Learning from Noisy Labels for Earth Observation

Prof. Dr. Begüm Demir

Remote Sensing Image Analysis (RSiM) Group, Faculty of EECS, TU Berlin

Big Data Analytics for Earth Observation (BigEarth) Group, BIFOLD

BIFOLD: Berlin Institute for the Foundations of Learning and Data

BBDC
BERLIN BIG DATA CENTER

SPONSORED BY THE
Federal Ministry of Education and Research

2014-2022

- Declarative ML in selected applications
- ML in heterogeneous datasets and data streams
- Scalable processing of heterogeneous, geo-distributed data streams
- Near real-time processing of millions of data sources

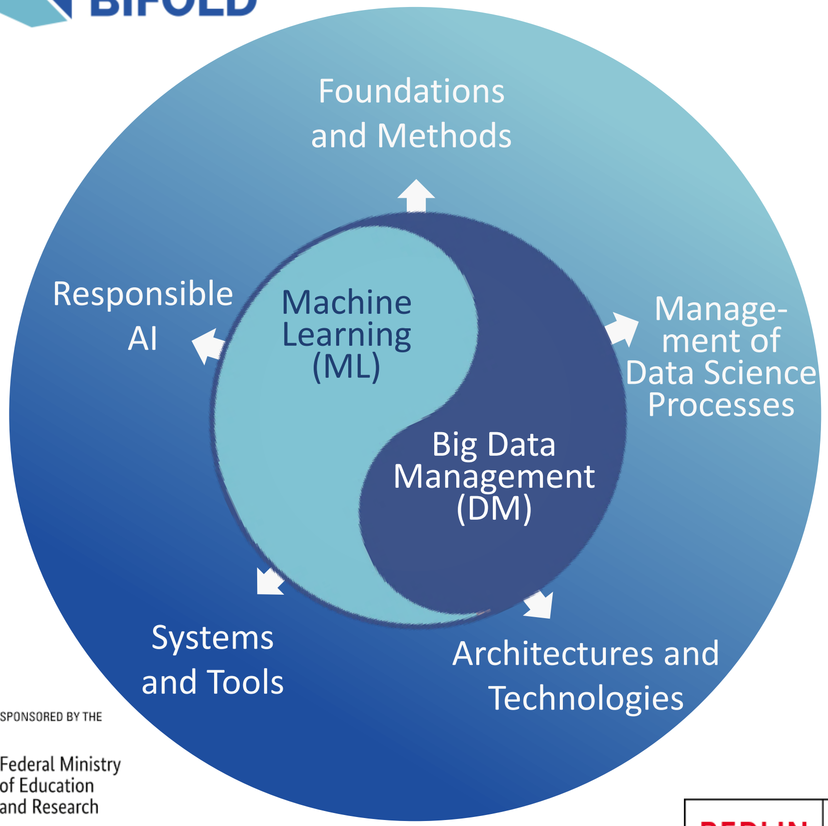
Berlin Center for Machine Learning

SPONSORED BY THE
Federal Ministry of Education and Research

2018-2022

- Theoretical and algorithmic foundations of ML
- New scientific and technical ML applications
- Usability of ML methods for business and science
- New interdisciplinary research in science and medicine

BIFOLD

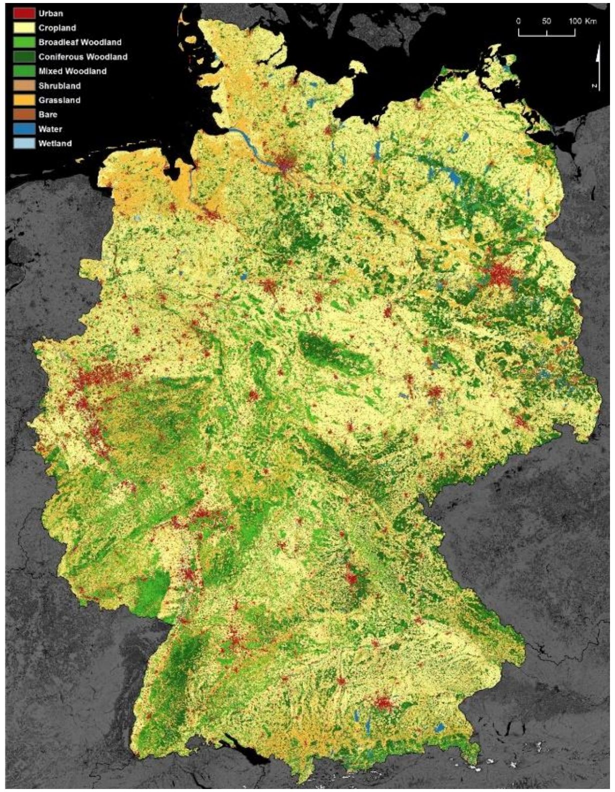


SPONSORED BY THE
Federal Ministry of Education and Research

BERLIN

Introduction

- The two tasks common to most of the EO applications are:
  - land-use land-cover (LULC) maps generation and updating



  - automatic scene understanding based on the LULC classes

Urban fabric, Arable land, Mixed forest, Land principally occupied by agriculture.

