

WAVELET-GUIDED DEEP NEURAL NETWORK FOR ROBUST ONE-CLASS CLASSIFICATION

Omid Ghozatlou¹, Miguel Heredia Conde², Mihai Datcu^{1,3}

¹Research Center for Spatial Information (CEOSpaceTech), University POLITEHNICA of Bucharest (UPB), Bucharest, Romania

²Center for Sensorsystems (ZESS), University of Siegen, Siegen, Germany

³Earth Observation Center (EOC), German Aerospace Center (DLR), Oberpfaffenhofen, Germany

ABSTRACT

This paper aims to provide a deep neural network (DNN) considering the statistical properties of data for robust one-class classification. To achieve that, we take advantage of the properties of Wavelet Scattering Transform (WST) to guide the DNN. WST is a translation-invariant image representation that retains high-frequency information for classification while being stable to rotation. The resulting stable and low-variance features make the clustering of data easier for DNN. The importance of WST in guiding the DNN for the classification of highly textured images is evaluated in terms of accuracy gain and robustness to outlier pollution. Superior robustness to both translation and rotation is also demonstrated. The method is not only evaluated in a standard computer vision dataset (CIFAR10), but the use of largely invariant features allows for coping with the more challenging case of satellite imagery (EuroSAT).

Index Terms— One-class classification, Deep learning, Wavelet scattering transform, Remote sensing, Sentinel-2 imagery.

1. INTRODUCTION

Deep learning has shown its huge potential in the field of image classification. However, most of the deep learning models heavily depend on the quantity of available training samples [1]. In this article, we propose a wavelet-guided deep neural network (DNN) to alleviate this issue by taking advantage of the properties of Wavelet Scattering Transform (WST) to extract invariant features for robust clustering. In the proposed method, the network is trained on samples from only one class in the training set and is evaluated on the test set including all classes. This strategy has the benefit of not requiring labeling for every class in the dataset.

In the first section, we focus on the one-class classification task and discuss the challenges and benefits of some methods

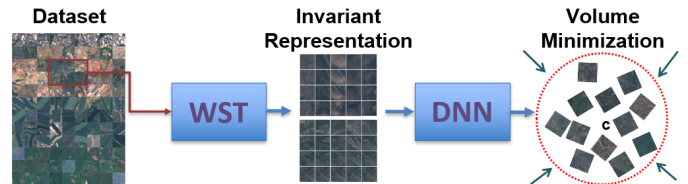


Fig. 1: For specified class of dataset, WST provides stable scattered features to aid the DNN in minimizing the volume of the hyper-sphere in latent space.

used in recent studies. One-class classification methods aim to directly learn a decision boundary with a low error when applied to unseen data [2]. Authors in [3] proposed a novel method, called one-class transfer learning. They took into account that in the field of computer vision, one can access labeled data from different domains that are not related to one-class classification datasets and benefit from using data from a different domain. Similarly, in [4], authors utilize the objective function inspired by information theory, which maximizes the distance between normal and anomalous data in terms of the joint distribution of images and their representation. In [5], authors proposed a novel self-learning technique, called GOAD, for classification-based anomaly detection, which unifies current methods that use only normal training data.

A prominent challenge of this task is the unsupervised nature of the problem. Therefore, unlike supervised deep learning, it is unclear what useful representation learning objectives for deep AD are. On the other hand, a Wavelet Scattering Transform (WST) network has been demonstrated being useful for image classification. This network computes image representations which are stable to rotation and preserves high-frequency information for classification [6]. There are many studies that exploited WST to improve the performance of the model [7]. For instance, in [8] a WST is used to extract reliable features that are stable to small deformation and rotation. The extracted features are used by a deep neural network (DNN) model to predict the location. In addition,

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860370.

authors in [9] proposed a 3D wavelet-domain convolutional neural network for change detection in hyper-spectral images.

