Hyperspectral Image Compression and Super-Resolution Using Tensor Decomposition Learning

A. Aidini*†, M. Giannopoulos*†, A. Pentari*†, K. Fotiadou*†, and P. Tsakalides*†
*Department of Computer Science, University of Crete
†Institute of Computer Science, Foundation for Research and Technology Hellas (FORTH)
Heraklion, 70013, Greece

***Abstract*—As the field of remote sensing for Earth Observation is rapidly evolving, there is an increasing demand for developing suitable methods to store and transmit the massive amounts of the generated data. At the same time, as multiple sensors acquire observations with different dimensions, super-resolution methods come into play to unify the framework for upcoming statistical inference tasks. In this paper, we employ a tensor-based structuring of multi-spectral image data and we propose a low-rank tensor completion scheme for efficient image-content compression and recovery. To address the problem of low-resolution imagery, we further provide a robust algorithmic scheme for super-resolving satellite images, followed by a state-of-the-art convolutional neural network architecture serving the classification task of the employed images. Experimental analysis on real-world observations demonstrates the detrimental effects of image compression on classification, an issued successfully addressed by the proposed recovery and super-resolution schemes.**

***Index Terms*—Multi-Spectral Image Classification, Compression, Tensor Unfoldings, Super Resolution, Alternating Direction Method of Multipliers**

I. Introduction

Remotely sensed images have been widely used in various Earth Observation applications, such as environmental monitoring, resource exploration, and disaster warning. Specifically, airborne and spaceborne Multispectral (MS) and Hyperspectral (HS) sensors can observe an extended region of the electromagnetic spectrum, providing valuable information that is utilized in order to detect and classify objects on Earth. Several Machine Learning (ML) algorithms have been used for land cover classification tasks [1], [2]. Particularly, Deep Learning (DL) architectures, like Convolutional Neural Networks (CNNs), train huge multi-layer networks ending up with impressive classification performance [3], [4].

However, the observed images generally have low-resolution due to the limitations of spaceborne imaging equipment and communication bandwidth that cannot meet the requirements of real satellite image analysis. Super-resolution (SR) techniques can overcome the above limitations, improving the spatial resolution of the images [5], [6].

In addition, satellite MS images that can be modeled as high dimensional data structures, known as tensors, introduce considerable challenges in terms of data storage and data transfer. Therefore, compression of the acquired MS images is of paramount importance in order to reduce bandwidth and increase system lifetime. To achieve higher compression rates, lossy compression algorithms can be employed, primarily by mapping input values from a large set to output values in a smaller set, a process known as quantization. However, compression affects the subsequent processing of MS imagery, even at low compression rates, as it is reported in [7], [8]. In order to improve classification accuracy, a tensor recovery algorithm can be applied. Although numerous algorithms have been presented for the recovery of the real-valued entries of a matrix from its quantized measurements [9], [10], no prior work has been presented for the recovery of a tensor from multi-level quantized observations, except for the case of binary measurements where methods based on various matricizations or tensor decomposition have been presented [11], [12].

In this paper, we examine the effect of compression and super-resolution on the classification task of satellite MS images, using CNNs. Specifically, a DL architecture is used to learn spatial features, and a Coupled Dictionary Learning (CDL) approach is proposed to obtain high spatial-resolution hypercubes from their low resolution acquired versions, using the Alternating Direction Method of Multipliers (ADMM). Furthermore, a constraint maximum likelihood estimation is presented for the recovery of the real-valued MS images from their quantized and possible corrupted entries. Experimental results on a recently released MS image dataset demonstrate that the super-resolution and the recovery of the acquired MS images are mandatory for an efficient classification, indicating the efficacy of the proposed method.

II. Proposed Method

Although classification of pristine high-resolution MS imagery is a well-studied problem, quite frequent is the case when compressed low-resolution MS images arise in practise, as illustrated in Figure 1.

