

# Semantic Segmentation with Multispectral Satellite Images of Waterfowl Habitat

Mateo Gannod<sup>1</sup>, Nicholas Masto<sup>1</sup>, Collins Owusu<sup>1</sup>, Cory Highway<sup>1</sup>,  
Katherine E. Brown<sup>1</sup>, Abigail Blake-Bradshaw<sup>1</sup>, Jamie C. Feddersen<sup>2</sup>,  
Heath M. Hagy<sup>3</sup>, Douglas A. Talbert<sup>1</sup>, Bradley Cohen<sup>1</sup>

<sup>1</sup>Tennessee Tech University, Cookeville, TN, USA

<sup>2</sup>Division of Wildlife and Forestry, Tennessee Wildlife Resources Agency, Nashville, TN, USA

<sup>3</sup>National Wildlife Refuge System, U.S. Fish and Wildlife Service, Stanton, TN, USA

{mcgannod42,nmmasto42,cowusu42,highway42,kebrow46,agblakebra42,dtalbert,bcohen}@tntech.edu,  
jamie.feddersen@tn.gov, heath\_hagy@fws.gov

## Abstract

Migratory waterfowl (i.e., ducks, geese, and swans) management relies on landscape bioenergetic models to inform on-the-ground habitat conditions and conservation practices. Therefore, conservation planners rely on accurate predictions of wetland habitats for waterfowl at regional scales. Unharvested flooded corn is a popular management tool on public and private lands that greatly increases landscape-level energy compared to other wetlands; thus, landscape bioenergetic models are particularly sensitive to these habitat features. Despite their importance to conservation planning and implementation, the abundance and distribution of unharvested flooded corn fields across North America is unknown. Furthermore, training data is difficult to collect and accurate predictions are challenging given their unique attributes and discreteness at landscape-level lens. Advances in multispectral imagery and deep learning algorithms may enable continuous and autonomous detection of these habitat features. Therefore, we conducted modeling experiments using training data of unharvested flooded corn fields in West Tennessee and multispectral imagery collected from Sentinel-2 satellite missions. We performed several experiments using individual band combination composites and/or vegetation indices to identify optimal bands using MRUNET architectures. We subsequently used 3 ensemble models of important individual networks. We found the use of multispectral bands was necessary and although the CIR composite and OSAVI index improved precision, the 12-band composite increased recall, the metric we were most interested in. Moreover, all ensembles exhibited poor performance. Here, we present results of our initial modeling experiments and suggest future modeling exercises including temporal image and vegetation index stacking using multi-modal and/or recurrent neural network architectures.

## Introduction

Federal management for migratory waterfowl (i.e., ducks, geese, and swans) during autumn, winter, and spring (hereafter, non-breeding periods) hierarchically step-down continental population objectives to regional geographies and subsequently implement bioenergetic models that translate species-specific abundance to needed landscape energy and

foraging habitat goals (U.S. Department of Interior 1986; Williams et al. 2014). Cultivation of high-energy croplands flooded during nonbreeding seasons for waterfowl allow state and federal biologists to efficiently meet target habitat objectives. Flooded unharvested corn fields are a particularly popular management tool on public and private lands that provide energy-dense waterfowl food resources (Highway 2022). Bioenergetic models are extremely sensitive to these habitat features. Despite the importance of these landscape habitat features to conservation planning, the abundance and distribution of flooded croplands are unknown.

Virtually no data on unharvested flooded corn field acreages exists because (1) they are difficult to identify from satellite imagery; (2) no reporting requirements exist; and (3), field data is costly to collect. However, novel geospatial software and technologies have enabled high-resolution spatiotemporal predictions of landcover, surface water, land-use change and other phenomenon of global importance (Brown et al. 2022; Owusu et al. 2022). For example, Sentinel-2 satellite missions continuously collect multispectral (12 channels) satellite imagery with a 5 day revisit frequency worldwide at 10 m granularity for visible and near infrared (NIR) light (Gorelick et al. 2017). Using Google Earth Engine (GEE), it is possible to quickly and freely retrieve satellite data continuously from around the world. Ideally, this data should enable continuous updates to flooded corn acreages for conservation planning and implementation purposes. However, digitizing these features on satellite imagery is expensive and time consuming.

