

Noise-Tolerant Hyperspectral Image Classification Using Discrete Cosine Transform and Convolutional Neural Networks

Conference Paper**Author(s):**

Voulodimos, Athanasios; Fokeas, K.; Doulamis, Nikolaos D.; Doulamis, Anastasios; Makantasis, Konstantinos

Publication date:

2020-08-14

Permanent link:

<https://doi.org/10.3929/ethz-b-000442465>

Rights / license:

[Creative Commons Attribution 4.0 International](#)

Originally published in:

International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLIII, <https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-1281-2020>

NOISE-TOLERANT HYPERSPECTRAL IMAGE CLASSIFICATION USING DISCRETE COSINE TRANSFORM AND CONVOLUTIONAL NEURAL NETWORKS

A. Voulodimos^{1*}, K. Fokeas², N. Doulamis³, A. Doulamis³, K. Makantasis⁴

¹ Department of Informatics and Computer Engineering, University of West Attica, Athens 12243, Greece – avoulod@uniwa.gr

² Department of Civil, Environmental and Geomatic Engineering, ETH Zurich, Zurich, Switzerland

³ School of Rural and Surveying Engineering, National Technical University of Athens, Athens 15773, Greece – {ndoulam, adoulam}@cs.ntua.gr

⁴ Institute of Digital Games, University of Malta, Msida, Malta - konstantinos.makantasis@um.edu.mt

KEY WORDS: Hyperspectral image classification, robustness to noise, Convolutional Neural Networks, Discrete Cosine Transform, noise tolerance, deep learning.

ABSTRACT:

Hyperspectral image classification has drawn significant attention in the recent years driven by the increasing abundance of sensor-generated hyper- and multi-spectral data, combined with the rapid advancements in the field of machine learning. A vast range of techniques, especially involving deep learning models, have been proposed attaining high levels of classification accuracy. However, many of these approaches significantly deteriorate in performance in the presence of noise in the hyperspectral data. In this paper, we propose a new model that effectively addresses the challenge of noise residing in hyperspectral images. The proposed model, which will be called DCT-CNN, combines the representational power of Convolutional Neural Networks with the noise elimination capabilities introduced by frequency-domain filtering enabled through the Discrete Cosine Transform. In particular, the proposed method entails the transformation of pixel macroblocks to the frequency domain and the discarding of information that corresponds to the higher frequencies in every patch, in which pixel information of abrupt changes and noise often resides. Experiment results in Indian Pines, Salinas and Pavia University datasets indicate that the proposed DCT-CNN constitutes a promising new model for accurate hyperspectral image classification offering robustness to different types of noise, such as Gaussian and salt and pepper noise.

1. INTRODUCTION

Hyperspectral imaging systems have been increasingly used in a wide area of applications including agriculture (Sahu et. al., 2019), surveillance (Freitas et. al., 2019), biomedical imaging (Offerhaus, Bohndiek & Harvey, 2019), and astronomy (Hege et. al., 2004). This is due to the fact that they efficiently combine the spatial information with the subtle differences in spectral signatures of various objects, making them valuable tools in tasks relevant to material detection and object recognition. These tasks can be viewed as classification problems, i.e. classify image pixels based on their spectral characteristics in order to recognize various materials, and objects.

The high dimensionality and complexity of hyperspectral information (HSI), makes classification an arduous task, that requires special learning architectures to tackle them. This prohibited the use of shallow neural network architectures, such as Feedforward Neural Networks, since the complexity of the data could result in the accumulation of errors, especially when the derivative is in the saturation regions, where small error changes affect the weights significantly. This was evident even when trying to classify vision based data with fewer dimensions than HSI, like RGB or Thermal, narrow-spectrum data.

The advent of deep learning brought a new era to hyperspectral image classification. Architectures such as Convolutional Neural Networks (CNN) tackled the aforementioned challenges by using

a number of convolution steps, extracting representative features at different hierarchies of resolutions from the input data. These features are forwarded to neural network architectures that drive the classification step. The feature extraction capabilities of CNN have significantly enhanced the performance of classifying visual based data.

However, the majority of these approaches significantly deteriorate in performance in the presence of noise in the hyperspectral data, thus indicating insufficient robustness to noise and outliers. In this paper, we propose a new model that effectively addresses the challenge of noise residing in hyperspectral images. The proposed model, which will be called DCT-CNN, combines the representational power of Convolutional Neural Networks with the noise elimination capabilities introduced by frequency-domain filtering enabled through the Discrete Cosine Transform.