Since full data transmission introduces considerable challenges in terms of data storage and data transfer, quantization plays an important role in the data acquisition pipeline of Remote Sensing (RS) systems. At the same time, super-resolution aims at tackling a fundamental problem being frequently the

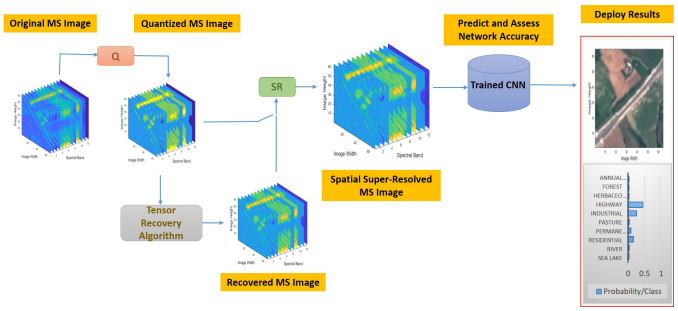

Fig. 1. Proposed pipeline for the classification task of compressed low-resolution MS images.

case in practice, where a deployed instrument measures image data in lower resolution than a ML model is trained with.

The naive straightforward approach for the classification task is to operate directly on the quantized super-resolved MS images, using a CNN classifier trained with pristine MS images. However, instead of classifying these images, we propose to apply a tensor-based recovery algorithm for the reconstruction of the real-valued images before the subsequent super-resolution and classification tasks take place. Experimental results on real satellite MS images demonstrate that direct processing using the quantized measurements, rather than recovering their real values, ends up with a significantly higher error, indicating the efficacy of the proposed approach.

A. Multispectral Image Prediction Modelling

The problem at hand can be casted as a classification-flavoured supervised learning one, tackled by employing second-order CNNs for efficient spatial feature learning of the image content. Instead of creating a network from scratch, in this study we focus our attention on architectures that have already learned a rich set of features, and behave remarkably well in practise for a *similar* task to the problem at hand (RGB instead of MS image classification). Due to the different nature of the problem though, fruitful intuition can be gained by obtaining only their architecture and not their weights as well. Based on this remark, the architecture of the CNN used throughout this paper is the one proposed in [13], inspired by an innovative work in image recognition.

Regarding the chosen residual architecture, we selected the ResNet-50 model. The network is 50 layers deep, comprising of about 25.6 million parameters (i.e. weights). In terms of the training process followed, as long as our intention is to classify MS images into significantly fewer than 1000 classes (for which ResNet-50 model is originally trained), we replaced the classification layers for the new MS image classification task and trained the network on the available *pristine* MS dataset. Assessment of the network's performance was conducted in different imagery occurring as test data (e.g. compressed, super-resolved, etc.).

As far as the optimization process during the training of the network is concerned, we employed Adagrad with a learning rate of 0.001. The latter one was selected (instead of the standard SGD one) due to its adaptive selection of increasing the learning rate of parameters which have small or infrequent updates and, conversely, reducing the learning rate of parameters with high gradients. Since the cross-entropy can be interpreted as the log-likelihood function of the training samples, categorical cross-entropy was used as the loss function, recasting in this way the network's outputs as probabilities.

B. Quantized Tensor Recovery

Let $\mathcal{M} \in \mathbb{R}^{I_1 \times .. \times I_N}$ be the N-th order tensor that models a high dimensional signal, like a multispectral image, whereby the order of the tensor is called the number of its dimensions, also known as ways or modes. Then, the quantized and corrupted observations can be modeled as

$$\mathcal{Y}_{i_1 \ldots i_N} = Q(\mathcal{M}_{i_1 \ldots i_N} + \epsilon_{i_1 \ldots i_N}), \tag{1}$$

where $\epsilon_{i_1..i_N}$ model the uncertainty on each measurement that follow either the logistic distribution with zero mean and unit scale, Logistic(0,1), or the standard normal distribution $\mathcal{N}(0,1)$. The function $\mathcal{Q}$ corresponds to a uniform quantizer, where the set of quantization bin boundaries is known a priori.

In terms of the likelihood of the observations $\mathcal{Y}_{i_1 \ldots i_N}$, the model in (1) can be written equivalently as

$$P(\mathcal{Y}_{i_1..i_N} | \mathcal{M}_{i_1..i_N}) = \Phi(\mathcal{U}_{i_1..i_N} - \mathcal{M}_{i_1..i_N}) - \Phi(\mathcal{L}_{i_1..i_N} - \mathcal{M}_{i_1..i_N}), \tag{2}$$

where $\mathcal{U}, \mathcal{L}$ contain the upper and lower bin boundaries of the measurements, and $\Phi(x)$ corresponds to an inverse link function ($\Phi_{\text{log}}(x) = \frac{1}{1+e^{-x}}$ for the logistic model and $\Phi_{\text{pro}}(x) = \int_{-\infty}^{x} \mathcal{N}(s \mid 0,1)\, ds$ for the probit model).

