Agriculture-Vision Challenge 2022 – The Runner-Up Solution for Agricultural Pattern Recognition via Transformer-based Models

Abstract

This paper explores the adaptation The Agriculture-Vision Challenge is one of the most famous and competitive challenges for global researchers to break the boundary between computer vision and agriculture sectors, aiming at agricultural pattern recognition from aerial images. In this paper, we propose our solution to the third Agriculture-Vision Challenge. We leverage a data pre-processing scheme and several Transformer-based models as well as data augmentation techniques to achieve a mIoU of 0.582, accomplishing the 2nd place in this challenge.

1 Introduction

This paper addresses the critical Computer vision applications in agricultural domain has become one of hot topics nowadays, especially using remote sensing satellite images and aerial images. With the rapid development of deep learning methods, numerous research studies have proposed pioneer and practical solutions to various computer vision problems in agriculture. Aside from fruitful research achievements, various algorithm challenges have been held at top-tier conferences for global researchers in recent years, in order to explore more effective algorithms to solve the specific problems. The Agriculture-Vision Challenge is one of most famous and competitive challenges in this inter-disciplinarity study. It aims at applying computer vision algorithms to agricultural pattern recognition from high-resolution aerial images. This year, holds the 3rd Agriculture-Vision Challenge, and we form our team to participate in this contest.

2 Related Work

This section reviews

3 Methodology

This section details of In this section, we elaborate on the given datasets, the pre-processing method, the proposed deep learning-based framework, and the test-time augmentation (TTA) strategy.

3.1 Description of Dataset

The challenge this year provides the entire Agriculture-Vision dataset. It contains 94,986 aerial farmland images collected throughout 2019 across the U.S. Each image has a size of 512×512 pixels and has 4 channels (RGB and NIR). A total of 9 label classes are manually labeled for every image. Table 1 shows the given amount of images in each class. Note that many images have multiple labels, and even have overlapped labels (one pixel has multiple labels).

Although the amount of the given training data is considerable, we still generate more data following the data augmentation scheme of the winner solution last year. They conducted an image mosaic

.

scheme to enable the model to have multi-scale views during the training. To fit the model input size, we create two new datasets using mosaicked images with down-sampling 2X (2 times) and down-sampling 3X. The down-sampling dataset has the same image size of 512×512 pixels that the recognition model can share the same network architecture among 1X, 2X, and 3X imagery.

3.2 Data Pre-Processing

We observe that the image counts in each category are uneven. For example, the image count of the background class is 25 times larger than the water class. To tackle the unbalance issue, we try to sample more images in the few-shot classes. The re-sampled image counts are listed in Table 1.

Table 1: Information of the given and resampled datasets for training and validation categories.

Class Index	Class Name	Original Amount (Train/Val)	Resampled Amount (Train/Val)
0	Background	56944 / 18334	75121 / 13642
1	Double Plant	6234 / 2322	10961 / 2294
2	Drydown	16806 / 5800	19320 / 3383
3	Endrow	4481 / 1755	8544 / 1858
4	Nutrient Deficiency	13308 / 3883	14859 / 2610
5	Planter Skip	2599 / 1197	5361 / 1015
6	Water	2155 / 987	4132 / 721
7	Waterway	3899 / 696	6024 / 1109
8	Weed Cluster	11111 / 2834	14423 / 2773

3.3 Framework

Fig. 1 shows our deep learning-based framework. SegFormer is a Transformer-based efficient segmentation model. It designs a hierarchical Transformer encoder with multi-level feature outputs. Unlike other cumbersome decoders, SegFormer's decoder adopts MLP layers to aggregate multi-scale feature outputs from different layers. One of the key advantages of SegFormer is that its model size is relatively small but the performance keeps outstanding. Therefore, SegFormer is suitable for this challenge due to the model size parameter limit of 150M.