In this paper, we exploit a WST network to guide the DNN for one-class classification. WST provides the invariant representation which are informative because of keeping high frequency of data. Using these properties, can cluster the normal samples and distinguish outliers. At first, WST extracts scattering coefficients from given images and then the network is applied to them (Fig. 1). This leads to obtaining more robust and accurate results. Furthermore, to address the problem of the insufficient amount of training set for DNN, an adequate initial feature extraction to ease the training task is essential. A thorough evaluation confirmed superior performance and robustness both to outliers (not normal samples) in the training set and to translation and rotation of test set. Furthermore, we investigate the relationship between the entropy of images and the guidance of WST for highly textured images. The methodology is explained in the next section. The third section presents the experiments. The results are discussed in section 4 and conclusions are drawn in section 5.

2. METHODOLOGY

The study aims to find the similar samples to the normal class and distinguish them apart from all other classes in the dataset. Typically, this is treated as an unsupervised learning problem where the anomalous samples are not known a priori. To achieve this, we exploit the WST as implemented by [10] to aid deep SVDD [11]. The WST network cascades wavelet transform convolutions with nonlinear modulus and averaging operators [6]. In the first step, we extract the scattering coefficients from data. For a given image X , the zeroth-order scattering coefficient is the local average given by $S[0]_X = |X * \phi_{2^J}|$ where ϕ_{2^J} is a low-pass filter and the integer $J > 0$ specifies the averaging scale 2^J of the filter. The first-order scattering coefficient is calculated by:

$$S[\lambda_1]_X = |X * \psi_{\lambda_1}| * \phi_{2^J} \quad (1)$$

where $*$ is convolution and ψ_{λ_1} is the first order wavelet filter bank. More informative structures are captured by decomposing $|X * \psi_{\lambda_1}|$ using the second-order filter bank ψ_{λ_2} . The second order scattering coefficient is defined by:

$$S[\lambda_1, \lambda_2]_X = ||X * \psi_{\lambda_1}| * \psi_{\lambda_2}| * \phi_{2^J} \quad (2)$$

According to [6], the energy of scattering wavelet transform is mainly concentrated on the less than third-order. Therefore, we do not use higher order of scattering transform. In order to obtain a locally translation-invariant scattering representation for a given image X , we concatenate the scattering coefficients as a vector:

$$S_X = [(S[0]_X)^\top, (S[\lambda_1]_X)^\top, (S[\lambda_1, \lambda_2]_X)^\top]^\top \quad (3)$$

In the second step, this scattering coefficients are used as the input of a DNN. The DNN aims to minimize the volume of the point cloud by finding a data-enclosing hyper-sphere of the smallest size and learns useful feature representations of the normal class samples. The network's objective is to optimize the network weights W by minimizing the mean distance of normal samples from the center of the hyper-sphere c . The loss function is defined by the following equation:

$$\min_W \frac{1}{n} \sum_{i=1}^n \|\phi(S_{X_i}; W) - c\|_2^2 + \frac{\lambda}{2} \sum_{k=1}^K \|W^k\|_F^2 \quad (4)$$

where S_{X_i} is the vector of scattering coefficients for i^{th} sample of normal class, $\phi(\cdot; W)$ is feature representation in the latent space of the network ϕ with K hidden layers, and W^k are the weights of layer k . The second term is a weight decay regularizer on the network parameters W with hyperparameter λ , where $\|\cdot\|_F$ denotes the Frobenius norm.

We must set a value for c during the initialization process. To achieve that, an auto-encoder is trained by maximizing reconstruction likelihood. Then the weights of the pre-trained encoder are transferred to the DNN. The center of hyper-sphere is obtained by feeding normal samples to the pre-trained DNN and then calculating the mean values of the network's output, as formulated in the following equation:

$$c = \frac{1}{n} \sum_{i=1}^n \phi(S_{X_i}; W) \quad (5)$$

To find the most relevant and most ambiguous samples, an similarity score D has been defined. For a given test sample X and the vector of scattering coefficients S_X , D is given by:

$$D(X) = \|\phi(S_X; W) - c\|_2^2 \quad (6)$$

The distance of each test sample from c in latent space can be used as a measure for the network to make a decision. The lowest score represents the most relevant (normal) sample and the highest score the most ambiguous (anomalous).

