IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium | 978-1-6654-2792-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/IGARSS46834.2022.9884880

ROBUST DISTRIBUTION-SHIFT AWARE SAR-OPTICAL DATA FUSION FOR MULTI-LABEL SCENE CLASSIFICATION

Jakob Gawlikowski[1,2], *Sudipan Saha*[2], *Julia Niebling*[1], *Xiao Xiang Zhu*[2,3]

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

ABSTRACT

Out-of-distribution (OOD) detection is an emerging research topic in remote sensing where existing works focus on single sensor analysis. However, many remote sensing works use multi-modal data to benefit from different characteristics of the sensors. Data that is in-domain for one sensor may be OOD for another sensor. In this work, we address such a scenario focusing on Synthetic Aperture Radar (SAR) and optical data fusion for multi-label scene classification. Besides data distribution shifts caused by unknown classes and snow, we also consider cases where only one modality is affected. Optical images acquired with significant cloud coverage are considered as OOD, while their corresponding SAR images can be in-distribution. We propose a weighted feature propagation strategy based on the in-distribution probabilities of the single modalities. We show, that we not only improve the prediction performance on the cloudy samples but also receive a higher predictive uncertainty when both modalities are OOD.

***Index Terms*—** Data Fusion, Out-of-Distribution, Uncertainty Quantification, Robustness, Remote Sensing

1. INTRODUCTION

In remote sensing, data fusion is a commonly used technique to benefit from the different modalities and at the same time gaining additional complementary information [1]. Taking earth observation data with different types of cloud coverage and illumination as an example, the fusion of SAR and optical data is beneficial as SAR is not impacted by those effects.

Recently, predictive uncertainty estimation and out-of-distribution detection have emerged as research topic in the machine learning community [2]. This topic has also gained attention in the remote sensing community [3]. Especially distribution shifts are common in remote sensing data, which can have unpredictable effects on the behaviour of a neural network that has never seen such shifted data before. This effect of distribution shift leads to epistemic uncertainty affecting the prediction [3, 4], since the network was never trained on how to handle such kind of data. While handling a distribution shift in a single-sensor setting is already challenging, handling it in a multi-sensor setting can be on one hand even more difficult, but on the other hand also offers new possibilities. This is because in the data fusion setting, not all sensors necessarily experience a distribution shift simultaneously. One prominent example for such a situation is the occurrence of clouds or the change in the illumination when working with optical and SAR images. While the optical images are affected by these changes in the environment, the SAR images do not experience any great effect. In other words, optical images suffer from distribution shift while SAR images do not. On the other hand, distribution shifts which will impact both sensors are also possible, for example due to unseen classes or snow and ice. Despite the relevance of the topic, there has been little research on uncertainty assessment in the context of remote sensing data fusion.

Existing predictive uncertainty estimation methods are designed to handle single input modalities [4]. On the contrary, most bi-sensor fusion methods use a two-stream network to process two inputs that are finally combined at the penultimate layer or one of the middle layers of the network [5, 6]. Thus it is not trivial to reuse existing predictive uncertainty methods for estimating uncertainty in remote sensing fusion. To circumnavigate this, we predict the in-distribution probability for each modality before realizing the fusion step. Following, we propagate the resulting estimated epistemic uncertainty to the fused prediction which is given as a predicted probability for each class. We train the network in a Siamese-like training [7], where the classifier is trained to give a prediction for each kind of modality combination (i.e., in our case, SAR and optical input, only SAR input, only optical input, no input). The case with no inputs leads to probabilities of 0.5 predicted for each class since no evidence for any class or against any class is available. Additional to the fusion network, we train an OOD detector for each modality based on in- and OOD-training data. We combine the components in such a way that the final approach is aware about the extent of uncertainty caused by distribution shifts affecting the single sensors. The contributions of this work are as follows:

Authorized licensed use limited to: Deutsches Zentrum fuer Luft- und Raumfahrt. Downloaded on January 13,2023 at 16:32:23 UTC from IEEE Xplore. Restrictions apply.

1. Introducing out-of-distribution detection in the data fusion setting on single modalities before the fusion step.

2. Proposing a propagation scheme to approximate the posterior predictive distribution including the in- distribution probabilities from the single branches.

3. Introducing a training and evaluation scheme for the adaptive fusion, based on the predicted in-distribution probabilities.

2. APPROACH

We train a ResNet50-based [8] fusion network consisting of two branches for the feature extraction of the optical Sentinel-2 and the Sentinel-1 SAR data, respectively. The two branches are fused by a concatenation based fusion layer and a final prediction is given by the ResNet50 classifier. In addition to the fusion layer we add an out-of-distribution detector directly to the embedded feature spaces of the two sensor branches. In contrast to a generic multi-class classification scenario, where each sample is predicted as exactly one out of multiple classes, multi-label classification gives a prediction for each class. The training and inference procedure is visualized in Fig. 1.