Arable land, Mixed forest, Urban fabric.

Urban fabric, Arable land, Coniferous forest, Transitional woodland/shrub, Land principally occupied by agriculture.

Urban fabric, Arable land, Land principally occupied by agriculture.

Urban fabric, Arable land, Coniferous forest, Mixed forest, Transitional woodland/shrub.

Urban fabric, Arable land, Pastures, Complex cultivation patterns

Coniferous forest, Mixed forest, Inland waters, Transitional woodland/shrub.

Urban fabric, Arable land, Coniferous forest, Mixed forest, Transitional woodland/shrub.

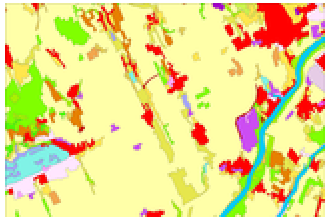
? Realization of these tasks requires a large amount of annotated training samples. Collecting annotations is costly in terms of human time/effort and needs expertise.

Thematic Products as Labelling Sources

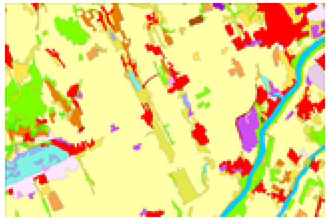
BIFOLD

Use the publicly available **thematic products** (digital maps).

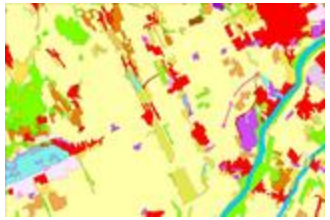
- Thematic products can be used to generate large scale training sets with zero-labeling-cost, covering wide areas, such as:
  - **Global scale,** e.g., Global Land Cover, the ESA CCI-LC product;
  - **Continental scale,** e.g., Corine Land Cover maps at European level;
  - **National scale,** e.g., American National Land Cover Dataset, DFD Land Use and Land Cover Product for Germany.



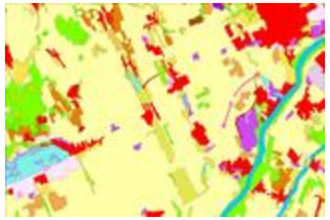
CLC 1990



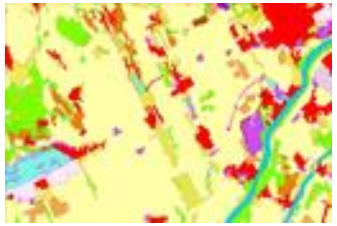
CLC 2000



CLC 2006



CLC 2012



CLC 2018

- G. Jaffrain, C. Sannier, A. Pennec, and H. Dufourmont, “Corine land cover 2012 - final validation report,” European Environment Agency, Tech. Rep., 2017.
- M. Hovenbitzer, F. Emig, C. Wende, S. Arnold, M. Bock, and S. Feigenspan. "Digital land cover model for Germany–DLM-DE," In Land Use and Land Cover Mapping in Europe, pp. 255-272. Springer, Dordrecht, 2014.
- ESA. Land Cover CCI Product User Guide Version 2. Technical Report, 2017.

Thematic Products as Labelling Sources

- Thematic maps can produce noisy labels due to:
  - errors in the map due to different the approaches used to generate the map;
  - changes in land-use/land-cover after the construction of the maps;
  - errors due to the misalignment between the map and the satellite image.
- Two types of label noise can be present in a training image:
  - missing labels;
  - wrong labels.

Training sets with noisy labels can distort the learning process. Methods that are robust to the label noise are required.