In order to recover the real values of the low-rank tensor $\mathcal{M}$ from its quantized observations, we unfold the measurement tensor $\mathcal{Y}$ along each mode and we obtain the recovered matrices $\mathbf{Z}_{(n)}, n = 1, ., N$ applying the following algorithm. Formally, the mode-n matricization or unfolding of $\mathcal{Y}$, $\text{unfold}_n(\mathcal{Y}) = \mathbf{Y}_{(n)} \in \mathbb{R}^{I_n \times \prod_{j \neq n} I_j}$, corresponds to a matrix with columns being the vectors obtained by fixing all indices of $\mathcal{Y}$ except the n-th index. Then, the recovered tensor can be estimated as the weighted sum of the folded recovered matrices, i.e., $\mathcal{M} \approx \sum_{n=1}^{N} a_n \cdot \text{fold}_n(\mathbf{Z}_{(n)})$ with weights a_n which depend on the fitting error and satisfy $\sum_n a_n = 1$.

The proposed algorithm can be regarded as an extension of the quantized matrix recovery [10] to the quantized tensor recovery. In particular, in order to recover the low-rank mode-n matricization $\mathbf{M}_{(n)}$ from its quantized measurements, one seeks to solve the constrained optimization problem:

$$\begin{aligned} \text{minimize}_{\ \mathbf{M}_{(n)}} \quad & -\sum_{j,k} \log P(\mathbf{Y}_{(n)jk} \mid \mathbf{M}_{(n)jk}) \\ \text{subject to} \quad & \|\mathbf{M}_{(n)}\|_* \leq \lambda. \end{aligned} \tag{3}$$

with respect to the nuclear norm constraint $\|\mathbf{M}_{(n)}\|_* \leq \lambda$ that promotes low-rankness on $\mathbf{M}_{(n)}$.

Starting with an initialization of the estimated matrix $\mathbf{Z}_{(n)}$ as a random matrix with entries between the corresponding quantization bin boundaries, the algorithm that solves the problem in (3) performs two steps at each iteration l. Both steps are repeated until a maximum number of iterations is

reached or the change in $\mathbf{Z}_{(n)}$ between consecutive iterations is below a given threshold.

The first step that aims at reducing the objective function $F_{\mathbf{Y}_{(n)}}(\mathbf{Z}_{(n)})$ of (3), is given by $\hat{\mathbf{Z}}_{(n)}^{l+1} \leftarrow \mathbf{Z}_{(n)}^{l} - \frac{c}{\sqrt{l}} \cdot \nabla F_{\mathbf{Y}_{(n)}}$, where c is the step-size ($c_{\log} = 4$ and $c_{\text{pro}} = 1$). The gradient of the objective function of each measurement is given by

$$[\nabla F_{\mathbf{Y}_{(n)}}]_{jk} = \frac{\Phi^{'}(\mathbf{L}_{(n)_{jk}} - \mathbf{Z}_{(n)_{jk}}) - \Phi^{'}(\mathbf{U}_{(n)_{jk}} - \mathbf{Z}_{(n)_{jk}})}{\Phi(\mathbf{U}_{(n)_{jk}} - \mathbf{Z}_{(n)_{jk}}) - \Phi(\mathbf{L}_{(n)_{jk}} - \mathbf{Z}_{(n)_{jk}})}, \tag{4}$$

where $\mathbf{L}_{(n)}, \mathbf{U}_{(n)}$ are the mode-n matricizations of $\mathcal{L}$ and $\mathcal{U}$, and the derivative of $\Phi(x)$ can be calculated as $\Phi^{'}_{\log}(x) = \frac{1}{2+e^{-x}+e^{x}}$ or $\Phi^{'}_{\text{pro}}(x) = \mathcal{N}(x \mid 0, 1)$ for each model. The second step aims to impose low-rankness on $\mathbf{Z}_{(n)}$ to make the solution satisfy the constraint $\|\mathbf{Z}_{(n)}\|_* \leq \lambda$. In order to achieve this, we take the SVD of $\hat{\mathbf{Z}}_{(n)}^{l+1}$ of the previous step, and we hold some of its singular values, depending on the parameter λ. Specifically, we keep 97% of the information of the singular values in our experiments.