SegFormer provides six versions with various settings of Transformer encoders, leading to different model sizes. These six models are named from B0 to B5, with the increased model size. To follow the policy, we select Mix Transformer (MiT) B3 and Mix Transformer B2 as our training models. Their model size information can be found in Table 7 "Mix Transformer Encoder". After obtaining the individual inference result from each model, the model ensemble is performed to predict the final segmentation results.

3.4 Test-Time Augmentation

Since our models are trained with 1X, 2X, and 3X down-sampling imagery, we conduct the same processing on the test dataset. In addition to the scale augmentation, we include image rotation and flip.

4 Results

This section presents the results

4.1 Evaluation Metric

The required evaluation metric is the average Intersection over Union metric (mIoU), which is defined as Eq. 1 to measure the performance.

$$mIoU = \frac{1}{c}\sum_{i=1}^{c}\frac{Area(P_c \cap T_c)}{Area(P_c \cup T_c)} \tag{1}$$

where c is the number of label classes (8 foreground classes + 1 background class for this challenge); P_c and T_c are the predicted label mask and ground truth label mask of the class c, respectively.

4.2 Experiment Results

Table 2 presents our results, the baseline provided by the host Agriculture-Vision organizers, and the results of other methods. Note that other baselines evaluate their performance on the validation set due to the unavailable test set. As we can see, while our single model baselines are competitive with other baselines, our proposed method effectively improves the single model performance. Even though some single models have peak performance in some classes (0.778 for "Background" and 0.782 for "Water"), our model ensemble enjoys the merits of multiple single models' strength to achieve the mIoU of 0.582. It also shows that our ensemble results significantly outperform other baselines and our implementation of various single models.

Table 2: Performance comparisons among various models. The bold font of numeric results indicates the best performance on the test set. BG: Background; DP: Double Plant; D: Drydown; E: Endrow; ND: Nutrient Deficiency; PS: Planter Skip; W: Water; WW: Waterway; WC: Weed Cluster. The number in the parentheses following the class name refers to the class index.

Models	mIoU	BG(0)	DP(1)	D(2)	E(3)	ND(4)	PS(5)	W(6)	WW(7)
(Other methods, on the val set)									
Agriculture-Vision baseline(RGBN)	0.434	0.743	0.285	0.574	0.217	0.389	0.336	0.736	0.344
MiT-B3(RGBN)	0.454	0.768	0.371	0.609	0.245	0.424	0.413	0.692	0.269
MiT-B5(RGB)	0.464	0.755	0.370	0.585	0.227	0.313	0.414	0.802	0.401
MiT-B5(RGBN)	0.490	0.762	0.373	0.618	0.246	0.428	0.420	0.813	0.437
(Our implementation, on the test set)									
HRNet-W48+OCR(RGB baseline)	0.413	0.717	0.316	0.567	0.233	0.269	0.283	0.718	0.289
MiT-B3(RGB baseline)	0.448	0.720	0.395	0.557	0.325	0.364	0.330	0.687	0.293
MiT-B2(RGBN+Our method)	0.554	0.778	0.483	0.632	0.476	0.570	0.403	0.768	0.410
MiT-B3(RGBN+Our method)	0.563	0.773	0.471	0.640	0.452	0.569	0.442	0.782	0.463
Model Ensemble(RGBN+Our method)	**0.582**	0.777	**0.485**	**0.646**	**0.481**	**0.573**	**0.471**	**0.779**	**0.547**

5 Conclusion

This paper presents a novel method In this paper, we propose our solution to the 3rd Agriculture-Vision Challenge. For data usage, we perform data pre-processing and test data augmentation schemes. Several SegFormer models are leveraged. We finally accomplish a mIoU of 0.582, achieving the 2nd place in this challenge.

Future Directions. The potential applications of our proposed algorithm include crop type identification in precision agriculture, agricultural asset estimation and agricultural insurance product design in the Environmental, Social, and Governance (ESG) domain. These future directions can illuminate the revitalization of rural areas and facilitate the service of inclusive finance in an eco-friendly way.