3. EXPERIMENTS

3.1. Datasets

We study the performance of the approach on the well-known CIFAR10 [12] dataset. We argue that the inclusion of the WST step is even more crucial for more challenging tasks like satellite scene classification. There are some reasons why satellite scene classification is more challenging: (1) big intra-class diversity, (2) high interclass similarity, (3) large variance of object/scene scales, and (4) coexistence of multiple ground objects [13]. To demonstrate this, we apply the method on a benchmarked remotely sensed dataset, called EuroSAT [14]. It has been collected by the Sentinel-2A satellite and is comprised of ten classes with a total of 27000 labeled and geo-referenced images. Fig. 2 shows one sample of each class, chosen to describe the dataset visually.



Fig. 2: A sample of each class of EuroSAT dataset. First row from left to right: Annual Crop (A.C.), Forest, Herbaceous Vegetation (H.V.), Highway, Industrial. Second row: Pasture, Permanent Crop (P.C.), Residential, River, Sea/Lake.

3.2. Scattering transform parameters

There are three parameters in WST network that play an important role on performance of classification. One parameter is the number of layers of scattering network, M . We used the second order (two layers) of the scattering ($M = 2$), as suggested in [6]; higher order transforms are not useful because the resulting scattering coefficients have negligible energy [15]. Another parameter is the number of orientations, L that plays an important role to extract a rotation-invariant representation. The maximum scale order, J , is used in the averaging filter. An analysis of the effect of these parameters on the overall performance is provided in the next section.

3.3. Evaluation

For both datasets, we have 10 one-class classification setups. In each setup, one of the classes is chosen as a normal class. Therefore, the network is separately trained using the normal class of each setup. We then evaluate performance on an independent test set, which contains samples from all classes, including normal and anomalous data. The model performance is then quantified using the area under the Receiver Operating Characteristic (ROC) curve metric (AUROC). In order to fairly compare results of each one-class classification setup, a random seed is fixed (set to 1) for all setups.

4. RESULTS

The results are presented in three experimental evaluations. The first one is a detailed performance evaluation in the absence of pollution. The second evaluates the robustness of the method to translations and rotations, while the third one explores the effect of outlier pollution on the performance.

4.1. Detailed evaluation in absence of pollution

The results of the first experimental evaluation on CIFAR10 and EuroSAT datasets are shown in Table 1 and Table 2, respectively. Each row represents AUROC values in % for a

Normal Class	L=6			L=8			No WST
	J=2	J=3	J=4	J=2	J=3	J=4	
Airplanes	<u>68.31</u>	50.41	40.12	69.01	57.19	54.69	58.08
Cars	66.78	58.51	52.98	<u>66.23</u>	59.57	56.68	62.85
Birds	57.1	58.64	54.27	52.61	59.16	52.81	48.96
Cats	55.79	56.42	53.26	57.93	53.53	<u>57.3</u>	57.19
Deer	<u>64.75</u>	51.85	46.1	68.1	64.04	46.7	57.58
Dogs	<u>60.56</u>	56.17	58.73	58.68	40.49	44.21	63.83
Frogs	77.51	66.42	60.27	<u>76.52</u>	61.11	61.38	58.73
Horses	65.47	62.56	57.33	<u>64.28</u>	61.73	58.02	61.43
Ships	<u>79.25</u>	72.69	62.33	81.81	73.63	58.35	76.88
Trucks	75.94	58.57	66.37	<u>73.16</u>	61.83	60.08	67.8
Average	67.146	59.224	55.176	<u>66.833</u>	59.228	55.022	61.333

Table 1: AUROC values in % on test set of CIFAR10. Columns 2-7 show results using WST with different parameter settings. The best and second-best results are bold and underlined, respectively.

