WASEDA University
Graduate School of Information, Production and Systems

Multi-Scanning Based Recurrent Neural Network for Hyperspectral Image Classification

Presenter: Weilian Zhou
From Kamata Lab.

Outline

- **Introduction**
  ① Hyperspectral image
  ② Two intrinsic properties
  ③ HSI classification
- **Existing works**
  ① Overview
  ② Classification results of examples
- **Observations**
- **Motivation-1**
- **Motivation-2**
- **Proposed method**
  ① Multi-scanning strategy
  ② Spectral-spatial structure of HSI for RNN
  ③ Whole framework of the method
- **Experiments**
- **Comparison**
- **Conclusion**

Introduction (Hyperspectral Image (HSI))

- **What is HSI?:** Hyperspectral sensors portrays the surface of the earth using hundreds of spectral bands in electromagnetic spectrum. **The generated image is called hyperspectral image (HSI) considered as a 3-D cube.**
- **Spectral dimension (1D): Capture the spectral signature of each material in different bands**. One pixel is a N-dimension feature vector representing the value of each band. (N denotes the number of bands)
- **Spatial dimension (2D): Record the locations of objects.**

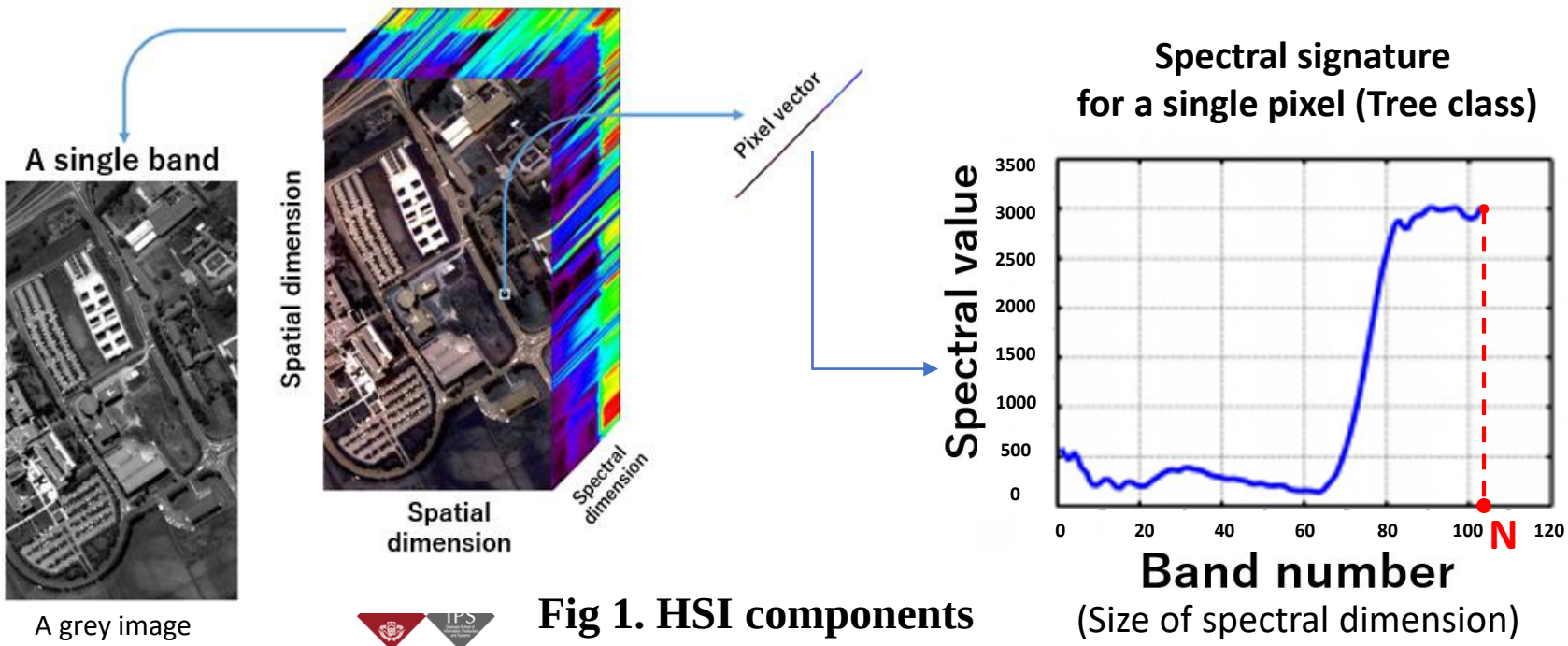


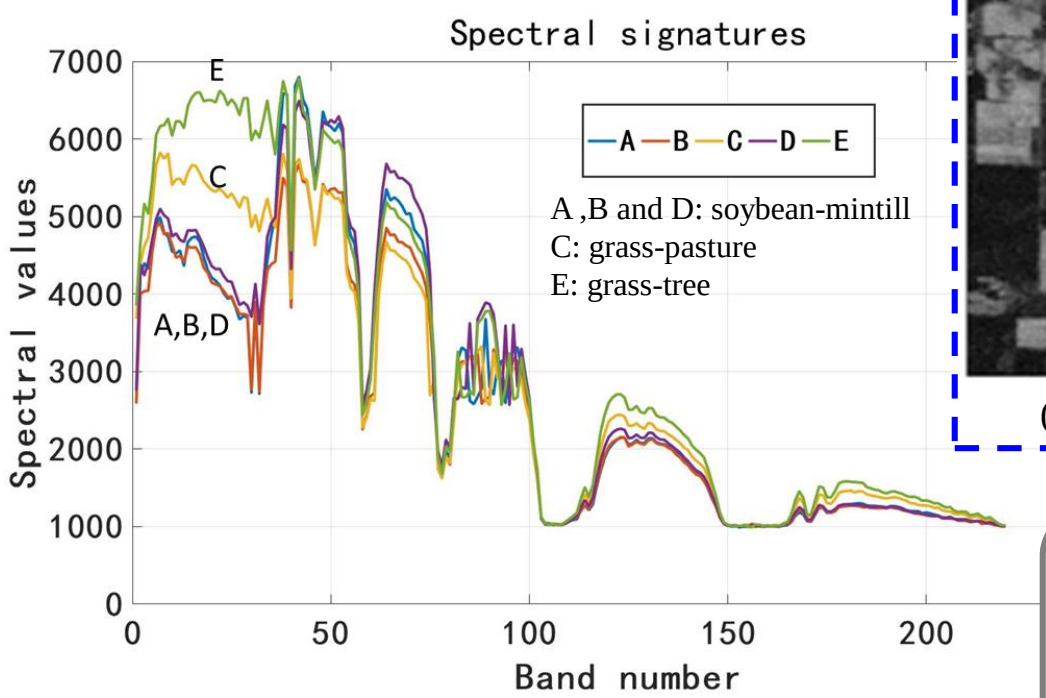
Fig 1. HSI components

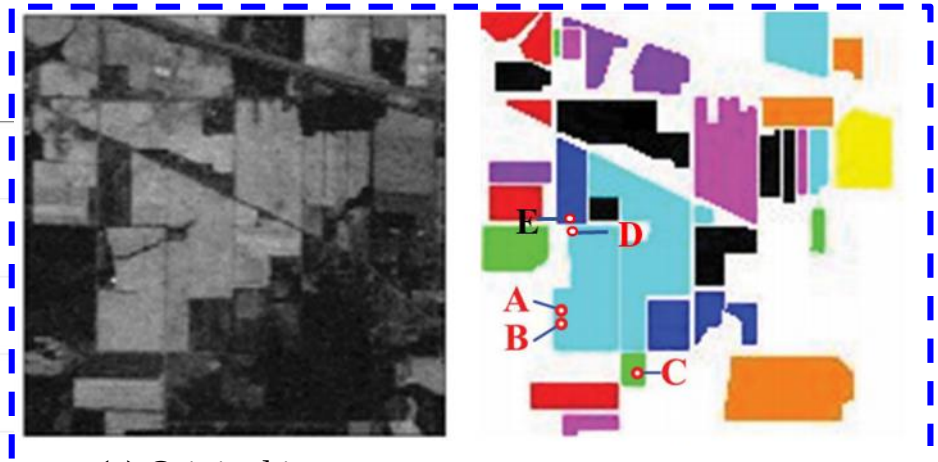
https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html

Introduction (Two intrinsic properties in HSI)

① The pixels that are close in the feature space (spectral dimension) are highly possible to have the same label. (A ,B and D)

② The pixels that are spatially nearby (spatial dimension) have the large possibilities to have the same label, but exception exists. (A and B, E and D)





(a) Original image. (b) Ground truth with A,B,C,D and E pixels

It's important to consider

- **Spatial (neighboring pixels)**
- **Spectral (hyperspectral signatures)**

information together to determine pixel's label.

Introduction (Hyperspectral Image Classification)

- **Target:** To assign a class label to each pixel.
- **Applications**: Environmental detection, urban planning and resource exploring, etc.
- **Recent situation:** Deep learning with this task got outstanding performances.

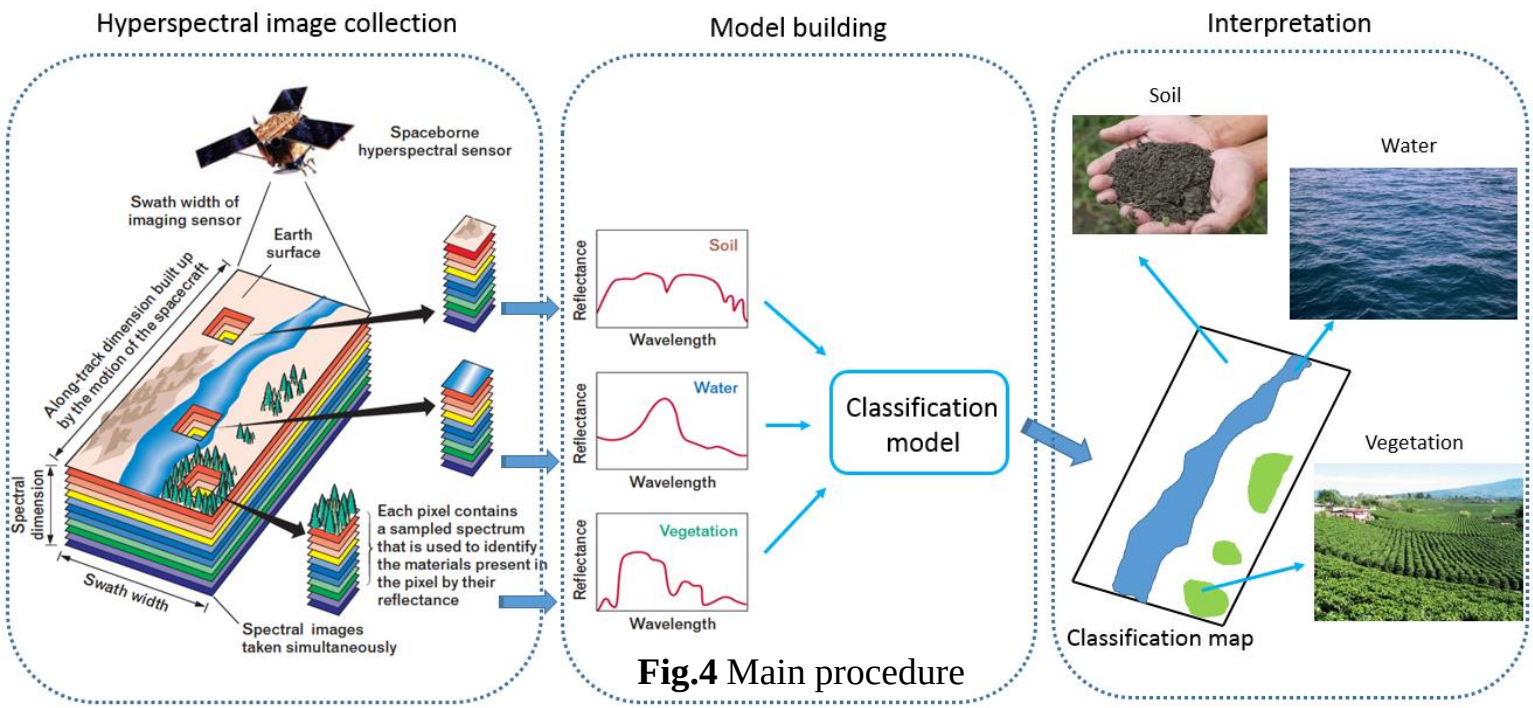


Fig.4 Main procedure

Why we require these applications?

1. Visualizing the object distributions of the ground surface.
2. Exploring what we want to discover on the ground.

https://ssl.jspacesystems.or.jp/project_hisui/

The better classification results will help improve the quality of these applications.

HSI classification is important!