Semantic image segmentation is an algorithmic approach to labelling an image's pixels (Liu, Deng, and Yang 2019). Since the advent of Fully Convolutional Networks (FCN) (Long, Shelhamer, and Darrell 2015), great strides have been made in Semantic Image Segmentation. For example, research in the medical field developed network architectures like U-Net (Ronneberger, Fischer, and Brox 2015) which has performed segmentation tasks with high levels of precision despite relatively low training data counts. U-Net has subsequently become popular for semantic segmentation tasks in other domains such as remote sensing (Chen et al. 2021; Shamsolmoali et al. 2019; Zhang, Liu, and Wang 2018).

Our overall goal is to create a deep learning model that can accurately assess the land cover of flooded unharvested corn. We plan to accomplish this by leveraging available

satellite imagery and deep learning algorithms to perform semantic segmentation. The model can then be used to create user-friendly software that enables continual and accurate predictions of unharvested flooded corn fields for conservation planning and implementation. Herein, we conducted modeling experiments to better understand salient features of different available datasets when detecting unharvested flooded corn. We also establish a baseline against which to compare as we further develop our model.

## Dataset



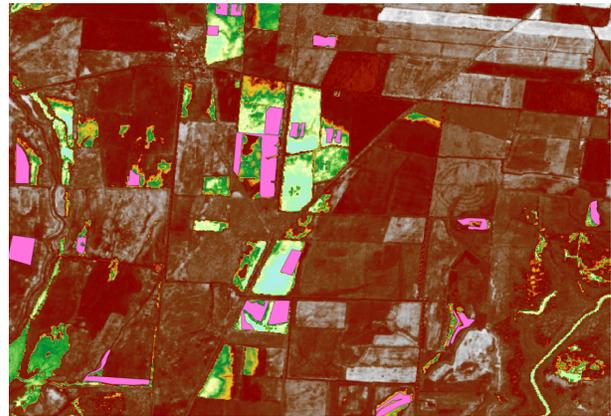
Figure 1: Our entire area of study in West Tennessee.

Our dataset consists of satellite imagery and a corresponding mask that shows where unharvested flooded corn is located. Our area of study is located in West Tennessee along the Mississippi river and contributing watersheds as shown in Figure 1. We gathered Sentinel 2 imagery from January 2022 with 10% minimal cloud coverage for our area using GEE (Gorelick et al. 2017). The data collected included all 12 spectral bands ranging between 442.3 nm to 2202.4 nm. We gathered 185 field locations from private and publicly locations by tracing the boundary with a Garmin eTrex 10 handheld GPS unit (Garmin International, Olathe, KS, USA) (Highway 2022). Individual patches of unharvested flooded corn are in close proximities; to prevent data leakage while training, we grouped patches within 0.5 km radius from each other resulting in 106 unique locations from which to sample. Next, we projected satellite images and field shapes to the same coordinate reference system to ensure geographical alignment. In addition to the 12-channel multi-spectral images, we also calculated the Optimized Soil Adjustment Vegetation Index, OSAVI. This index takes the multispectral bands and applies a mathematical transformation to better highlight plant life and vitality. We selected the OSAVI because it was designed to mute soil brightness (or in our case water reflectance) (Fern et al. 2018). The formula for

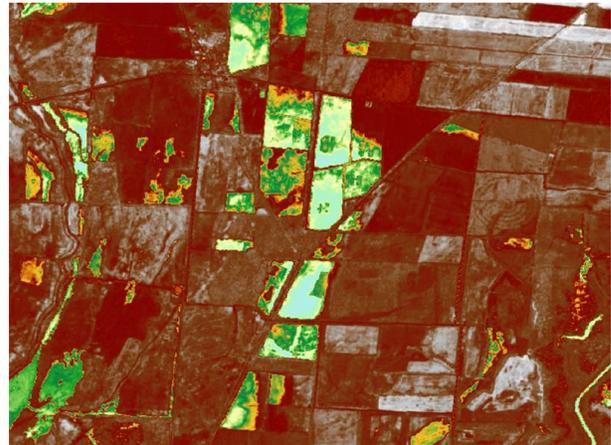
the OSAVI index is as follows:

