# Robust Deep Metric Learning for Remote Sensing Images with Label Noise

**Jian Kang**[1], Ruben Fernandez Beltran[2], Puhong Duan[3], Xudong Kang[3], Antonio Plaza[4]

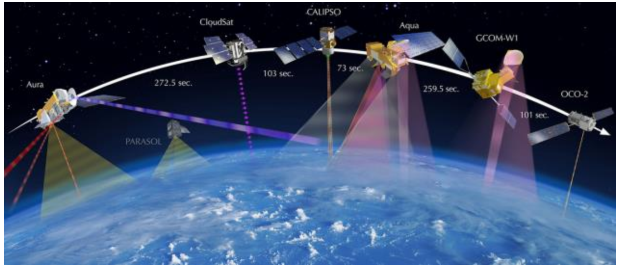
1. School of Electronic and Information Engineering, Soochow University, Suzhou 215006, China
2. Institute of New Imaging Technologies, University Jaume I, E-12071 Castellon, Spain
3. College of Electrical and Information Engineering, Hunan University, 410082 Changsha, China
4. Hyperspectral Computing Laboratory, University of Extremadura, E-10003 Caceres, Spain

# Outline

- Introduction
- Motivation
- Background Knowledge
- Robust Normalized Softmax Loss (RNSL)
- Experiments
- Conclusion

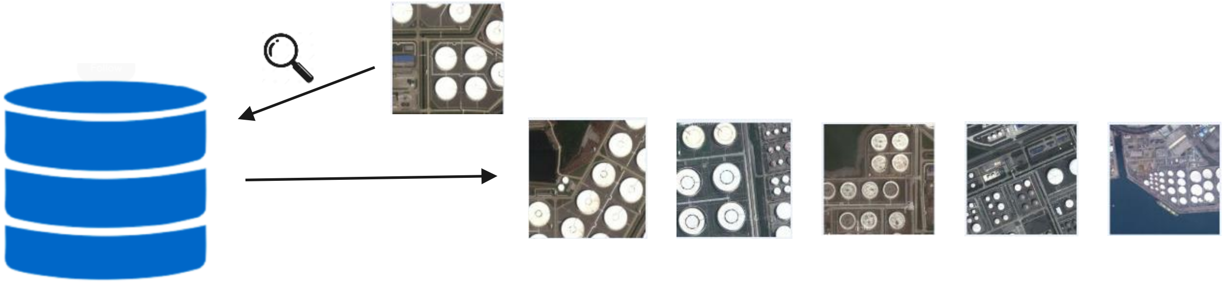
# Introduction

- Remote Sensing (RS) technology development meets big Earth Observation (EO) data



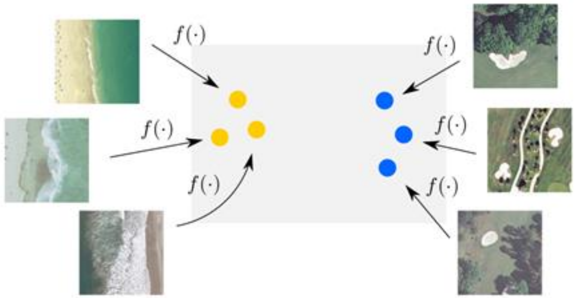

Credit: Wikipedia

- Retrieving interested contexts from big EO data is a basic task in RS



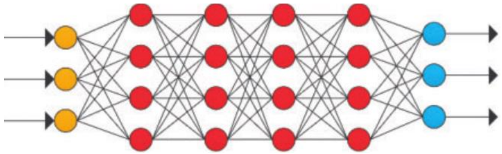
## Introduction

- Characterizing the contexts of RS images with low-dimensional features is the key for achieving image retrieval




- Deep learning has been a workhorse for learning those features

$$f(\cdot) \longrightarrow$$

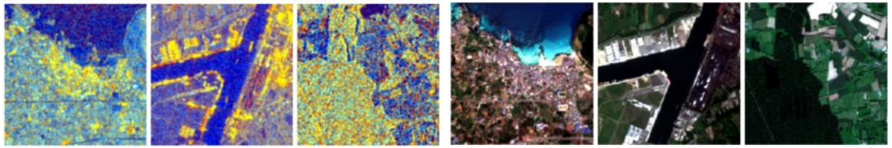


# Introduction

- Labeling RS scene datasets for developing advanced deep metric learning algorithms
  - Human experts: AID[1] , NWPU-RESISC45[2]
  - Crowd-sourcing data: SEN12MS[3]