Existing work (Overview)

- Recent HSI classification works with deep learning can be divided into three categories. [12]

1. (1D) Spectral-based 2. (2D) Spatial-based 3. Spectral and Spatial-based

Categories	Existing Works	Methods	Problems
1. Spectral	Hu et al.[1] (*J. Sensors*)	1-D CNN for analyzing spectral signature	Just input the raw spectral signatures into models; None of them consider the 'redundancy' between adjacent bands and complementarity between non-adjacent bands.
	Mou et al.[2] (*IGARSS*)	Deep RNN for finding the relationship between spectral information	
2. Spatial	Li et al.[7] (*IGARSS*)	Fully CNN with three principle components of PCA to analyze	None of them consider the complete spectral information. And neighboring inferring pixels may exist.
	Feng et al.[5] (*Neurocomputing*)	First PCA component as input for LSTM	
3. Spectral-Spatial	Leng et al.[8] (*ICTAI*)	A novel Cube-CNN-SVM. Target pixel with spectral information and its neighbors are organized into 3D-CNN	3D-CNN is suitable for the data structure of the HSI. Though most of the local space areas are homogeneous in an HSI, casually selecting all the pixels' spectral information in the local aera (patch) incurs multiplication and addition from different classes during the training.
	Li et al.[9] (*Remote sensing*)	3D-CNN without relying on any pre-processing or post-processing to extract the in-depth combined features.	

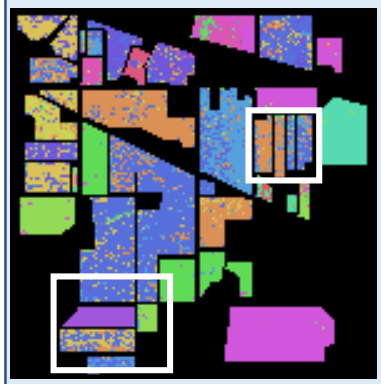
Today's topic

Next slide

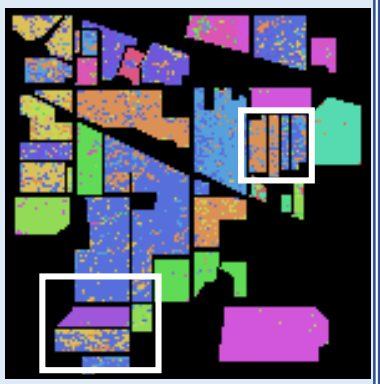
Existing work (classification results on three categories)

- The classification results on **Indian Pines dataset**. (145*145*200)

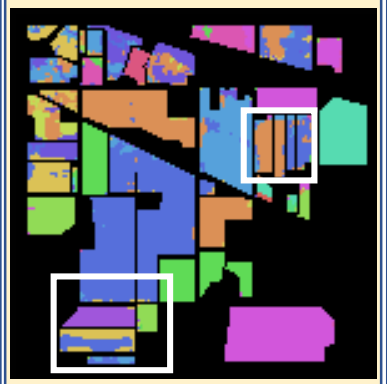
Spectral-based | **Spatial-based** | **Spectral-Spatial-based**



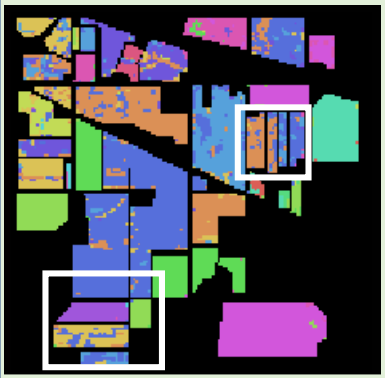
Hu[1]: 75.882%



Mou[2]: 76.175%



Liu[13]: 80.804%



Li[9]: 82.642%

Leng[8]: 90.488%

Drawback:

① Noise-like misclassification everywhere.

② One pixel's neighboring pixels are mess.

Merit:

① Less noise-like misclassification;

② neighboring pixels are considered.

③ More accurate and smooth in flat region

Drawback:

① Object-like misclassifications.

② Misclassifications mainly on boundaries.

Merit:

① Accuracy is higher.

② Misclassification regions are smaller and smoother than spatial-based method.

③ Less noise-like misclassification than spectral-based method.

④ The boundaries are cleaner.

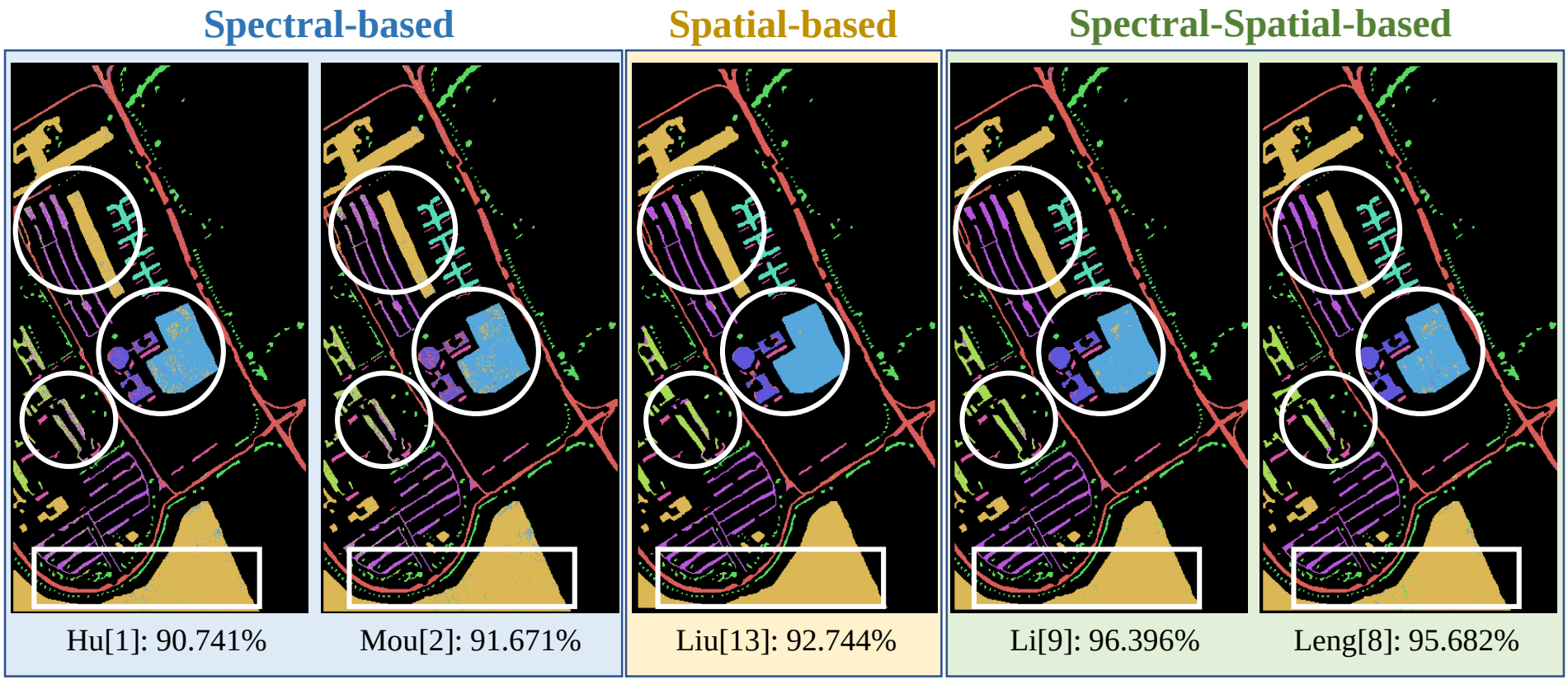
Drawback:

① Computational burden with long training time.

② Still has noise-like and object-like misclassification.

Existing work (classification results on three categories)

- The classification results on **Pavia Uni. dataset** (610*340*103)



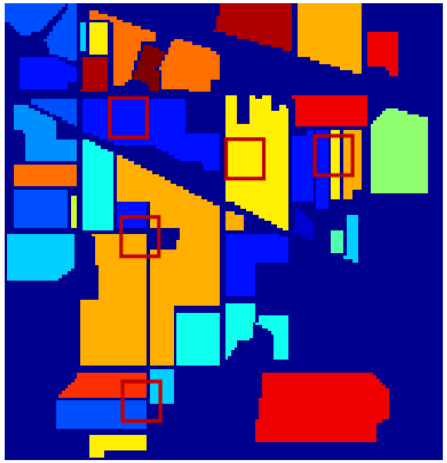
It is better to combine the spatial and spectral information together.

but

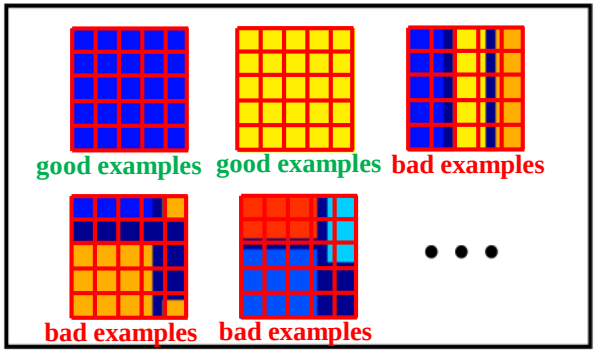
Observation (spatial information)

- **Assumption**: in spatial domain, adjacent pixels are highly likely to have same labels.
- **Real situation**: at edge of each land-cover area, cropped HSI cubes usually contain several pixels whose labels are different from the center pixel. (**interfering pixels**)
- **Problem**: the pixels in the local patch may have different appearance variations, and this probably affects the final classification accuracy.

Original image



Ground truth

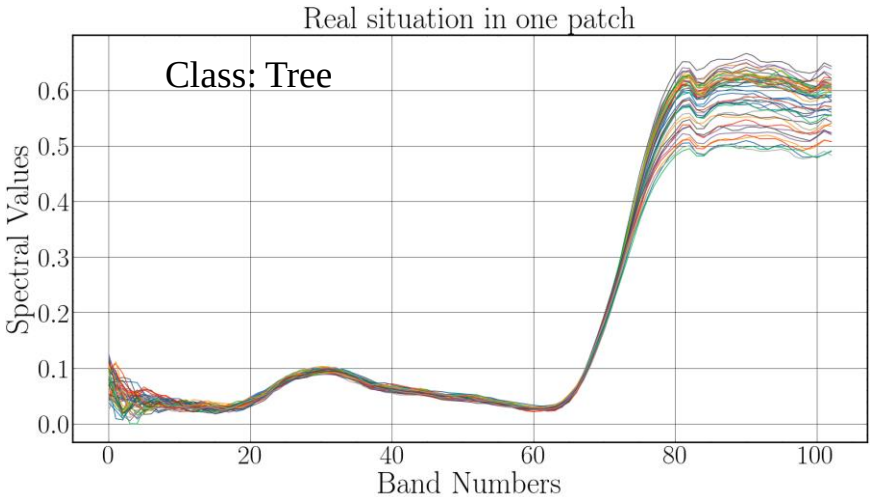


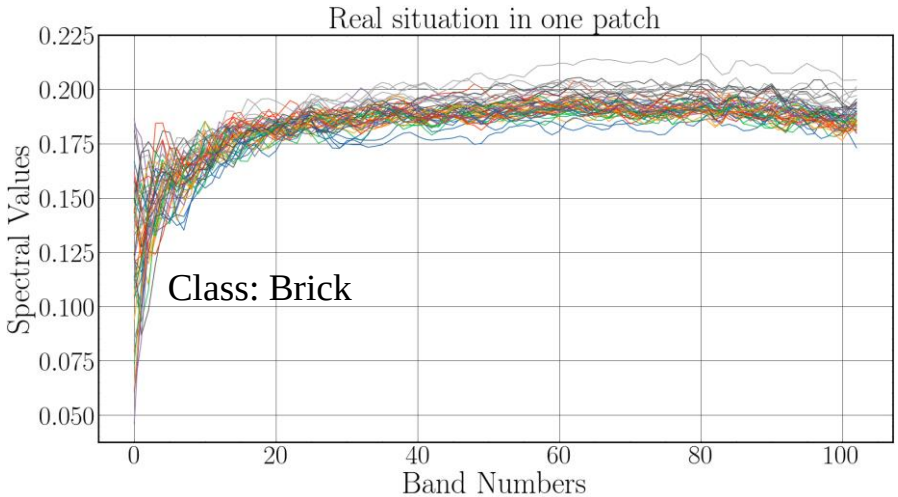
Cropped patch(5*5)