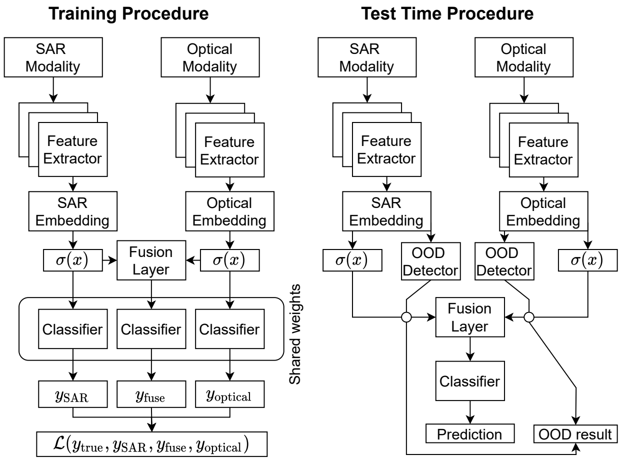

Fig. 1: The considered network structure for the training (left) and the testing (right) time. At training time, the network receives the groundtruth and gives a prediction based on the fused features and the single modalities, predicted with shared weights. The embeddings are pointwise mapped to the interval of $[-1, 1]$ by $\sigma(\cdot)$. At test time, an OOD detector is used to propagate distributional uncertainty onto the fused prediction.

2.1. Data Fusion Module

For each batch the optimization step is based on three forward passes, one with the SAR data, one with the optical data, and one with the concatenation of the extracted SAR and optical features. Doing so, the network is trained to give proper predictions when there is only SAR or optical data available. After training the network this way, we use hold-out OOD training data to train OOD detection heads for each modality in order to predict whether a sample is in-distribution or out-of-distribution.

We consider two encoder branches $f_{\text{opt}}(\cdot)$ and $f_{\text{SAR}}(\cdot)$ which encodes optical and SAR data, x_{opt} and x_{SAR}, respectively. The fusion layer takes the encoded features $f_{\text{opt}}(x_{\text{opt}})$ and $f_{\text{SAR}}(x_{\text{SAR}})$ as inputs and is defined as

$$f_{\text{fuse}}(y_{\text{opt}}, y_{\text{SAR}} | \lambda_{\text{opt}}, \lambda_{\text{SAR}}) := [\lambda_{\text{opt}} \cdot \sigma\left(f_{\text{opt}}(x_{\text{opt}})\right), \lambda_{\text{SAR}} \cdot \sigma\left(f_{\text{SAR}}(x_{\text{SAR}})\right)] \quad (1)$$

where $\lambda_{\text{opt}}, \lambda_{\text{SAR}} \in \{0, 1\}$, $[\cdot, \cdot]$ is the concatenation of the features and $\sigma(\cdot)$ maps the input element-wise to the interval of $(-1, 1)$ and is defined as $\sigma(x) := 2 \cdot (\text{sigmoid}(x) - 0.5)$. We apply σ in order to receive a more balanced contribution among the different modalities. The classifier part of our network is defined as $f_{\text{c}}(y_{\text{opt}}, y_{\text{SAR}} | \lambda_{\text{opt}}, \lambda_{\text{SAR}})$ and concatenates the above defined fusion layer and the classifier part of the ResNet50.

The classification part is trained with shared network parameters and fusion hyper parameters $(\lambda_{\text{opt}}, \lambda_{\text{SAR}}) = (1.0, 0.0)$, $(\lambda_{\text{opt}}, \lambda_{\text{SAR}}) = (0.0, 1.0)$, and $(\lambda_{\text{opt}}, \lambda_{\text{SAR}}) = (1.0, 1.0)$, respectively. The three parallel forward passes represent the three different cases of classification based on the pure optical features, the pure SAR features and the fused features of the optical and the SAR modality.

As a loss function we sum up the the binary cross-entropy loss $\mathcal{L}_{\text{BCE}}$ applied to the groundtruth y_{true} and the predictions $\hat{y}_{\text{SAR}}, \hat{y}_{\text{opt}}, \hat{y}_{\text{fuse}}$ received from the three output branches:

$$\begin{aligned}\mathcal{L}(y_{\text{true}}, \hat{y}_{\text{SAR}}, \hat{y}_{\text{opt}}, \hat{y}_{\text{fuse}}) := \; & w_{\text{fuse}} \cdot \mathcal{L}_{\text{BCE}}(y_{\text{true}}, \hat{y}_{\text{fuse}}) \\ & + w_{\text{SAR}} \cdot \mathcal{L}_{\text{BCE}}(y_{\text{true}}, \hat{y}_{\text{SAR}}) \quad (2) \\ & + w_{\text{opt}} \cdot \mathcal{L}_{\text{BCE}}(y_{\text{true}}, \hat{y}_{\text{opt}}) \,,\end{aligned}$$

where $w_{\text{fuse}}, w_{\text{opt}}, w_{\text{SAR}} > 0$ are scalar hyper-parameters for weighting the single cases. We choose $w_{\text{fuse}} = 2, w_{\text{opt}} = 1, w_{\text{SAR}} = 1$ but found the method to be robust on the choice.

