COMPUTER VISION IN AGRICULTURE

Vid Chan, Jon Kang, Anderson Monken, and Gabriella Zakrocki

December 16, 2020

I. Abstract

Agriculture is vitally important to our collective existence on this planet. With a constant rise in global population, greater food availability with a more efficient use of resources to produce it is more critical than ever. Improvement in agricultural technology is thought as a solution to achieve this goal. There has been considerable advancement in computer vision which includes image and data processing in agriculture over the last decade. However, the lack of credible image dataset remains a challenge for many researchers in this field. A dataset provided by the Agricultural-Vision Challenge which contains 21,061 aerial farmland images captured throughout 2019 across the US is used in this study. Different techniques have been applied to the dataset from Convolutional Encoder-Decoder, Convolutional Encoder-Decoder with batch normalization, concatenated NIR Convolutional Encoder-Decoder, and concatenated RGB convolutional Encoder-Decoder models. To evaluate the results the mean intersection-over-union (mIoU) - one of the most commonly used measures in semantic segmentation, is used.

The best result comes from the convolutional encoder-decoder with mIoU of 29.7%. All results from our models are quite adequate against other teams in the Agricultural-Vision Challenge.

II. Introduction

Deep learning in visual recognition has become increasingly popular in the field of agriculture. However, one of the biggest challenges in conducting such research is the lack of a credible image dataset, which has slowed progress in this research field. The objective of this study is to harness convolutional neural networks to perform image segmentation and detection on satellite agricultural images. Better models for analyzing the status of crop fields will lead to more efficient use of resources and greater food availability.

III. Related Work

Patricio and Rieder (2018) [1] conducted a systematic review of computer vision and

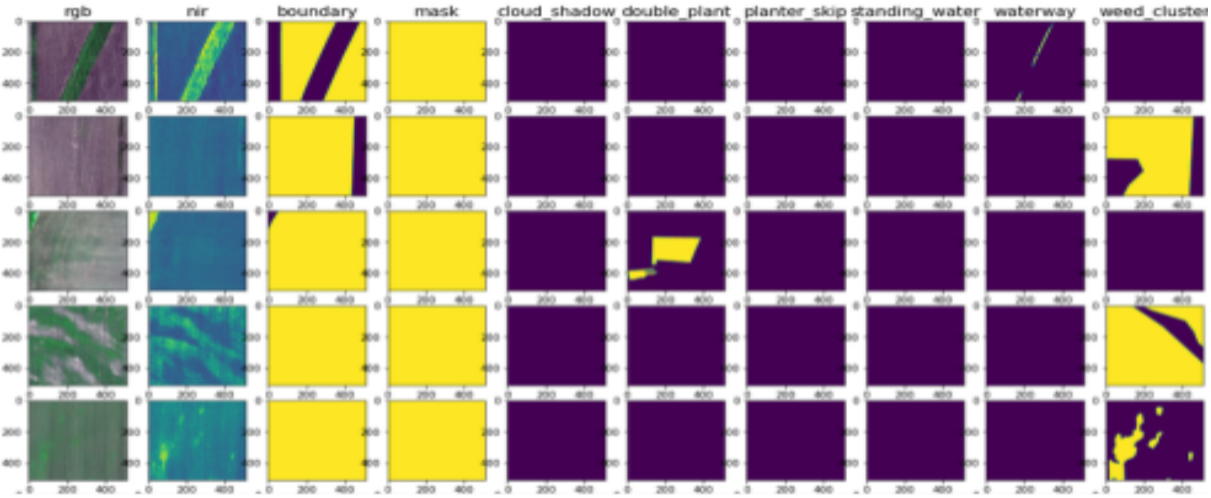

Figure 1

artificial intelligence in precision agriculture for production of the five most produced grains in the world: maize, rice, wheat, soybean, and barley. The study highlights important computer vision solutions combined with AI algorithms such as DBN (Deep Belief Networks) that help solve pattern detection issues in agricultural images. Barbedo (2013) [2] used detection, classification, and quantification techniques such as support vector machines, fuzzy logic, and neural networks to detect, quantify, and classify plant diseases from digital images. Sabanci et al. (2017) [3] used artificial neural networks which depend on multilayer perceptron (MLP) to classify wheat grains into bread or durum. The images of 100 bread and 100 durum wheat grains are taken and subjected to pre-processed before inputting into the ANN model. Using the correlation-based "CfsSubsetEval" algorithm to simplify the ANN model, seven input parameters such as length, length and width ratio, green, blue, green ratio, homogeneity, and entropy are determined as most effective in classifying the results.

IV. Datasets

The Dataset contains 21,061 aerial farmland images captured throughout 2019 across the US. Each image consists of four 512x512 color channels, which are RGB and Near Infra-red (NIR). This dataset contains six types of annotations: Cloud shadow, Double plant, Planter skip, Standing Water, Waterway and Weed cluster, stored as binary masks. Data processing involved combining all of the input channels together into the four channel input image, and the 6 classification channels, the invalid pixel channel, and a computed background pixel channel together into an output image of eight channels.