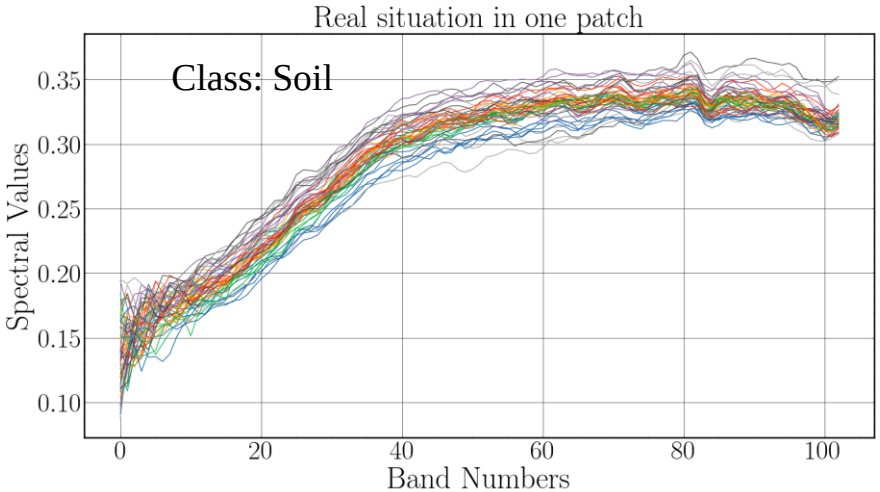
Patch: The main idea is to associate each pixel within an adaptive neighborhoods (to combine the spatial information)

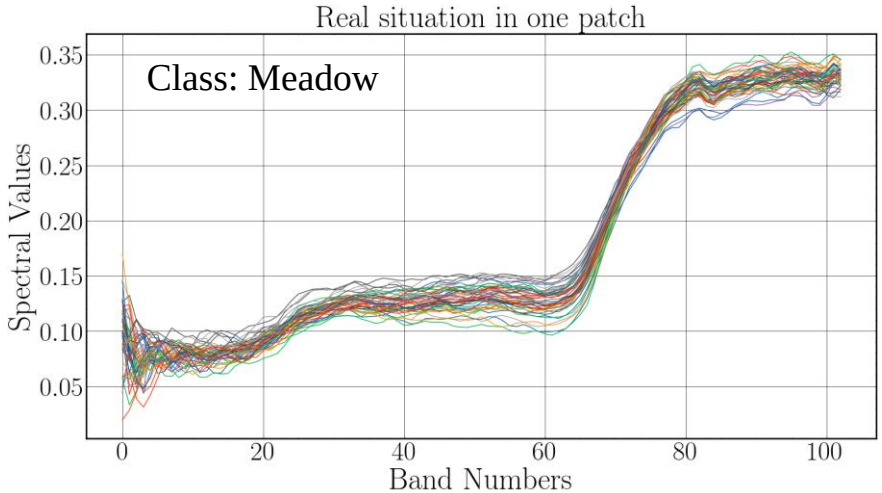
Observation (good examples in a patch)

- Real situation in experiments about cropped patch. (**good examples**)
- Pixels in this patch will have same spectral signatures distribution.



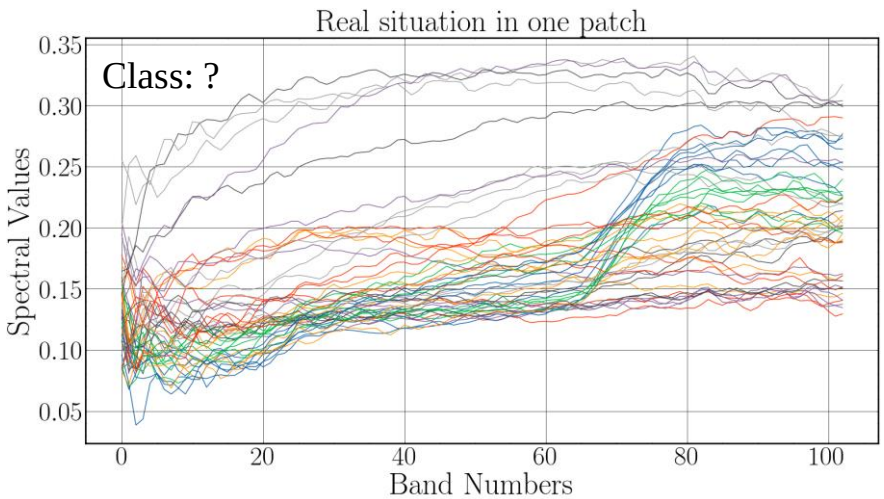


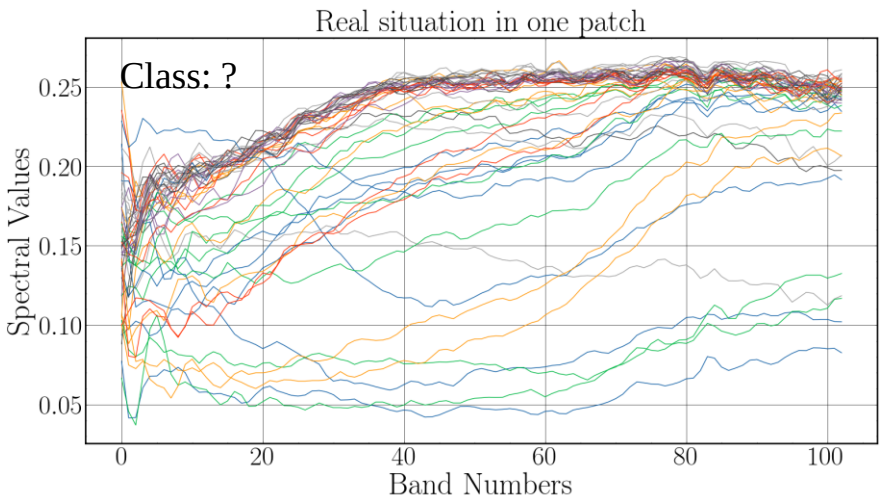


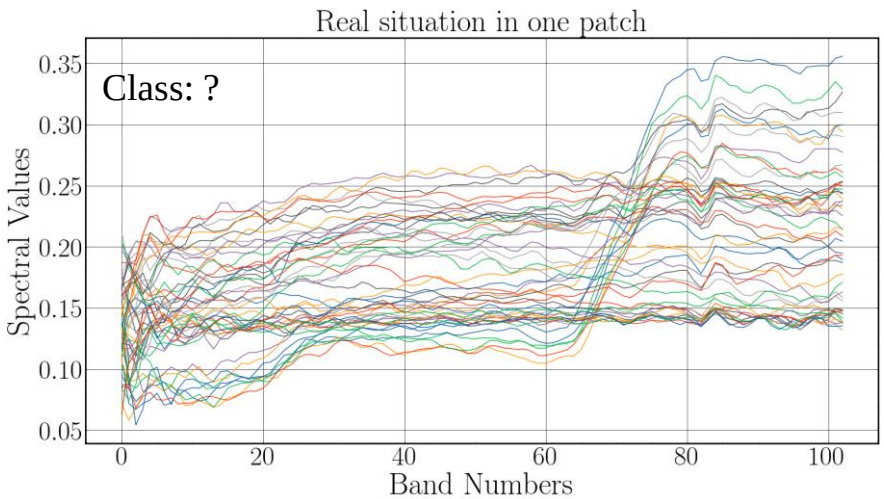


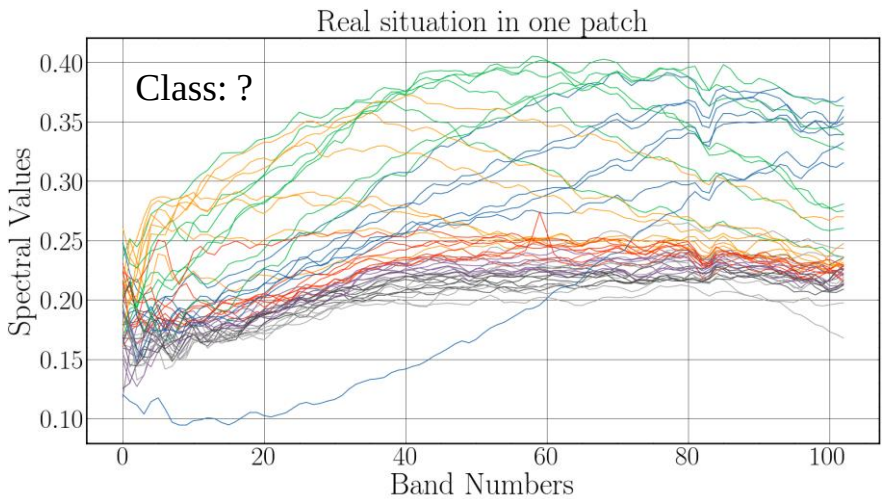
Observation (bad examples in a patch)

- Real situation in experiments about cropped patch. (**bad examples**)
- Pixels in this patch will have big variations on spectral signatures. (hard to decide a class)


Real situation in one patch


Real situation in one patch

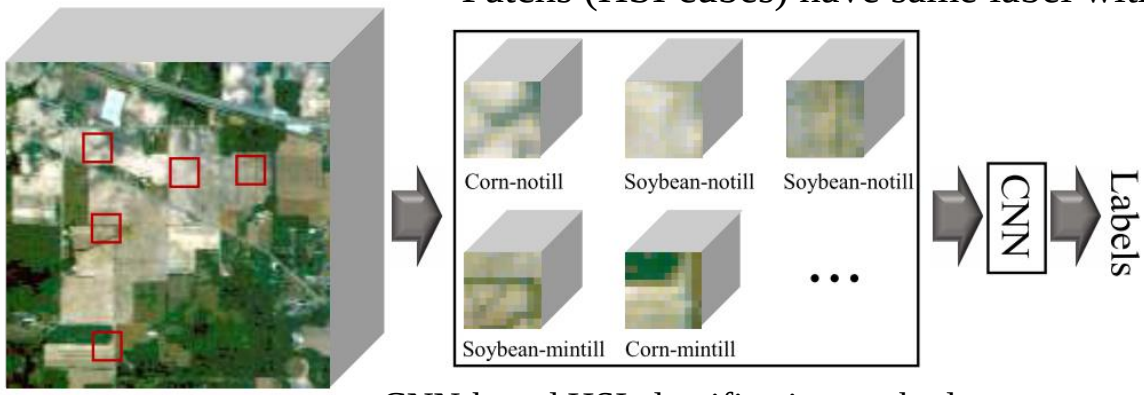

Real situation in one patch


Real situation in one patch

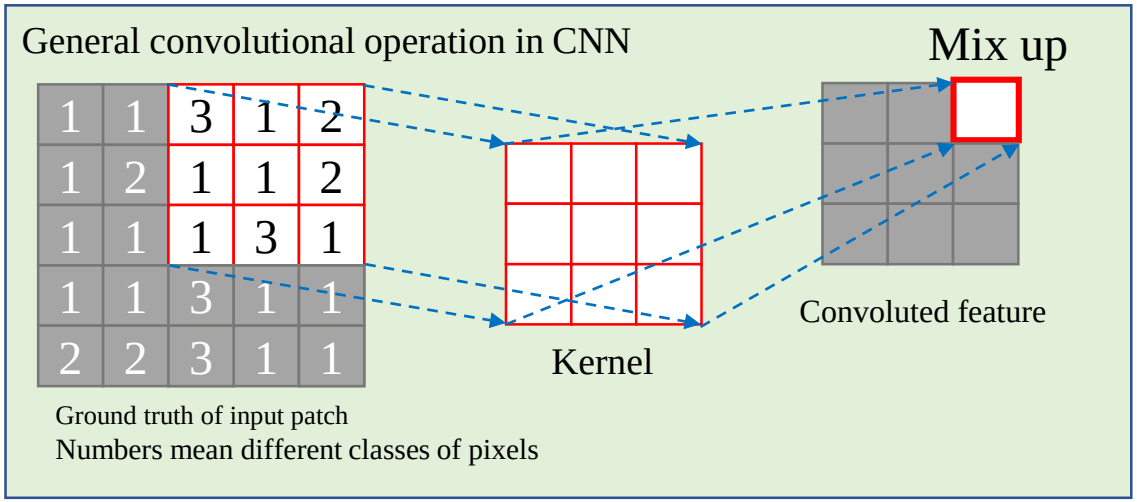
Motivation-1 (complex situation in a patch)

- **Problem-1**: directly use cropped patchs (HSI cubes) as input into network (such as CNN), interfering pixels will give bad influence with the extraction of features.