Avg. Ent.	Normal Class	L=6			L=8			No WST
		J=2	J=3	J=4	J=2	J=3	J=4	
5.781	A.C	43.45	34.81	<u>59.96</u>	44.26	58.78	56.33	60.42
3.698	Forest	92.1	95.9	<u>95.61</u>	92.38	93.84	93.71	90.13
5.655	H.V	42.53	<u>65.3</u>	63.89	44.57	66.95	63.54	40.14
5.942	Highway	43.77	59.71	60.19	46.06	61.34	<u>60.59</u>	46.28
6.836	Industrial	87.4	84.26	78.32	84.04	<u>84.65</u>	79.83	49.22
4.907	Pasture	72.82	<u>75.23</u>	74.56	73.27	75.76	71.64	72.08
6.189	P.C	53.49	63.05	<u>72.61</u>	63.88	72.82	69.63	38.46
6.126	Residential	79.99	85	<u>85.36</u>	86.87	77.52	74.81	38.9
5.391	River	58.25	67.78	62.91	56.36	70.53	<u>69.44</u>	60.73
2.384	Sea/Lake	<u>95.37</u>	86.77	67.18	93.52	91.59	24.31	96.13
	Average	66.917	71.781	<u>72.059</u>	68.521	75.38	66.383	59.249

Table 2: AUROC values in % for a each normal class of EuroSAT. The first column compares average entropy of each class. The columns 3-8 show results using WST with different parameter settings. The results for the classes with the high entropy have been highlighted. The best and second-best results are bold and underlined, respectively.

specific normal class and the best and second-best results are bolded and underlined, respectively. The first column of Table 1 provides the normal classes of CIFAR10 dataset. The next six columns show results of the model using WST with different maximum order of scale: $J = 2$, $J = 3$, and $J = 4$ when the number of orientation is fixed at $L = 6$ and $L = 8$, respectively. The last column compares the results of the original Deep SVDD [11]. Looking at the last row, it is clear that the performance improves when a WST stage is added. When we examine the values for each class more closely, we can see that the majority of the best and second-best values are presented in columns of $J = 2$. This means that increasing the maximum order of scale reduces the AUROC values for this dataset. To better understand the method's behavior, it is necessary to evaluate its performance on another dataset too.

Since EuroSAT dataset has a high diversity of classes and multiple ground objects in each sample (see Fig.1), we calculate Shannon's entropy [16] of classes. In information theory, Shannon's entropy quantifies the amount of information in a variable and it used for the study of the theoretical foundation of deep learning [17]. The first column in Table 2 compares

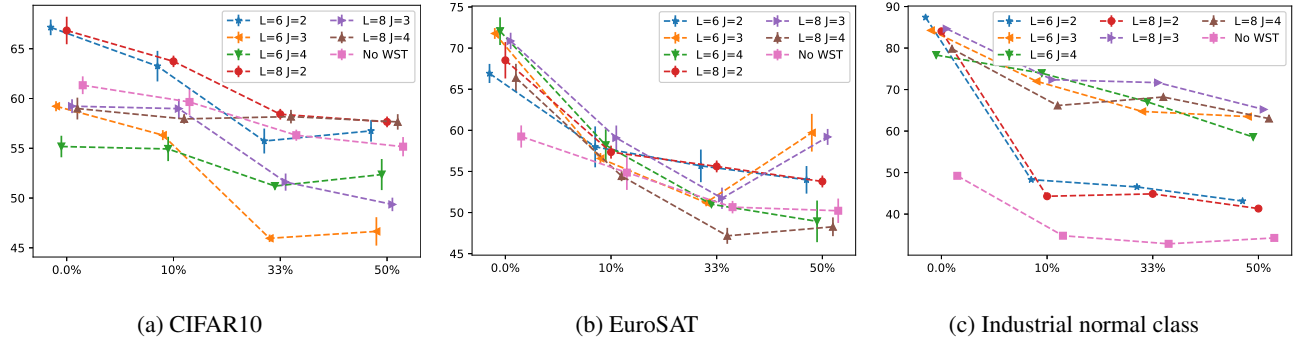


Fig. 3: Average of AUROC in % using WST with different parameters and without WST versus percentage of pollution.