Discontinuous urban fabric
Coniferous forest
Mixed forest
Industrial or commercial units

missing label

Discontinuous urban fabric
Industrial or commercial units
Non-irrigated arable land
Coniferous forest

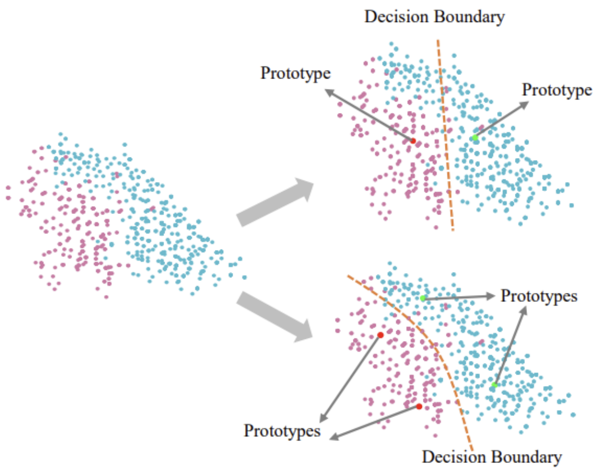
wrong label

Learning from Noisy Labels

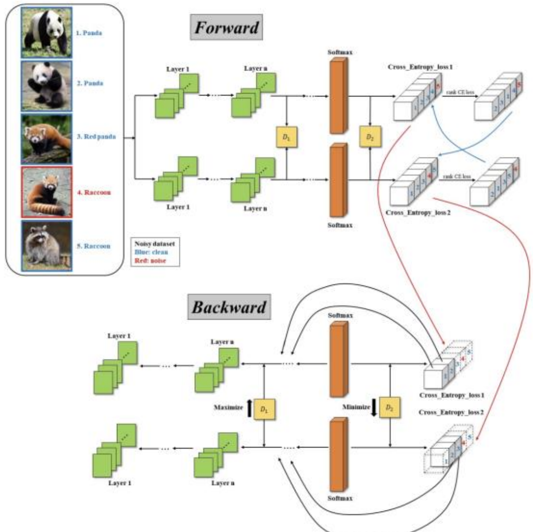
- There are several label-noise robust methods, relying on:
  - estimation of a noise transition matrix;
  - noise robust loss functions;
  - building class prototypes;
  - student-teacher networks;
  - data augmentation techniques.

Most of these works target the wrong labels only and thus not directly suitable for multi-label scene analysis problems in RS.

Song et al. 'Learning from Noisy Labels with Deep Neural Networks: A Survey', in IEEE Transactions on Neural Networks and Learning Systems, doi: 10.1109/TNNLS.2022.3152527, 2022.



Han et al, Deep Self-Learning From Noisy Labels, CVPR 2019.



Han et al, Learning from Noisy Labels via Discrepant Collaborative Training, WACV 2020.

Multi-Label Classification in Remote Sensing

- Automatic scene understanding based on the LULC classes

Urban fabric,
Arable land,
Mixed forest, Land principally occupied by agriculture.

Arable land,
Mixed forest,
Urban fabric.

Urban fabric,
Arable land, Coniferous forest,
Transitional woodland/shrub,
Land principally occupied by agriculture.

Urban fabric,
Arable land,
Land principally occupied by agriculture.

Urban fabric,
Arable land,
Coniferous forest,
Mixed forest, Transitional woodland/shrub.