$$OSAVI = (1 + 0.16) \frac{NIR - Red}{NIR + RED + 0.16}$$

Where *NIR* is the Sentinel 2 band for Near Infrared light, and *Red* is the Sentinel band for red light. We expected these infrared bands and indices to improve detection and segmentation of unharvested corn and water (Figure 2). Additionally, we used two subsets of the full 12-channel images. The first uses the red, green, and blue channels (RGB), and the second uses the NIR, red, and green channels (CIR).



(a) Visualization of OSAVI index with fields in pink.



(b) OSAVI index visualization without corn shapes.

Figure 2: Visualizations of the OSAVI index.

We generated labels for training data by converting GPS corn shapes to a label mask for the study area using Rasterio in Python (Gillies and others 2013). Each pixel is labeled 0, background, or 1, unharvested flooded corn Figure 3.

Often in remote sensing, segmentation tasks for specific landscape features encounter extreme class imbalances. Figure 4 shows the label mask for our entire area of study. In Figure 4 it is apparent that a small area of our study region as a whole has been labeled. Secondly, while in Figure 3 the foreground class is relatively common, more often than not, the foreground class is scarce. Figure 5 shows the massive class imbalance present in the dataset.



Figure 3: Example label mask. Background class is in purple, unharvested flooded corn is shown in yellow. Each pixel is 10x10 meters.



Figure 5: Example label mask with massive class imbalance. Background class is in purple, unharvested flooded corn is shown in yellow. Each pixel is 10x10 meters.

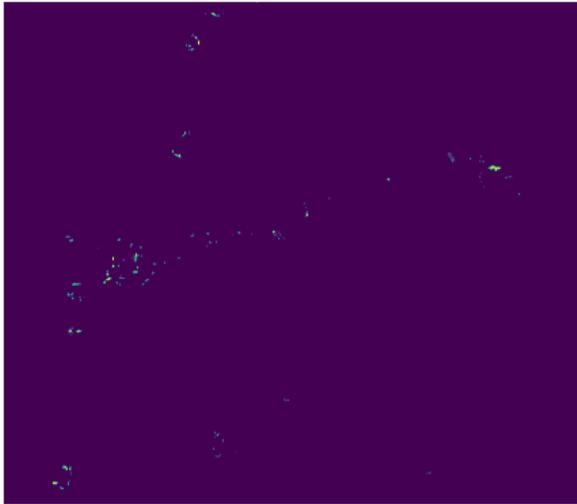


Figure 4: Label mask for our study region. Foreground class is in yellow with a 10x10 meter pixel resolution

To deal with the small sample size challenges (Figure 4), we applied image augmentation to artificially increase the number of available training samples. We found that this performed better than no augmentations, due to mitigation of over fitting. Whenever an image is loaded during training, we generated a centroid for each field. A random pixel is then selected from a 64x64 pixel area around the centroid and used to extract a 128x128 pixel image. Next, the image is flipped about the x and/or y axis and rotated in a range of  $\pm 1-45$  degrees. Finally, the image is extracted to be 128x128 pixels. With these augmentations ( $64^2$  translations \* 4 flips \* 90 rotations) we increased the available sample space by a factor of 1 million. When training on a batch size of 20 for 5 steps across 200 epochs, we used roughly 20,000 of the images.

## Model

### Architecture

We implemented our model in Python 3.7 using Tensorflow 2.0 (Abadi et al. 2015). For our model we chose the MultiResUNet (MRUNet) architecture, which builds upon U-Net by

