TRANSFER DEEP LEARNING FOR REMOTE SENSING DATASETS: A COMPARISON STUDY

Itza Hernandez-Sequeira, Ruben Fernandez-Beltran, Filiberto Pla

Institute of New Imaging Technologies, University Jaume I, E-12071 Castellón de la Plana, Spain

ABSTRACT

Remote sensing is also benefiting from the quick development of deep learning algorithms for image analysis and classification tasks. In this paper, we evaluate the classification performance of a well-known Convolutional Neural Network (CNN) models, such as ResNet50, using a transfer learning approach. We compare the performance when using vector-features acquired from general purpose data, such as the ImageNet [1], versus remote sensing data like BigEarthNet [2], UCMerced [3], RESISC45 [4] and So2Sat [5]. The results show that the model pre-trained on RESISC-45 data achieved the highest accuracy when classifying the Eurosat [6] testing dataset. This was followed by the model pre-trained on Imagenet with 95.94% and BigEarthNet with 95.93%. When presented with diverse remote sensing data, the classification improved in regards to large quantities of general-purpose data. The experiments carried out also show, that multi modal (co-registered synthetic aperture radar and multispectral) did not increase the classification rate with respect to using only multispectral data. The source codes of this work are available for reproducible research at https://github.com/itzahs/CNN-RS.

Index Terms— deep learning, transfer learning, remote sensing, Keras, Tensorflow

1. INTRODUCTION

Deep learning (DL) in remote sensing is often linked to image processing tasks such as image fusion, segmentation, and registration; and to classification tasks that include land cover and land use classification (LCLU), scene classification, and object detection [7]. Convolutional Neural Networks (CNN) are at the core of supervised image classification tasks using DL. Proven architectures like ResNet50 [4] are available as ready-to-use models in DL libraries (e.g. Keras [8]).

One of the common approaches is to implement transfer learning techniques to exploit large quantities of general-purpose data that is publicly available data (e.g. ImageNet [9]) in order to extract the main image features like borders, textures, and so forth. Then, instead of starting from a random initialization of the model parameters and training the model from the scratch, it is possible to leverage the features previously learned with other data and only calibrate the final layers in a small set of remote sensing images annotated for the specific problem to solve.

However, the main limitation of these general pre-trained models with respect to remote sensing is that benchmark archives contain RGB images while satellite imagery is usually composed of multiple bands (multispectral and hyperspectral) with different resolutions. In some cases, they are co-registered using multiple sensors (e.g. Optical and Synthetic Aperture Radar (SAR) [5]). Besides, in general image recognition problems, images are annotated by a single high-level category label whereas land cover is a pixel-wise classification problem [2].

In this paper, we aim to evaluate the transferability of proven architectures such as ResNet50 by comparing the results of a classification using feature vectors trained on general-purpose data matched up against remote sensing-specific data. In the first approach, we used a model pre-trained on Imagenet whereas in the second one we used transfer learning from standardized remote sensing models available in Tensorflow [10]. In section 2, first, we introduce the remote sensing datasets that were used to train the considered models. Then, we conduct the classification using transfer learning into the chosen benchmark data set: the Eurosat [6]. In section 3., we show the comparative results based on accuracy and compare them to the results reported by [6] on the Eurosat databse when training a ResNet50 from the scratch. Finally in Section 4., we provide conclusions and limitations regarding the transferability of standardized DL architectures in remote sensing.

In an effort to contribute to reproducible science, all the scripts and source code can be run directly in Google Colab Pro and are available in the following repository: https://github.com/itzahs/CNN-RS.

2. DATA AND METHODS

In this section, we describe the transfer learning implementation process using the Keras [8] library for DL. First, we present the previous-trained models available, and then the

This work was supported by the Ministry of Science, Innovation and Universities of Spain (RTI2018-098651-B-C54), and the University Jaume I (PREDOC/2020/50).

preprocessing of the input data to perform the classification.