1.1 Related Work

Earlier works on classification of HSI, consist of two main steps, the computation of hand-crafted features from the raw data and the use of such features to train classifiers, such as Support Vector Machines (SVM) and Neural Networks (NN) (Camps-Valls & Bruzzone, 2009). Camps-Valls et. al. (2014) employs statistical learning methods to classify high-dimensional data, with a few training samples available. These approaches however

* Corresponding author

assume a priori knowledge of the important features for the classification, which is usually not-known.

More efficient approaches employ the paradigm of pattern recognition and machine learning (Lecun et. al., 1998, Hinton and Salakhutdinov, 2006, Hinton et. al. 2006, Bengio et. al., 2007), that automate the feature extraction step, by building high-level features from low-level ones. Moreover, when in possession of bigger training sets, such as large images with very high spatial and spectral resolutions, deep learning approaches seem more adequate for the classification problem (Chen et. al., 2014). Techniques based on deep learning have already been shown promising results for the detection (classification) of specific objects (Han et. al., 2018), materials (Mnih and Hinton, 2012), vehicles (Montavon et. al. 2012), actions (Bakalos et. al., 2019a), and HSI (Chen et. al., 2014). Chen et. al., (2014) utilises a greedy layer based training framework (Bengio et. al., 2007) and Autoencoders, which are used as the basic building block for spectral feature extraction that are in turn combined with spatial information and fed to a logistic regression classifier. However, it is necessary to employ various adaptations of simple deep learning techniques in classifying high dimensional data. This is evident even in application scenarios where the dimensionality of data is much lower than HSI such as Bakalos et. al., (2019b).

Typical deep learning architectures contain a high number of tunable parameters implying that a large number of samples is also needed to accurately train the network, and high complexity is introduced in the training process. Furthermore, such approaches are often not sufficiently tolerant to the presence of noise within the hyperspectral information. The proposed approach in this paper introduces a pre-processing step based on the Discrete Cosine Transform in a CNN framework for HSI.

2. THE PROPOSED DCT-CNN FOR NOISE-TOLERANT HYPERSPECTRAL IMAGE CLASSIFICATION

The proposed model, DCT-CNN, combines the representational power of Convolutional Neural Networks in pixel-level classification of HSI with the noise elimination capabilities

introduced by frequency-domain filtering enabled through the Discrete Cosine Transform. In particular, the proposed method entails the transformation of pixel macroblocks to the frequency domain and the discarding of information that corresponds to the higher frequencies in every patch, in which pixel information of abrupt changes and noise often resides. We hereby present the steps of our proposed technique: we first describe the plain CNN approach without the DCT pre-processing, and the explain the additional DCT step.

The deep learning approach presented in this work employs CNN for a pixel level classification of a hyper-spectral image. However, the training of such architectures requires the convolution of hundreds of channels along the network inputs. It is understood that training of deep CNN architectures requires the convolution of hundreds of channels along the network inputs. This increases the computational costs of training and prediction. Also through a statistical analysis of spectral responses of pixels that belong to the same class, one easily observes that the variance of responses is small. This suggests that pixels that belong to the same class tend to have very similar values at every channel. At the same time, pixels that belong to different classes present different spectral properties.

Based on the above, to achieve the best performance, a pre-processing step that uses a Randomised Principal Component Analysis (R-PCA) is employed. R-PCA is applied among the spectral information, without affecting the spatial information. The number of principal components that are retained after the application of R-PCA, is appropriately set, in order to keep at least 99.9% of initial information. During the experimentation process on widely-used hyperspectral datasets, this amount of information is preserved by using the first 10 to 30 principal components, reducing this way up to 15 times dimensionality of the raw input.