Patchs (HSI cubes) have same label with the central pixel



CNN-based HSI classification methods



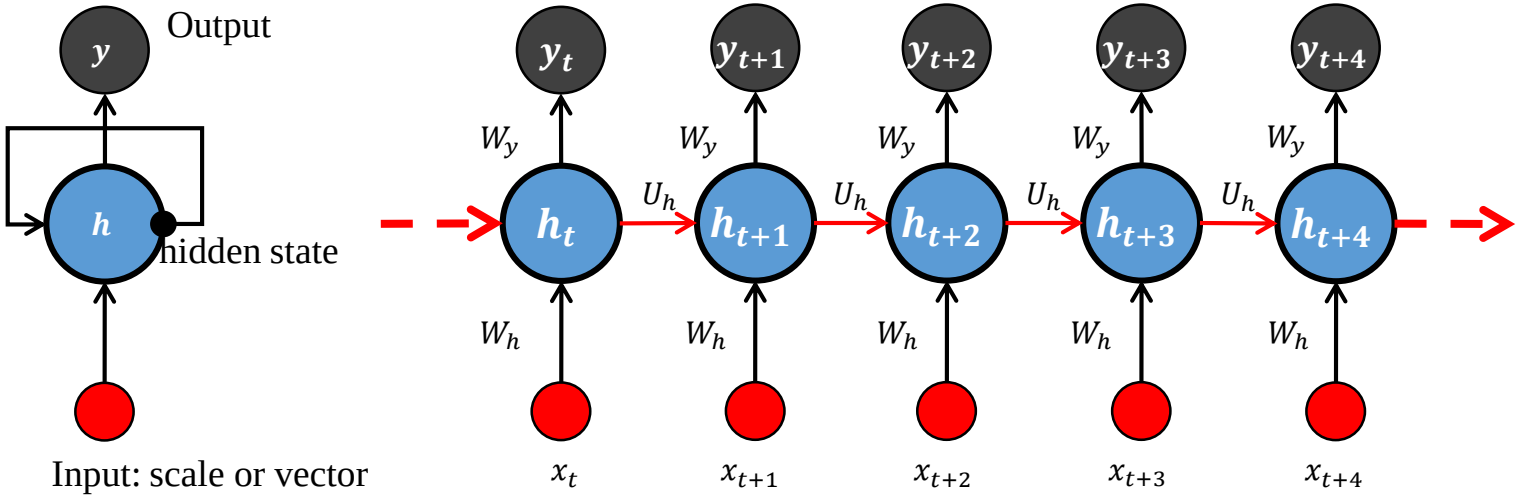
Idea: to find a method to deal with the local pixels one by one individually instead of convolving them together

↓

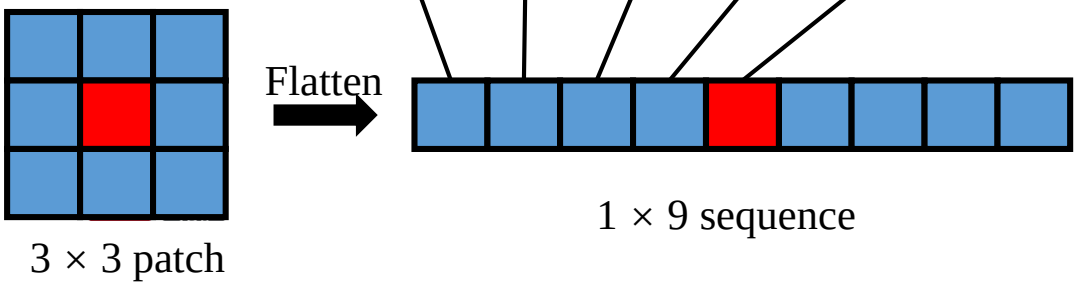
Recurrent neural network (RNN) is good at dealing with sequential data with time steps.

Motivation-1

- Recurrent Neural Network (RNN)
- Learning the relationship between inputs step by step .



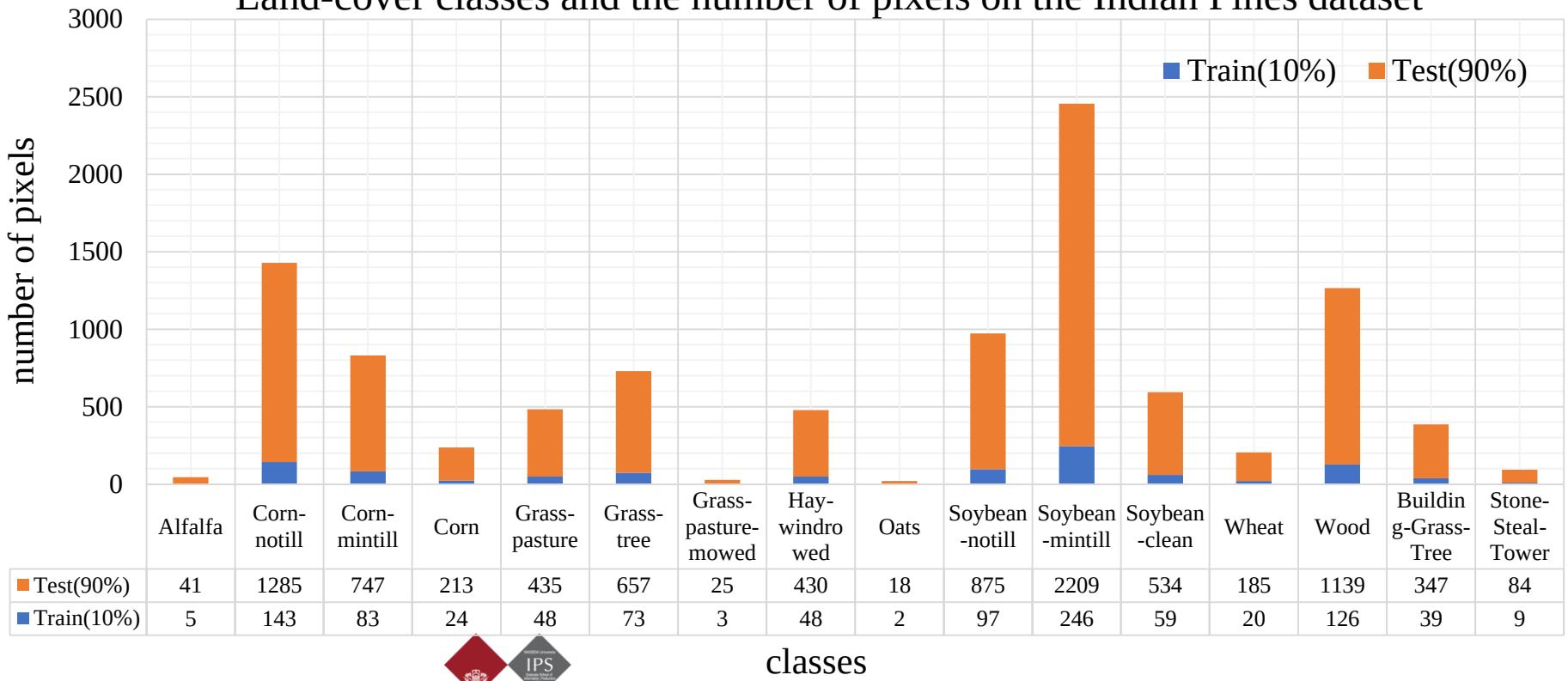
- RNN for a patch. Transfer it into a sequence.
- Each pixel or step is a feature vector



Motivation-2 (limited labeled pixels)

- **Problem-2**: In the HSI, unbalanced the number of pixels are severe
- In the real HSI, labeled pixels are taken in a limited set of conditions. But the target may exist in a variety of conditions, such as different location, scale, brightness etc.

Land-cover classes and the number of pixels on the Indian Pines dataset



	Alfalfa	Corn-notill	Corn-mintill	Corn	Grass-pasture	Grass-tree	Grass-pasture-mowed	Hay-windrowed	Oats	Soybean-notill	Soybean-mintill	Soybean-clean	Wheat	Wood	Building-Grass-Tree	Stone-Steal-Tower
Test(90%)	41	1285	747	213	435	657	25	430	18	875	2209	534	185	1139	347	84
Train(10%)	5	143	83	24	48	73	3	48	2	97	246	59	20	126	39	9

Motivation-2

- Classification results on existing works with each class.(Indian Pines dataset)

classes	Hu[1]	Mou[2]	Liu[13]	Li[9]	Leng[8]
Alfalfa	57.61%	78.31%	42.33%	57.12%	85.22%
Corn-notill	72.23%	74.90%	78.41%	77.98%	88.51%
Corn-mintill	54.23%	61.02%	65.63%	69.45%	81.71%
Corn	57.92%	46.92%	63.66%	74.11%	84.70%
Grass-pasture	89.51%	88.66%	85.70%	89.46%	96.00%
Grass-trees	93.77%	93.41%	96.11%	97.48%	98.23%
Grass-pasture-mowed	61.54%	71.88%	57.12%	86.41%	93.77%
Hay-windrowed	95.49%	96.31%	96.33%	96.33%	98.61%
Oats	42.91%	45.22%	10.52%	43.59%	75.72%
Soybean-notill	61.96%	67.81%	72.23%	76.88%	85.00%
Soybean-mintill	73.00%	73.15%	82.69%	81.70%	90.01%
Soybean-clean	72.38%	76.99%	71.23%	64.65%	81.76%
Wheat	98.11%	96.54%	99.26%	99.58%	97.91%
Woods	93.45%	92.51%	96.87%	95.26%	95.80%
Building-Grass-Tree	68.66%	66.53%	66.11%	75.74%	81.99%
Stone-Steel-Tower	87.32%	85.97%	91.66%	100%	96.96%
Overall accuracy	76.92%	78.77%	81.99%	83.51%	90.32%

red parts: accuracy lower than 80%
green parts: overall accuracy

Motivation-2

- **Idea**: using data augmentation methods for these limited labeled pixels.
- **Data augmentation**: To get more data. Popular data augmentation methods: flip, rotation, scale, crop, translation, reflection, mirror, noise, radiation, GAN.

Original data

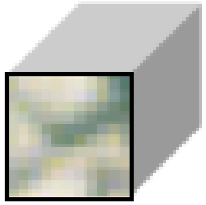
augment

Augmented data

Network

Class: dog

Inspired



Original patch(5*5)
label: Corn-notill

3	3	2	1	1
2	1	1	1	2
1	1	1	1	3
1	2	1	2	1
1	1	2	1	3

ground truth

augment

Augmented data

3	3	2	1	1
2	1	1	1	2
1	1	1	1	3
1	2	1	2	1
1	1	2	1	3

original

3	2	1	1	1
3	1	1	2	1
2	1	1	1	2
2	1	1	2	1
1	2	3	1	3

mirroring

1	2	3	1	3
1	1	1	2	1
2	1	1	1	2
3	1	1	2	1
3	2	1	1	1

rotation

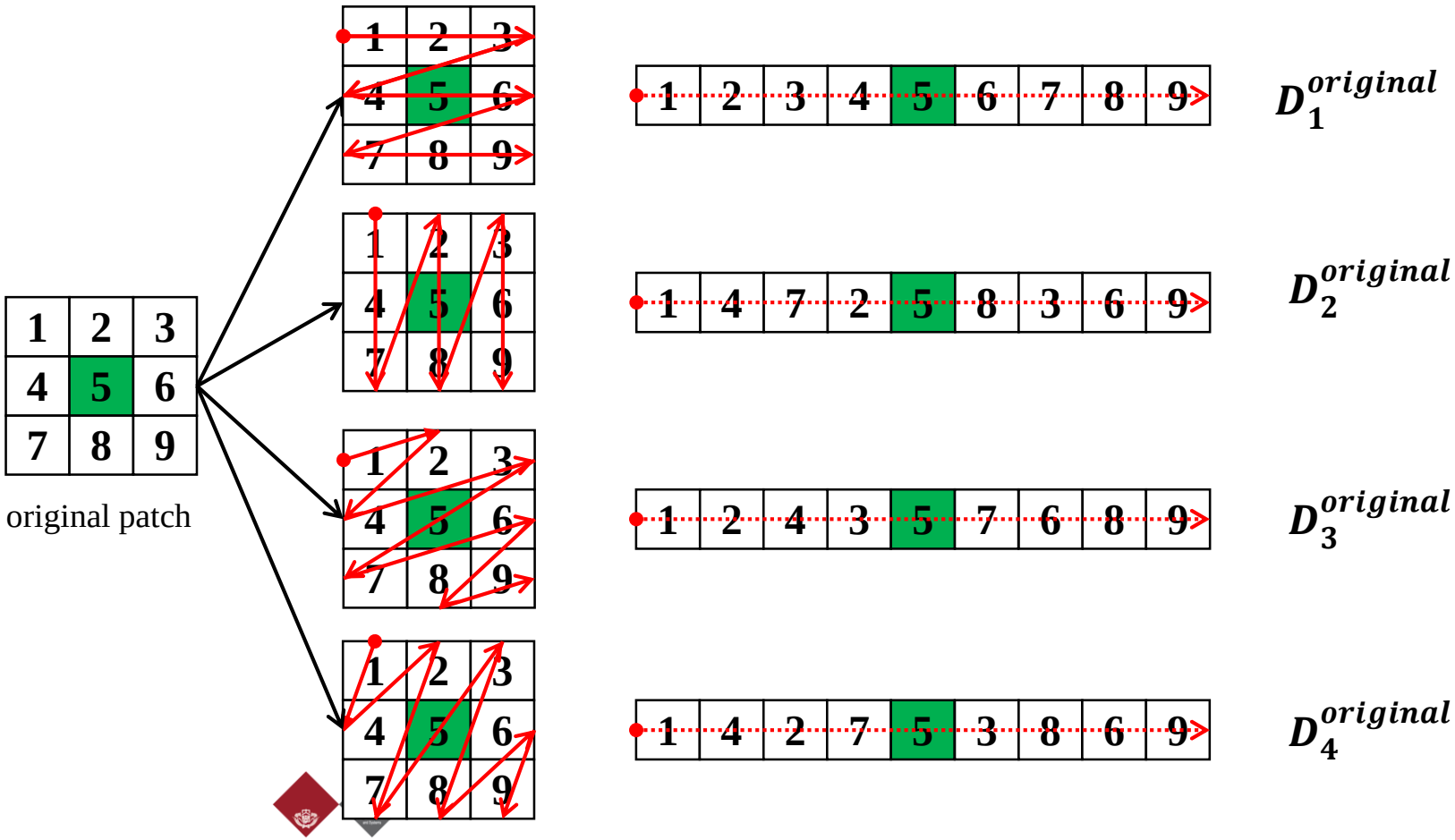
1	1	2	1	3
1	2	1	2	1
1	1	1	1	3
2	1	1	1	2
3	3	2	1	1

flipping

number means class.