C. Super-Resolution via Coupled Dictionary Learning

1) Spatial Super-Resolution: The proposed approach synthesizes high-spatial resolution hypercubes from their low resolution acquired versions by capitalizing on the *Sparse Representations* learning framework [14]. According to this framework, image examples extracted from multispectral scenes, can be represented as sparse linear combinations of elements from learned over-complete dictionaries. Each low resolution hyper-pixel $\mathbf{s}_l \in \mathbb{R}^P$ can thus be expressed as a sparse linear combination of elements from a dictionary matrix, $\mathbf{D}_l \in \mathbb{R}^{P \times N}$, composed of hyper-pixel atoms from low resolution training data-cubes, according to $\mathbf{s}_l = \mathbf{D}_l \mathbf{w}$, where $\mathbf{w} \in \mathbb{R}^N$. Recovery of the sparse coding vector $\mathbf{w}$ is accomplished by solving the l_1-minimization problem

$$\min_{\mathbf{w}} ||\mathbf{w}||_1 \quad \text{subject to} \quad ||\mathbf{s}_l - \mathbf{D}_l \mathbf{w}||_2^2 < \epsilon, \tag{5}$$

where the constraint of the ℓ_1-norm, leads to robust solutions and efficient optimization.

To obtain the high-resolution signal, the optimal sparse code $\mathbf{w}^\star$ from (5), is directly mapped onto the high resolution dictionary $\mathbf{D}_h \in \mathbb{R}^{M \times N}$, to synthesize the high resolution image region, according to $\mathbf{s}_h = \mathbf{D}_h \mathbf{w}^\star$. The concatenation of all the recovered high-spectral resolution hyper-pixels synthesizes the high-spatial resolution three-dimensional hypercube, as shown in Fig. 2.

2) ADMM for Coupled Dictionary Learning: CDL refers to the problem of identifying two dictionary matrices standing for two different signal representations, for instance high-spectral and low-spectral resolution hypercubes [6], [15]. In this paper we propose an interesting perspective towards this challenging problem. The proposed CDL algorithm relies on generating coupled dictionaries which jointly encode two coupled feature spaces, namely, the observation low-resolution $\mathbf{S}_l \in \mathbb{R}^{P \times K}$, and the latent high-resolution $\mathbf{S}_h \in \mathbb{R}^{M \times K}$. Formally, the ideal pair of coupled dictionaries $\mathbf{D}_l$ and $\mathbf{D}_h$ can be estimated by

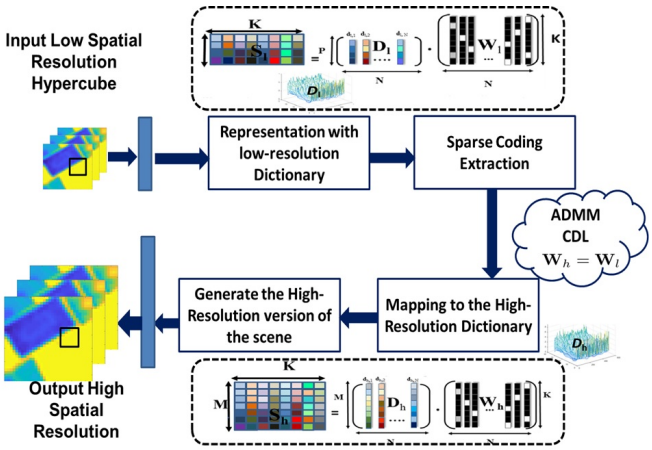

Fig. 2. Proposed-Method-Super-Resolution.

solving the following un-constrainted decomposition problem, using the ADMM [16], as:

$$\begin{aligned} &\underset{\mathbf{D}_h, \mathbf{W}_h, \mathbf{D}_\ell, \mathbf{W}_\ell}{\operatorname{argmin}} ||\mathbf{S}_h - \mathbf{D}_h \mathbf{W}_h||_F^2 + ||\mathbf{S}_\ell - \mathbf{D}_\ell \mathbf{W}_\ell||_F^2 \\ &+ \lambda_\ell ||\mathbf{Q}||_1 + \lambda_h ||\mathbf{P}||_1 \\ &\text{subject to } \mathbf{P} - \mathbf{W}_h = 0, \mathbf{Q} - \mathbf{W}_\ell = 0, \mathbf{W}_h - \mathbf{W}_\ell = 0, \\ &||\mathbf{D}_h(:,i)||_2^2 \leq 1, ||\mathbf{D}_\ell(:,i)||_2^2 \leq 1 \end{aligned} \tag{6}$$