The hyperspectral image can be expressed as a 3D tensor, $p_1 \times p_2 \times p_3$, p_1 and p_2 representing the spatial information and p_3 representing the spectral bands. We analyse the captured hyperspectral images in $s - size$ square patches. After the R-PCA method, in the traditional CNN approach, each patch is a tensor of $s \times s \times C_{PCA}$ size, where parameter C_{PCA} is the number

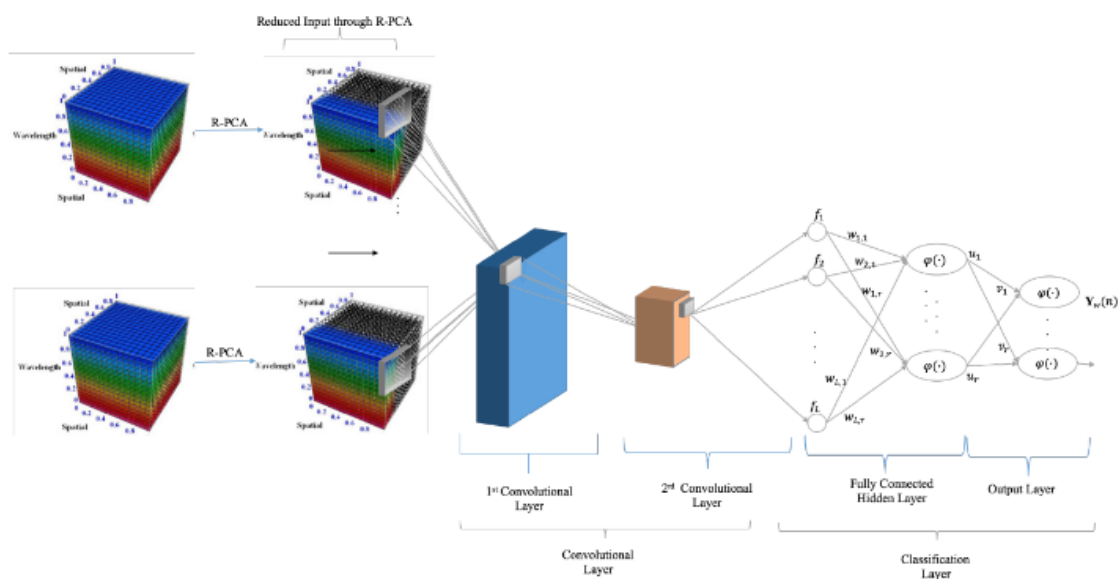


Figure 1: Deep Learning Architecture for HSI

of principal components extracted during the R-PCA method. During the performance evaluation the selected batch size, s can be tried for different values. Increasing the size of s also increases the number of neighbours considered during the classification. Thus, it is evident that increasing the size of s , also increases the algorithms complexity. The selection of the batch-size is a critical step in the training process, as it optimises the accuracy/complexity ratio of the algorithm. Selecting batch-sizes bigger than 13 usually deteriorate the performance of the model. The CNN structure design includes the aforementioned data input and preprocessing (R-PCA) layers. These layers are followed by two convolutional layers of 3×3 and $3 \times C_{PCA}$ size. During training, the standard backpropagation algorithm was employed, i.e. minimizing the negative log-likelihood of the datasets under the model parameters. The CNN architecture can be seen in Figure 1.

When assuming that hyperspectral data include noise, then we leverage the Discrete Cosine Transform. Transforming the information to the frequency domain, the core information is retained in low-frequency coefficients, whereas high-frequency coefficients tend to represent abrupt changes, edges, outliers and noise, hence eliminating this information can potentially provide robustness to noise. Based on this, the noisy hyperspectral data are dissembled into patches of a fixed size (different patch sizes have been investigated in the experimental evaluation, as noted above) and a Discrete Cosine Transform is applied to every patch. Thus, passing to the frequency domain, each patch is now represented by amplitude coefficients of increasingly high frequencies. In our method, the coefficients corresponding to the higher frequencies of each patch are turned into zero. More specifically, in the evaluation we have experimented with various percentages of zeroed coefficient values, ranging from 0% to 64% of the coefficients, always starting from the highest frequency, i.e. from the lower right corner of the patch.

3. EXPERIMENTAL EVALUATION

3.1 Experimental Setup

In our study, we experimented and validated the developed framework with three publicly available datasets. In particular, we employed i) the Indian Pines dataset, which consists of 145 x 145 pixels and 224 spectral reflectance bands in the wavelength range 0.4 to 2.510... meters. ii) the Salinas dataset, which is an 224-band hyperspectral image, characterized by high resolution and iii) the Pavia university dataset, whose number of spectral bands account for 102.

