Band-wise Multi-scale CNN Architecture for Remote Sensing Image Scene Classification

Jian Kang and Begüm Demir

Remote Sensing Image Analysis (RSiM) Group, TU Berlin

jian.kang@tu-berlin.de, demir@tu-berlin.de

Outline

- Introduction
- Motivation
- Band-wise multi-scale CNN architecture
- Experiments
- Conclusion

RSiM

Introduction

- Scene classification of Remote Sensing(RS) images
  - Characterization of remote sensing images based on single-label or multi-label land-use or land-cover classes
  - Existing large-scale scene classification datasets: e.g., AID[1] and BigEarthNet[2]

AID: Aerial images
30 scene classes
Single-label
RGB bands

Bareland
Beach
Farmland
River

BigEarthNet:
Sentinel-2 images
43 scene classes
Multi-labels
Multi-spectral bands

Discontinuous urban fabric; Land principally occupied by agriculture; Non-irrigated arable land.
Broad-leaved forest; Complex cultivation patterns; Discontinuous urban fabric; Non-irrigated arable land; Pastures.
Broad-leaved forest; Complex cultivation patterns; Land principally occupied by agriculture; Transitional woodland/shrub
Agro-forestry areas; Broad-leaved forest; Transitional woodland/shrub; Water bodies.

[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] Sumbul, Gencer, et al. "Bigearthnet: A large-scale benchmark archive for remote sensing image understanding." *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2019.

Introduction

- Scene classification of RS images
  - Deep learning has achieved state-of-the-art classification performance
  - Most of the proposed methods for scene classification are based on the pre-trained convolutional neural network (CNN) architectures on the large-scale computer vision archives (e.g., ImageNet).
  - However, these pre-trained CNN architectures cannot be directly applied on the scene classification with high-dimensional RS images (e.g., multi-spectral images)
- Motivation of this work:
  - Characterization of semantic contents for high-dimensional RS images based on a novel CNN architecture

RSiM

Band-wise multi-scale CNN architecture

- Standard 2D convolutional layer (bias is omitted)

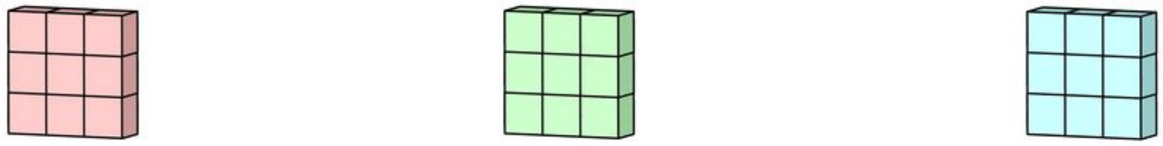


Image Credit[3]

[3] https://towardsdatascience.com/intuitively-understanding-convolutions-for-deep-learning-1f6f42faee1

Band-wise multi-scale CNN architecture

- Standard 2D convolutional layer (bias is omitted)

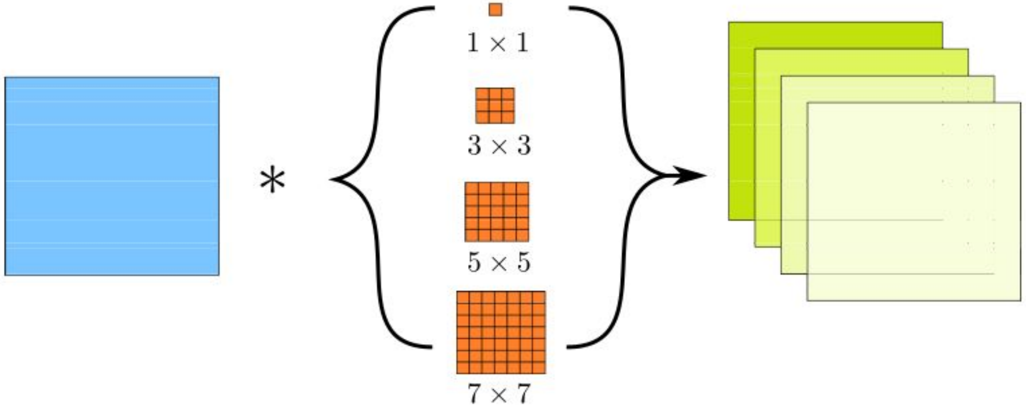
$$\mathbf{u}_m = \sigma(\mathbf{w}_m * \mathbf{x}_i) = \sigma(\sum_{c=1}^{C} \mathbf{w}_m^c * \mathbf{x}_i^c)$$

  - Through such operation, the spectral features may not be optimally extracted, since the process for the spectral feature extraction is entangled within the summation of the spatial convolution results.

