TOWARDS OUT-OF-DISTRIBUTION DETECTION FOR REMOTE SENSING

Jakob Gawlikowski[1,2], *Sudipan Saha*[1], *Anna Kruspe*[1], *Xiao Xiang Zhu*[1,3]

[1]Data Science in Earth Observation, Technical University of Munich, Taufkirchen/Ottobrunn, Germany
[2]Institute of Data Science, DLR, Jena, Germany
[3]Remote Sensing Technology Institute, DLR, Weßling, Germany

ABSTRACT

In remote sensing, distributional mismatch between the training and test data may arise due to several reasons, including unseen classes in the test data, differences in the geographic area, and multi-sensor differences. Deep learning based models may behave in unexpected manners when subjected to test data that has such distributional shifts from the training data, also called out-of-distribution (OOD) examples. Vulnerability to OOD data severely reduces the reliability of deep learning based models. In this work, we address this issue by proposing a model to quantify distributional uncertainty of deep learning based remote sensing models. In particular, we adopt a Dirichlet Prior Network for remote sensing data. The approach seeks to maximize the representation gap between the in-domain and OOD examples for a better identification of unknown examples at test time. Experimental results on three exemplary test scenarios show that the proposed model can detect OOD images in remote sensing.

Index Terms— Out-of- distribution, open set recognition, robustness, remote sensing

1. INTRODUCTION

Following the trend in computer vision and other fields, deep learning has revolutionized the field of remote sensing in the last few years. Deep learning based approaches have been successfully applied in various remote sensing tasks, including classification [1], change detection [2], and image fusion [3]. Most of these approaches assume that test data is similarly distributed as the training data on which the model is trained. These similarities contain for example geographical characteristics, identical sets of classes, and the types of sensors used.

Remote sensing deals with a large number of sensors, operating across a variety of different geographies, and often distinguishes between a large number of classes. Considering this variation, the above assumptions often do not hold. There are a few works related to domain adaptation that try to align the target distribution with the source distribution [4]. However, such methods are only effective when the domain shift between the source and target is small [4]. Moreover, they do not consider the presence of unseen and open-set classes. Deep learning models are likely to fail or behave in an unexpected way when faced with open-set classes. A deep model trained on images from a forest area will for example fail when asked to predict urban images consisting of residential complexes and parking lots. Similarly, deep models behave in unexpected way when fed with data from seen classes but with a considerable geographic variation. For example, a model trained on European urban area (where skyscrapers are rare) will fail when asked to predict for images from most Asian urban areas. When deep learning based systems fail, they do not provide sufficient clue to the user and can give a wrong prediction, yet with high confidence. To address this issue, predictive uncertainty estimation has recently emerged as a research topic in the machine learning community [5]. Uncertainty estimation informs users about the confidence on a prediction, thus gives a hint on the reliability of such systems and possible weaknesses.

Deep learning based classification models are prone to predictive uncertainties from three different sources [5]: model or epistemic uncertainty, data or aleatoric uncertainty, and distributional uncertainty. Epistemic uncertainty is uncertainty caused by the model parameters while aleatoric uncertainty arises from complexities related to the data distribution, e.g. class overlap in the data. Distributional uncertainty can be also seen as a special case of epistemic uncertainty and stems from a distributional mismatch between the training and the test data. In remote sensing distributional uncertainty may arise due to various reasons, as unseen classes, geographic differences, and sensor differences. Considering its high relevance in remote sensing, our work focuses on the distributional uncertainty [6].

Our work is based on a Dirichlet Prior Network (DPN) that separately models different aforementioned uncertainty types. The Dirichlet distribution is a distribution over the categorical distribution, i.e. it can model uncertainty on a soft-max output of a classification model. DPNs separate in-distribution and OOD examples by producing sharp Dirichlet distributions for in-domain examples (low deviation in the soft-max output) while producing flat Dirichlet distributions for OOD ones (high deviation in the soft-max output) [5]. In particular, we base our work on an extension of the DPN

classifier [7] that focuses on increasing the representation gap between in-domain and OOD examples. We experimentally show that the proposed approach is able to detect OOD examples in remote sensing images, thus improving the reliability and robustness of deep learning based models in remote sensing. To the best of our knowledge this is the first work that specifically addresses out-of-distribution detection in remote sensing.