expanding the receptive field of the network and correcting for possible semantic inconsistencies (Ibtehaz and Rahman 2020). U-Net is a fully convolutional network for image segmentation that has seen great success in both the medical and remote sensing fields (Ronneberger, Fischer, and Brox 2015; Chen et al. 2021; Shamsolmoali et al. 2019; Zhang, Liu, and Wang 2018). U-Net has three main features, the encoder, the decoder, and the skip connections. The encoder has multiple tiers of two convolutional layers with kernel size 3x3 followed by a max pooling layer with a pooling window of size 2x2. The convolutional layers generate a specified number of filters that represent some feature of their input. Next, the max pooling layer downsamples the generated filters. At each tier of the network, the filter dimensionality decreases, and the number of filters generated is doubled. Then, at the lowest tier, the filters are passed to the decoder where this operation is mirrored. Instead of max pooling, the filters are transposed up a tier before the two convolutional layers. Mirroring the encoder, as the dimensionality of the filters increases, the filters decrease. Many semantic segmentation networks follow this design, with the intuition that, over time, deeper features are extracted in order to identify objects (Liu, Deng, and Yang 2019). The main contribution of U-Net was introducing the skip connections from the encoder to the decoder section. At each matching tier, the output of the encoder section is passed to the decoder section. The max pooling layer takes features for a section of the image and chooses the max value. So, it is possible that some nuance of the image could be lost. By implementing these skip connections, spatial features that may have been lost throughout the pooling in the network are preserved.

MRUNet extends upon U-Net based on a few insights. First, objects are rarely of the same scale, and it is important for the network to be able to handle objects of different sizes. MRUNet handles this by introducing the MultiRes block to the network (Ibtehaz and Rahman 2020). Instead of doing two convolutional layers, it has a series of three 3x3 convolutions, with each set of output filters appended to the output. This approximates a 3x3, 5x5, and 7x7 convolution in parallel. The three layers are then added to a residual filter generated from a 1x1 convolution. This broadens the model's receptive field, or the spatial information it can process at one time. The MultiRes block therefore pro-

vides additional information that a 3x3 kernel may not incorporate. The second insight is a bridge for the “semantic gap” between the encoder and decoder. As the features get deeper and deeper into the network, they are continuously processed by the convolutional layers. When merging the filters from the encoder and decoder there is a different level of processing and therefore incompatible data. MRUNet builds on top of the skip connections by passing the skip connection through a series of 3x3 convolutions added with a residual 1x1 convolution for each tier. As a result, the filters in the skip connection are processed in a way that matches the corresponding decoder section and closes gap between the encoder and decoder.

We also implemented unified focal loss in order to mitigate class imbalances (Yeung et al. 2022). Unified focal loss works by combining two other loss functions (focal loss and Tversky focal loss) into a single metric. Both functions assign a weight to individual classes, in our case, the background class and the foreground class. They are also weighted by how well the model performed for a given training example. If the model produces a good prediction, it is not carried into the backpropagation as heavily as poor predictions. As a result of using focal loss and focal Tversky loss, the model maximizes both a good distribution of predictions and similarity between the prediction and the ground truth. Both the focal and Tversky focal loss functions are parameterized by the variables  $\alpha$  and  $\gamma$  which control the class weightings (foreground and background) and difficulty weight respectively. Unified focal loss combines them into a single  $\alpha$  and  $\gamma$  for simplicity.

## Experiments

For all experiments we set our loss function parameters to  $\alpha=0.7$ ,  $\gamma=0.5$ . The parameter  $\alpha$  was set to 0.7 to encourage the model to favor the foreground class and  $\gamma$  was set to 0.5 so that difficult training corn field examples were more heavily weighted. This setting for  $\gamma$  is suggested to be optimal for the function (Yeung et al. 2022). We performed 5-fold cross validation, generating stratified training and validation sets of size 85 and 21 respectively. During training, the 85 training images are then augmented as previously described. For each fold we trained for 150 epochs with an initial learning rate of 0.0005 and the Adam optimizer available in TensorFlow (Abadi et al. 2015).

For these experiments we tested the following 3 questions: (1) Is there a subset of channels (e.g., color-infrared) from the full 12 channel multi-spectral images that performs better than the full 12?; (2) Does the OSAVI agricultural index perform better than the full 12?; and (3) Does an ensemble across all inputs improve performance?