AID:

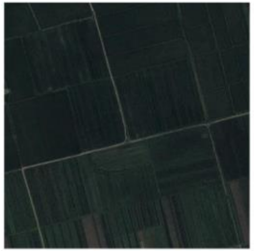
SEN12MS:



[1] Xia, Gui-Song, et al. "AID: A benchmark data set for performance evaluation of aerial scene classification." *IEEE Transactions on Geoscience and Remote Sensing* 55.7 (2017): 3965-3981.
[2] Cheng, Gong, Junwei Han, and Xiaoqiang Lu. "Remote sensing image scene classification: Benchmark and state of the art." *Proceedings of the IEEE* 105.10 (2017): 1865-1883.
[3] Schmitt, Michael, et al. "SEN12MS--A Curated Dataset of Georeferenced Multi-Spectral Sentinel-1/2 Imagery for Deep Learning and Data Fusion." *arXiv preprint arXiv:1906.07789* (2019).

# Introduction

- Labeling based on crowd-sourcing data may contain noise
  - geo-location/registration errors
  - land-cover changes
  - low-quality Volunteered Geographic Information (VGI)

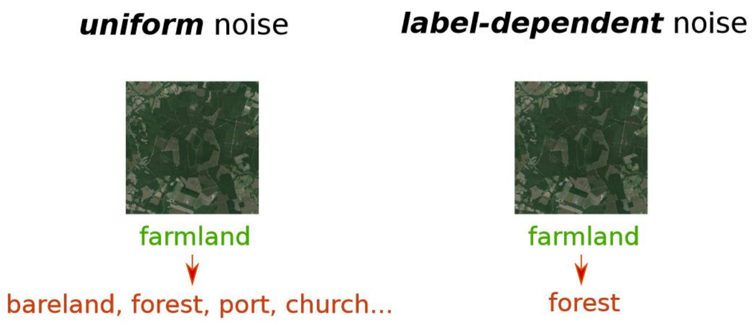


Bare land → Desert    Church → Downtown    Farmland → Meadow

# Motivation and Background Knowledge

- **Extracting deep embeddings of RS images in a robust manner**

- Noise type:
  - Uniform noise:
    a true label is flipped into other labels with equal probability
  - Label-dependent noise:
    a true label is more likely to be mistakenly labeled with a particular class

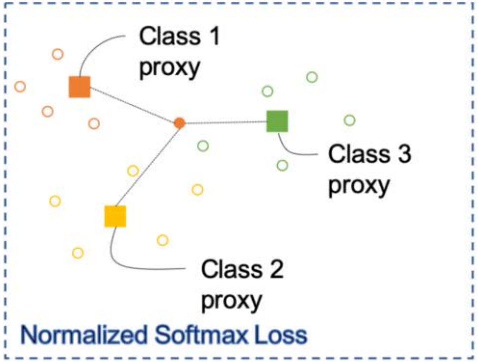


# Background Knowledge

- Normalized Softmax Loss (NLS) [5]

$$L_{\text{NSL}} = -\frac{1}{N}\sum_i \sum_c y_i^c \log\left(\frac{\exp(\mathbf{w}_c^T f(\mathbf{x}_i)/\sigma)}{\sum_k \exp(\mathbf{w}_k^T f(\mathbf{x}_i)/\sigma)}\right)$$

- Objectives of NLS:
  - Learning the normalized center embedding for each class
  - Pulling the features of each class to their associated center embeddings in latent space




Credit: [5]

[5] Zhai, Andrew, and Hao-Yu Wu. "Classification is a strong baseline for deep metric learning." *arXiv preprint arXiv:1811.12649* (2018).

# Robust Normalized Softmax Loss (RNSL)

- Gradients of $L_{\text{NSL}}$ with respect to $\mathbf{w}_c$ :

$$\frac{\partial L_{\text{NSL}}}{\partial \mathbf{w}_c} = -\frac{1}{N}\sum_i \sum_c \frac{y_i^c}{p_i^c}\frac{\partial p_i^c}{\partial \mathbf{w}_c}$$

Classification probability $p_i^c = \frac{\exp(\mathbf{w}_c^T f(\mathbf{x}_i)/\sigma)}{\sum_k \exp(\mathbf{w}_k^T f(\mathbf{x}_i)/\sigma)}$