V. Methods

Our initial Convolutional Encoder-Decoder model involved an encoder block with four steps of 2D convolution, 2D max pooling (2x2), and dropout (0.05), with progressively smaller size and channels, as well as a decoder block with four steps of 2d convolution, upsampling (2x2), and dropout (0.05). The dimensionality of the tensor went from 512x512x4 *(W X H X Channels)* down to 32x32x32 at the smallest point and back up to 512x512x8. See the model figure on page 2 using model visualization software Bäuerle et. al. (2019) [6]. Instead of outputting a flat output image using a softmax activation, we output an eight channel image, one for each class plus invalid pixels, using a sigmoid activation and binary cross entropy loss. After developing our initial model, we applied batch normalization to the Convolutional Encoder-Decoder. Unfortunately due to GPU RAM limitations, not enough images could be run in a batch for normalization to be effective in improving results. Due to mistakes in classifying waterways, which can be seen in

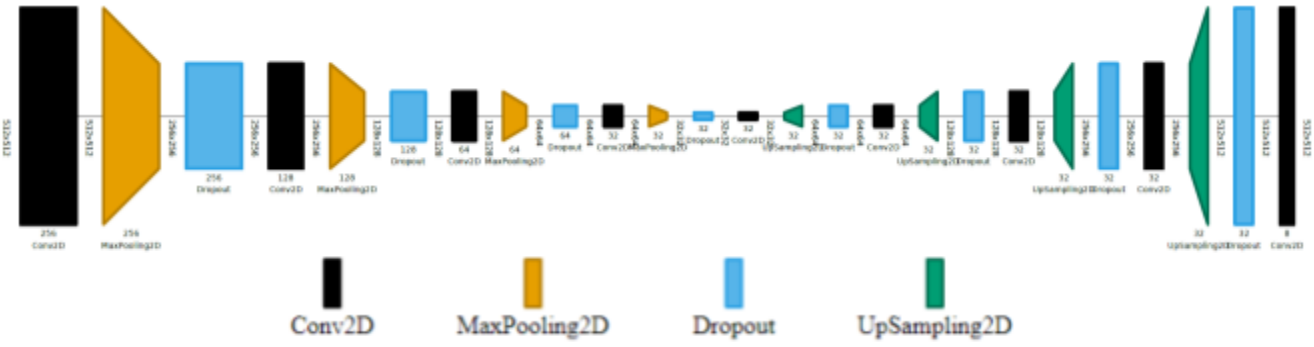

Figure 2

the prediction images on page 4, attempts were made to concatenate a smaller Convolutional Encoder-Decoder for only some of the input channels to amplify the signal from those channels. Concatenating a Convolutional Encoder-Decoder for three RGB channels was attempted along with a similar architecture for the NIR channel only.

VI. Results

The evaluation metric used to quantify our results is the mean intersection-over-union (mIoU):

$$mIoU = \frac{1}{c} \sum \frac{Area\,(P_c \bigcap T_c)}{Area\,(P_c \bigcup T_c)}$$

Where c is the number of annotation types (c = 7 in our dataset, with 6 patterns + background), Pc and Tc are the predicted mask and ground truth mask of class c respectively.

mIoU is one of the most commonly used measures in semantic segmentation. The mIoU accommodates overlapping labels by categorizing predictions of any label in a pixel as a correct prediction. For pixels with multiple labels, a prediction of either label will be counted as a correct pixel classification for that label, and a prediction that does not contain any ground truth labels will be counted as an incorrect classification for all ground truth labels.

Our best model, the Convolutional Encoder-Decoder, had an mIoU value of 29.7%. Figure 3 compares our other models' performances.

VII. Discussion of Results

Our results were rather adequate against the Agriculture Vision Challenge leaderboard. In the table below we see that for our different methods there are trade-offs. For example, the encoder-decoder with batch normalization had a double plant pattern and no planter-skip pattern. Whereas, the regular encoder-decoder had a substantial cloud shadow pattern but no double plant or planter-skip pattern. Figure 4 is an actual image from the dataset split by patterns versus the predicted image by patterns. We can see that the model believed a portion of the highlighted land in the image to be a waterway when in fact it is not.

The team with the highest mIoU recorded thus far for the Ag-Vision Challenge resulted with a 63.9%.The lowest mIoU recorded was a 10.3%. Looking at previous works related to semantic segmentation, there were various methods taken to solve this problem. The first related work, *Farm Parcel Delineation Using Spatio-temporal Convolutional Networks* (2020) [4], utilizes various U-Net methodology to segment farm parcel areas and boundaries in satellite images. The U-Net differs from our model by coupling the encoder-decoder rather than separating both. Thus meaning the output relies on the input. We did not implement the u-net method in our models, but it would be something to try for further work.