Supervised training was conducted using the ground truth images of aforementioned datasets. In particular, we split the tagged parts of images into two sets, training and testing data, with a split ratio 5:95 for Salinas and Pavia datasets, whilst for the Indian Pines dataset we used a different ratio accounted for 25:75. The different ratio concerned the last-mentioned image, has to do with the high variability of the number of samples between classes. The splitting was done randomly, and each experiment was conducted 10 times and the results demonstrated at the next section, are the average number of them. All models have been quantitatively validated in terms of classification accuracy.

In the experimental evaluation phase, two different types of noise were introduced. Gaussian noise, one with std = 300 and mean = 0 and one with std = 100 and mean = 0. We also applied the well-known Salt and Pepper (SnP) noise. In particular, the overall amount of imported noise was 10% and 80% respectively, with an analogy between salt and pepper of 50:50.

		CNN			SVM-Linear			K-nn		
		No DCT	N = 3	N = 4	No DCT	N = 3	N = 4	No DCT	N = 3	N = 4
Gaussian noise	Patch Size									
	5x5	68.48	77.66	82.03	50.75	53.97	61.36	45.38	70.78	85.82
	7x7	75.89	82.40	88.93	54.90	55.44	56.22	27.34	59.75	73.36
	9x9	86.17	86.76	91.32	61.68	59.78	60.94	23.39	55.61	68.70
	11x11	91.70	90.33	92.05	67.57	70.16	70.69	18.46	52.15	64.13
	13x13	94.08	91.96	93.55	74.01	78.67	79.18	16.17	50.83	59.72
	15x15	94.13	91.92	93.91	80.77	84.56	85.23	15.43	49.42	55.43
Salt and peper amount	Patch Size									
	5x5	40.07	43.45	47.44	20.03	22.95	29.67	18.67	37.31	59.33
	7x7	42.10	42.33	48.52	21.64	25.54	27.40	17.30	42.97	50.92
	9x9	48.39	50.24	55.04	23.38	21.76	25.84	10.81	28.80	47.31
	11x11	53.92	53.91	61.77	24.74	28.41	30.54	8.22	28.00	46.88
	13x13	60.79	55.81	61.03	25.80	37.53	38.60	8.16	22.74	46.68
	15x15	64.54	58.12	60.92	28.75	43.05	46.20	8.20	22.34	46.75

Figure 2: Indian Pines Dataset: Classification accuracy rates for CNN, SVM and k-NN based approaches, for Gaussian and Salt & Pepper noise, for different patch sizes and different numbers of zeroed high-frequency DCT coefficients. Yellow-coloured rates show the best performing configuration per technique, and green-coloured rates the best performing one among all techniques for the given patch size.