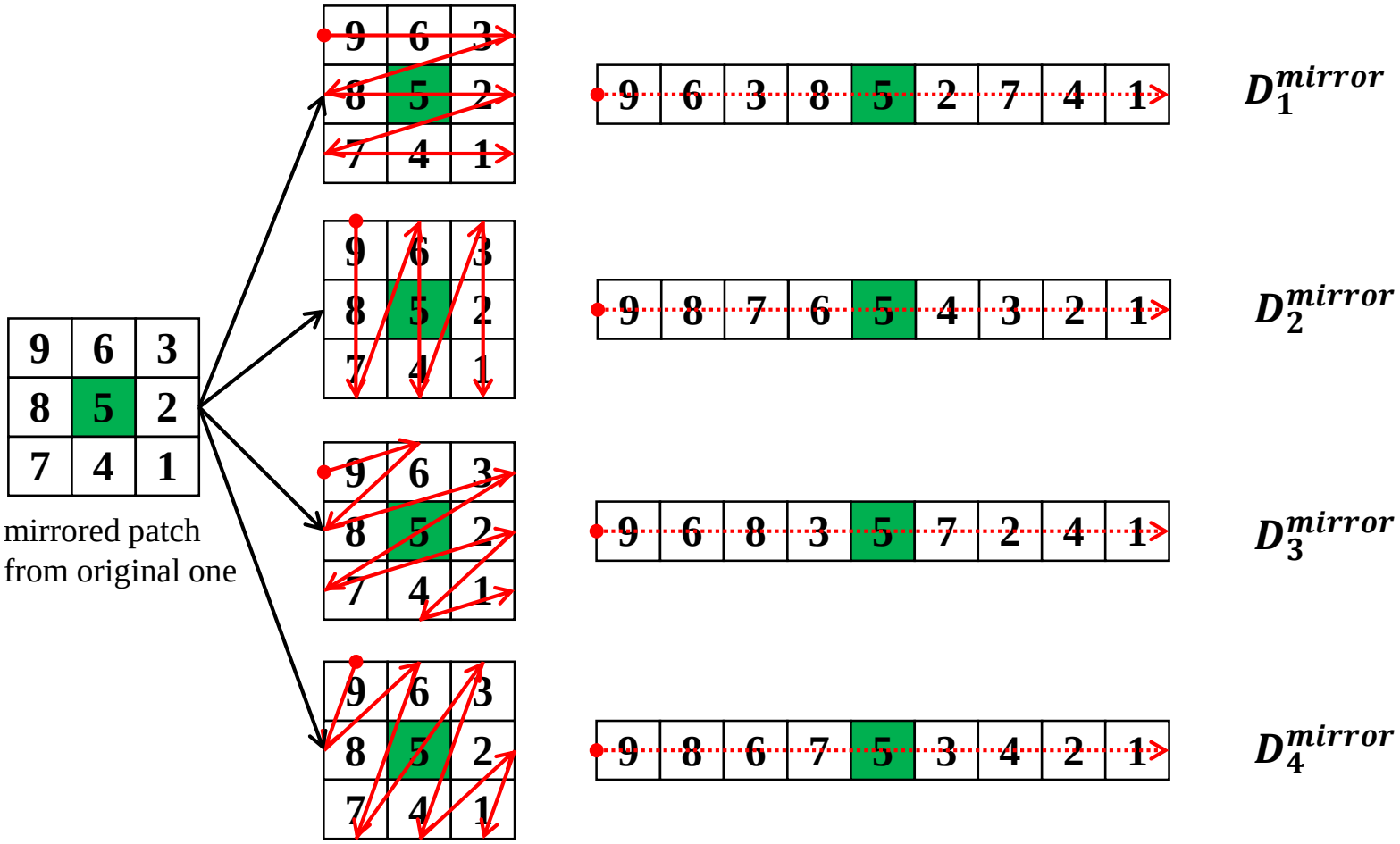
Proposed method (Multi-scanning strategy)

- In order to feed a local patch into RNN, we must transfer the patch into the sequence. But there's no fixed scanning direction exists.
- Take 3*3 patch (9 pixels) as an example. Start from first pixel.



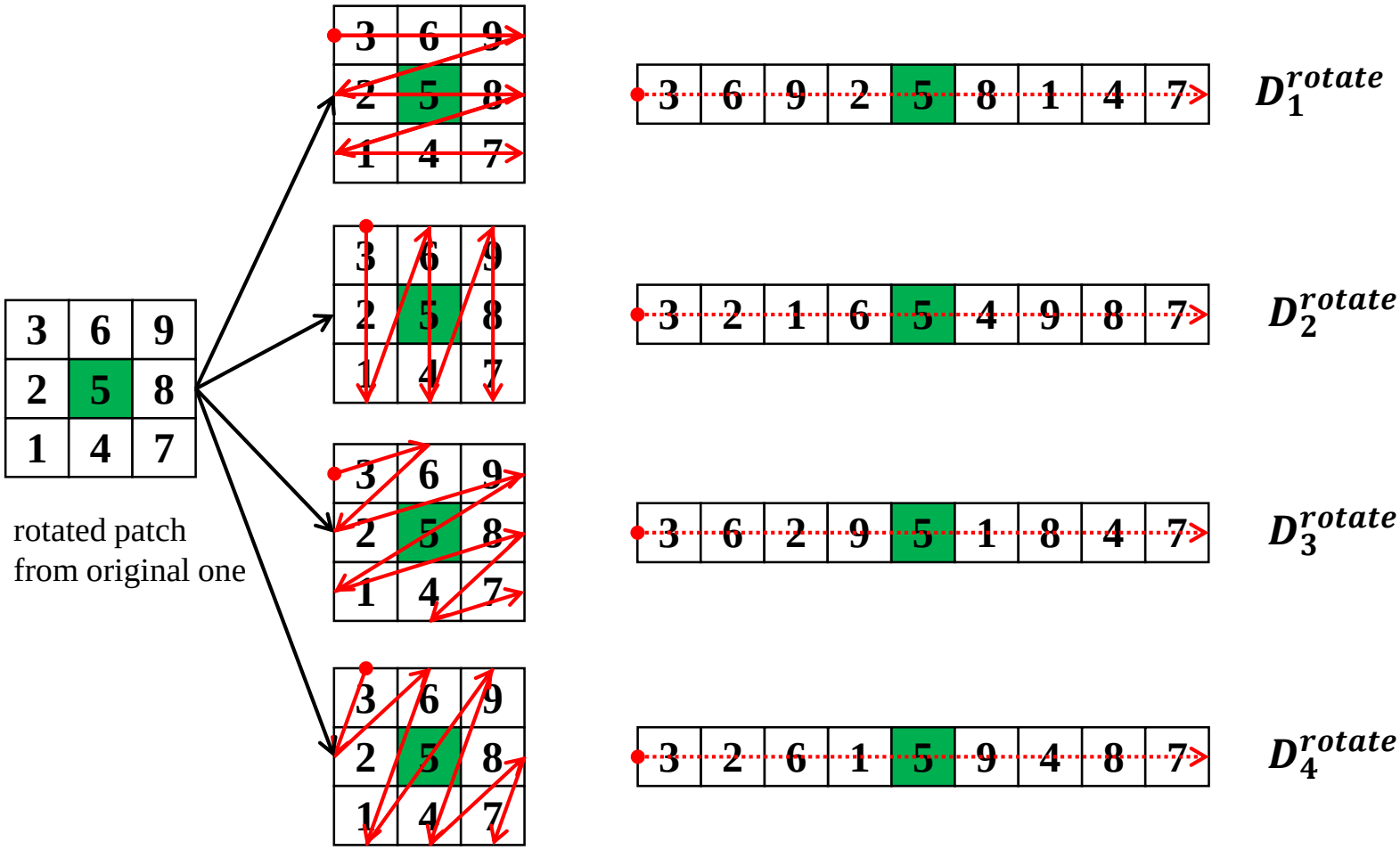
Proposed method (Multi-scanning strategy)

- After augmenting (**mirroring**) the patch.



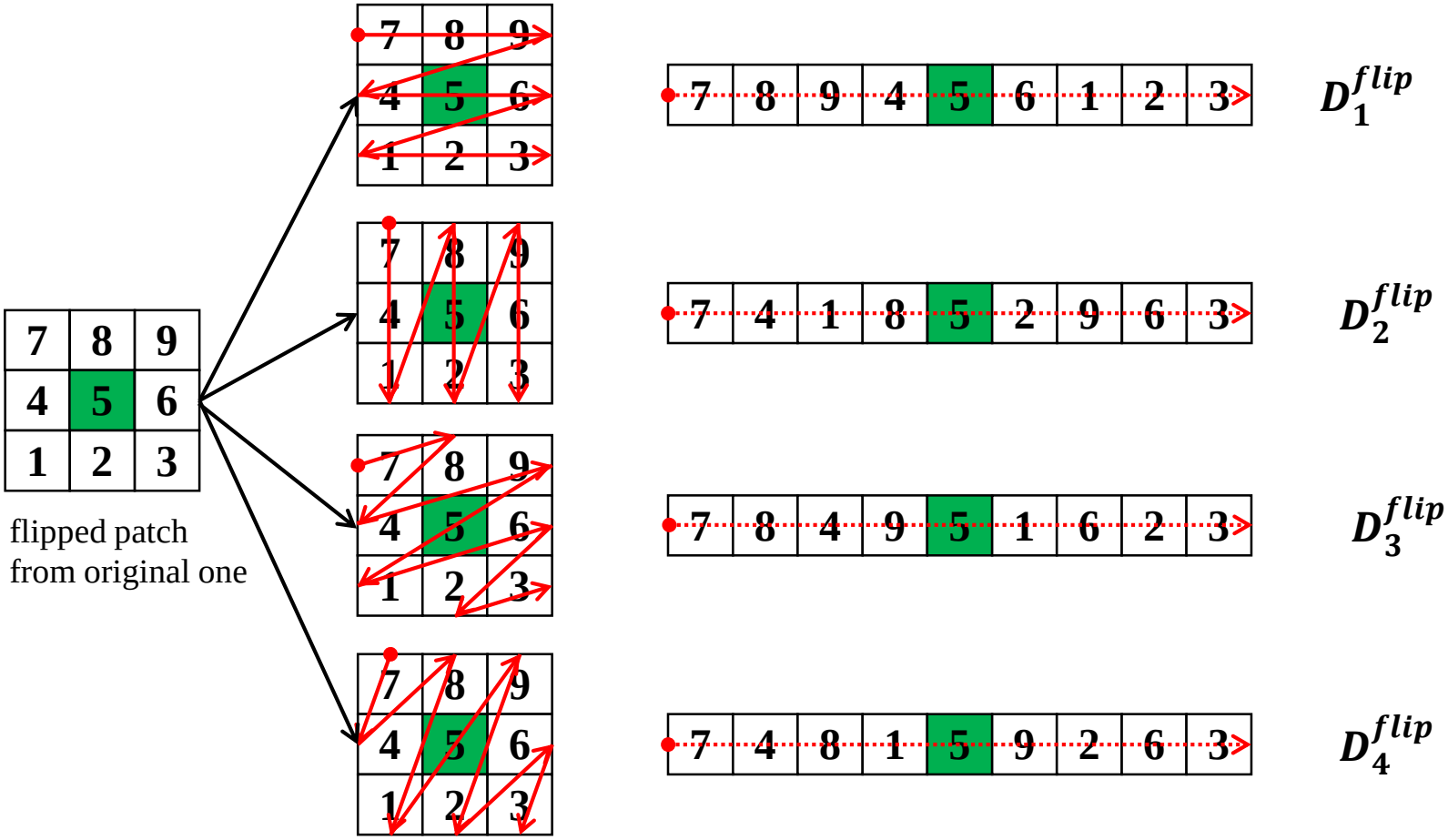
Proposed method (Multi-scanning strategy)

- After augmenting (**rotation**) the patch.