Model	mIoU (%)	Background	Cloud Shadow	Double Plant	Planter Skip	Standing Water	Waterway	Weed cluster
Encoder-decoder	29.7	74.6	28.6	0.0	0.0	13.9	46.4	44.0
w/ batch normalization	29.5	80.3	0.5	18.0	0.0	35.2	32.8	40.0
concat NIR e-d on e-d	17.4	65.7	0.3	0.0	0.3	5.4	13.9	36.5
concat RGB e-d on e-d	29.3	79.7	0.3	18.0	0.1	42.8	25.1	39.3

Figure 3

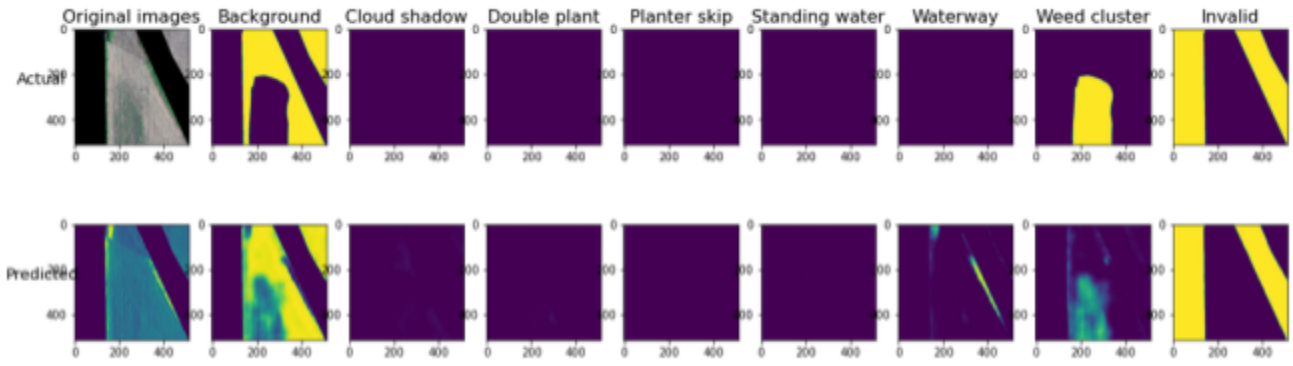

Figure 4

Along with the first related work, the second work is beneficial because it is from the same Ag-Vis Challenge dataset. *The 1st Agriculture-Vision Challenge: Methods and Results* (2020) [5] writes about a few notable methods that were submitted to the Ag-Vis Challenge and takes a deeper dive into some teams' research. The team with the highest mIoU performed a Residual DenseNet with Squeeze-and-Excitation blocks (RD-SE). RD-SE is based on U-Net architecture but it has an encoder/decoder format, like our model. Looking at the models created by a number of different teams helps to increase our awareness of future work and possibilities to improve our next models.

VIII. Conclusions

Our project worked successfully with a basic encoder-decoder model resulting in an mIoU of 29.7%. The other models we tried were not as successful, although the Convolutional Encoder-Decoder with batch normalization came close. Future directions we would plan to take would be to use other image data augmentation techniques such as rotation, zoom, brightness, blur, and sheer. We also would like to expand our model types by adding skip connections and trying out transfer learning using VGGNet or ResNet as the base. Overall, a convolutional encoder-decoder was a decent model for this dataset, but we believe that our future work would bring even better results.

IX. Reference

[1] Diego Inacio Patricio, and Rafael Reider Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review
ScienceDirect, volume 153 (2018), p. 69-81
<https://www.sciencedirect.com/science/article/abs/pii/S0168169918305829?casa_token=Wu38g4EdmDwAAAAA:zB0rC36VGIPDve4tBT2NJNOUXL-EJu-4qJYc9aHL_RGrfNZoGrgOFR651hkRsu2otERlvKyi7OY>

[2] J.G.A. Barbedo
Digital image processing techniques for detecting, quantifying and classifying plant diseases
SpringerPlus, 2 (1) (2013), p. 660, 10.1186/2193-1801-2-660
<http://springerplus.springeropen.com/articles/10.1186/2193-1801-2-660>

[3] K. Sabanci, A. Kayabasi, A. Toktas
Computer vision-based method for classification of wheat grains using artificial neural network
J. Sci. Food Agric., 97 (8) (2017), pp. 2588-2593, 10.1002/jsfa.8080

[4] Aung, Han Lin, et al. "Farm Parcel Delineation Using Spatio-Temporal

Convolutional Networks.” *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2020, doi:10.1109/cvprw50498.2020.00046.

[5] Chiu, Mang Tik, et al. “The 1st Agriculture-Vision Challenge: Methods and Results.” *Http://Openaccess.thecvf.com/*, Computer Vision Foundation, 1 Jan. 2020, openaccess.thecvf.com/content_CVPRW_2020/html/w5/Chiu_The_1st_Agriculture-Vision_Challenge_Methods_and_Results_CVPRW_2020_paper.html.

[6] Bäuerle, Alex, Christian van Onzenoodt, and Timo Ropinski. "Net2Vis: transforming deep convolutional networks into publication-ready visualizations." *arXiv preprint arXiv:1902.04394* (2019). https://arxiv.org/abs/1902.04394