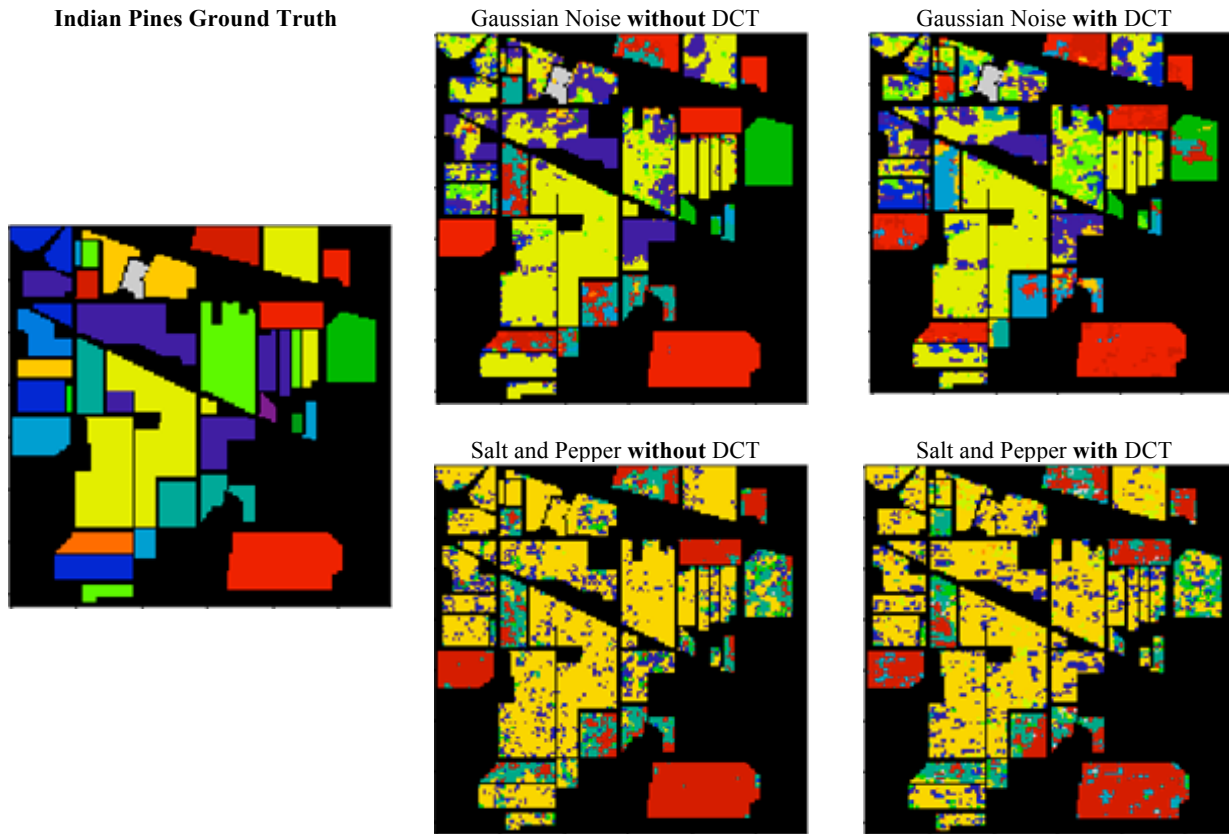
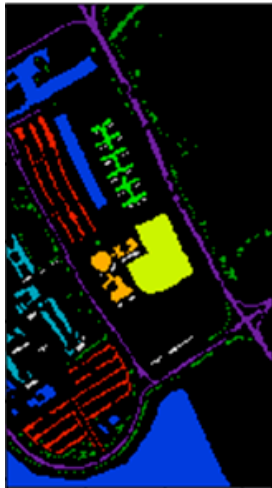


Figure 3: Indian Pines Dataset: Classification maps for indicative CNN-based configurations.

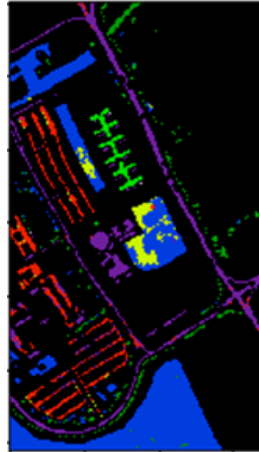
		CNN			SVM-Linear			K-nn		
		No DCT	N = 3	N = 4	No DCT	N = 3	N = 4	No DCT	N = 3	N = 4
Gaussian noise	Patch Size									
	5x5	79.60	86.66	85.83	75.00	74.40	78.61	81.99	84.87	86.95
	7x7	84.60	90.04	88.93	72.42	72.59	74.76	83.76	87.81	89.94
	9x9	88.68	90.36	90.39	75.57	72.88	73.93	78.99	88.79	90.00
	11x11	90.39	94.32	94.29	78.98	75.13	74.76	78.26	83.18	89.06
	13x13	93.39	95.89	95.62	81.88	74.87	75.83	75.71	84.34	85.98
	15x15	94.43	95.41	95.57	83.34	79.50	79.31	76.27	81.36	82.27
Salt and peper	Patch Size									
	5x5	67.35	72.31	72.71	59.45	60.43	62.70	37.99	45.81	62.48
	7x7	71.90	74.66	75.95	49.17	54.57	58.42	31.96	37.16	53.00
	9x9	76.07	77.99	78.29	47.37	49.27	52.72	35.22	37.57	43.86
	11x11	78.62	79.82	80.88	52.01	47.09	49.16	34.83	31.65	38.68
	13x13	80.92	81.98	82.04	54.98	43.75	47.39	30.91	33.39	32.47
	15x15	83.04	83.38	83.09	56.29	46.19	45.18	33.19	28.88	32.44

Figure 4: Pavia University Dataset: Classification accuracy rates for CNN, SVM and k-NN based approaches, for Gaussian and Salt & Pepper noise, for different patch sizes and different numbers of zeroed high-frequency DCT coefficients. Yellow-coloured rates show the best performing configuration per technique, and green-coloured rates the best performing one among all techniques for the given patch size.