Normal Class	Rotation		Translation		Rot. & Tran.	
	SVDD	WST	SVDD	WST	SVDD	WST
Airplanes	48.09	59.44	57.15	66.88	47.21	57.55
Cars	55.67	60.24	57.88	62.4	54.52	59.44
Birds	47.24	45.87	47.89	50.01	49.35	45.94
Cats	57.43	54.81	56.2	54.15	55.74	53.14
Deer	57.15	57.39	57.65	56.96	58.23	56.8
Dogs	58.97	56.41	59.35	56.66	58.09	54.29
Frogs	58.56	61.6	59.71	64.77	58.05	62.18
Horses	51.43	55.52	60.08	59.65	50.87	53.77
Ships	67.74	68.33	74.04	75.37	63.7	62.72
Trucks	52.63	63.97	64.12	72.83	53.7	62.83
Average	55.49	58.358	59.41	61.97	54.95	56.87

Table 3: AUROC values (in %) of CIFAR10 test set corrupted by rotation, translation, and combined them.

Normal Class	Rotation		Translation		Rot. & Tran.	
	SVDD	WST	SVDD	WST	SVDD	WST
A.C.	53.72	53.42	55.59	53.85	53.28	53.44
Forest	90.18	94.76	90.24	94.6	90.28	94.7
H.V.	39.17	60.65	40.25	62.49	39.51	61.84
Highway	44.8	61.27	43.63	59.91	43.86	59.87
Industrial	44.04	79.93	59.06	84.16	54.5	80.55
Pasture	71.3	78.76	71.18	78.03	71.51	79.41
P.C.	36.33	66.99	36.56	68.34	37.79	67.03
Residential	35.35	78.26	37.74	77.89	37.81	79.39
River	60.17	67.59	60.2	70.57	60.68	69.37
Sea/Lake	95.09	92.28	95.69	92.02	96.01	92.2
Average	57.02	73.39	59.01	74.19	58.523	73.78

Table 4: AUROC values (in %) of EuroSAT test set corrupted by rotation, translation, and combined them.

the average entropy of the samples of each class. The last row displays the average of all classes' AUROC values for each method. The fact that the all values in last row are higher than last column demonstrates the importance of WST. WST compensates for DNN's shortcomings on classes with a larger average entropy. Classes with a high entropy such as Industrial, Permanent Crop (P.C.), and Residential got more than 30% improvement.

4.2. Evaluation of the robustness to transformation

In order to evaluate the robustness to transformations, we employ rotation, translation, and combination of them. The trained models have been tested on transformed images of the test set. The performance of the method (AUROC in %) on CIFAR10 and EuroSAT is described in Table 3 and Table 4, respectively. In both cases WST demonstrates the highest robustness against all transformations. As EuroSAT has not enough training samples (about 1800 each normal class), WST aids the SVDD to learn more invariant features. The higher differences of average AUROC for SVDD and WST demonstrate it. Similar to Table 2, highly-textured classes take more advantage of WST because of stable features.

4.3. Evaluation of the effect of pollution

Fig.3a and Fig.3b describe the results of the third experimental evaluation on CIFAR10 and EuroSAT datasets, respectively. The error bar shows average and standard deviation of AUROC values of all classes in % versus different ratios of pollution. The results demonstrate the benefit of WST while the training set is polluted by anomalies. Increasing the ratio of pollution reduces the AUROC values of each method. However, models using WST with appropriate parameters are more robust compared to the original Deep SVDD (No WST). This improvement is seen especially for the more complex EuroSAT dataset. Fig. 3c shows the results of AUROC only for Industrial class (highest entropy) versus the ratio of pollution. All models using WST outperform the original Deep SVDD. We also witness small degradation of performance for $J = 3$ and $J = 4$ while increasing the ratio of pollution.

5. CONCLUSIONS

In this paper we have proposed guiding DNNs for image one-class classification with a wavelet scattering stage. The results demonstrate the improvement brought by WST, especially on

more complex image data such as EuroSAT. The proposed method alleviates the need for a large training set for the DNN because it leverages WST to achieve more stable features (see Table 4). The robustness to pollution in normal training set and also robustness to transformation were discussed. In addition, it was observed that WST compensates for the limited performance of the DNN for classes with high entropy (see Fig.3c and the highlighted row in Table 2).