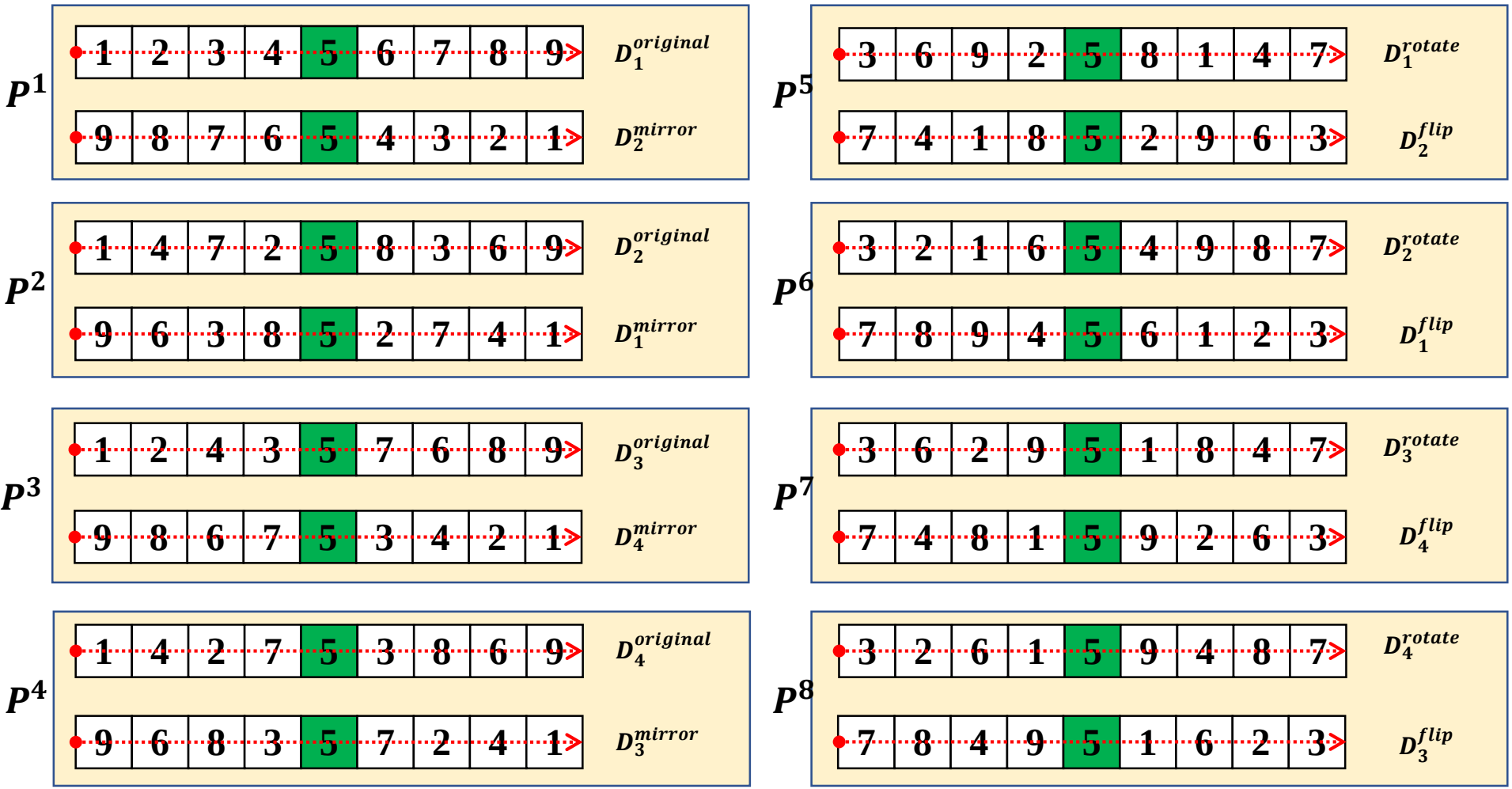
Proposed method (Multi-scanning strategy)

- After augmenting (**flipping**) the patch.



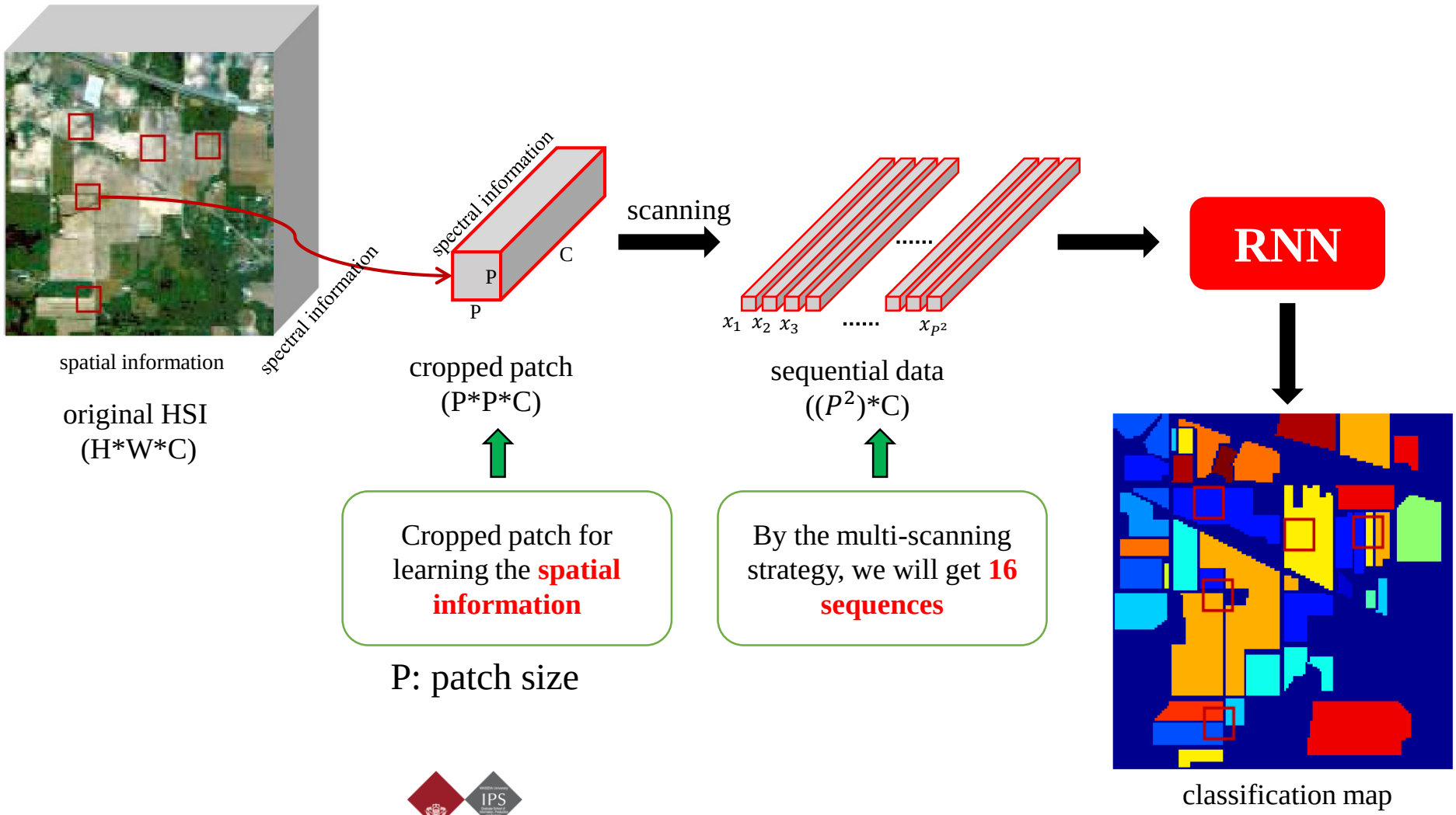
Proposed method (Multi-scanning strategy)

- In 16 directions, each of them is paired with another one.
- Forward and back-forward direction sequences are better to construct a bidirectional-RNN (Bi-RNN).
- Bidirectional-RNN has wider reception steps and consider the whole sequence.



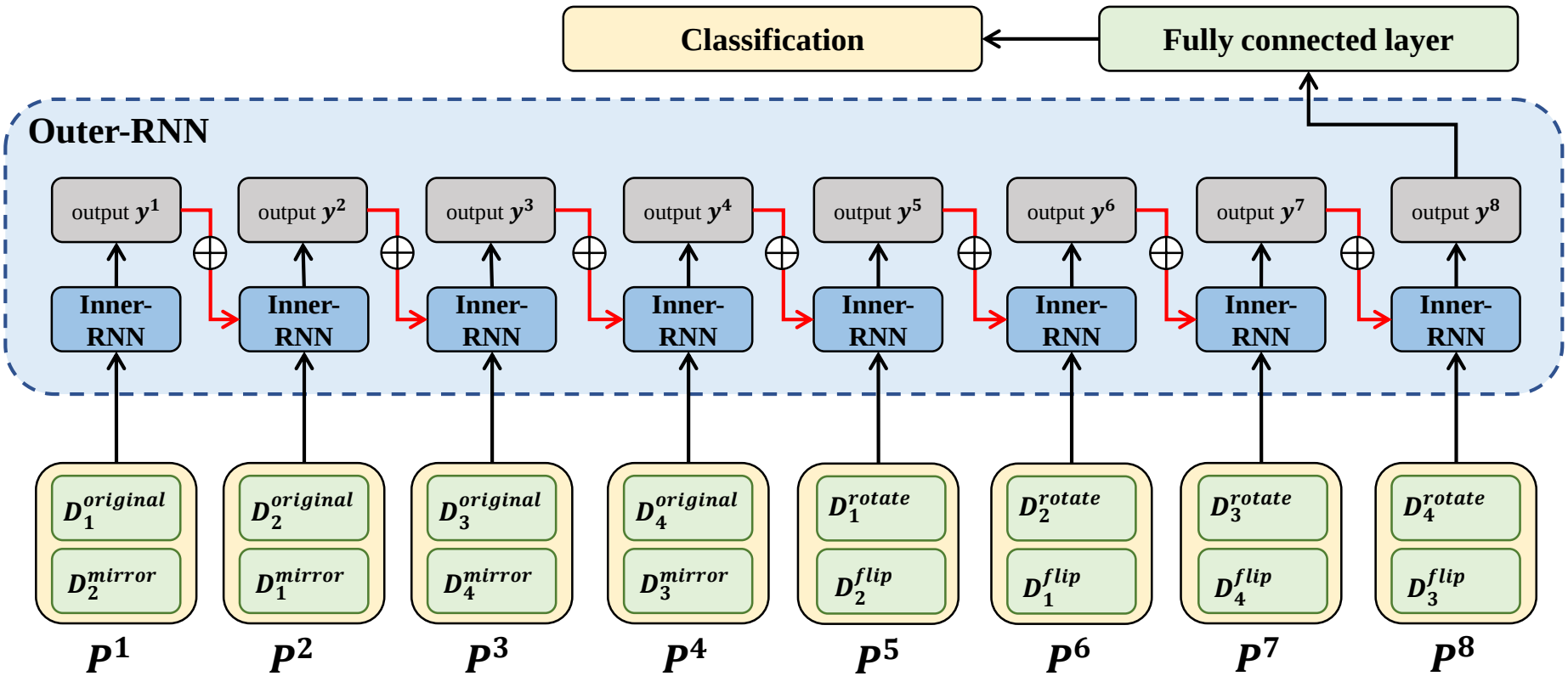
Proposed method (spectral-spatial structure)

- Spectral-spatial structure for HSI with RNN.



Proposed method (whole framework)

- 16 scanning directions are complementary for one patch.
- I learned their complementary features by designing an Outer-RNN.



Inner-RNN: Bidirectional RNN, has 25 or 49 steps (patch * patch). **Outer-RNN**: has 8 steps (from P^1 to P^8).
P^i: a pair of sequences
y^i: output feature from Bi-RNN.

Experiments with a single direction

$D_1^{original}$
91.503%

D_2^{flip}
89.844%

$D_3^{original}$
90.863%

$D_4^{original}$
90.527%

D_2^{mirror}
90.789%

D_1^{rotate}
90.072%

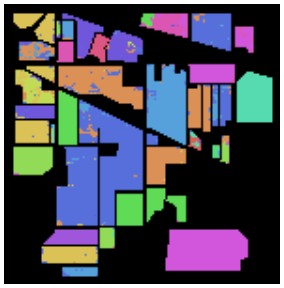
D_4^{mirror}
89.649%

D_3^{mirror}
90.744%

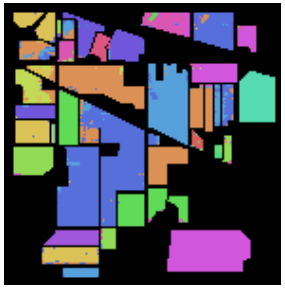
$D_2^{original}$
90.484%

D_2^{rotate}
90.939%

D_3^{rotate}
89.451%



D_4^{rotate}
91.087%



D_1^{mirror}
91.839%

D_1^{flip}
89.877%

D_4^{flip}
89.548%

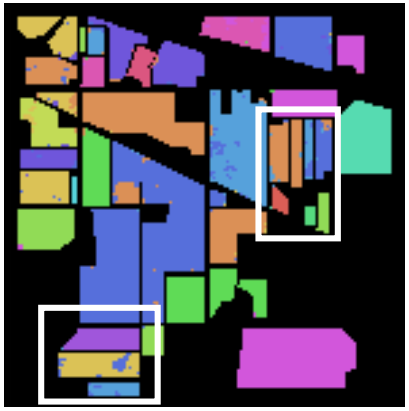
D_3^{flip}
90.570%

Experiments with several directions

- Classification results with 1, 4, 8 and 16 directions on **Indian Pines dataset**.