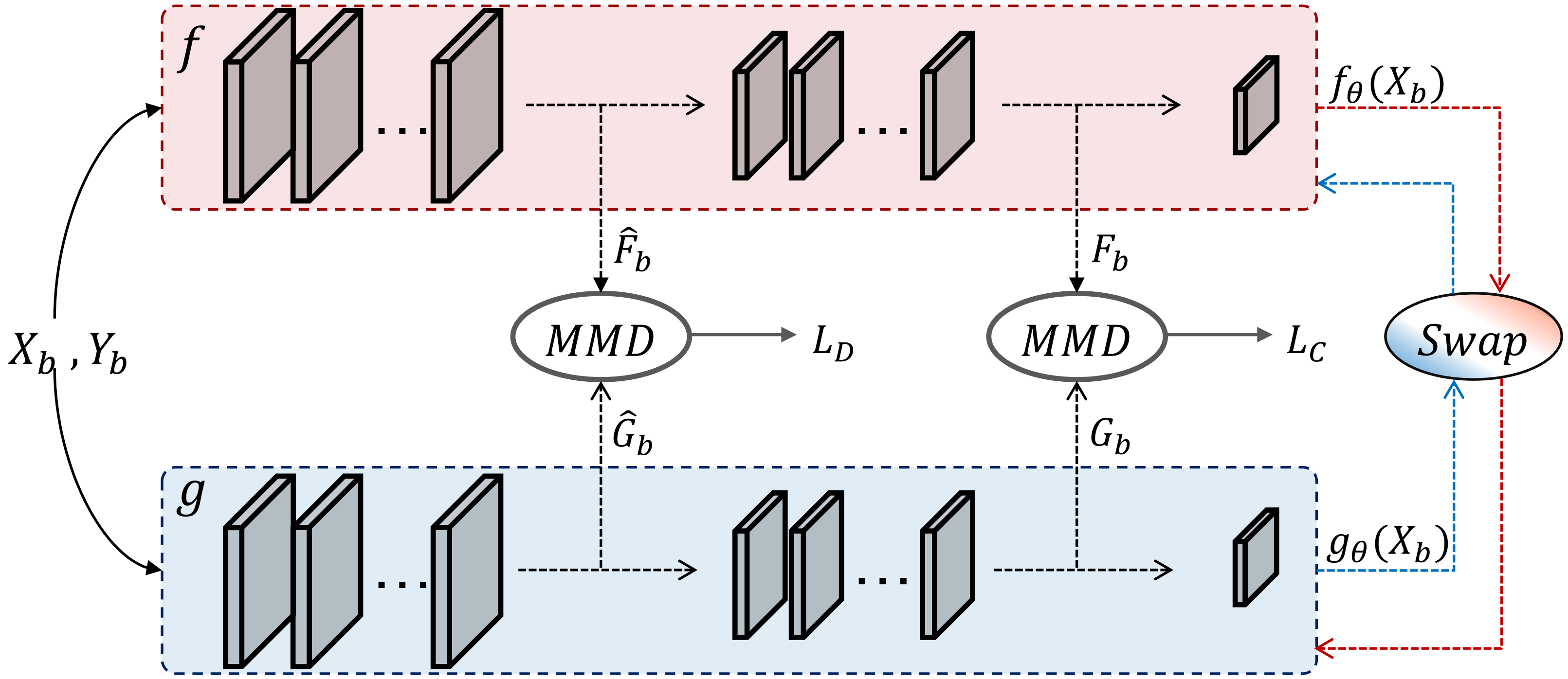
Urban fabric,
Arable land,
Pastures,
Complex cultivation patterns

Coniferous forest,
Mixed forest,
Inland waters,
Transitional woodland/shrub.

Urban fabric,
Arable land,
Coniferous forest,
Mixed forest,
Transitional woodland/shrub.

Solution: We have developed a robust consensual collaborative learning method (RCML).

RCML: Robust Collaborative Multi-Label Learning Method

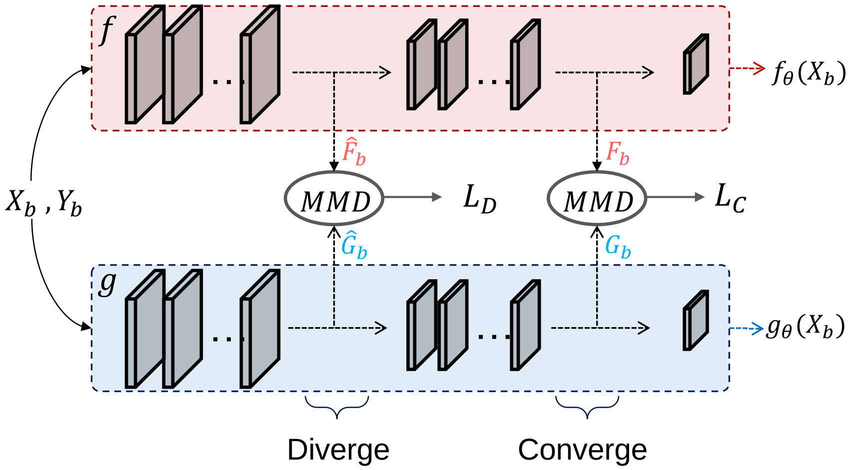


A. K. Aksoy, M. Ravanbakhsh, T. Kreuziger, B. Demir, "A Consensual Collaborative Learning Method for Remote Sensing Image Classification Under Noisy Multi-Labels", IEEE International Conference on Image Processing, Alaska, USA, Sep. 2021.

A. K. Aksoy, M. Ravanbakhsh, B. Demir, "Multi-Label Noise Robust Collaborative Learning Model for Remote Sensing Image Classification", IEEE Transactions on Neural Networks and Learning Systems, in minor revision, 2022 (Arxiv Preprint- arxiv.2012.10715).