  - In addition, the convolution layer with a fixed size filter $\mathbf{w}_m$ may not sufficiently extract the spatial features, especially for different land-use or land-cover objects with different spatial sizes.

Band-wise multi-scale CNN architecture

- Band-wise multi-scale convolution
  - Sufficiently characterizing the multi-scale spatial features in a band-wise manner

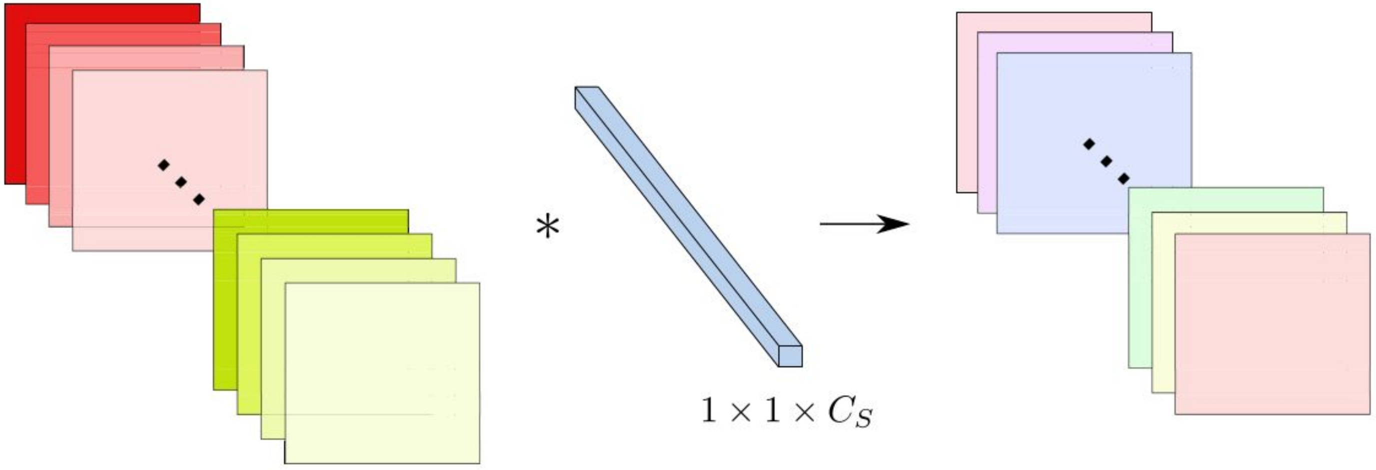


$$\hat{\mathbf{X}}_i^{c_s} = \mathbf{W}^{c_s} * \mathbf{X}_i^{c}$$

RSiM

Band-wise multi-scale CNN architecture

- Pixel-wise convolution
  - Learning the spectral information fusion in a pixel-wise manner