1-direction: 91.839%

4-directions: 92.531%



8-directions: 93.106%

16-directions: 93.989%

Multi-scanning:

1. Increasing the number of directions, the accuracy also increases.
2. It is more accurate and smoother in the flat aera
3. The boundary or outlines of regions are clearer.

Experiments with several directions

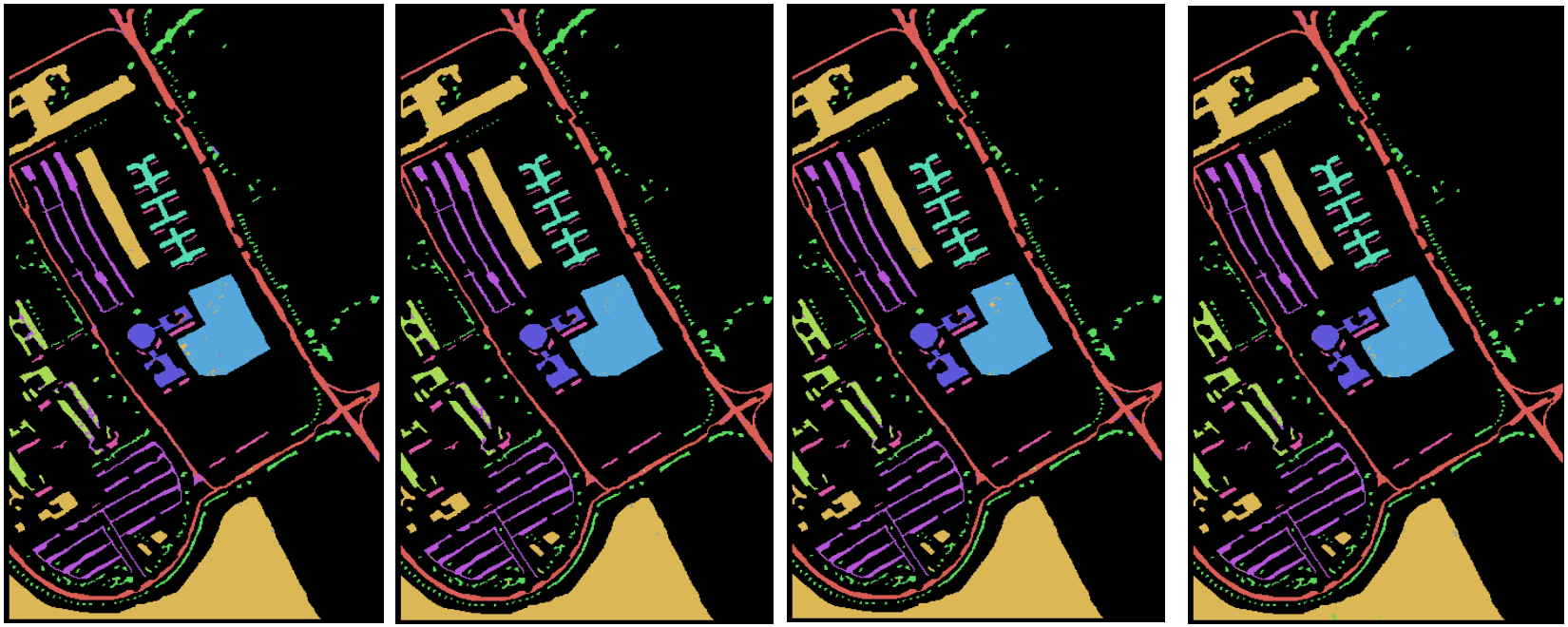
- Classification results on each class. (**Indian Pines**)

	1 direction	4 direction	8 direction	16 direction	3D-CNN Li[9]	3D-CNN Leng[8]
Alfalfa	58.15%	95.12%	95.79%	91.26%	57.12%	85.22%
Corn-notill	89.33%	92.23%	93.44%	96.11%	77.98%	88.51%
Corn-mintill	83.55%	86.99%	90.29%	91.33%	69.45%	81.71%
Corn	73.02%	84.42%	84.01%	90.87%	74.11%	84.70%
Grass-pasture	94.68%	94.21%	95.59%	94.18%	89.46%	96.00%
Grass-trees	98.66%	97.49%	98.77%	97.21%	97.48%	98.23%
Grass-pasture-mowed	91.30%	95.78%	84.21%	92.77%	86.41%	93.77%
Hay-windrowed	97.11%	99.81%	99.11%	99.54%	96.33%	98.61%
Oats	66.69%	97.21%	96.88%	100%	43.59%	75.72%
Soybean-notill	89.59%	87.10%	94.51%	88.71%	76.88%	85.00%
Soybean-mintill	90.55%	92.14%	94.12%	94.15%	81.70%	90.01%
Soybean-clean	85.29%	91.99%	91.83%	93.18%	64.65%	81.76%
Wheat	99.59%	98.89%	99.17%	100%	99.58%	97.91%
Woods	96.08%	96.51%	97.34%	97.85%	95.26%	95.80%
Building-Grass-Tree	79.35%	81.29%	80.89%	79.89%	75.74%	81.99%
Stone-Steel-Tower	98.28%	98.20%	98.71%	98.59%	100%	96.96%
Overall accuracy	91.84%	92.53%	93.11%	93.99%	83.51%	90.32%

yellow parts: obvious improvements by my method. Green parts: overall accuracy.
red parts: lower accuracy relatively in 3D-CNN.

Experiments with several directions

- Classification results with 1, 4, 8 and 16 directions on **PaviaU. dataset**.



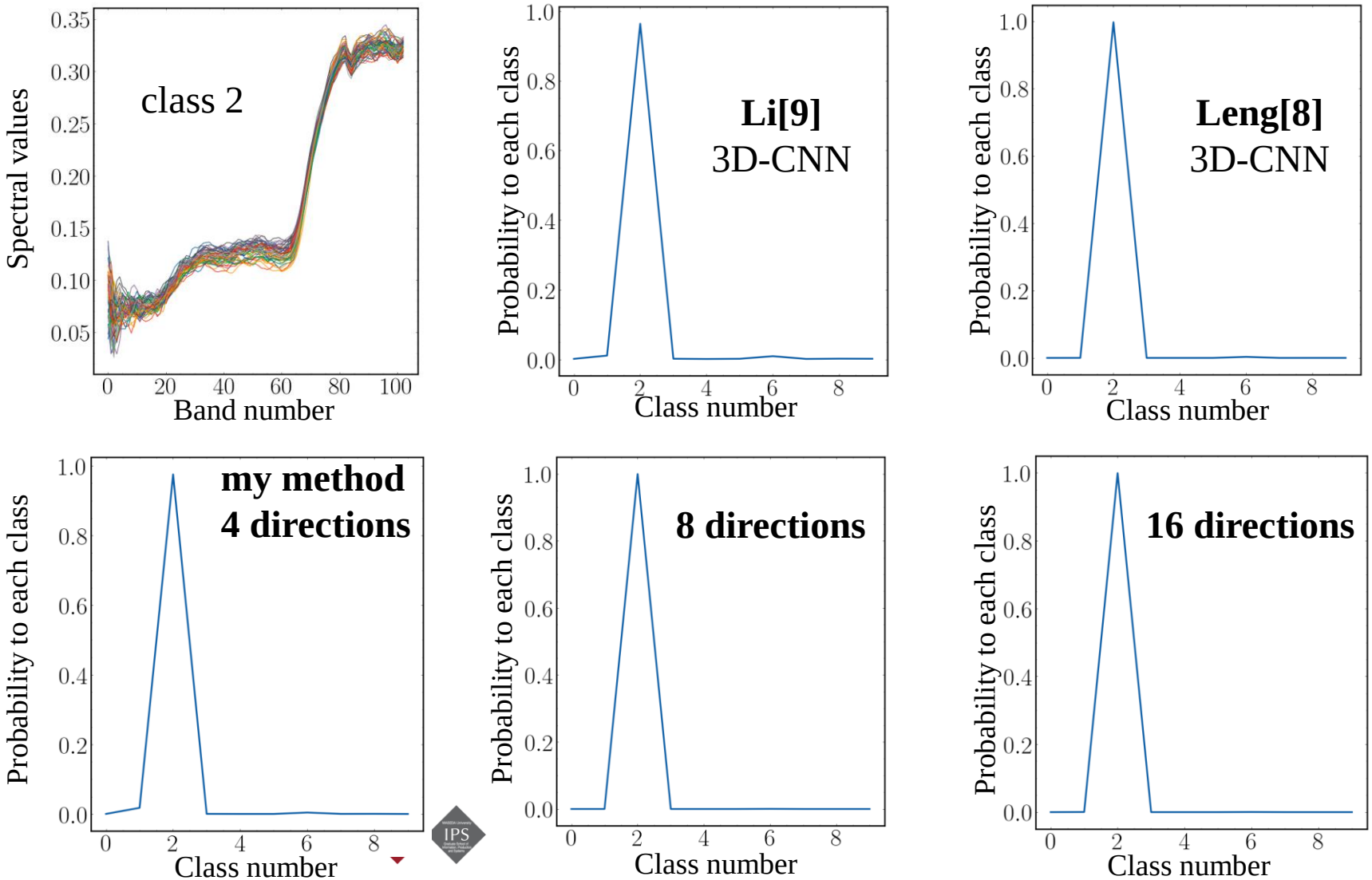
1-direction: 95.315% 4-directions: 96.380% 8-directions: 96.583% 16-directions: 96.861%

Multi-scanning:

1. Increasing the number of directions, the accuracy also increases.
2. It is more accurate and smoother in the flat aera
3. The boundary or outlines of regions are clearer.

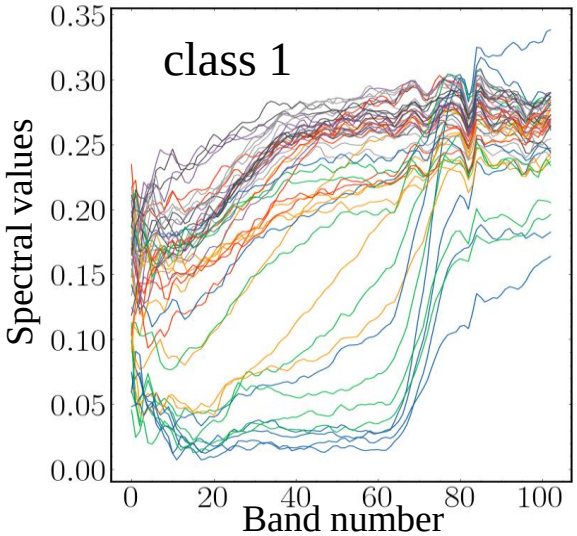
Comparison

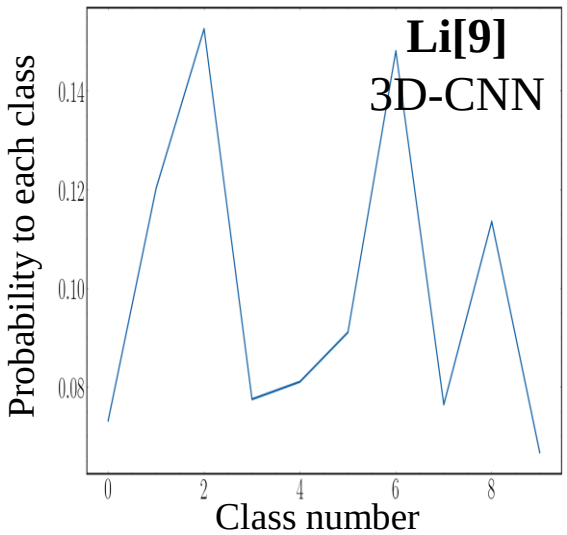
- Classification probability for **good examples**.

