pp. 609–616.
8. Wang, N.; Yeung, D.Y. Learning a deep compact image representation for visual tracking. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 5–10 December 2013; pp. 809–817.
9. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
10. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.



11. Sermanet, P.; Eigen, D.; Zhang, X.; Mathieu, M.; Fergus, R.; LeCun, Y. Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv* **2013**, arXiv:1312.6229.
12. Perez, D.; Banerjee, D.; Kwan, C.; Dao, M.; Shen, Y.; Koperski, K.; Marchisio, G.; Li, J. Deep learning for effective detection of excavated soil related to illegal tunnel activities. In Proceedings of the IEEE Ubiquitous Computing, Electronics and Mobile Communication Conference, New York, NY, USA, 19–21 October 2017.
13. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
14. Oguslu, E.; Islam, K.; Perez, D.; Hill, V.; Bissett, W.; Zimmerman, R.; Li, J. Detection of seagrass scars using sparse coding and morphological filter. *Remote Sens. Environ.* **2018**, *213*, 92–103. [[CrossRef](#)]
15. Perez, D.; Lu, Y.; Kwan, C.; Shen, Y.; Koperski, K.; Li, J. Combining Satellite Images with Feature Indices for Improved Change Detection. In Proceedings of the IEEE Ubiquitous Computing, Electronics & Mobile Communication Conference, New York, NY, USA, 8–10 November 2018.
16. Lu, Y.; Perez, D.; Dao, M.; Kwan, C.; Li, J. Deep Learning with Synthetic Hyperspectral Images for Improved Soil Detection in Multispectral Imagery. In Proceedings of the IEEE Ubiquitous Computing, Electronics & Mobile Communication Conference, New York, NY, USA, 8–10 November 2018; pp. 8–10.
17. Hoque, M.R.U.; Islam, K.; Perez, D.; Hill, V.; Schaeffer, B.; Zimmerman, R.; Li, J. Seagrass Propeller Scar Detection using Deep Convolutional Neural Network. In Proceedings of the IEEE Ubiquitous Computing, Electronics & Mobile Communication Conference, New York, NY, USA, 8–10 November 2018.
18. Yang, J.; Zhao, Y.Q.; Chan, J.C.W. Learning and transferring deep joint spectral–spatial features for hyperspectral classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 4729–4742. [[CrossRef](#)]
19. Hinton, G.; Deng, L.; Yu, D.; Dahl, G.E.; Mohamed, A.R.; Jaitly, N.; Senior, A.; Vanhoucke, V.; Nguyen, P.; Sainath, T.N.; et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Process. Mag.* **2012**, *29*, 82–97. [[CrossRef](#)]
20. Yu, D.; Deng, L. *Automatic Speech Recognition*; Springer London limited: London, UK, 2016.
21. Chen, C.; Seff, A.; Kornhauser, A.; Xiao, J. Deepdriving: Learning affordance for direct perception in autonomous driving. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 2722–2730.
22. Huval, B.; Wang, T.; Tandon, S.; Kiske, J.; Song, W.; Pazhayampallil, J.; Andriluka, M.; Rajpurkar, P.; Migimatsu, T.; Cheng-Yue, R.; et al. An Empirical Evaluation of Deep Learning on Highway Driving. *arXiv* **2015**, arXiv:1504.01716.
23. Ning, R.; Wang, C.; Xin, C.; Li, J.; Wu, H. Deepmag: Sniffing mobile apps in magnetic field through deep convolutional neural networks. In Proceedings of the 2018 IEEE International Conference on Pervasive Computing and Communications (PerCom), Athens, Greece, 19–23 March 2018; pp. 1–10.
24. Chowdhury, M.M.U.; Hammond, F.; Konowicz, G.; Xin, C.; Wu, H.; Li, J. A few-shot deep learning approach for improved intrusion detection. In Proceedings of the 2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), New York, NY, USA, 19–21 October 2017; pp. 456–462.
25. Perez, D.; Li, J.; Shen, Y.; Dayanghirang, J.; Wang, S.; Zheng, Z. Deep Learning for Pulmonary Nodule CT Image Retrieval—An Online Assistance System for Novice Radiologists. In Proceedings of the 2017 IEEE International Conference on Data Mining Workshops (ICDMW), New Orleans, LA, USA, 18–21 November 2017; pp. 1112–1121.
26. Niu, Z.; Zhou, M.; Wang, L.; Gao, X.; Hua, G. Ordinal regression with multiple output cnn for age estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 4920–4928.
27. Yuan, J.; Ni, B.; Kassim, A.A. Half-CNN: A general framework for whole-image regression. *arXiv* **2014**, arXiv:1412.6885.
28. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
29. Gidaris, S.; Komodakis, N. Object detection via a multi-region and semantic segmentation-aware cnn model. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1134–1142.