6. REFERENCES

- [1] Renlong Hang, Feng Zhou, Qingshan Liu, and Pedram Ghamisi, "Classification of Hyperspectral Images via Multitask Generative Adversarial Networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 2, pp. 1424–1436, 2021.
- [2] Lukas Ruff, Jacob R. Kauffmann, Robert A. Vandermeulen, Gregoire Montavon, Wojciech Samek, Marius Kloft, Thomas G. Dietterich, and Klaus Robert Müller, "A Unifying Review of Deep and Shallow Anomaly Detection," *Proceedings of the IEEE*, vol. 109, no. 5, pp. 756–795, 2021.
- [3] Pramuditha Perera and Vishal M. Patel, "Learning Deep Features for One-Class Classification," *IEEE Transactions on Image Processing*, vol. 28, no. 11, pp. 5450–5463, 2019.
- [4] Fei Ye, Huangjie Zheng, Chaoqin Huang, and Ya Zhang, "Deep unsupervised image anomaly detection: An information theoretic framework," in *2021 IEEE International Conference on Image Processing (ICIP)*, 2021, pp. 1609–1613.
- [5] Liron Bergman and Yedid Hoshen, "Classification-based anomaly detection for general data," *ArXiv*, vol. abs/2005.02359, 2020.
- [6] Joan Bruna and Stéphane Mallat, "Invariant scattering convolution networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1872–1886, 2013.
- [7] Axel Vierling, Charu James, Karsten Berns, and Nikoletta Katsaouni, "Provable translational robustness for object detection with convolutional neural networks," in *2021 IEEE International Conference on Image Processing (ICIP)*, 2021, pp. 694–698.
- [8] Bedionita Soro and Chaewoo Lee, "A wavelet scattering feature extraction approach for deep neural network based indoor fingerprinting localization," *Sensors (Switzerland)*, vol. 19, no. 8, 2019.
- [9] Xianghai Wang, Chengdi Xing, Yining Feng, Ruoxi Song, and Zhenhua Mu, "A Novel Hyperspectral Image Change Detection Framework Based on 3D-Wavelet Domain Active Convolutional Neural Network," pp. 4332–4335, 2021.
- [10] Mathieu Andreux, Tomás Angles, Georgios Exarchakis, Roberto Leonarduzzi, Gaspar Rochette, Louis Thiry, John Zarka, Stéphane Mallat, Joakim Andén, Eugene Belilovsky, Joan Bruna, Vincent Lostanlen, Muawiz Chaudhary, Matthew J. Hirn, Edouard Oyallon, Sixin Zhang, Carmine Cella, and Michael Eickenberg, "Kymatio: Scattering transforms in python," *Journal of Machine Learning Research*, vol. 21, no. 2012, pp. 2012–2017, 2020.
- [11] Lukas Ruff, Robert A. Vandermeulen, Nico Görnitz, Lucas Deecke, Shoaib A. Siddiqui, Alexander Binder, Emmanuel Müller, and Marius Kloft, "Deep one-class classification," *35th International Conference on Machine Learning, ICML 2018*, vol. 10, pp. 6981–6996, 2018.
- [12] Alex Krizhevsky, "Learning multiple layers of features from tiny images," Tech. Rep., 2009.
- [13] Gong Cheng, Xingxing Xie, Junwei Han, Lei Guo, and Gui-Song Xia, "Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 3735–3756, 2020.
- [14] 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, 2019.
- [15] Rosemberg Rodriguez, Eva Dokladalova, and Petr Dokladal, "Rotation invariant CNN using scattering transform for image classification," in *2019 IEEE International Conference on Image Processing (ICIP)*, 2019, pp. 654–658.
- [16] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.
- [17] Lukas Ruff, Robert A. Vandermeulen, Nico Görnitz, Alexander Binder, Emmanuel Müller, Klaus-Robert Müller, and Marius Kloft, "Deep Semi-Supervised Anomaly Detection," 2019.