Report: Agri-vision Challenge 2021

I. Team details

- **Challenge Track**: Supervised Track
- **Team name**: AInnovation-SmileLab
- **Username on Codalab**: smilelab2021
- **Team leader email**: sun.bi@northeastern.edu
- **Team members (and affiliation)**:
  Bin Sun[2], Songyao Jiang[2], and Yun Fu[1,2],
  1. AInnovation Labs Inc.
  2. Department of Electrical and Computer Engineering, Northeastern University, Boston, MA
- **Team website URL (if any)**:
  https://www.ainnovationlab.com/
  https://web.northeastern.edu/smilelab/

II. Contribution details

A. Introduction and Motivation

Ensemble of multiple models is an effective way to improve performance. Since some top-rank methods in the previous challenge have their codes public, it is straightforward to use their codes for training. Therefore, we choose two public methods for this challenge. One is MSCGNet [1] which uses a graph Network for image segmentation. The other one is DeepLabv3 [2] which uses ResNet101 as the backbone, as switchable normalization as its batch normalization, and multi-loss combination as its loss function. Besides, we also design a modified Jaccard loss [3] based on the modified canonical mIoU metrics [4], [5], and the unbalance of training data.

B. Detailed Method Description

In this subsection, we will first introduce our modified Jaccard loss and data re-balance strategy. Then we will introduce our ensemble strategy. At last, we will show some details about the training process.

1) Modified jaccard loss and data re-balance: We can learn from the AgriVision-Challenge 2020 [4], [5] that the evaluation metric of the challenge considers the overlap of different classes and increase the reward and penalty of the overlap region. Therefore, we similarly modify the Jaccard loss to stress the region of overlap. Besides, we also add weights to different labels due to the unbalance of different labels in training data. The original Jaccard loss can be written as:

$$L = 1 - \frac{pred * gt}{pred + gt - pred * gt}, \tag{1}$$

where $pred$ means the vector of prediction, gt means the vector of ground truth. The modified Jaccard loss can be

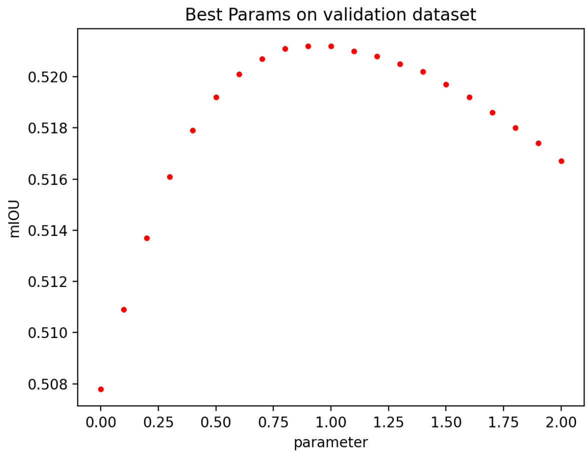

Fig. 1. Best hyper-parameter for ensemble through experiment on the validation dataset. The parameter is multiplied on the output of MSCGNet.

written as:

$$\begin{aligned} newgt &= \sum_c gt_c * gt + gt * weight, \\ L &= 1 - \frac{pred * newgt}{pred + newgt - pred * newgt}, \end{aligned} \tag{2}$$

where $newgt$ means the modified ground truth vector in the loss, c means class, $weight$ means the weight vector to re-balance the data. We set $weight$ as $[0, 1, 0, 1, 0, 1, 0, 0, 0, 1]$, which means the weights of the labels lack of data (label 1,3,5,8) are set to 1 and the others' weights are set to 0. Besides modifying loss, we also repeat the data of weak labels to re-balance the training data.

2) Ensemble Strategy: After training the MSCGNet and deeplabv3, we sum the output of the two models to get better results. To decide the best parameter for ensemble, we test the parameter from 0 to 2 on the validation dataset with stride as 0.1. The result is shown Figure 1. The best parameter is 0.9.

3) Details of training process: For MSCGNet, we follow the provided training process [1]. For DeepLabv3, we mainly follow the provided training process [2], except we use our modified Jaccard loss to re-train the model with 20 epochs. While re-training, the optimizer is SGD, the weight decay is set to 0, and the learning rate is set to 0.005. The dataset we use is original training data plus the data with multi-labels and the endrow data. We use multiple loss combinations to fine-tune the model. Before the test phase, we use validation

[1] https://github.com/samleoqh/MSCG-Net
[2] https://github.com/LAOS-Y/AgriVision

data and training data to fine-tune the model for 20 epochs with the learning rate 0.001. We use ResNet101 training on ImageNet as our pre-trained model.

C. Challenge Results and Final Remarks

Our approaches are evaluated on the testing server using the modified mIoU. As shown in Table I, our ensemble model improves the performance of each single model and achieves higher mIoU (0.507). The total number of parameters are 91M, which is within the 150M limits as required by the challenge.

TABLE I

RESULTS OBTAINED BY THE PROPOSED APPROACHES ON TEST SET.

Model	**Backbone**	**# Params**	**mIoU**
MSCG-Net	Resnet-101	31M	0.464
DeepLabv3	Resnet-101	60M	0.494
Ensemble	DeepLabv3+0.9MSCGNet	91M	**0.507**

REFERENCES

[1] Q. Liu, M. C. Kampffmeyer, R. Jenssen, and A.-B. Salberg, "Multi-view self-constructing graph convolutional networks with adaptive class weighting loss for semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 44–45.

[2] S. Yang, S. Yu, B. Zhao, and Y. Wang, "Reducing the feature divergence of rgb and near-infrared images using switchable normalization," in *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020.

[3] L. Hamers *et al.*, "Similarity measures in scientometric research: The jaccard index versus salton's cosine formula." *Information Processing and Management*, vol. 25, no. 3, pp. 315–18, 1989.

[4] M. T. Chiu, X. Xu, Y. Wei, Z. Huang, A. G. Schwing, R. Brunner, H. Khachatrian, H. Karapetyan, I. Dozier, G. Rose, D. Wilson, A. Tudor, N. Hovakimyan, T. S. Huang, and H. Shi, "Agriculture-vision: A large aerial image database for agricultural pattern analysis," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[5] M. T. Chiu, X. Xu, K. Wang, J. Hobbs, N. Hovakimyan, T. S. Huang, and H. Shi, "The 1st agriculture-vision challenge: Methods and results," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 48–49.