Additionally, we impose the constraint that the two sparse representations $\mathbf{W}_\ell$ and $\mathbf{W}_h$, should be the same, directly into the optimization. The ADMM scheme relies on the minimization of the augmented Lagrangian function:

$$\begin{aligned} &\mathcal{L}(\mathbf{D}_h, \mathbf{D}_\ell, \mathbf{W}_h, \mathbf{W}_\ell, \mathbf{P}, \mathbf{Q}, Y_1, Y_2, Y_3) = \frac{1}{2}||\mathbf{D}_h \mathbf{W}_h - \mathbf{S}_h||_F^2 + \\ &\frac{1}{2}||\mathbf{D}_\ell \mathbf{W}_\ell - \mathbf{S}_\ell||_F^2 + \lambda_h ||\mathbf{P}||_1 + \lambda_\ell ||\mathbf{Q}||_1 + < Y_1, \mathbf{P} - \mathbf{W}_h > \\ &+ < Y_2, \mathbf{Q} - \mathbf{W}_\ell > + < Y_3, \mathbf{W}_h - \mathbf{W}_\ell > + \frac{c_1}{2}||\mathbf{P} - \mathbf{W}_h||_F^2 + \\ &\frac{c_2}{2}||\mathbf{Q} - \mathbf{W}_\ell||_F^2 + \frac{c_3}{2}||\mathbf{W}_h - \mathbf{W}_\ell||_F^2 \end{aligned} \tag{7}$$

where $\mathbf{Y}_1$, $\mathbf{Y}_2$, and $\mathbf{Y}_3$ stand for the Lagrange multiplier matrices, while $c_1 > 0$, $c_2 > 0$, and $c_3 > 0$ denote the step size parameters. Following the general algorithmic strategy of the ADMM scheme, we seek for the stationary point solving iteratively for each one of the variables while keeping the others fixed. The overall algorithm for learning the coupled dictionaries can be found in [6].

III. Experimental Data and Results

In order to evaluate the proposed scheme, we report on the EUROSAT MS image dataset [17], [18], comprising Sentinel 2A MS image patches measuring 64×64 pixels. Each one of them corresponds to spatial resolution of 10 meters/pixel, across 13 different spectral bands (in the 443nm-2190nm wavelength range), ending up with several thousands of sample satellite images for each one of the 10 different classes gathered in total.

Aiming to define two non-overlapping training-test sets for our study, while at the same time avoiding class imbalance problems, we employed 20000 MS sample images from the EUROSAT dataset (assuming that each class consists of 2000

samples). We split each class into 90%-5% training/test sets, keeping at the same time another 5% of each class for validation purposes during the training process of the network. In an attempt to observe how the prediction accuracy changes by increasing the number of epochs the network is trained for, training was performed using a constant number of up to 100 epochs.

To quantify the performance of the proposed system from a ML perspective, we report the obtained *Classification Accuracy* metric. However, in order to further examine the effect of the quantization on the classification task, we quantized the test MS image samples (i.e. 100 of each class, 1000 in total) of our dataset to various numbers of bits (lower than the nominal ones of the original MS images, which employ 12 bits per pixel). Attempting to evaluate the effects of super-resolution in the classification process, we further down-sampled the quantized test MS images by a factor of 2 across each spatial dimension, and super-resolved them right away via the proposed method (SR-CDL) as well as the *Bicubic Interpolation* (SR-BI) one. Then, we compared the classification performance given the pristine, the quantized & super-resolved and the recovered & super-resolved images-obtained via the proposed super-resolution and recovery algorithms. The reconstruction quality is evaluated by computing the *Peak-Signal-to-Noise-Ratio* (PSNR) between the pristine and the estimated MS images.

A. Effect of Super-Resolution on Low-Resolution MS Imagery

In a first set of experiments, we aim at quantifying the effect of super-resolving low-resolution MS images. To do so, we employ the proposed SR-CDL method as well as the baseline SR-BI one, in order to up-sample previously down-sampled pristine MS images.

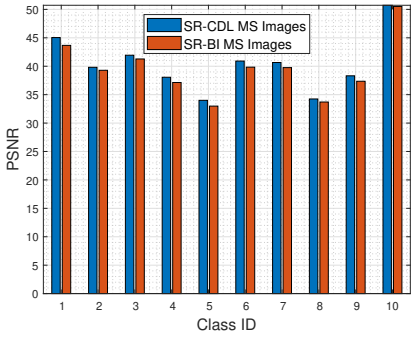

Fig. 3. Mean PSNR w.r.t. different class.