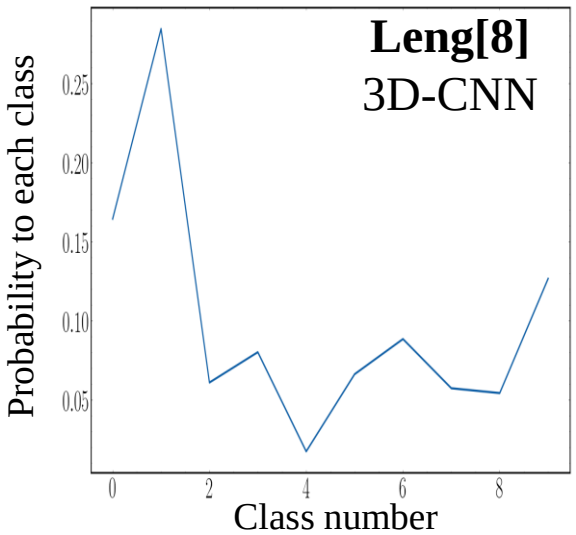


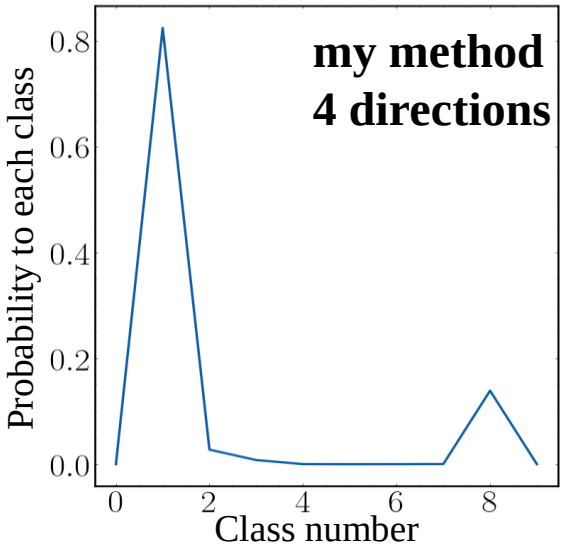
Comparison

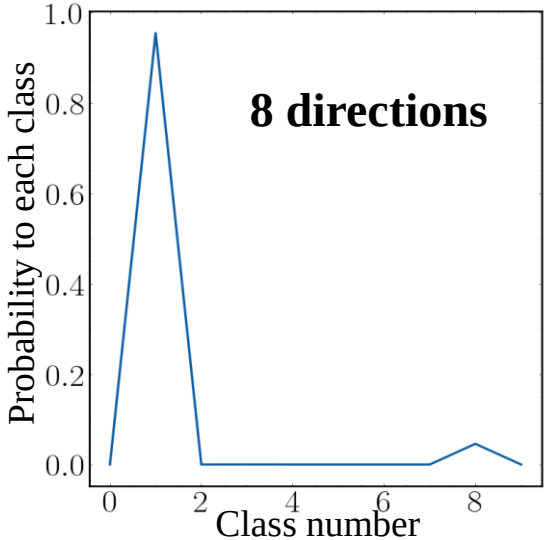
- Classification probabilities for **bad examples**.

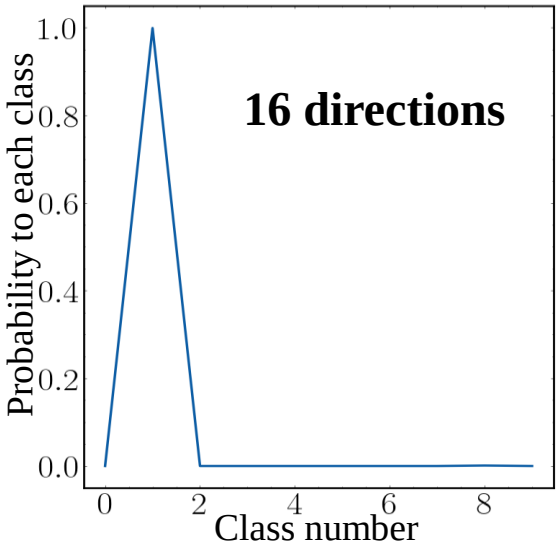






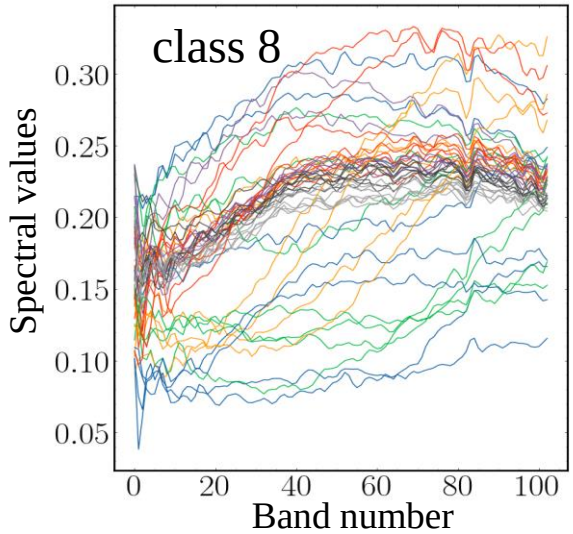


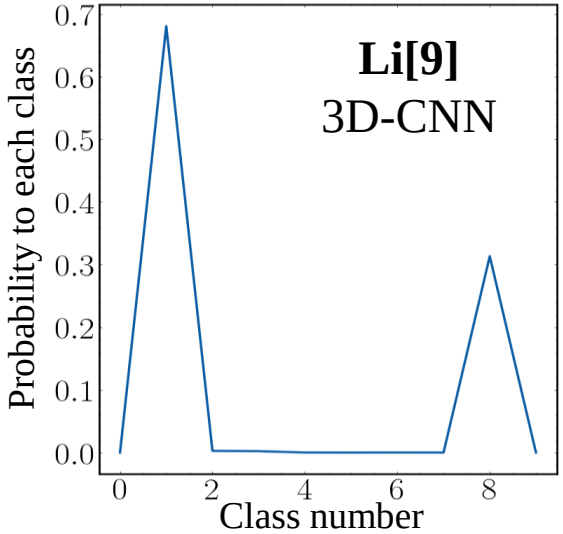


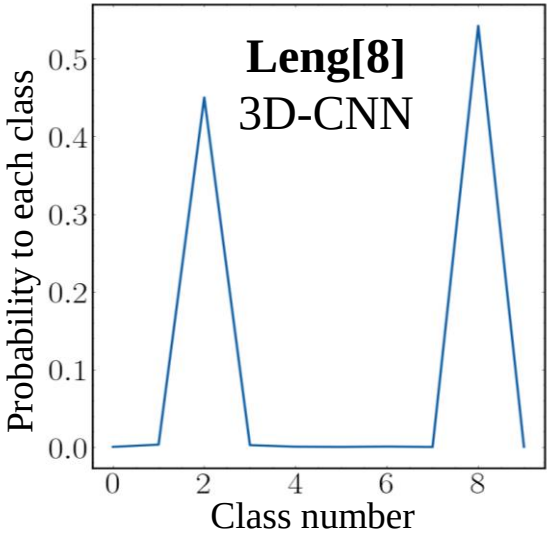


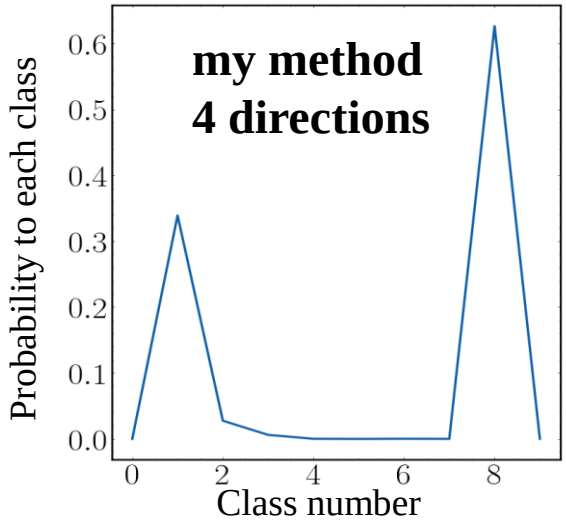
Comparison

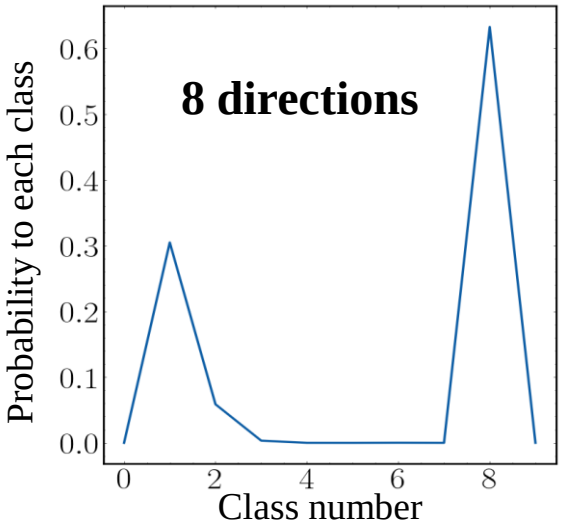
- Classification probabilities for **bad examples**.

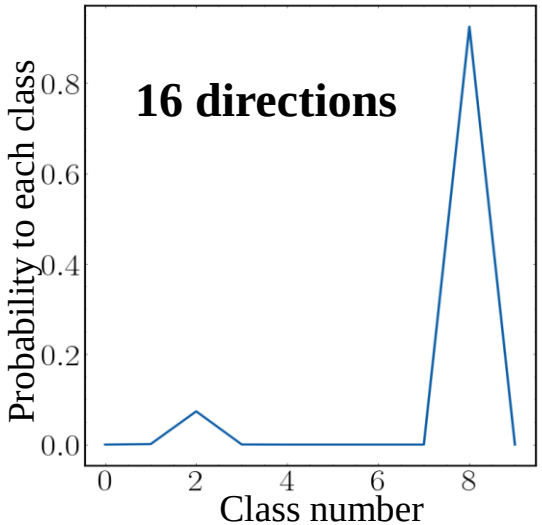












Conclusion

① My method can fully explore the spatial and spectral information together to discriminate each pixel.

② The results by my method become more accurate and smoother in the flat region.

③ The boundary or outlines of regions are more clear.

④ We hope it is a viable alternative for using CNN-based method in this task.

Reference

[1]W. Hu, Y. Huang, L. Wei, F. Zhang, H. Li, "**Deep convolutional neural networks for hyperspectral image classification**", *J. Sensors*, vol. 2015, pp. 258619-1-258619-12, Jan. 2015.

[2]L. Mou, P. Ghamisi and X. X. Zhu, "**Deep Recurrent Neural Networks for Hyperspectral Image Classification**," in IEEE Transactions on Geoscience and Remote Sensing, vol. 55, no. 7, pp. 3639-3655, July 2017. (1-D RNN)

[3]Hang, Renlong et al. "**Cascaded Recurrent Neural Networks for Hyperspectral Image Classification**." IEEE Transactions on Geoscience and Remote Sensing 57.8 (2019): 5384–5394.

[4]H. Sun, X. Zheng, X. Lu and S. Wu, "**Spectral–Spatial Attention Network for Hyperspectral Image Classification**," in IEEE Transactions on Geoscience and Remote Sensing, vol. 58, no. 5, pp. 3232-3245, May 2020

[5]F. Zhou, R. Hang, Q. Liu and X. Yuan, **Hyperspectral image classification using spectral-spatial LSTMs**, Neurocomputing, Volume 328, 2019, Pages 39-47.

[6]Y. Xu, L. Zhang, B. Du and F. Zhang, "**Spectral–Spatial Unified Networks for Hyperspectral Image Classification**," in IEEE Transactions on Geoscience and Remote Sensing, vol. 56, no. 10, pp. 5893-5909, Oct. 2019

[7]Li. Jiaojiao, Zhao. Xi, Li. Yunsong, et al. **Classification of Hyperspectral Imagery Using a New Fully Convolutional Neural Network**. IEEE Geoscience and Remote Sensing Letters, 2018, pp. 1-5

[8]Leng, Jiabing, Tao Li, Gang Bai, Qiankun Dong and Han Dong. **Cube:CNN-SVM: A Novel Hyperspectral Image Classification Method[C]**, 2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI), 2016: 1027-1034.

[9]Li Y, Zhang and H, Shen Q. **Spectral–Spatial Classification of Hyperspectral Imagery with 3D Convolutional Neural Network**, Remote Sensing. 2018, 9(1):67.

[10]Ahmed, Ayman M. et al. "**Hyperspectral bands prediction based on inter-band spectral correlation structure**." *Electronic Imaging* (2013).

[11]X. Jia, B. Kuo and M. M. Crawford, "**Feature Mining for Hyperspectral Image Classification**," in Proceedings of the IEEE, vol. 101, no. 3, pp. 676-697, March 2013

[12]Lv, Wenjing & Wang, Xiaofei. (2020). **Overview of Hyperspectral Image Classification**. Journal of Sensors. 2020

[13] Bing Liu, Xuchu Yu, Pengqiang Zhang, Xiong Tan, Anzhu Yu & Zhixiang Xue (2018) **A semi-supervised convolutional neural network for hyperspectral image classification**, Remote Sensing Letters, 8:9, 839-848