2.2. Out-of-Distribution Detector

Building up on the modality-wise data feature encoding branches described above, we train out-of-distribution sample detectors for each modality. For that we train simple two-layer binary classifiers that should learn to distinguish the resulting embeddings of in-distribution samples from out-of-distribution samples. For this purpose, OOD samples are needed at training time.

2.3. OOD Aware Prediction

Based on the out-of-distribution detector and the fusion network we compute the expected predictions based on

Authorized licensed use limited to: Deutsches Zentrum fuer Luft- und Raumfahrt. Downloaded on January 13,2023 at 16:32:23 UTC from IEEE Xplore. Restrictions apply.

the in-distribution information. For this we consider λ_{SAR} and λ_{opt} as Bernoulli-distributed random variables with $\lambda_{\text{SAR}} \sim \text{Ber}(p_{\text{s}})$ and $\lambda_{\text{opt}} \sim \text{Ber}(p_{\text{o}})$, where p_{s} and p_{o} are the (predicted) probabilities that the SAR and the optical modality are in-distribution. Assuming that the two random variables are independent, this leads to the following prediction:

$$\begin{aligned} y &= \mathbb{E}_{\lambda_{\text{opt}} \sim \text{Ber}(p_{\text{s}}), \lambda_{\text{SAR}} \sim \text{Ber}(p_{\text{o}})} \left[f_{\text{c}} \left(y_{\text{opt}}, y_{\text{SAR}} | \lambda_{\text{SAR}}, \lambda_{\text{opt}} \right) \right] \\ &= (1 - p_{\text{s}}) \cdot (1 - p_{\text{o}}) \cdot f_{\text{c}} \left(y_{\text{opt}}, y_{\text{SAR}} | 0, 0 \right) \\ &\quad + p_{\text{s}} \cdot (1 - p_{\text{o}}) \cdot f_{\text{c}} \left(y_{\text{opt}}, y_{\text{SAR}} | 0, 1 \right) \\ &\quad + (1 - p_{\text{s}}) \cdot p_{\text{o}} \cdot f_{\text{c}} \left(y_{\text{opt}}, y_{\text{SAR}} | 1, 0 \right) \\ &\quad + p_{\text{s}} \cdot p_{\text{o}} \cdot f_{\text{c}} \left(y_{\text{opt}}, y_{\text{SAR}} | 1, 1 \right) . \end{aligned} \tag{3}$$

.

3. EXPERIMENTS

3.1. Data and Setup

For our experiments we use the multi-modal version of the BigEarthNet dataset [9]. In addition to clean Sentinel-1 and Sentinel-2 samples from different regions within Europe representing 19 different classes, where we used 12 classes as in-distribution and 7 classes as unknown OOD classes. For the training of the OOD detectors, we used as OOD training data all patches which only contain the classes 1 and 2 (Urban Fabric and Industrial). Similarly, as OOD test data we used patches containing only classes 15-19 (sand/dunes and all water related classes). Patches that contain in-distribution and OOD classes at the same time were removed from the dataset. The dataset also contains a set of samples affected by clouds and seasonal snow which we will use as OOD testing data. Further, we sample 200 patches with full cloud cover by hand for testing.

We train our approach for 100 epochs and save the parameters for the best performances on the given validation split. We set the loss parameters to $w_{\text{opt}} = 1$, $w_{\text{SAR}} = 1$ and $w_{\text{fuse}} = 2$. Following this, we train the OOD detectors for 5 epochs on the optical and the SAR branch. We apply intensity augmentations on the OOD-data in order to get a better sensitivity to distribution shifts. We compare the proposed fusion process with considering in-distribution probabilities (adaptive fusion) against the proposed training procedure without in-distribution probability (non-adaptive fusion) and the performance of optical only, SAR only and a baseline approach trained equivalently to the once proposed by [9]. We differentiate between different settings of in-distribution data, left-out-classes, and different effects of clouds and shadows, snow, and handpicked fully cloud-covered samples.

For the evaluation we consider the F1 and F2 score for multi-label classification to measure the classification performance and the average entropy and confidence over the single class predictions to represent the certainty the different approaches give on their predictions.