We ran individual models using the RGB bands, OSAVI, CIR, and 12 channels, as well as an ensemble using the OSAVI, CIR, and 12 channel models. For our experiments, the RGB and 12 channel models served as baselines. We collected and calculated the confusion matrix for each individual model across the 5 folds from cross validation. We then calculated the metrics described below. Last, we combined the CIR, OSAVI, and 12 band composite into ensemble models. Ensembles often improve performance by tak-

ing information from each model (Ma et al. 2021). We performed three methods of ensembling. (1) We averaged the percentage across all the models that each pixel is part of the foreground class; (2) We trained more convolutional layers on each model’s prediction; and (3) We trained a decision tree to predict the class using the output of the three models. We also attempted to increase the recall of our models by increasing  $\alpha$  in unified focal loss (Ma et al. 2021). We experimented with values of 0.8 and 0.9 for  $\alpha$ .

## Metrics

Performance metrics used were the Intersection over Union (IoU), precision, recall and F1 score. The four equations are

$$IoU = \frac{TP}{TP + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall}$$

IoU is a popular metric in image segmentation (Minaee et al. 2021). It measures how similar two label masks are by computing the area of their intersection, and dividing by their union. This is a better metric than accuracy due to our class imbalance, which can still have high values, even if no foreground pixels were guessed correctly. The confusion matrix was generated for all five folds and then summed. The resulting metrics were taken from the summed confusion matrices.

## Results

OSAVI				
Model	Precision	Recall	IoU	F1
RGB	0.17	0.03	0.03	0.05
OSAVI	0.23	0.22	0.13	0.22
CIR	0.25	0.18	0.12	0.21
12-channel	0.17	0.32	0.13	0.23
Ensemble	0.29	0.08	0.07	0.12

Table 1: Metrics for each model across 5 folds.

Table 1 shows the metrics of our models. RGB has the lowest scores in Recall, IoU and F1, 0.03, 0.03, and 0.05 respectively. RGB does however, retain a precision score similar to the full 12 band model, 0.17. The OSAVI, CIR, and 12 channel models share similar IoU and F1 scores, however, when compared to the OSAVI index and CIR composite, the 12-band composite performed 0.06 percentage points lower in precision than OSAVI, but 0.10 points higher than OSAVI in recall.

For the ensemble model we listed the model that used the average percentage chance without the increased recall in Table 1. Our other ensemble models did not perform as well, scoring  $< .05$  and  $.10$  for IoU and F1 respectively. The ensemble had the highest precision, 0.29, but performed poorly in every other metric.

## Discussion



(a) Visualization of CIR composite with fields in orange.



(b) CIR composite visualization without corn shapes.

Figure 6: Visualizations of the CIR composite.

One key observation is that the RGB bands alone are unable to detect the important features of flooded corn, namely water and dead corn stalks. Conversely, any imagery containing infrared light performed better. Figures 2 and 6 show water and vegetation strongly because both reflect visible and near infrared light (Lillesaeter 1982). Comparing 2a to 2b, the flooded areas are all green, open water are in light blue hues, and non-flooded and apparently harvested agriculture or bare dirt appear brown. However, many, but not all green pixels are flooded corn. Most shapes in Figure 2a are surrounded by light blue water. Therefore, identifying green areas alone in the OSAVI index may be insufficient to identify flooded corn fields. Instead, a combination or hierarchical approach of first identifying green and then blue reflectance in the OSAVI index might improve performances. Likewise, in the CIR composite, the flooded areas are blue, and the surrounding dormant vegetation is pink. As mentioned, the infrared bands detected water reflectance and delineated water from dry ground during January exceptionally well. However, our experiments could not reliably identify the standing corn stalks within the water. In some cases, the models appeared to identify inundated corn by searching for dark green and orange pixels; however, in other cases, an en-

tire flooded area was dark green, but the entire area was not necessarily unharvested flooded corn. These landscape feature discrepancies of vegetation within water was clearly a challenge for individual segmentation models to overcome.

It appears that there is no distinct advantage that distinguishes the OSAVI Index and CIR composite. The reason for this may be that while they are different in their construction, their components are similar. Each uses the red and NIR bands, and the infrared bands appear to be the most important for finding water at least. So the differences between the two may not be enough to provide any substantial benefit over the other.