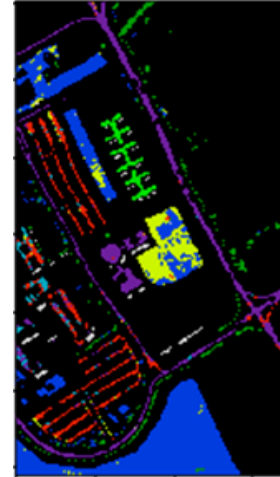
Pavia University Ground Truth



Gaussian Noise **without** DCT



Gaussian Noise **with** DCT



Salt and Pepper **without** DCT



Salt and Pepper **with** DCT

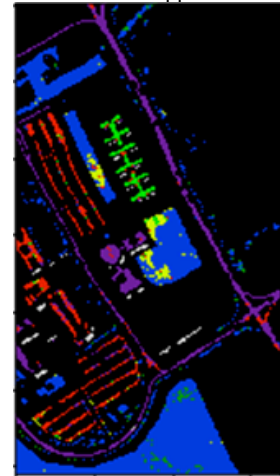


Figure 5: Pavia University Dataset: Classification maps for indicative CNN-based configurations.

		CNN			SVM-Linear			K-nn		
		No DCT	N = 3	N = 4	No DCT	N = 3	N = 4	No DCT	N = 3	N = 4
Gaussian noise	Patch Size									
	5x5	74.34	82.97	82.99	75.57	75.65	76.24	81.56	83.78	85.63
	7x7	80.41	87.58	86.94	72.03	72.68	76.46	84.69	87.34	88.54
	9x9	87.68	91.48	87.86	71.69	70.08	73.68	86.08	89.63	90.91
	11x11	95.78	93.40	93.85	71.95	65.80	69.61	87.04	89.85	91.61
	13x13	96.60	94.70	95.21	73.01	81.74	80.95	88.07	91.53	92.02
15x15	97.08	94.69	95.64	75.94	83.10	83.68	87.76	91.30	92.49	
Salt and peper	Patch Size									
	5x5	58.90	60.15	60.45	40.11	40.41	45.94	16.36	28.20	49.57
	7x7	59.11	63.56	65.17	32.13	35.94	41.05	10.34	20.45	35.39
	9x9	62.55	68.65	66.09	29.46	30.82	30.16	13.01	19.08	28.88
	11x11	66.32	67.91	72.37	30.12	28.90	29.88	13.11	17.40	26.61
	13x13	69.72	72.71	75.56	33.84	37.38	38.02	10.94	15.04	20.10
15x15	77.97	76.31	77.67	36.65	38.64	39.17	12.44	14.84	20.08	

Figure 6: Salinas Dataset: Classification accuracy rates for CNN, SVM and k-NN based approaches, for Gaussian and Salt & Pepper noise, for different patch sizes and different numbers of zeroed high-frequency DCT coefficients. Yellow-coloured rates show the best performing configuration per technique, and green-coloured rates the best performing one among all techniques for the given patch size.