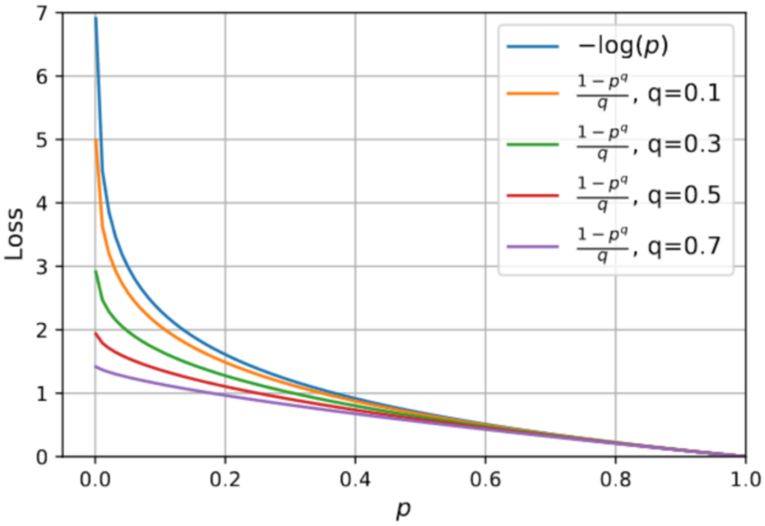
- **Hard samples are given more attention than the ones which are easily classified**
- **When label noise exists, $L_{\text{NSL}}$ can lead the trained models overfitting to noisy samples**

# Robust Normalized Softmax Loss (RNSL)

- RNSL exploits negative Box-Cox transformation with the form [6,7]:

$$L_{\text{RNSL}} = \frac{1}{N}\sum_i \sum_c y_i^c \frac{(1-(p_i^c)^q)}{q}, \quad q \in (0,1)$$

- With different values of $q$, the loss changes with respect to the classification probability $p$





[6] Zhang, Zhilu, and Mert R. Sabuncu. "Generalized cross entropy loss for training deep neural networks with noisy labels." arXiv preprint arXiv:1805.07836 (2018).
[7] Kang, Jian, et al. "Robust Normalized Softmax Loss for Deep Metric Learning-Based Characterization of Remote Sensing Images With Label Noise." IEEE Transactions on Geoscience and Remote Sensing (2020).

# Robust Normalized Softmax Loss (RNSL)

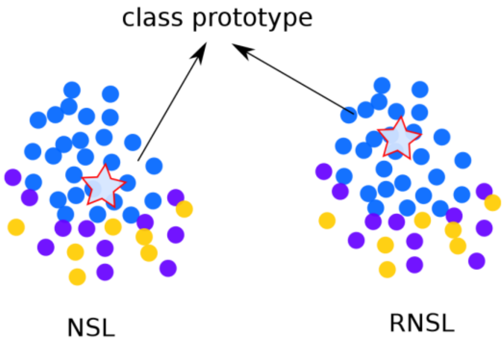
- Gradients of $L_{\mathrm{RNSL}}$ with respect to $\mathbf{w}_c$ :

$$\frac{\partial L_{\mathrm{RNSL}}}{\partial \mathbf{w}_c} = \frac{1}{N}\sum_i \sum_c y_i^c (p_i^c)^q \boxed{\left(-\frac{1}{p_i^c}\frac{\partial p_i^c}{\partial \mathbf{w}_c}\right)}$$

$$\frac{\partial L_{\mathrm{NSL}}}{\partial \mathbf{w}_c} = -\frac{1}{N}\sum_i \sum_c \frac{y_i^c}{p_i^c}\frac{\partial p_i^c}{\partial \mathbf{w}_c}$$

- Downweighting effects of $(p_i^c)^q$ , which can reduce the influence of noisy samples on learning the parameters

$$\lim_{q \to 0} L_{\mathrm{RNSL}} = L_{\mathrm{NSL}}$$



# Robust Normalized Softmax Loss (RNSL)

- To further improve the robustness of RNSL when heave noisy labels exist, a truncated version of RNSL is introduced:

$$\mathcal{L}_{\text{t-RNSL}} = \frac{1}{N}\sum_i \sum_c y_i^c \begin{cases} \frac{1-k^q}{q}, & \text{if } p_i^c \leq k \\ \frac{1-(p_i^c)^q}{q}, & \text{if } p_i^c > k \end{cases}$$