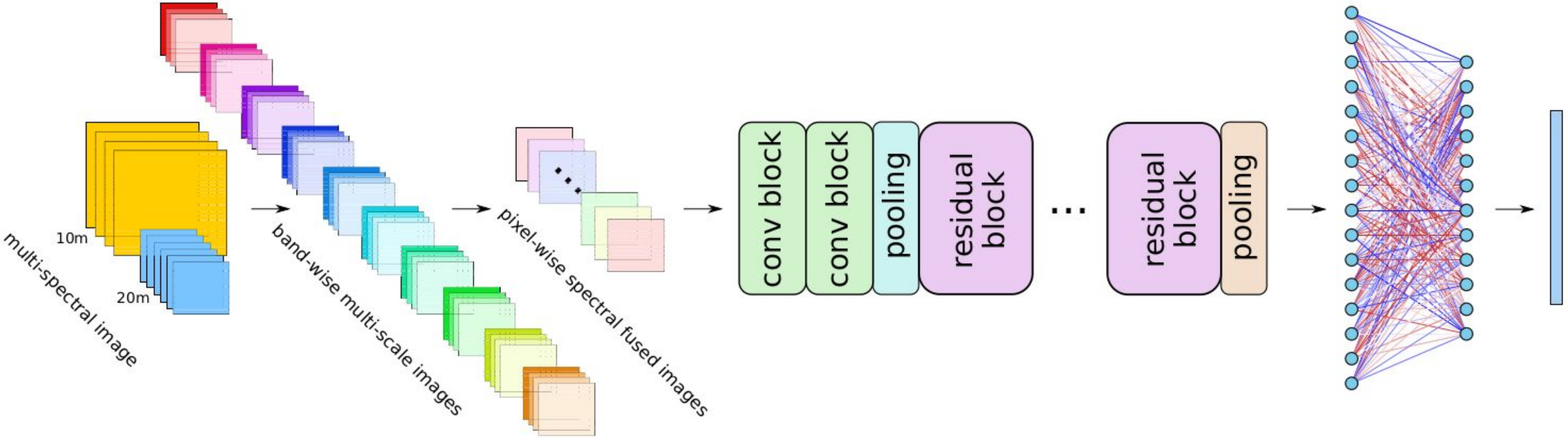


$$u_m(i,j) = \sigma\left(\sum_{c_s} \hat{w}_m^{c_s} \hat{x}_i^{c_s}(i,j)\right)$$

RSiM

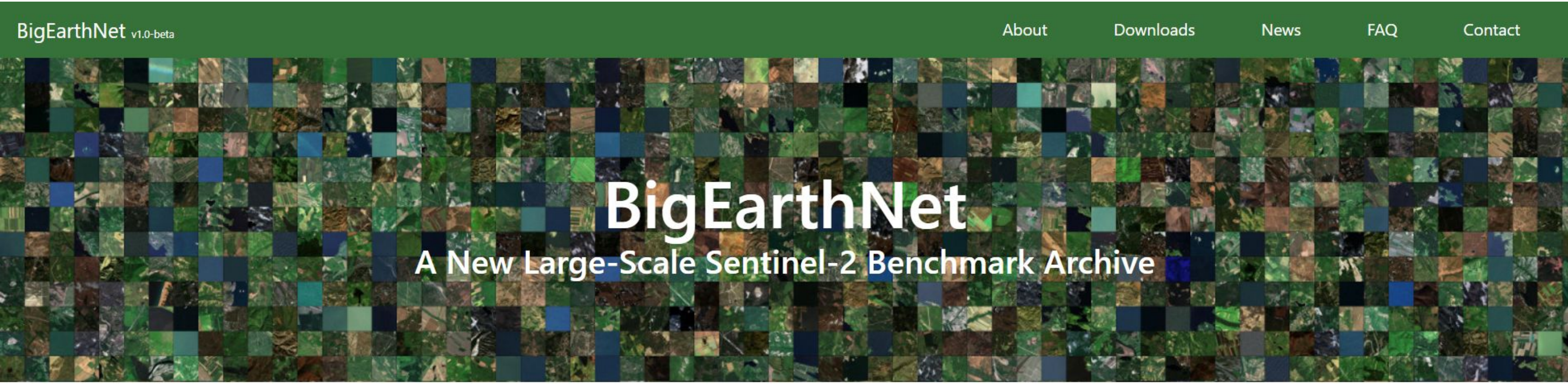
Band-wise multi-scale CNN architecture

- Standard 2D residual blocks for learning high-level semantic information



BWMS CNN architecture

Experiments: Dataset Description



BigEarthNet is utilized for evaluating the performance of the proposed CNN architecture in the task of multi-label classification, where 10m and 20m bands are exploited and the training, validation and test images are following [4].

[4] Sumbul, Gencer, et al. "BigEarthNet Deep Learning Models with A New Class-Nomenclature for Remote Sensing Image Understanding." *arXiv preprint arXiv:2001.06372* (2020).