In a similar vein, our ensemble had high precision, but failed to identify and segment as many fields as the single model, which was the primary objective of this approach. As mentioned above, perhaps the similarity between the OSAVI and CIR failed to give the ensemble enough information to pick out important aspects of unharvested flooded corn. Another possible explanation is the single models do not perform well enough on their own for the ensemble itself to perform well. As a result, the ensemble doesn't have enough to learn off of. This would explain why simply averaging the percentage chance for each pixel outperformed more complicated neural network or decision tree methods.

## Conclusion and Future Work

In this work, we compared the performance of satellite bands, their subsets, a vegetation index used for agriculture, and combinations thereof. We found that the use of multi-spectral bands and agricultural indices are vital to identifying water and to some degree, vegetation, but ultimately insufficient to segment flooded agriculture. While a composite of all 12 bands can perform well, using a simpler form of the data, such as a subset or agricultural index, like the CIR and OSAVI, improves precision. However, the 12 channel model did perform better in recall, the performance metric we are most interested in given class imbalances, indicating additional informative but untested channels in our experiments.

Nevertheless, we demonstrate initial possibilities to segment unharvested flooded corn but considerable improvements are necessary. For example, one unreported experiment used slope and annual inundation frequency as inputs, assuming corn fields are not planted on steep inclines and that inundation frequency would aid in identifying areas that were partially flooded during a given year. These inputs proved to not be useful in prediction, however, slope and inundation frequency may be helpful in a pre-processing step thereby removing extraneous locations. By reducing the overall search space, we can reduce the time spent searching large areas and ensure the final tool is as efficient as possible. Future improvements may also include additional channel transformations and temporal image stacking (Qiao et al. 2021). For example, the OSAVI appears to delineate water edges and green features (often corn) well. A normalized difference moisture index (NDMI) has been shown to be useful for identifying water stress in crops such as corn and sunflowers (Mimić et al. 2022), which would be the case for flooded corn in January. Last, the features of interest change from October-January; thus, multi-temporal image

stacking in addition to vegetation index stacking may prove useful in the future and could be accomplished using multimodal or recurrent networks (Qiao et al. 2021). Although of little consequence here, expanding the size of these training datasets will be important for conservation planners to consider. One such endeavor is ongoing where an aerial observer has been taking GPS locations of flooded unharvested corn fields across the Southeast (Hagy H and Others 2022). Another option may be to utilize semi-supervised learning with partially labeled data to increase the number of training images.

## Limitations

We acknowledge there are some limitations in what we have covered in this work. Our dataset is relatively small for the image processing space, and we intend to address that by adding more labels and utilizing semi-supervised learning techniques. We are also considering different datasets and models to use for transfer learning. We also acknowledge that there may be better models other than those based on U-Net that we have yet to try.

## Acknowledgments

We appreciate support and funding from Tennessee Wildlife Resources Agency, U.S. Fish and Wildlife Service, and the Center for the Management, Protection, and Utilization of Water Resources (Water Center) at Tennessee Tech. Any use of trade, product, or firm names are for descriptive purposes only and do not imply endorsement by the U.S. Government. The views expressed in this article are the authors' own and do not necessarily represent the views of the U.S. Fish and Wildlife Service.