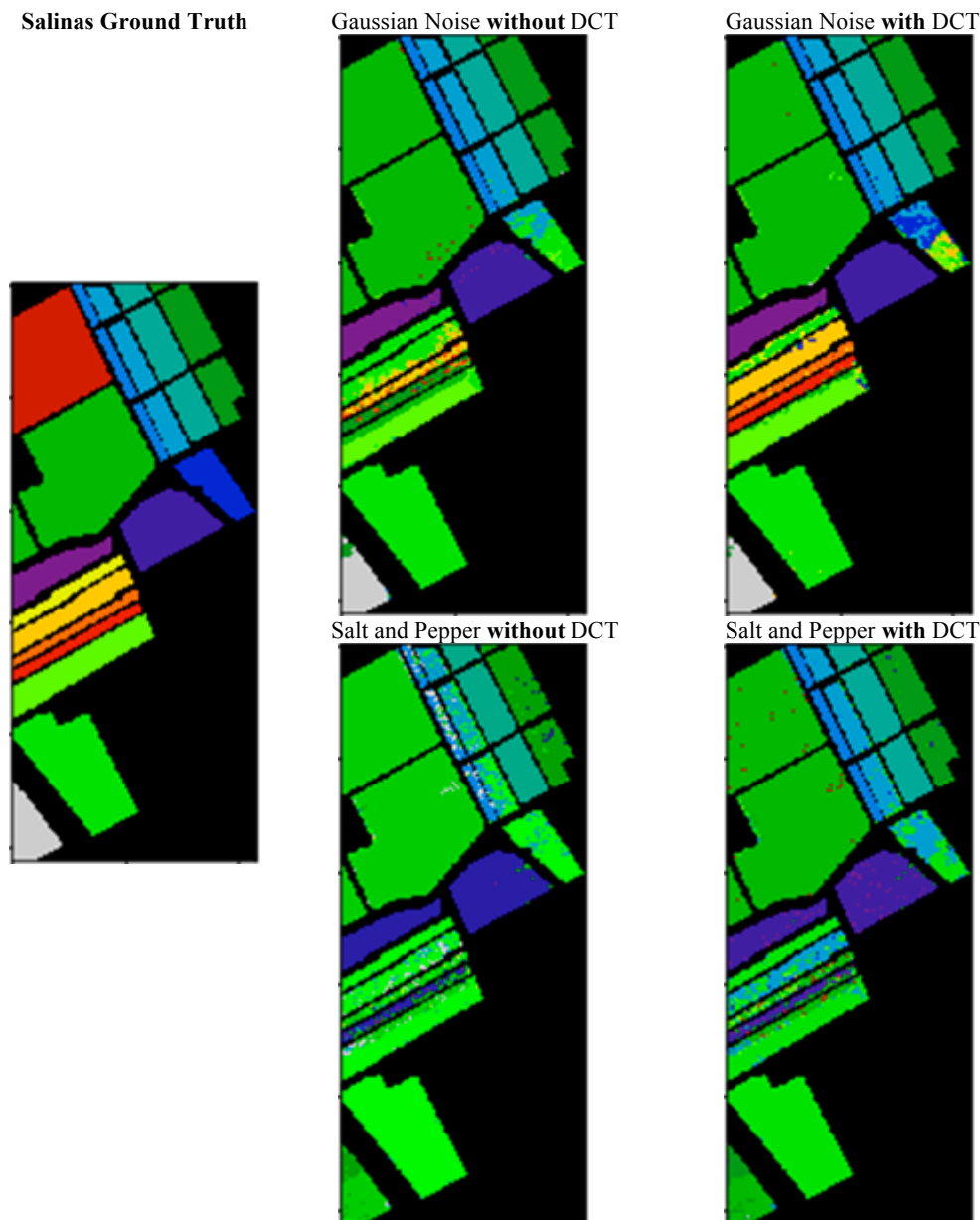


Figure 7: Salinas Dataset: Classification maps for indicative CNN-based configurations.

3.2 Results

Experimental results are provided for various patch sizes, ranging from 5x5 up to 15x15. Furthermore, we have experimented with different ratios of zeroed high-frequency DCT coefficients. The results tables indicate the yielded classification accuracy of CNN, SVM and kNN models, for different patches and different ratios of zeroed high-frequency DCT coefficients (i.e., no coefficients zeroed, a sub-patch of size 3x3 in the lower-right corner zeroed and a sub-patch of size 5x5 in the lower-right corner zeroed). Figures 2, 4, and 6 present the accuracy rates for the Indian Pines, Salinas, and Pavia University datasets respectively, whereas Figures 3, 5, and 7 provide the hyperspectral classification maps for indicative cases of noisy data handled with plain CNN and DCT-CNN.

The results indicate that the approach based on DCT-CNN handles noisy hyperspectral data in a more effective manner than plain CNN approaches. As expected, the noise-tolerance technique based on CNN outperforms similar techniques applied on other learning models such as SVM and k-NN, although in those cases the benefits of DCT for noise robustness can be seen as well. It can also be observed that generally when the patch size increases, a higher number of zeroed high-frequency DCT coefficients tends to provide better performance rates.

4. CONCLUSION

In this paper, we have proposed DCT-CNN, a model that combines the representational power of Convolutional Neural

Networks with the noise elimination capabilities introduced by frequency-domain filtering enabled through the Discrete Cosine Transform. The presented method involves the transformation of pixel macroblocks to the frequency domain and the discarding of information that corresponds to the higher frequencies in every patch, in which pixel information of abrupt changes and noise often resides. Experiment results in Indian Pines, Salinas and Pavia University datasets indicate that the proposed DCT-CNN constitutes a promising new model for accurate hyperspectral image classification offering robustness to different types of noise, such as Gaussian and Salt and Pepper noise. As future directions of our work, we aim at investigating the efficacy of other transformations, such as wavelets, in conjunction with deep learning, as well as tensor-based machine learning which have been shown to require a lower amount of annotated data for training.