2. PROPOSED METHOD

In remote sensing image classification, images x and their corresponding labels y can be characterized using their distribution $p(x, y)$. In practice, we only have a finite dataset $\mathcal{D} = \{x_j, y_j\}_{j=1}^{N}$ corresponding to the distribution $p(x, y)$. Since the training data is a random subset and the training process is also affected by randomness, Bayesian neural networks model the parameters θ of a neural network as a random variable. For a classifier with parameters θ, the predictive uncertainty on a prediction ω is then given by $p(y = \omega|x^*, \mathcal{D}) = \int p(y = \omega|x^*, \theta)p(\theta|D)d\theta$.
The sources of predictive uncertainty [5] can be broadly categorized into the following three categories:

1. *Epistemic or model uncertainty* characterizes the uncertainty caused by the network parameters and structure, trained on a finite training dataset. Additional training data can reduce epistemic uncertainty.

2. *Aleatoric or data uncertainty* arises from the complexity in the data distribution, e.g. class overlap and label noise, as data having different values of y may have very similar representations in x.

3. *Distributional uncertainty* arises from a mismatch between the training and the test data distribution. In this case, the test data is distributed by $p'(x, y) \neq p(x, y)$.

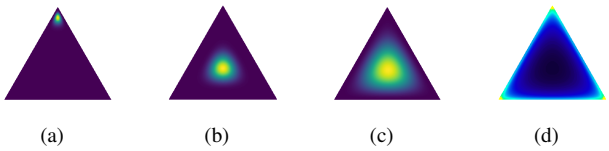
Approaches as Bayesian Neural Networks and deep ensembles consider the distributional uncertainty as part of the epistemic uncertainty. These approaches seek to explicit predict the aleatoric uncertainty and to quantify the epistemic uncertainty by performing several predictions with different model parameters [8].

2.1. Dirichlet Prior Network

Dirichlet distributions are popularly used as a prior distribution in Bayesian learning [9]. Motivated by this, Malinin and Gales [5] proposed Dirichlet Prior Networks (DPNs). DPNs are deterministic neural networks that efficiently mimic the behavior of Bayesian Neural Networks by parameterizing a Dirichlet distribution over the categorical distribution given by a soft-max classification output. Convenient to remote sensing applications, any neural network with a soft-max activation can be considered as a DPN. A Dirichlet distribution over K classes is characterized by concentration parameters $\{\alpha_1, ..., \alpha_K\} > 0$. For a DPN, the concentration is given by the exponentials of the network's logit values z,

$$\alpha_k = exp(z_k(x^*)) \; . \tag{1}$$

The sum of the concentrations $\alpha_0 = \alpha_1 + ... + \alpha_K$ is called the precision of the distribution. The larger the precision, the sharper is the Dirichlet distribution.



(a) (b) (c) (d)

Fig. 1. Different desired predictive uncertainties shown over the unit simplex (cf. [7]): (a) In-domain confident, (b) In-domain aleatoric uncertainty, (c) OOD (with DPN [5]), (d) OOD (with DPN$^-$ [7]).

For in-domain samples where the classifier is confident, DPNs aim to produce an uni-modal distribution at the corner of the solution simplex with the correct class (Figure 1(a)) [5]. For in-domain samples with high data uncertainty DPNs aim to produce a sharp distribution at the center (Figure 1(b)) and for OOD data a flat distribution (Figure 1(c)). However, for in-domain examples with high aleatoric uncertainty among multiple classes, DPNs could also produce flat Dirichlet distributions [7], what often leads to representations which are indistinguishable from OOD examples. To overcome this, Nandy et al. [7] proposed the DPN$^-$ approach. DPN$^-$ aims at learning a sharp multi-modal distribution ($\alpha_0 << 1$) instead of a flat uni-modal distribution for OOD examples. Additional, Nandy et al. chose the DPN parameters in a way, that the loss simplifies to the cross-entropy plus a precision regularization term.