- The training strategy:
  - Within the first T epochs, the models are trained with RNSL
  - After T epochs, the loss function is switched to the truncated version

# Experimental Setup

| | |
|---|---|
| **Dataset** | AID; NWPU-RESISC45 |
| **Noise Type** | Uniform; Label-dependent |
| **Noise Level** | 0.1; 0.3; 0.5; 0.7 |
| **Data Splitting** | Train:0.7, Val:0.1, Test:0.2 |
| **Tasks** | KNN classification; Clustering; Image retrieval |
| **Metrics** | OA; NMI; ACC; MAP |

# Experimental Setup

- Compared methods:
  - D-CNN [8]
  - Triplet [9]
  - SNCA [10]
  - NSL [5]
  - ArcFace [11]


[8] Cheng, Gong, et al. "When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative CNNs." IEEE transactions on geoscience and remote sensing 56.5 (2018): 2811-2821.

[9] Schroff, Florian, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

[10] Wu, Zhirong, Alexei A. Efros, and Stella X. Yu. "Improving generalization via scalable neighborhood component analysis." Proceedings of the European Conference on Computer Vision (ECCV). 2018.

[5] Zhai, Andrew, and Hao-Yu Wu. "Classification is a strong baseline for deep metric learning." *arXiv preprint arXiv:1811.12649* (2018).

[11] Deng, Jiankang, et al. "Arcface: Additive angular margin loss for deep face recognition." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.

# Experimental Results

- KNN classification

| | AID | | | | | | | | NWPU-RESISC45 | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Uniform | | | | Label-dependent | | | | Uniform | | | | Label-dependent | | | |
| | 0.1 | 0.3 | 0.5 | 0.7 | 0.1 | 0.3 | 0.5 | 0.7 | 0.1 | 0.3 | 0.5 | 0.7 | 0.1 | 0.3 | 0.5 | 0.7 |
| D-CNN | 92.40 | 86.80 | 75.25 | 60.10 | 92.45 | 88.95 | 84.40 | 84.25 | 90.05 | 81.92 | **88.62** | 40.59 | 90.06 | 85.70 | 80.00 | 77.19 |
| Triplet | 91.80 | 85.35 | 77.15 | 55.35 | 93.30 | **90.50** | 85.80 | **85.60** | 86.92 | 75.19 | 61.57 | 50.68 | 90.06 | 87.76 | 83.41 | **79.25** |
| NSL | 89.35 | 84.35 | 75.60 | 63.90 | 90.20 | 87.25 | 85.15 | 84.15 | 87.46 | 78.73 | 65.57 | 45.38 | 88.27 | 84.14 | 80.19 | 78.00 |
| SNCA | 90.65 | 82.70 | 64.55 | 42.90 | 90.40 | 81.95 | 75.40 | 74.45 | 87.90 | 76.08 | 59.75 | 30.03 | 87.17 | 77.97 | 68.02 | 63.44 |
| ArcFace | 90.30 | 79.95 | 87.30 | **82.30** | 90.80 | 81.35 | **86.55** | 83.60 | 87.75 | 88.68 | 85.52 | **80.71** | 87.16 | 71.25 | 80.35 | 78.25 |
| MAE | 82.15 | 82.70 | 79.90 | 80.65 | 83.45 | 81.05 | 82.25 | 81.15 | 78.69 | 78.53 | 75.43 | 74.03 | 77.94 | 77.65 | 76.97 | 77.17 |
| RNSL | 93.25 | 91.15 | 81.10 | 54.25 | 93.15 | 85.65 | 79.65 | 76.10 | 92.25 | 88.92 | 80.54 | 49.63 | 91.30 | 84.32 | 74.35 | 68.21 |
| t-RNSL | **94.05** | **91.80** | **89.50** | 78.25 | **93.80** | **90.50** | 86.05 | 81.55 | **92.30** | **90.76** | 88.56 | 79.84 | **92.03** | **89.30** | **84.84** | 76.84 |