ACKNOWLEDGEMENTS

This research has been co-financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (project code: T1EDK-01169).

REFERENCES

- Bakalos, N., Voulodimos, A., Doulamis, N., Doulamis, A., Ostfeld, A., Salomons, E., Caubet, J., Jimenez, V. and Li, P., 2019. Protecting water infrastructure from cyber and physical threats: Using multimodal data fusion and adaptive deep learning to monitor critical systems. *IEEE Signal Processing Magazine*, 36(2), pp.36-48.
- Bakalos, N., Voulodimos, A., Doulamis, A. and Doulamis, N., 2019, May. A Deep-narma Filter for Unusual Behavior Detection from Visual, Thermal and Wireless Signals. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 4005-4009). IEEE.
- Bengio, Y., Lamblin, P., Popovici, D. and Larochelle, H., 2007. Greedy layer-wise training of deep networks. In *Advances in neural information processing systems* (pp. 153-160).
- Camps-Valls, G. and Bruzzone, L. eds., 2009. *Kernel methods for remote sensing data analysis*. John Wiley & Sons.
- Camps-Valls, G., Tuia, D., Bruzzone, L. and Benediktsson, J.A., 2013. *Advances in hyperspectral image classification: Earth monitoring with statistical learning methods*. *IEEE signal processing magazine*, 31(1), pp.45-54.
- Chen, Y., Lin, Z., Zhao, X., Wang, G. and Gu, Y., 2014. Deep learning-based classification of hyperspectral data. *IEEE Journal of Selected topics in applied earth observations and remote sensing*, 7(6), pp.2094-2107.
- Freitas, S., Silva, H., Almeida, J.M. and Silva, E., 2019. Convolutional neural network target detection in hyperspectral imaging for maritime surveillance. *International Journal of Advanced Robotic Systems*, 16(3), p.1729881419842991.
- Han, J., Zhang, D., Cheng, G., Liu, N. and Xu, D., 2018. Advanced deep-learning techniques for salient and category-specific object detection: a survey. *IEEE Signal Processing Magazine*, 35(1), pp.84-100.
- Hege, E.K., O'Connell, D., Johnson, W., Basty, S. and Dereniak, E.L., 2004, January. Hyperspectral imaging for astronomy and space surveillance. In *Imaging Spectrometry IX* (Vol. 5159, pp. 380-391). International Society for Optics and Photonics.
- Hinton, G.E., Osindero, S. and Teh, Y.W., 2006. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7), pp.1527-1554.
- Hinton, G.E. and Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks. *science*, 313(5786), pp.504-507.
- LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P., 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), pp.2278-2324.
- Mnih, V. and Hinton, G.E., 2012. Learning to label aerial images from noisy data. In *Proceedings of the 29th International conference on machine learning (ICML-12)* (pp. 567-574).
- Montavon, G., Orr, G. and Müller, K.R. eds., 2012. *Neural networks: tricks of the trade* (Vol. 7700). springer.
- Offerhaus, H.L., Bohndiek, S.E. and Harvey, A.R., 2019. Hyperspectral imaging in biomedical applications. *Journal of optics*, 21(1).
- Sahu, A., Dante, H., Haase, E. and Morris, J.W., 2019, April. Hyperspectral imaging using CCD imager and broadband energy source for agricultural grading: implementation. In *Wide Bandgap Power and Energy Devices and Applications III* (Vol. 10754, p. 107540G). International Society for Optics and Photonics.
- Tan, X., Zhang, Y., Tang, S., Shao, J., Wu, F. and Zhuang, Y., 2012, October. Logistic tensor regression for classification. In *International Conference on Intelligent Science and Intelligent Data Engineering* (pp. 573-581). Springer, Berlin, Heidelberg.
- Zhou, H., Li, L. and Zhu, H., 2013. Tensor regression with applications in neuroimaging data analysis. *Journal of the American Statistical Association*, 108(502), pp.540-552.