The precision regularization is achieved by introducing a bounded regularization term

$$\alpha_0' = \frac{1}{K} \sum_{k=1}^{K} \text{sigmoid}(z_k(x))$$

as a regularizer along with the cross-entropy loss. This gives the following two loss formulations for in-domain and OOD examples:

$$\mathcal{L}_{in}(\theta; \lambda_{in}) := \mathbb{E}_{P_{in}(x,y)} \left[-\log p(y|x, \theta) - \lambda_{in}\alpha_0' \right] \tag{2}$$

and

$$\mathcal{L}_{out}(\theta; \lambda_{out}) := \mathbb{E}_{P_{out}(x,y)} \left[\mathcal{H}_{ce}(\mathcal{U}; p(y|x, \theta)) - \lambda_{out}\alpha_0' \right] \; . \tag{3}$$

$\mathcal{U}$ denotes the uniform distribution over all classes, $\mathcal{H}_{ce}$ denotes the cross-entropy function, and the precision is controlled by two hyper-parameters $\lambda_{in} > 0$ and $\lambda_{out} < 0$. The combined loss-function is given by

$$\mathcal{L}(\theta; \gamma, \lambda_{in}, \lambda_{out}) = \mathcal{L}_{in}(\theta; \lambda_{in}) + \gamma\mathcal{L}_{out}(\theta; \lambda_{out}), \tag{4}$$

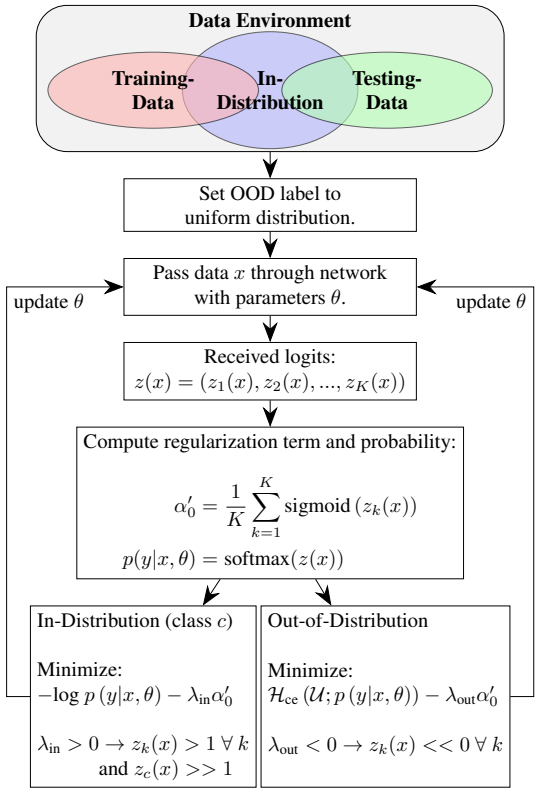


Fig. 2. A visualization of the training procedure for the considered DPN$^-$ approach.

where the in-domain and OOD samples are balanced by a factor $\gamma > 0$.

For in-domain examples which are confidently predicted, the cross-entropy loss maximizes the logit value of the correct class. However, for in-domain samples with aleatoric uncertainty, the optimizer maximizes sigmoid$(z_k(x))$ for all classes k, thus yielding a sharp but centered distribution. By choosing $\lambda_{out} < 0$, DPN$^-$ produces uniform negative values for $z_k(x^*)$ for an OOD example x^*. This leads to $\alpha_k << 1$ for all $k = 1, ..., K$, and thus an OOD sample yields a sharp multimodal Dirichlet distribution with uniform weights at each corner of the simplex (Fig 1(d)). Figures 1(b) and 1(d) are more distinct over the simplex, making the OOD samples easier to distinguish from the in-domain ones. In Figure 2, a visualization of the training process of DPN$^-$ is given.