RCML Modules

- The Discrepancy Module aims at forcing the two networks to learn diverse features, while achieving consistent predictions.
- It includes: 1) Disparity loss (L_D); and 2) Consistency loss (L_C).



distribution P distribution Q

$$\underbrace{MMD(P,Q)}_{\text{maximum mean discrepancy}} = \|\mu_P - \mu_Q\|_H$$

mean of P mean of Q

- The Group Lasso Module:
  - identify potentially noisy labels by using the predictions of the two networks;
  - identify the type of label noise by computing a sample-wise ranking loss as:

assigned labels unassigned labels ranking error

$$\mathrm{Lasso}_f(x_i) = \underbrace{\alpha \sum_{\hat{c}=a+1}^{m} \sqrt{\sum_{c=1}^{a} \epsilon}}_{\text{aggregated loss based on missing labels}} + \underbrace{\beta \sum_{c=b+1}^{m} \sqrt{\sum_{\hat{c}=1}^{b} \epsilon}}_{\text{aggregated loss based on wrong labels}}$$

$$E_{c,\hat{c}}(\mathbf{x}_i) = \max(0, 2[f_{\hat{c}}(\mathbf{x}_i) - f_c(\mathbf{x}_i)] + 1)$$

Missing label	Wrong label	$f_{\hat{c}}$	f_c	$E_{c,\hat{c}}$
✗	✗	0	1	0
☑	✗	0	0	+1
✗	☑	1	1	+1
☑	☑	1	0	+3

RCML Modules

- The Swap Module :
  - exchanges the ranking information between the networks;
  - takes the Binary Cross Entropy and ranking losses into consideration to eliminate the detected noisy samples from back-propagation.

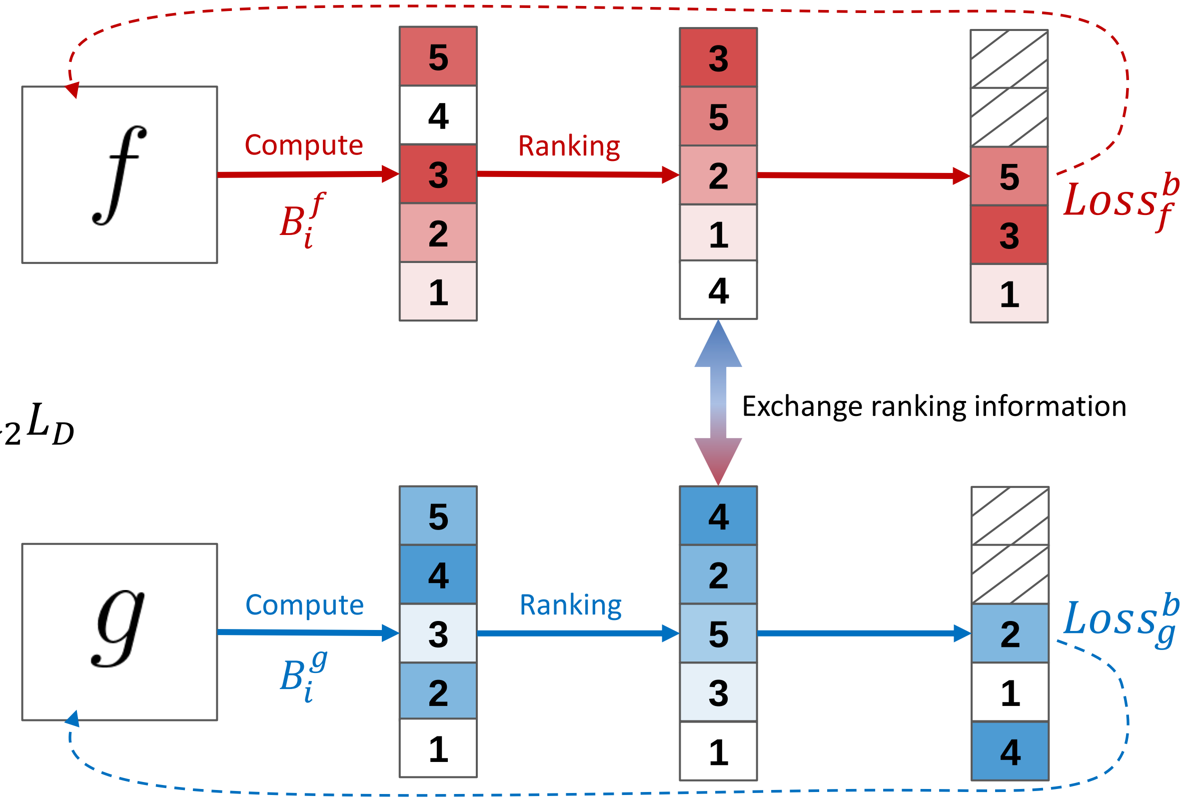
BCE loss

Sample from mini-batch

$$B_i^f = L_f(\mathbf{x}_i) + \gamma Lasso_f(\mathbf{x}_i)$$

$$Loss_f^b = \frac{1}{R}\sum_{i=1}^{R} L_f(\mathbf{x}_i) + \lambda_1 L_c + \lambda_2 L_D$$