Figure 3 depicts the obtained mean PSNR value for each class, by super-resolving 100 test samples intended to be used for testing purposes of the trained CNN. For every class of the employed dataset, the proposed SR-CDL method outperforms the conventional SR-BI by a margin of approximately 1 dB, indicating the efficacy of the followed approach. Based on this argument, in the rest of this study, super-resolution is performed via our method in order to assess its contribution in the context of the classification task described in Fig. 1.

B. Effect of Training Set Size & Super-Resolution on Classification

In this set of experiments, our objective is to assess the performance of the employed CNN relative to the size of the training set, for each of the employed methods for super-resolution. A random selection of up to 200 training examples per class was initially considered, a number which was augmented by 200 at each step until containing all available training examples of each class (i.e. 2000 per class). The number of test examples was 100 per class (i.e. 1000 in total), as mentioned before.

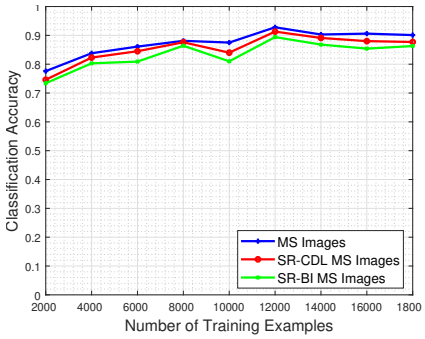

Fig. 4. Classification accuracy w.r.t. the number of training examples.

Figure 4 illustrates the performance of the CNN as a function of the number of the training examples. The results in Fig. 4 demonstrate that increasing the number of training examples has a positive effect on the generalization capacity of the CNN, both when tested on pristine MS images or on super-resolved ones. It should be noted that the CNN classifier achieves better performance in all cases when applied on super-resolved MS images via the proposed SR-CDL method as compared to the conventional SR-BI ones.

C. Effect of Quantization & Super-Resolution on Classification

In a third set of experiments, we investigate the destructive effects of the quantization process on the classification potential of a DL system, and whether these effects can be alleviated via super-resolution. For that purpose, the trained CNN model is evaluated to a test set that has been previously quantized to a specific number of bits and super-resolved right away via the proposed SR-CDL method, in order to obtain intuition about the system response when faced with corrupted image samples.

Figure 5 shows the performance with respect to the number of training examples for different quantization bit values imposed on the test set image samples, in the case of adopting the proposed SR-CDL method for super-resolving the quantized measurements. The results clearly demonstrate that, whatever the quantization level, the classification performance severely deteriorates even if the system is trained with the maximum permitted number of training examples.

D. Effect of Recovery & Super-Resolution on Classification

The objective of this experiment is to further investigate the enhancing effect that our proposed recovery and super-

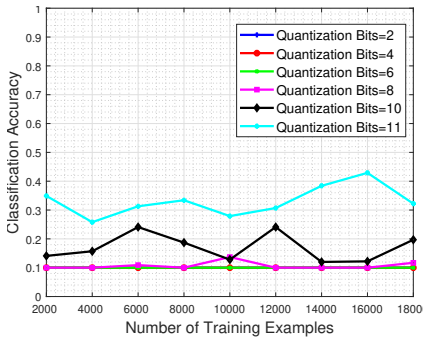

Fig. 5. Classification accuracy w.r.t. the number of training examples, for several quantization levels, followed by the SR-CDL method for super-resolution.

resolution processes have in the classification performance of the system. Towards that goal, the trained CNN model is evaluated using a recovered by our approach test set (which was previously quantized to a specific number of bits), which is super-resolved via the proposed SR-CDL method.

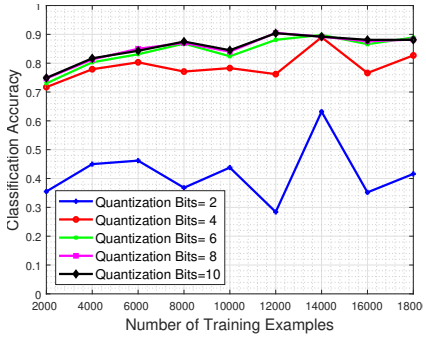

Fig. 6. Classification accuracy w.r.t. the number of training examples, for various levels of quantized images subsequently recovered and super-resolved using the proposed methods.