2.1. Remote sensing DL models

The Tensorflow Hub repository contains a collection of five pre-trained models for remote sensing focused on satellite and airborne imagery [11]. These models use standardized datasets for classification tasks and share the same ResNet50 V2 architecture [1]. We used the models trained on benchmark data for remote sensing available in TensorFlow datasets, which are summarized in the table 1.

Table 1. Remote Sensing datasets in Tensorflow

Pre-training datasets	Source	Size	Image size
BigEarthNet [2]	Sentinel-2	590k	120x120*
Eurosat [6]	Sentinel-2	27k	64x64
RESISC45 [4]	Aerial	31.5k	256x256
So2Sat [5]	Sentinel-1/2	376k	32x32
UC Merced [3]	Aerial	2.1k	256x256

*Image size varies depending on resolution, from 120x120 to 60x60 and to 20x20.

The first two of these datasets are based on Sentinel-2 satellite imagery and incorporate multispectral information, including analogous bands for RGB. The largest one is the BigEarthNet [2] with 590,326 Sentinel-2 image patches distributed over 43 classes. In comparison, the Eurosat [6] has a total of 10 classes with 27,000 labeled and geo-referenced images. So2Sat [5] is a co-registered dataset with Sentinel-1 and Sentinel-2. The last two datasets are based on aerial imagery. They correspond to the RESISC45 [4], a collection of 31,500 RGB images for 45 scene classes as well as the UC Merced [3] which covers 21 land use classes with 100 images per class.

In addition to the RS models, we used a ResNet50 model pre-trained with Imagenet weights. To date (01/17/2022), Imagenet contains 14,197,122 images from 21, 841 categories.

2.2. Data preprocessing

In order to measure the classification performance of all the pre-trained models, we established a reference benchmark dataset. We decided to use the Eurosat for training, validation, and testing with a 60-10-30 percent split respectively. This dataset is relatively large and balanced, containing between 2000 to 3000 images per class. Furthermore, it has a reported classification accuracy of 95.32% for 70-30 training-test splits using a ResNet-50 architecture when training that model from the scratch [6].

Since the image patches have a size of 64x64 pixels, we resized them to 224 x 224 as this is the standard input size of all RS TensorFlow models. This was also necessary to implement the ResNet50-Imagenet in Keras. Nevertheless, we normalized the images between $[0, 1]$, taking into account the dynamic range of each image type, u8int values for RGB images and 16uint for multispectral. Concerning the inputs for the ResNet50 model with Imagenet weights, we used the preprocessing_input module. This component converts the RGB to BGR format and zero-centers each color channel with respect to the ImageNet dataset. Finally, no image augmentation was applied as it affects the intensity of the values and should be discarded [11].

Our approach was to carry out a feature extraction from a previously trained ResNet50 [1] architecture and freeze it to repurpose the feature maps learned into another data domain as seen in figure 1.