## References

- Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G. S.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Goodfellow, I.; Harp, A.; Irving, G.; Isard, M.; Jia, Y.; Jozefowicz, R.; Kaiser, L.; Kudlur, M.; Levenberg, J.; Mané, D.; Monga, R.; Moore, S.; Murray, D.; Olah, C.; Schuster, M.; Shlens, J.; Steiner, B.; Sutskever, I.; Talwar, K.; Tucker, P.; Vanhoucke, V.; Vasudevan, V.; Viégas, F.; Vinyals, O.; Warden, P.; Wattenberg, M.; Wicke, M.; Yu, Y.; and Zheng, X. 2015. TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.
- Brown, C. F.; Brumby, S. P.; Guzder-Williams, B.; Birch, T.; Hyde, S. B.; Mazzariello, J.; Czerwinski, W.; Pasquarella, V. J.; Haertel, R.; Ilyushchenko, S.; et al. 2022. Dynamic world, near real-time global 10 m land use land cover mapping. *Scientific Data* 9(1):251.
- Chen, Z.; Wang, C.; Li, J.; Xie, N.; Han, Y.; and Du, J. 2021. Reconstruction bias u-net for road extraction from optical remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14:2284–2294.
- Fern, R. R.; Foxley, E. A.; Bruno, A.; and Morrison, M. L. 2018. Suitability of ndvi and osavi as estimators of green biomass and coverage in a semi-arid rangeland. *Ecological Indicators* 94:16–21.
- Gillies, S., et al. 2013. Rasterio: geospatial raster i/o for Python programmers.
- Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; and Moore, R. 2017. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*.
- Hagy H and Others. 2022. Waterfowl monitoring plan for national wildlife refuges in the southeast.
- Highway, C. J. 2022. Spatiotemporal factors influencing mallard foraging and activity dynamics.
- Ibtehaz, N., and Rahman, M. S. 2020. Multiresunet: Rethinking the u-net architecture for multimodal biomedical image segmentation. *Neural networks* 121:74–87.
- Lillesaeter, O. 1982. Spectral reflectance of partly transmitting leaves: laboratory measurements and mathematical modeling. *Remote sensing of Environment* 12(3):247–254.
- Liu, X.; Deng, Z.; and Yang, Y. 2019. Recent progress in semantic image segmentation. *Artificial Intelligence Review* 52:1089–1106.
- Long, J.; Shelhamer, E.; and Darrell, T. 2015. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3431–3440.
- Ma, T.; Zhang, H.; Ong, H.; Vora, A.; Nguyen, T. D.; Gupta, A.; Wang, Y.; and Sabuncu, M. R. 2021. Ensembling low precision models for binary biomedical image segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 325–334.
- Mimić, G.; Živaljević, B.; Blagojević, D.; Pejak, B.; and Brdar, S. 2022. Quantifying the effects of drought using the crop moisture stress as an indicator of maize and sunflower yield reduction in serbia. *Atmosphere* 13(11):1880.
- Minaee, S.; Boykov, Y. Y.; Porikli, F.; Plaza, A. J.; Kehtarnavaz, N.; and Terzopoulos, D. 2021. Image segmentation using deep learning: A survey. *IEEE transactions on pattern analysis and machine intelligence*.
- Owusu, C.; Snigdha, N. J.; Martin, M. T.; and Kalyanapu, A. J. 2022. Pygee-swtoolbox: A python jupyter notebook toolbox for interactive surface water mapping and analysis using google earth engine. *Sustainability* 14(5):2557.
- Qiao, M.; He, X.; Cheng, X.; Li, P.; Luo, H.; Zhang, L.; and Tian, Z. 2021. Crop yield prediction from multi-spectral, multi-temporal remotely sensed imagery using recurrent 3d convolutional neural networks. *International Journal of Applied Earth Observation and Geoinformation* 102:102436.
- Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III* 18, 234–241. Springer.
- Shamsolmoali, P.; Zareapoor, M.; Wang, R.; Zhou, H.; and Yang, J. 2019. A novel deep structure u-net for sea-land

segmentation in remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12(9):3219–3232.

U.S. Department of Interior. 1986. North american waterfowl management plan. u.s. department of the interior and environment canada, washington, d.c., usa.

Williams, C. K.; Dugger, B.; Brasher, M.; Collucy, J.; Cramer, D. M.; Eadie, J. M.; Gray, M.; Hagy, H. M.; Livolsi, M.; McWilliams, S. R.; Petrie, M.; Soulliere, G. J.; Tirpak, J.; ; and Webb., L. 2014. Estimating habitat carrying capacity for migrating and wintering waterfowl: considerations, pitfalls and improvements. *Wildfowl (Special Issue No.4)* 407–435.

Yeung, M.; Sala, E.; Schönlieb, C.-B.; and Rundo, L. 2022. Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation. *Computerized Medical Imaging and Graphics* 95:102026.

Zhang, Z.; Liu, Q.; and Wang, Y. 2018. Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters* 15(5):749–753.