3. EXPERIMENT AND RESULTS

We want to test the performance of DPN$^-$ for OOD detection on remote sensing data. In order to evaluate the gap between in-domain and OOD samples we use the same measures as in [7], namely mutual information, maximum probability, and the precision α_0. The general performance is characterized by *area under the receiver operating characteristic* (AUROC) scores based on these three measures.

Test dataset: We use the So2Sat LCZ42 dataset [10] for evaluating the OOD detection performance. The dataset consists of local climate zone (LCZ) labels of approximately half a million Sentinel-1 and Sentinel-2 patches. The local climate zones are described by 17 classes, 1-10 corresponding to urban areas, A-F corresponding to non-urban areas, and G corresponding to water body. We perform our experiments on the Sentinel-2 data and consider the following three combinations:

1. Urban classes as in-domain data, non-urban ones as OOD data during training, and water body as OOD data during testing.

2. Urban classes as in-domain data, water body as OOD data during training, and non-urban classes as OOD data during testing.

3. Red channels (corresponding to all 17 classes) as in-domain, green channels as OOD during training, and blue channels as OOD during testing.

Table 1. Resulting AUROC scores (times 100) of the proposed DPN$^-$ and the compared DPN$_{\text{forw}}$ and ENN classifiers. The scores are based on maximum probability, mutual information, and precision for the DPN$^-$. The results are given as mean and standard deviation of five runs.

		Max. Prob.	Mutual Info.	α_0
Test Case 1	DPN$^-$	98.58 ± 0.89	99.35 ± 0.29	99.34 ± 0.29
	DPN$_{\text{forw}}$	95.87 ± 2.28	54.74 ± 9.97	50.59 ± 10.48
	ENN	75.64 ± 5.70	75.33 ± 4.46	76.75 ± 2.84
Test Case 2	DPN$^-$	78.65 ± 0.61	89.53 ± 0.53	89.67 ± 0.54
	DPN$_{\text{forw}}$	44.65 ± 5.09	34.21 ± 7.84	33.09 ± 7.47
	ENN	71.76 ± 0.90	71.75 ± 0.35	68.80 ± 2.23
Test Case 3	DPN$^-$	91.79 ± 0.20	95.52 ± 0.29	95.52 ± 0.38
	DPN$_{\text{forw}}$	71.89 ± 4.53	12.26 ± 3.10	11.80 ± 2.71
	ENN	58.89 ± 0.70	58.17 ± 1.23	56.83 ± 2.02

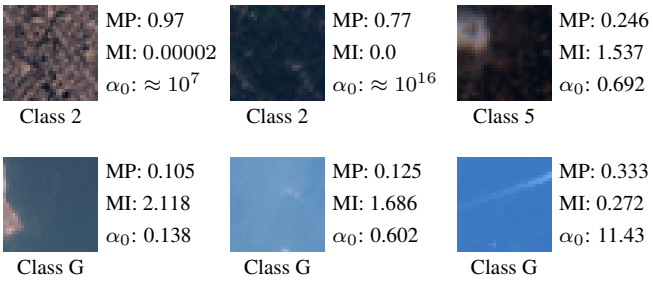


Fig. 3. A visualization of six example samples from the left out 30% of the training set of the So2Sat LCZ42 data set. The results are based on the DPN$^-$ network trained on urban (in-distribution) and vegetation (out-of-distribution) samples. One can clearly see the differences in the metrics. The two examples on the right side do not fit well into our assumptions, possibly caused by the clear edge in the water image and the blur in the urban image.

Deep architecture: For all experiments, we use the baseline approach for the LCZ42 dataset as proposed in [11], however without the multilevel feature fusion.

Comparison methods: We compare the proposed method to the DPN with a forward KL-divergence loss [5] (DPN_{forw}). This method uses OOD samples during training, similar to the proposed approach. We also compare to the Evidential Neural Network (ENN) with expected KL-divergence as loss and a precision regularization as proposed in [12]. This method does not need OOD samples during training. We evaluate the performance on a left-out 30% subset of the training set (evaluation on seen regions) in order to avoid OOD effects caused by a region shift.