# Experimental Results

- Clustering
  - ACC

| | AID | | | | | | | | NWPU-RESISC45 | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Uniform | | | | Label-dependent | | | | Uniform | | | | Label-dependent | | | |
| | 0.1 | 0.3 | 0.5 | 0.7 | 0.1 | 0.3 | 0.5 | 0.7 | 0.1 | 0.3 | 0.5 | 0.7 | 0.1 | 0.3 | 0.5 | 0.7 |
| D-CNN | 91.65 | 80.90 | 65.50 | 28.85 | 87.70 | 79.50 | 73.35 | 64.50 | 86.86 | 77.86 | **85.00** | 17.41 | 85.71 | 78.65 | 66.40 | 59.78 |
| Triplet | 89.20 | 79.40 | 65.25 | 34.20 | 89.70 | 85.20 | 70.95 | 65.30 | 79.21 | 60.40 | 42.54 | 25.30 | 84.41 | 80.94 | 72.60 | 57.54 |
| NSL | 85.10 | 71.85 | 45.30 | 21.75 | 87.25 | 77.95 | 70.70 | **68.55** | 80.97 | 69.41 | 43.81 | 10.75 | 83.19 | 74.95 | 69.35 | **65.65** |
| SNCA | 90.00 | 81.75 | 60.20 | 26.30 | 90.15 | 79.05 | 63.80 | 59.55 | 88.00 | 76.06 | 58.59 | 24.62 | 86.94 | 76.21 | 60.03 | 47.02 |
| ArcFace | 90.55 | 79.25 | 78.50 | 68.95 | 87.90 | 77.00 | 65.30 | 62.00 | 87.70 | 78.48 | 75.41 | 68.38 | 86.87 | 68.49 | 52.73 | 48.62 |
| MAE | 63.85 | 66.45 | 53.30 | 57.75 | 67.15 | 57.20 | 56.45 | 56.65 | 52.13 | 52.97 | 50.70 | 46.46 | 53.33 | 50.13 | 51.84 | 49.37 |
| RNSL | 90.90 | 90.45 | 72.60 | 23.75 | 90.70 | 76.75 | 62.90 | 54.85 | 89.41 | 88.11 | 75.59 | 26.86 | 88.33 | 79.68 | 64.14 | 53.41 |
| t-RNSL | **94.00** | **91.40** | **86.60** | **69.55** | **93.50** | **86.65** | **76.55** | 63.20 | **91.48** | **90.27** | 84.98 | **78.16** | **89.22** | **86.63** | **75.10** | 61.22 |