Figure 6 depicts the system performance as a function of the number of training examples, for test images quantized to 2, 4, 6, 8, and 10 bits, subsequently recovered and super-resolved using our proposed methods. The results show that our approach clearly ameliorates the classification accuracy of the system, even when it faces recovered and super-resolved MS images that were previously quantized to significantly fewer bits than the nominal ones.

IV. Conclusion

In this work, we proposed a DL architecture for the problem of classifying MS image data, based on residual networks for learning efficient spatial feature representations. Furthermore, we designed, tested and evaluated with real-world data two novel techniques to address the real life scenarios of dealing with quantized and low-resolution MS imagery. Based on our experimental findings on a recently released MS image dataset, we demonstrated that improved classification accuracy is feasible, even when data observations are quantized to one-third of their nominal bits-per-pixel, using state-of-the-art CNNs. Clearly, adopting a tensor recovery algorithm followed by a super-resolution one proved to be a good strategy for efficient classification of previously quantized, as well as low-resolution MS data, demonstrating the efficacy of the proposed approach.

V. Acknowledgement

The research work was funded by Greece and the European Union (European Social Fund) in the context of the Youth Employment Initiative through the Operational Program for Human Resources Development, Education and Lifelong Learning, under grant no. MIS 5004457.

References

[1] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: A review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 3, pp. 247–259, 2011.

[2] M. Belgiu and L. Drăguţ, "Random forest in remote sensing: A review of applications and future directions," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 114, pp. 24–31, 2016.

[3] M. Castelluccio, G. Poggi, C. Sansone, and L. Verdoliva, "Land use classification in remote sensing images by convolutional neural networks," *arXiv preprint arXiv:1508.00092*, 2015.

[4] G. Tsagkatakis, A. Aidini, K. Fotiadou, M. Giannopoulos, A. Pentari, and P. Tsakalides, "Survey of deep-learning approaches for remote sensing observation enhancement," *Sensors*, vol. 19, no. 18, p. 3929, 2019.

[5] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, vol. 1. IEEE, 2004, pp. I–I.

[6] K. Fotiadou, G. Tsagkatakis, and P. Tsakalides, "Spectral super resolution of hyperspectral images via coupled dictionary learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 5, pp. 2777–2797, 2018.

[7] F. Garcia-Vilchez, J. Muñoz-Marí, M. Zortea, I. Blanes, V. González-Ruiz, G. Camps-Valls, A. Plaza, and J. Serra-Sagristà, "On the impact of lossy compression on hyperspectral image classification and unmixing," *IEEE Geoscience and remote sensing letters*, vol. 8, no. 2, pp. 253–257, 2010.

[8] A. Marsetic, Z. Kokalj, and K. Ostir, "The effect of lossy image compression on object based image classification-worldview-2 case study," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 38, no. 4/W19, 2011.

[9] M. A. Davenport, Y. Plan, E. Van Den Berg, and M. Wootters, "1-bit matrix completion," *Information and Inference: A Journal of the IMA*, vol. 3, no. 3, pp. 189–223, 2014.

[10] A. S. Lan, C. Studer, and R. G. Baraniuk, "Matrix recovery from quantized and corrupted measurements," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 4973–4977.

[11] A. Aidini, G. Tsagkatakis, and P. Tsakalides, "1-bit tensor completion," *Electronic Imaging*, vol. 2018, no. 13, pp. 261–1, 2018.

[12] N. Ghadermarzy, Y. Plan, and O. Yilmaz, "Learning tensors from partial binary measurements," *IEEE Transactions on Signal Processing*, vol. 67, no. 1, pp. 29–40, 2018.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[14] M. Elad, "Prologue," in *Sparse and Redundant Representations*. Springer, 2010, pp. 3–15.

[15] K. Fotiadou, G. Tsagkatakis, and P. Tsakalides, "Spectral resolution enhancement of hyperspectral images via sparse representations," in *Proceedings) Proc. Computational Imaging, IS&T Int. Symposium on Electronic Imaging*, 2016.

[16] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[17] P. Helber, B. Bischke, A. Dengel, and D. Borth, "Introducing eurosat: A novel dataset and deep learning benchmark for land use and land cover classification," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2018, pp. 204–207.

[18] ——, "Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification," *arXiv preprint arXiv:1709.00029*, 2017.