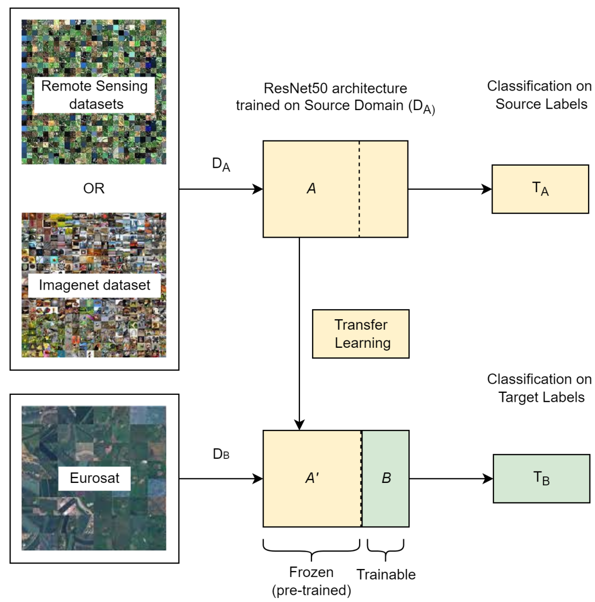

Fig. 1. Eurosat classification with transfer learning

We used the weights of Imagenet, BigEarthNet, RESISC45, So2Sat, and UCMerced and keep the auto-encoder layers frozen or non-trainable. We also added new trainable layers on top corresponding to Eurosat RGB (3 bands and 10 classes) or All (13 bands and 10 classes). Since all the input data has to be in 3 channels only, we use a 2D convolution layer for dimensional reduction when using multispectral data for the pre-trained RGB-based models (table 2).

Afterwards, we trained the classifier on top of the pre-trained model, followed by flattening the ResNet50 base-model output, and by a fully connected ReLu layer with 512 activation nodes. A Dropout layer with dropout nodes was set to 0.2 and passed to a fully connected (Dense) layer with 10 possible classes and softmax activation.

We carried out each experiment for 100 epochs while monitoring the validation accuracy and loss. We initially used a batch size of 128, but due to hardware limitations it couldn't be performed when using all multispectral bands.

Table 2. Model summary for the multispectral classification

Layer type	Output shape	# Param
Conv2D	(None, 224, 224, 3)	42
Keraslayer*	(None, 2048)	23,633,899
Flatten	(None, 2048)	0
Dense_1	(None, 512)	1,049,088
Dropout	(None, 512)	0
Dense_2	(None, 10)	5,130
Total parameters : 24,688,159		
Trainable parameters: 1,054,260		
Non-trainable parameters: 23,633,899		

*ResNet-50 models pre-trained on ImageNet or RS datasets.

In order to keep a straightforward implementation, no fine-tuning was performed.

All the experiments for RGB images have been conducted in Google Colab Pro with 16GB GPU and 25GB RAM. Whereas the classification with all 13 bands was executed on Ubuntu 20.04 x64 server with 11GB GPU and 112GB RAM.

3. EXPERIMENTAL RESULTS

In this section, we summarize the results and make a comparison of the models performance based on the achieved accuracy on the reference dataset chosen for the experiments, that is Eurosat dataset.

The highest accuracy was achieved by the features extracted from the RESISC45 with a 97.23% for a 64 batch size and 97.10% for a 128 batch size. One of the featured strengths of the reference dataset is that it has a balanced number of images (700 per class), with high within-class diversity and between-class similarity for increased complexity [4]. It is also the only dataset holding different spatial resolutions, from scenes classifying airplanes and rectangular farmlands to clouds.

Table 3. Transfer-learning from a ResNet50 model pre-trained on benchmark datasets and tested on Eurosat

	128 batch size		64 batch size	
Model	Loss	Acc	Loss	Acc
Imagenet[1]	0.4819	0.9594	0.5782	0.9593
So2Sat[5]	0.3302	0.8970	0.3831	0.8922
UC Merced[3]	0.2327	0.9486	0.2822	0.9475
RESISC45[4]	0.1345	**0.9710**	0.1554	**0.9723**
BEN-RGB[2]	0.1729	0.9593	0.2007	0.9586
BEN-ALL[2]	-*	-*	0.2915	0.9406

*Experiment could not be performed due to hardware limitations. BEN stands for BigEarthNet.

The other two models that overcome the classification benchmark of 95.32% in Eurosat are the Imagenet with 95.94% and the BigEarthNet-RGB with 95.93%. The more general Imagenet performs similarly to the remote sensing specific BigEarthNet. Both models were trained with datasets containing a large number of images, allowing them to reach high generalization ability. A possible drawback for BigEarthNet is that this dataset contains imbalanced classes, which can lower overall accuracy. For example, the mixed forest has over 217k images, whereas continuous urban fabric 10k and burnt areas 328. Furthermore, it is worth mentioning that, in BigEarthNet all tiles were atmospherically corrected, unlike the Eurosat that contains images with a color cast due to atmospheric effects [6] [2].

