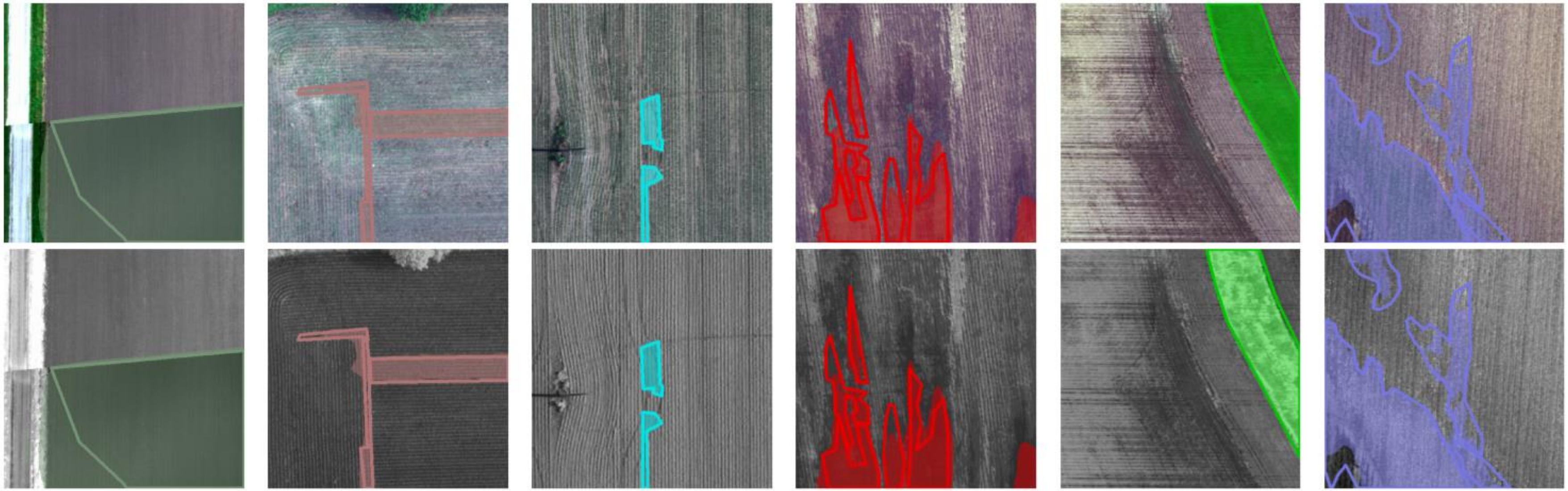
Reducing the feature divergence of RGB and near-infrared images using Switchable Normalization

Agriculture-Vision Workshop @ CVPR 2020

Siwei Yang* , Shaozuo Yu* , Bingchen Zhao* , Yin Wang
Tongji University, Shanghai, China

* Equal contribution

Multi-modality Nature of Agriculture Data



(a) cloud shadow (b) double plant (c) planter skip (d) standing water (e) waterway (f) weed cluster

Figure 3: For each class, top: RGB image; bottom: NIR image; line: ground truth; mask: prediction. All predictions are given by IBN-Net101-s trained with BCE+Dice+Lovasz as loss function.

Feature Divergence

We use symmetric KL divergence to measure the difference of between features on RGB and NIR images.

$$KL(F_{iA}\|F_{iB}) = log\frac{\sigma_{iA}}{\sigma_{iB}} + \frac{\sigma_{iA}^2 + (\mu_{iA} - \mu_{iB})^2}{2\sigma_{iB}^2} - \frac{1}{2} \tag{1}$$

$$D(F_{iA}\|F_{iB}) = KL(F_{iA}\|F_{iB}) + KL(F_{iB}\|F_{iA}) \tag{2}$$

$$D(L_A\|L_B) = \frac{1}{C}\sum_{i=1}^{C} D(F_{iA}\|F_{iB}) \tag{3}$$

μ_A and σ_A stand for mean and standard deviation of the output of a certain channel from a certain layer over modality A.

Check out our paper for more details.

Feature Divergence

	mIoUs(%)	
	Train on RGB	Train on NIR
Test on RGB	46.05	23.39
Test on NIR	18.29	44.22
Perf. Decay	27.76	20.83

Table 1: 'Perf. Decay' stands for "Performance decay". When models are trained on NIR images alone, NIR images with a single channel will be duplicated channel-wise to transform into a three-channel images. All the other settings about the experiment regarding this table are the same as ones in Section 5. Testing on different modalities from training can reduce the performance by a huge margin. This phenomenon suggests that the feature divergence between RGB and NIR can be huge.