BCE loss for R samples associated with low loss values identified by *g*



BigEarthNet: A Large-Scale Benchmark Archive for EO

BIFOLD

In the experiments we used our **BigEarthNet** dataset that:

- consists of 590,326 Sentinel-1&2 images.
- opens up promising directions to support studies for the analysis of large-scale EO data archives.

Available also at:

Google Earth Engine

TensorFlow

Radiant MLHub
EARTH IMAGERY FOR IMPACT

BigEarthNet: A Large-Scale Benchmark Archive for Remote Sensing Image Understanding. IEEE IGARSS, Yokohama, Japan, 2019.

BigEarthNet-MM: A Large Scale Multi-Modal Multi-Label Benchmark Archive for Remote Sensing Image Classification and Retrieval. IEEE GRSM, 2021.

Experimental Setup

BIFOLD

- The ResNet50 architecture was considered as backbone for our RCML.
- We report the results of training obtained after 100 epochs.
- The model training and further experiments were conducted on a Tesla V100 GPU with 32 GB RAM.
- In our experiments, the ratio of the introduced noise was varied from 10% to 50%.
- The results of the experiments were provided in terms of: 1) mean average precision (mAP); and 2) F1 score.
- We compared our proposed method with four baseline methods:

T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, “Focal loss for dense object detection,” in Proceedings of the IEEE international conference on computer vision, Venice, Italy, 2017, pp. 2980–2988..

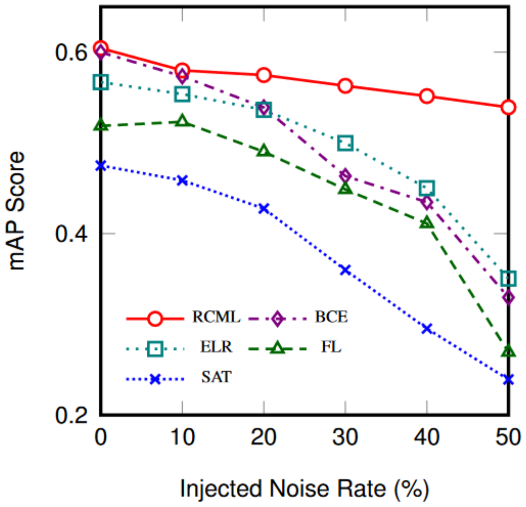
S. Liu, J. Niles-Weed, N. Razavian, and C. Fernandez-Granda, “Early-learning regularization prevents memorization of noisy labels,” in Advances in Neural Information Processing Systems, vol. 33, Vancouver, Canada, 2020.

L. Huang, C. Zhang, and H. Zhang, “Self-adaptive training: beyond empirical risk minimization,” Advances in Neural Information Processing Systems, vol. 33, 2020.