Experimental Design

- Pytorch implementation PyTorch

- The class probabilities are obtained by applying *sigmoid* activation function

- Binary Cross-Entropy (BCE) loss is utilized

- Adam optimizer with learning rate(LR) of 10^{-3}

- LR is decayed by a factor of 0.5 in every 30 epochs

- *Leaky ReLU* is exploited inside the proposed architecture

- ResNet18, ResNet50, and 3D-CNN are regarded as baseline methods

RSiM

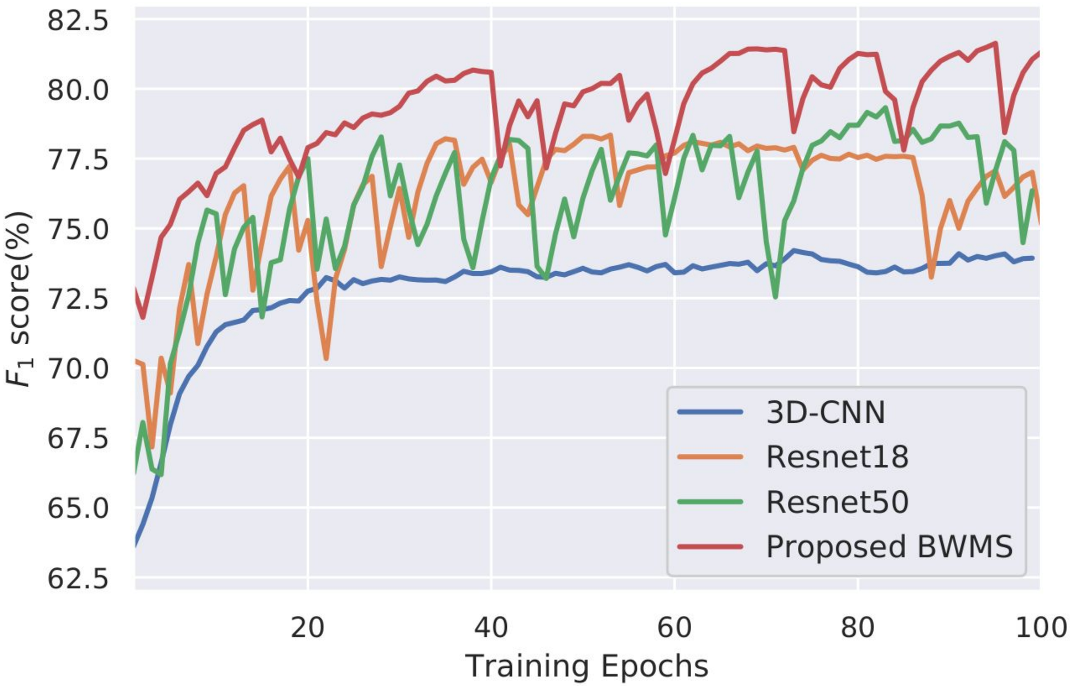
Experimental Design

- Metrics for the evaluation [5]:
  - F1 score, an integrated metric of sample precision and recall
  - Accuracy (Acc), the degree of sample-wise correctness
  - Hamming Loss (HL), evaluates the fraction of misclassified labels
  - Ranking Loss (RL), evaluates the fraction of reversely ordered label pairs

[5] Zhang, Min-Ling, and Zhi-Hua Zhou. "A review on multi-label learning algorithms." *IEEE transactions on knowledge and data engineering* 26.8 (2013): 1819-1837.

Experimental Results

- Learning curves of all the considered CNN architectures:



RSiM

Experimental Results

- Classification performances (%) under all the metrics and the numbers of parameters (#) for all the considered methods:

Architectures	F1	Acc	HL	RL	#para
3D-CNN	74.67	64.73	7.74	4.51	1.75M
ResNet18	78.68	69.38	6.74	3.52	11.2M
ResNet50	81.05	72.13	6.18	2.87	23.6M
Proposed BWMS	**81.84**	**73.07**	**5.97**	**2.70**	11.3M

Conclusion

- A novel CNN architecture for accurately capturing the spectral-spatial information content present in high-dimensional RS images.

- The proposed architecture is composed of:
  - A convolutional layer for extracting band-wise multi-scale spatial features
  - A convolutional layer for extracting pixel-wise spectral features
  - Standard 2D convolutional and residual blocks for learning the high-level semantic features

- The proposed convolutional layers can improve the classification performance by sufficiently extracting spectral-spatial features

- They can be also integrated into other high-dimensional RS image classification network, such as hyperspectral images.

Thank you very much for your attention!

RSiM