SOOCHOW UNIVERSITY

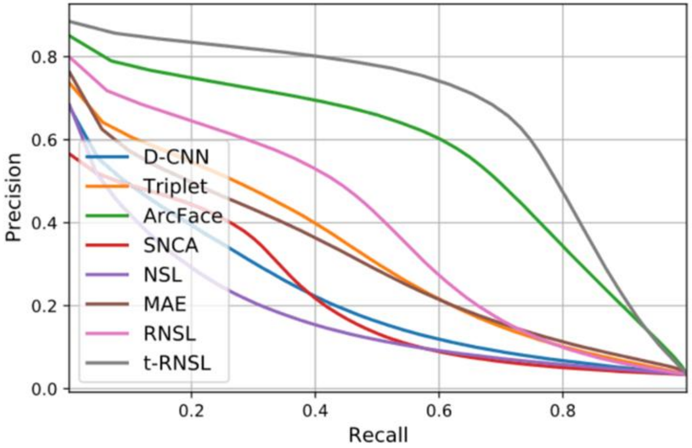
# Experimental Results

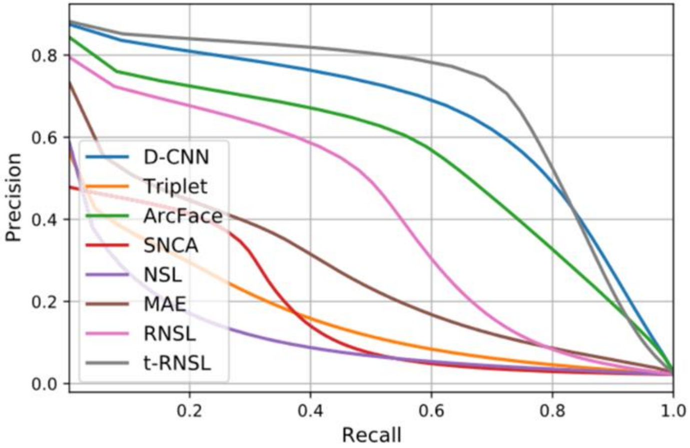
- Image Retrieval
  - MAP@20

| | AID | | | | | | | | NWPU-RESISC45 | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Uniform | | | | Label-dependent | | | | Uniform | | | | Label-dependent | | | |
| | 0.1 | 0.3 | 0.5 | 0.7 | 0.1 | 0.3 | 0.5 | 0.7 | 0.1 | 0.3 | 0.5 | 0.7 | 0.1 | 0.3 | 0.5 | 0.7 |
| D-CNN | 93.25 | 84.63 | 69.17 | 50.98 | 93.44 | 87.28 | 81.46 | 78.61 | 91.60 | 80.94 | 90.28 | 41.53 | 91.02 | 85.24 | 77.80 | 73.44 |
| Triplet | 93.13 | 85.87 | 74.96 | 56.34 | 93.79 | 90.64 | 86.44 | **83.56** | 87.99 | 75.10 | 61.50 | 51.79 | 90.95 | 89.57 | 84.42 | **78.85** |
| NSL | 90.61 | 81.01 | 67.96 | 55.26 | 90.40 | 85.50 | 81.84 | 78.96 | 88.71 | 76.53 | 62.30 | 44.44 | 89.39 | 83.42 | 78.14 | 75.52 |
| SNCA | **96.81** | 89.12 | 68.92 | 48.05 | **95.41** | 85.72 | 78.64 | 76.99 | 96.09 | 85.49 | 67.07 | 42.06 | 92.71 | 83.24 | 72.64 | 69.38 |
| ArcFace | 96.04 | 85.55 | 85.41 | 77.27 | 94.72 | 83.88 | 84.47 | 81.73 | **96.17** | 88.59 | 85.20 | 78.30 | 93.84 | 77.83 | 79.69 | 77.37 |
| MAE | 79.18 | 80.09 | 74.04 | 74.77 | 81.88 | 76.29 | 77.04 | 76.27 | 76.03 | 75.57 | 73.13 | 69.74 | 81.88 | 76.29 | 77.04 | 76.27 |
| RNSL | 95.53 | 93.28 | 82.23 | 51.48 | 94.78 | 85.72 | 76.92 | 71.12 | 95.26 | 92.43 | 84.49 | 53.21 | 94.11 | 86.24 | 73.25 | 66.62 |
| t-RNSL | 96.04 | **93.71** | **90.55** | **77.41** | 95.38 | **91.37** | **86.53** | 77.79 | 95.10 | **94.11** | **91.33** | **83.29** | **94.83** | **91.32** | **85.35** | 76.23 |

# Experimental Results

- Image Retrieval
  - PR curve

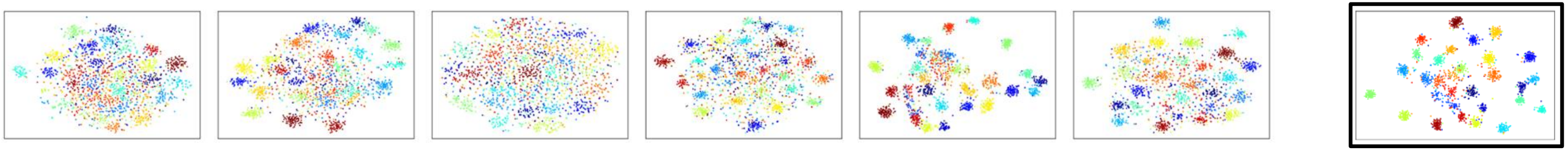


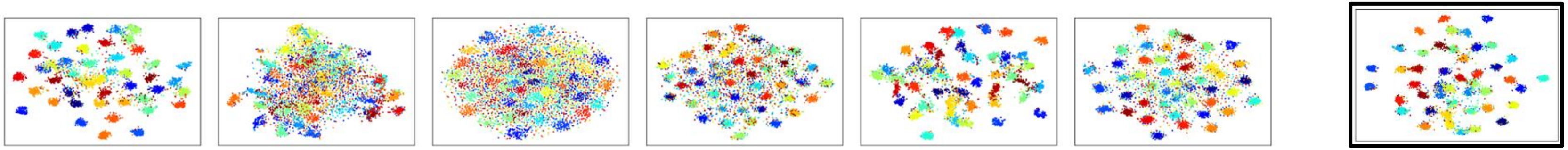
# Experimental Results

- Feature visualization (noise level:0.5)
  - t-SNE

AID



NWPU-RESISC45



# Conclusion

- A novel robust loss function is proposed for deep embedding of RS images

- Compared to other state-of-the-art methods, RNSL achieves significant performance improvement in several tasks when the noisy labels exist

# Thank you for your attention!

https://jiankang1991.github.io/