Results: In Table 1 the results based on five runs for each setting are presented and in Figure 3 six examples are shown. The DPN^{-} clearly outperforms the other methods in separating in-domain and OOD examples and yields significantly more homogeneous results. On the contrary, optimizing towards a specific target concentration as done for DPN_{forw} shows unstable performance. The usage of the mutual information or the precision value α_0 contributes to the increment of the AUROC scores for the DPN^{-} approach for all test instances. For the other approaches, which do not aim at minimizing α_0 for OOD samples, this is not the case. Among the different test cases, separating urban and water classes with vegetation classes as OOD samples during training is clearly the easiest task, while separating urban and vegetation classes with only water as OOD training samples is more difficult.

4. CONCLUSION

In this paper, we proposed a method for distributional uncertainty quantification in deep learning models for remote sensing. The method ingests in-domain and OOD images during the training process and is subsequently used for OOD detection in the test images. It adopts a Prior Network to estimate different types of uncertainty by producing sharp Dirichlet distributions for in-domain examples and multi-modal Dirichlet distributions for OOD examples. The need of OOD examples at training time represents the largest restriction of the method. We tested the method on the So2Sat LCZ42 dataset considering open set classes and selected bands as OOD. In the future, we will perform extensive experiments on different geographic areas, considering one geographic area as in-domain while treating the other as OOD. In such settings, the effect of classes like water as in-domain and OOD examples at the same time is an open but relevant question. Furthermore, we plan to extend the proposed method to multi-sensor domain and multi-sensor data fusion.

5. REFERENCES

[1] S. Roy, E. Sangineto, N. Sebe, and B. Demir, "Semantic-fusion GANs for semi-supervised satellite image classification", in *2018 25th IEEE Int. Conf. on Image Processing (ICIP)*. IEEE, 2018, pp. 684–688.

[2] S. Saha, F. Bovolo, and L. Bruzzone, "Building change detection in VHR SAR images via unsupervised deep transcoding", *IEEE Transactions on Geoscience and Remote Sensing*, 2020.

[3] J. Gawlikowski, M. Schmitt, A. Kruspe, and X.X. Zhu, "On the fusion strategies of sentinel-1 and sentinel-2 data for local climate zone classification.", in *Proc. IGARSS*, 2020.

[4] N. Courty, R. Flamary, D. Tuia, and A. Rakotomamonjy, "Optimal transport for domain adaptation", *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 9, pp. 1853–1865, 2016.

[5] A. Malinin and M. Gales, "Predictive uncertainty estimation via prior networks", in *Advances in Neural Information Processing Systems*, 2018, pp. 7047–7058.

[6] Y. Gal, "Uncertainty in deep learning", *University of Cambridge*, vol. 1, no. 3, 2016.

[7] J. Nandy, W. Hsu, and M.L. Lee, "Towards maximizing the representation gap between in-domain & out-of-distribution examples", *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[8] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles", *Advances in neural information processing systems*, vol. 30, pp. 6402–6413, 2017.

[9] D. Geiger and D. Heckerman, "A characterization of the dirichlet distribution with application to learning bayesian networks", in *Maximum entropy and Bayesian methods*, pp. 61–68. Springer, 1996.

[10] X.X. Zhu, J. Hu, C. Qiu, Y. Shi, J. Kang, L. Mou, H. Bagheri, M. Häberle, Y. Hua, R. Huang, et al., "So2sat lcz42: A benchmark dataset for global local climate zones classification", *IEEE Geoscience and Remote Sensing Magazine*, 2020.

[11] Chunping Qiu, Xiaochong Tong, Michael Schmitt, Benjamin Bechtel, and Xiao Xiang Zhu, "Multilevel feature fusion-based cnn for local climate zone classification from sentinel-2 images: Benchmark results on the so2sat lcz42 dataset", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 2793–2806, 2020.

[12] Murat Sensoy, Lance M. Kaplan, and Melih Kandemir, "Evidential deep learning to quantify classification uncertainty", in *NeurIPS*, 2018, pp. 3183–3193.