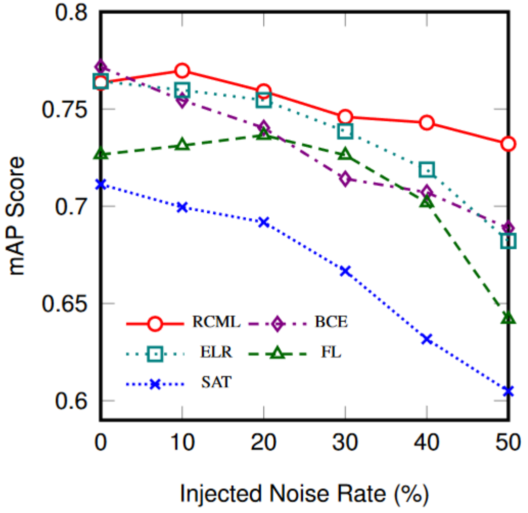
RCML: Results on BigEarthNet

BIFOLD

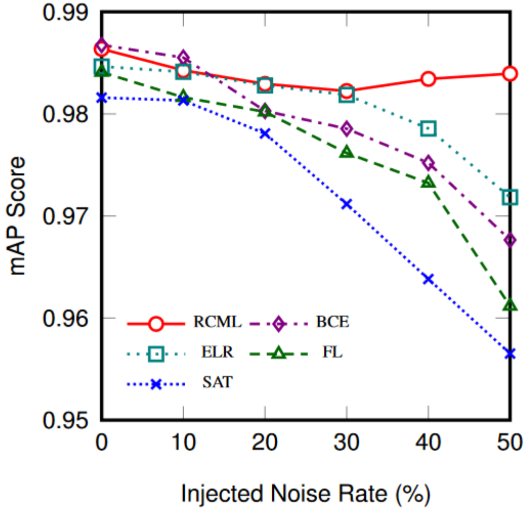
Urban fabric

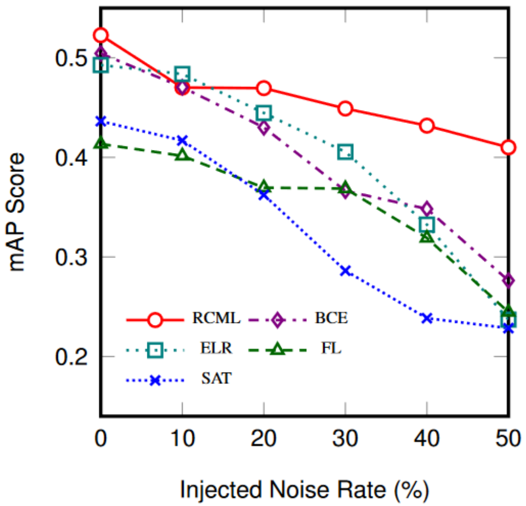


Arable lands

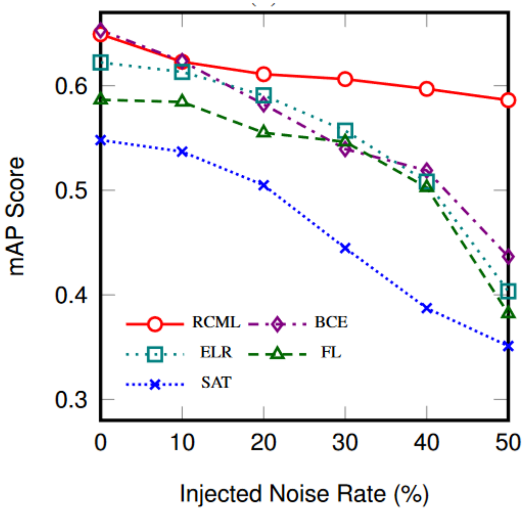


Pasture

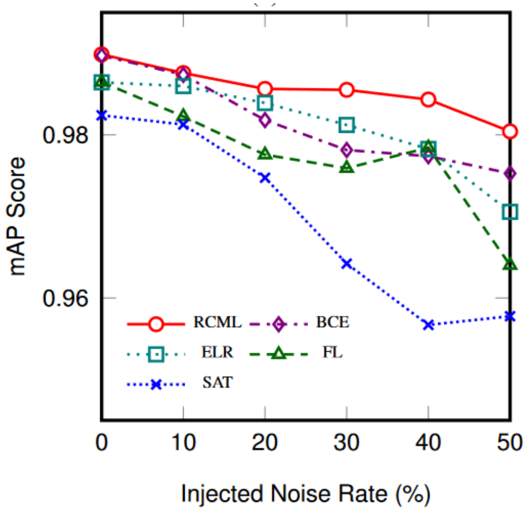




Moors, heathland and sclerophyllous vegetation



Inland wetlands



Marine Waters

FL: focal loss

SAT: self-adaptive training

ELR: early-learning regularization

BCE: standard binary cross-entropy

RCML: Results on BigEarthNet

Noise Rate	F_1 (%)					mAP (%)				
	FL	SAT	ELR	BCE	RCML	FL	SAT	ELR	BCE	RCML
0%	64.0	66.0	66.2	**72.5**	72.3	45.5	43.8	47.7	49.6	**49.8**
10%	62.4	65.0	65.2	71.0	**72.3**	45.1	43.0	47.0	47.5	**47.8**
20%	55.9	62.4	62.3	67.9	**71.7**	43.8	41.0	45.5	45.5	**47.2**
30%	48.5	56.2	57.0	61.4	**70.8**	42.7	38.3	43.6	42.8	**46.6**
40%	39.0	48.3	48.4	56.3	**70.3**	40.7	36.0	41.4	41.7	**46.5**
50%	30.7	34.0	35.1	37.5	**67.8**	35.3	34.2	37.3	37.9	**45.3**