The UCMerced and the BigEarthNet models with all the spectral bands performed below the benchmark with 94.75% and 94.06% respectively in a 64 batch size. Although initial CNN layers for dimensionality reduction is well-suited for processing multiband remote-sensing data [7] and several studies have found that including all bands outperforms RGB bands [6] [2] it was not the case in this study. This is mainly due to fact that the BigEarthNet model available was trained only in the RGB representation and not on full multispectral bands datasets.

The poorest performance corresponds to the So2Sat features with 89.22% for a 64 batch size. The dataset used to train this model is multi-modal, meaning that contains images from the synthetic aperture radar and multispectral (Sentinel-1 and Sentinel-2 respectively). The RGB image models might not have benefited completely from this, whereas it is possible that other datasets such as the recently released BigEarthNet-MM [12] might take advantage of this feature.

4. CONCLUSIONS

In the comparison study here presented, we analyzed the performance of 5 models corresponding to a pre-trained architecture of ResNet50 on remote sensing models and the well-known Imagenet. Overall, transfer learning from both general images and remote sensing performed better than training the models from the scratch with random initialization of the parameters according to the accuracy reported in Eurosat.

As a result of this evaluation, the RESISC-45 model, using the RGB band combination outperformed the rest of the models used in this study, with an overall classification accuracy of 97.10% in a 128 batch size and 97.23% in a 64 batch size. Thus, the main conclusion from this work is that for classifying remote sensing datasets it is more relevant to have more diverse representations from remote sensing data than training the models on an all-purpose dataset such as Imagenet.

5. REFERENCES

[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 770–778.

[2] Gencer Sumbul, Marcela Charfuelan, Begüm Demir, and Volker Markl, "Bigearthnet: A Large-Scale benchmark archive for remote sensing image understanding," in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, July 2019, pp. 5901–5904.

[3] Yi Yang and Shawn Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, New York, NY, USA, Nov. 2010, GIS '10, pp. 270–279, Association for Computing Machinery.

[4] Gong Cheng, Junwei Han, and Xiaoqiang Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proceedings of the IEEE*, vol. 105, no. 10, pp. 1865–1883, Oct 2017.

[5] Xiao Xiang Zhu, Jingliang Hu, Chunping Qiu, Yilei Shi, Jian Kang, Lichao Mou, Hossein Bagheri, Matthias Haberle, Yuansheng Hua, Rong Huang, Lloyd Hughes, Hao Li, Yao Sun, Guichen Zhang, Shiyao Han, Michael Schmitt, and Yuanyuan Wang, "So2Sat LCZ42: A benchmark data set for the classification of global local climate zones [software and data sets]," *IEEE Geoscience and Remote Sensing Magazine*, vol. 8, no. 3, pp. 76–89, Sept. 2020.

[6] Patrick Helber, Benjamin Bischke, Andreas Dengel, and Damian Borth, "Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 7, pp. 2217–2226, July 2019.

[7] Lei Ma, Yu Liu, Xueliang Zhang, Yuanxin Ye, Gaofei Yin, and Brian Alan Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 152, pp. 166–177, 2019.

[8] François Chollet et al., "Keras," 2015, Software available from https://keras.io.

[9] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, June 2009, pp. 248–255.

[10] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, Software available from tensorflow.org.

[11] Maxim Neumann, Andre Susano Pinto, Xiaohua Zhai, and Neil Houlsby, "In-domain representation learning for remote sensing," *arXiv preprint arXiv:1911.06721*, Nov. 2019.

[12] Gencer Sumbul, Arne de Wall, Tristan Kreuziger, Filipe Marcelino, Hugo Costa, Pedro Benevides, Mário Caetano, Begüm Demir, and Volker Markl, "Bigearthnet-mm: A large-scale, multimodal, multilabel benchmark archive for remote sensing image classification and retrieval [software and data sets]," *IEEE Geoscience and Remote Sensing Magazine*, vol. 9, no. 3, pp. 174–180, Sep. 2021.