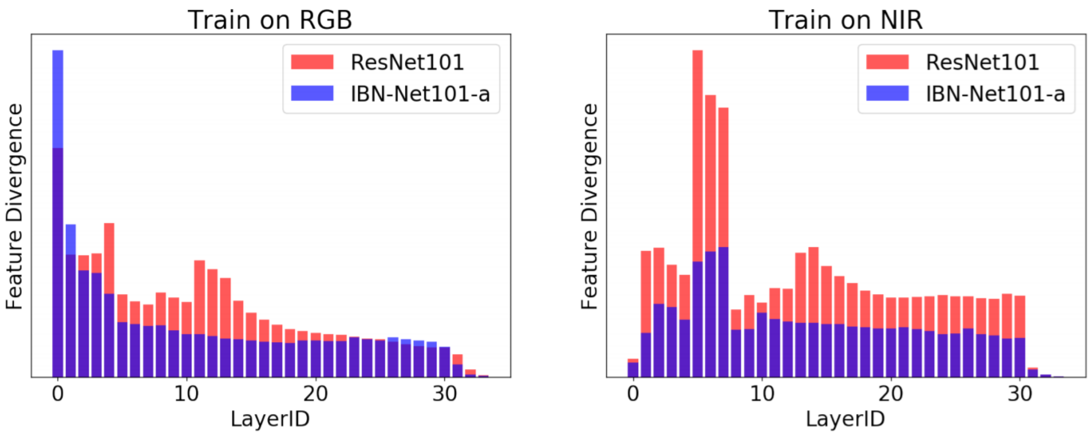


Figure 2: Both figures represent feature divergence between RGB images and NIR images. The vertical axis indicates the feature divergence between RGB images and NIR images. IBN-Net101-a reduces the feature divergence dramatically when trained on RGB images or NIR images.

Methodology

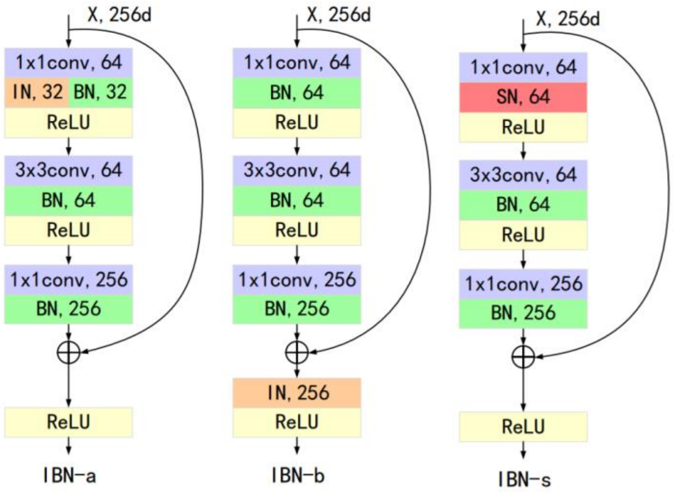


Figure 1: Instance-batch normalization (IBN) block. IBN-a and IBN-b are the variants of IBN-Net proposed in [16], we change the first BN to Switchable-Normlization for better performance.

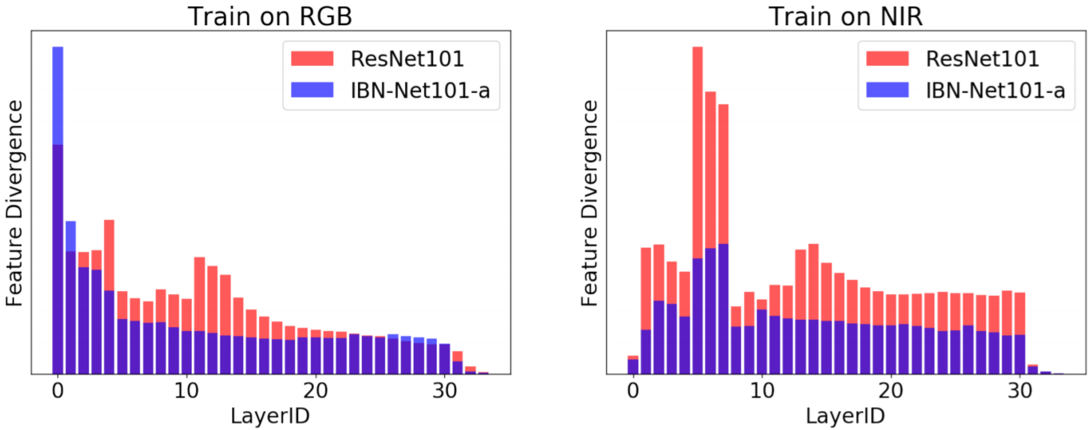


Figure 2: Both figures represent feature divergence between RGB images and NIR images. The vertical axis indicates the feature divergence between RGB images and NIR images. IBN-Net101-a reduces the feature divergence dramatically when trained on RGB images or NIR images.