FL: focal loss

SAT: self-adaptive training

ELR: early-learning regularization

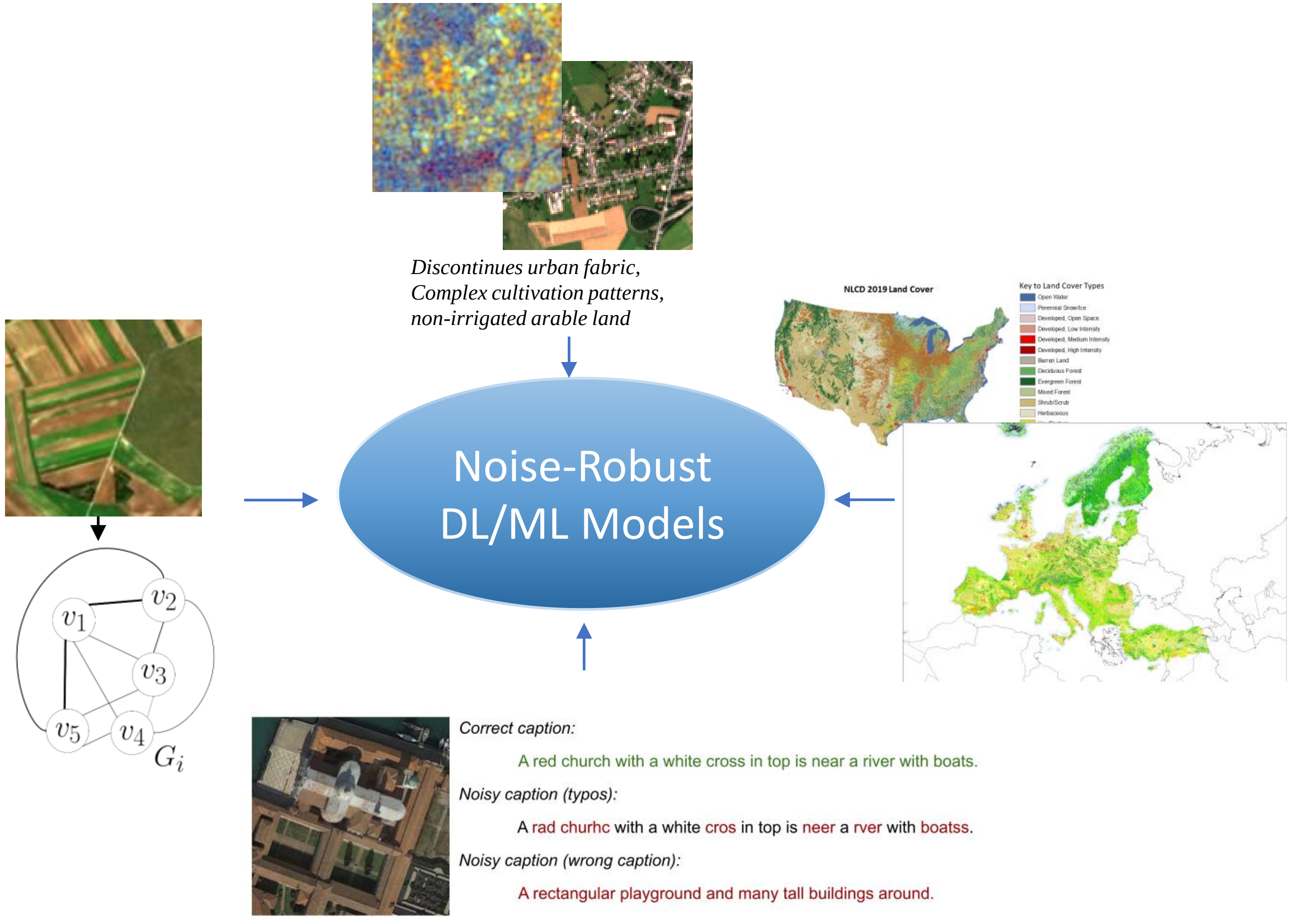
BCE: standard binary cross-entropy

Concluding Remarks

- Developing efficient techniques for handling label noise is becoming more and more important.

- RCML is promising since it:
  - is able to automatically identify two different types of noise without making any prior assumption;
  - is architecture-independent, and thus can be used within different network architectures;
  - is applicable in different EO applications (e.g., large scale image retrieval, etc.).

Outlook



- The codes developed and maintained at our group are publicly available:

  https://rsim.berlin/software

  Follow us in twitter: @BigEarthERC