30. Li, R.; Zhang, W.; Suk, H.I.; Wang, L.; Li, J.; Shen, D.; Ji, S. Deep learning based imaging data completion for improved brain disease diagnosis. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Boston, MA, USA, 14–18 September 2014; pp. 305–312.
31. Sabour, S.; Frosst, N.; Hinton, G.E. Dynamic routing between capsules. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 3857–3867.
32. Sabour, S.; Frosst, N.; Hinton, G. Matrix capsules with EM routing. In Proceedings of the 6th International Conference on Learning Representations (ICLR), Vancouver, BC, Canada, 30 April–3 May 2018; pp. 1–15.
33. Xi, E.; Bing, S.; Jin, Y. Capsule Network Performance on Complex Data. *arXiv* **2017**, arXiv:1712.03480.
34. Afshar, P.; Mohammadi, A.; Plataniotis, K.N. Brain tumor type classification via capsule networks. *arXiv* **2018**, arXiv:1802.10200.
35. Shen, Y.; Gao, M. Dynamic routing on deep neural network for thoracic disease classification and sensitive area localization. In Proceedings of the International Workshop on Machine Learning in Medical Imaging, Granada, Spain, 16 September 2018; pp. 389–397.
36. Qiao, K.; Zhang, C.; Wang, L.; Yan, B.; Chen, J.; Zeng, L.; Tong, L. Accurate reconstruction of image stimuli from human fMRI based on the decoding model with capsule network architecture. *arXiv* **2018**, arXiv:1801.00602.
37. Andersen, P.A. Deep reinforcement learning using capsules in advanced game environments. *arXiv* **2018**, arXiv:1801.09597.
38. LaLonde, R.; Bagci, U. Capsules for Object Segmentation. *arXiv* **2018**, arXiv:1804.04241.
39. Islam, K.; Perez, D.; Hill, V.; Schaeffer, B.; Zimmerman, R.; Li, J. Seagrass Detection in Coastal Water through Deep Capsule Networks. In Proceedings of the Chinese Conference on Pattern Recognition and Computer Vision, Guangzhou, China, 23–26 November 2018.
40. Pérez, D.; Islam, K.; Hill, V.; Zimmerman, R.; Schaeffer, B.; Li, J. Deepcoast: Quantifying seagrass distribution in coastal water through deep capsule networks. In Proceedings of the Chinese Conference on Pattern Recognition and Computer Vision (PRCV), Guangzhou, China, 23–26 November 2018; pp. 404–416.
41. Phinn, S.; Roelfsema, C.; Dekker, A.; Brando, V.; Anstee, J. Mapping seagrass species, cover and biomass in shallow waters: An assessment of satellite multi-spectral and airborne hyper-spectral imaging systems in Moreton Bay (Australia). *Remote Sens. Environ.* **2008**, *112*, 3413–3425. [[CrossRef](#)]
42. Short, F.T.; Coles, R.G. *Global Seagrass Research Methods*; Elsevier: Amsterdam, The Netherlands, 2001; Volume 33.
43. Yang, D.; Yang, C. Detection of seagrass distribution changes from 1991 to 2006 in Xincun Bay, Hainan, with satellite remote sensing. *Sensors* **2009**, *9*, 830–844. [[CrossRef](#)]
44. Pu, R.; Bell, S.; Meyer, C.; Baggett, L.; Zhao, Y. Mapping and assessing seagrass along the western coast of Florida using Landsat TM and EO-1 ALI/Hyperion imagery. *Estuar. Coast. Shelf Sci.* **2012**, *115*, 234–245. [[CrossRef](#)]
45. Dierssen, H.M.; Zimmerman, R.C.; Leathers, R.A.; Downes, T.V.; Davis, C.O. Ocean color remote sensing of seagrass and bathymetry in the Bahamas Banks by high-resolution airborne imagery. *Limnol. Oceanogr.* **2003**, *48*, 444–455. [[CrossRef](#)]
46. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [[CrossRef](#)]
47. Donahue, J.; Jia, Y.; Vinyals, O.; Hoffman, J.; Zhang, N.; Tzeng, E.; Darrell, T. Decaf: A deep convolutional activation feature for generic visual recognition. In Proceedings of the International Conference on Machine Learning, Beijing, China, 21–26 June 2014; pp. 647–655.
48. Yosinski, J.; Clune, J.; Bengio, Y.; Lipson, H. How transferable are features in deep neural networks? In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada 8–13 December 2014; pp. 3320–3328.
49. Xie, M.; Jean, N.; Burke, M.; Lobell, D.; Ermon, S. Transfer learning from deep features for remote sensing and poverty mapping. *arXiv* **2015**, arXiv:1510.00098.
50. Jun, G.; Ghosh, J. An efficient active learning algorithm with knowledge transfer for hyperspectral data analysis. In Proceedings of the 2008 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2008), Boston, MA, USA, 7–11 July 2008; Volume 1, pp. 1–52.
51. Hu, F.; Xia, G.S.; Hu, J.; Zhang, L. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sens.* **2015**, *7*, 14680–14707. [[CrossRef](#)]

52. Banerjee, D.; Islam, K.; Mei, G.; Xiao, L.; Zhang, G.; Xu, R.; Ji, S.; Li, J. A Deep Transfer Learning Approach for Improved Post-Traumatic Stress Disorder Diagnosis. In Proceedings of the 2017 IEEE International Conference on Data Mining (ICDM), New Orleans, LA, USA, 18–21 November 2017; pp. 11–20.
53. Li, B.; Shen, C.; Dai, Y.; van den Hengel, A.; He, M. Depth and surface normal estimation from monocular images using regression on deep features and hierarchical CRFs. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1119–1127.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).