Methodology

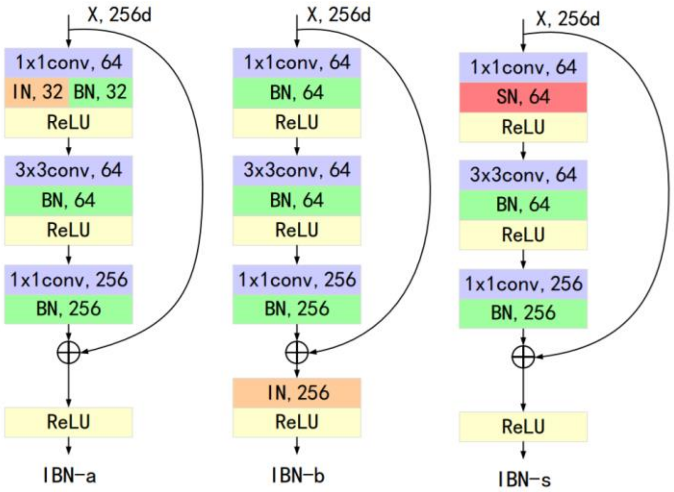


Figure 1: Instance-batch normalization (IBN) block. IBN-a and IBN-b are the variants of IBN-Net proposed in [16], we change the first BN to Switchable-Normlization for better performance.

Backbone	IoUs(%)							
	Background	Cloud shadow	Double plant	Planter skip	Standing Water	Waterway	Weed Cluster	Mean IoU
ResNet50	79.95	45.75	36.98	1.18	59.67	58.03	48.58	47.16
IBN-Net50-a	80.51	**53.67**	40.87	1.83	64.44	**61.67**	49.71	50.39
IBN-Net50-b	80.41	52.48	39.32	**4.19**	62.82	57.91	49.36	49.50
IBN-Net50-s	**80.82**	53.52	**43.63**	3.59	**65.44**	60.20	**50.84**	**51.15**
ResNet101	79.32	47.95	39.77	0.98	62.47	61.17	49.36	48.72
IBN-Net101-a	80.79	52.64	38.27	2.72	**67.52**	61.96	48.52	50.35
IBN-Net101-b	**80.88**	52.05	40.75	3.19	64.21	59.88	**51.05**	50.29
IBN-Net101-s	80.78	**52.69**	**44.53**	**3.34**	66.26	**62.26**	50.39	**51.46**

Table 2: Different backbones setting with DeepLabV3+ architecture and BCE+Dice as loss function on NRGB images. The detailed loss defintion is given in 4. IBN-Net101-s achieves highest mIoU of 51.46% on val set.

Loss

$$L_{hybrid} = \frac{1}{1+\lambda_1+\lambda_2}(L_{BCE} + \lambda_1 L_{Dice} + \lambda_2 L_{Lovasz})$$

λ_1	λ_2	IoUs(%)							
		Background	Cloud shadow	Double plant	Planter skip	Standing Water	Waterway	Weed Cluster	Mean IoU
0	0	80.49	51.09	40.98	3.77	63.75	57.01	49.75	49.55
1	0	80.79	52.64	38.27	2.72	**67.52**	61.96	48.52	50.35
0	1	**81.35**	55.43	40.33	1.42	65.70	**64.30**	48.83	51.05
1	1	81.06	**57.56**	**46.30**	**12.45**	65.11	60.63	**51.55**	**53.52**

Table 3: Different loss settings with IBN-Net101-a+DeepLabV3+ on NRGB images. (λ_1, λ_2)=(1, 1) achieves highest mIoU of 53.52% on val set.

Tricks used in competition

Train-time Augmentation

Scale Ratio ~ U(-1.0, -1.0)

Aspect Ratio ~ U(-1.0, -1.0)

Shift Ratio ~ U(-1.0, -1.0)

Vertical Flip

Horizontal Flip

Rotate 90°

Test-time Augmentation

Vertical Flip

Horizontal Flip

Rotate 90°

IoU on Test Set

Class	IoU	Rank
Background	0.799	12
Cloud shadow	0.366	24
Double plant	0.548	6
Planter skip	0.414	8
Standing water	0.698	16
Waterway	0.669	4
Weed cluster	0.520	15
Mean	**0.574**	**9**

CVPR SEATTLE WASHINGTON JUNE 16-18 2020

Thank you

Link to Github:

https://github.com/LAOS-Y/AgriVision