3.2. Results

The results are shown in Table 1. For the clear testing data the performance of our approach is slightly below the performance of the baseline model but also improves the performance compared to predictions based on single modalities. For the OOD classes one can see, that only the adaptive fusion gives a significant change in the average entropy and confidence on the final output and hence expressing uncertainty. The performance on data that is shifted from the original dataset by including clouds and shadows in the optical version results in a slightly worse prediction performance in all settings but the SAR only setting, where the performance even improved compared to the clear test dataset. The performance drop is significantly larger for the baseline and the optical only approaches trained with the proposed training procedure. For the handpicked cloudy test case again all approaches but the SAR only approach result in worse classification performance. The proposed adaptive fusion shows the smalles drop while the baseline method and the optical only approach show the largest ones. Only for the adaptive fusion the average entropy increases and the average confidence decreases. For the snow and ice test case, a decrease of the classification performance appears only for the baseline approach and the SAR only approach.

4. DISCUSSION

The presented results underline, that distributional uncertainty quantification on single modalities can improve the performance especially in the case, that one modality experiences a significant change in the data distribution, resulting in a significant drop in the classification performance on this modality. The experiments on the clear OOD samples based on left-out classes show the capability of detecting unknown classes while keeping the classification performance high on the in-distribution samples. Even though for the clear in-distribution test data the classification performance of the adaptive fusion is a little bit below the one of the considered baseline, the method is significantly more robust against distribution shifts. The weaker performance might be explained by the very simple OOD detectors and a possibly weaker performance of the OOD detector on the SAR modality. The increase in the classification performance in the SAR-only model when comparing the clear and the cloudy data can be explained by a different class distribution in the two datasets and is negligible at this point. Besides the classification results, also the predictive uncertainty is of interest for a user and indicated by the average minimum entropy and the average confidence of the single class predictions. Here, the experiments show that the predictive uncertainty

Authorized licensed use limited to: Deutsches Zentrum fuer Luft- und Raumfahrt. Downloaded on January 13,2023 at 16:32:23 UTC from IEEE Xplore. Restrictions apply.

Table 1: Overview over the performance on the classification and the OOD detection task for the considered baseline, the optical branch, the SAR branch and the proposed training procedure without adaptive fusion and OOD detection and the proposed adaptive fusion approach.

Testing dataset		Baseline Fusion	Non-Adaptive Fusion	Adaptive Fusion	Optical Only	SAR Only
Clear	F1 Score	77.90	76.86	76.52	75.29	67.28
	F2 Score	77.56	77.59	77.07	75.29	66.60
	Avg. Entropy	0.318	0.207	0.302	0.303	0.408
	Avg. Confidence	0.903	0.938	0.903	0.909	0.872
OOD Classes (clear)	Avg. Entropy	0.306	0.237	0.932	0.347	0.337
	Avg. Confidence	0.900	0.922	0.577	0.888	0.895
Clouds and Shadows	F1 Score	71.62	75.48	75.25	68.14	70.54
	F2 Score	71.62	76.23	75.94	68.21	70.50
	Avg. Entropy	0.343	0.190	0.302	0.320	0.339
	Avg. Confidence	0.896	0.943	0.906	0.904	0.894
Cloudy Handpicked	F1 Score	38.92	62.25	69.56	38.41	70.50
	F2 Score	37.27	61.61	70.52	36.68	71.75
	Avg. Entropy	0.310	0.194	0.411	0.348	0.195
	Avg. Confidence	0.908	0.940	0.866	0.895	0.897
Snow and Ice	F1 Score	72.43	76.24	75.85	76.26	62.99
	F2 Score	70.13	77.30	76.89	76.89	61.84
	Avg. Entropy	0.224	0.203	0.255	0.305	0.368
	Avg. Confidence	0.926	0.939	0.922	0.905	0.882

received from the adaptive fusion approach, seam to draw an improved representation of the predictive uncertainty when the classification is affected by distribution changes.

5. CONCLUSION

In this work we presented an approach for fusing optical and SAR data while taking the distribution of the known training data into account. The prediction is given as an expected prediction, taking the in-distribution probabilities of the single modalities into account. The proposed method has shown great potential in making data fusion approaches more robust to distribution changes. In the future, we want to evaluate the performance in a much broader setup with a special focus on region shifts, try out new out-of-distribution detection approaches and evaluate the applicability of including other sources of uncertainties (e.g., aleatoric uncertainty) during the fusion process. Further the evaluation of contradictive information from the different modalities and different fusion layers will be part of future research.

6. REFERENCES

[1] Michael Schmitt and Xiao Xiang Zhu, "Data fusion and remote sensing: An ever-growing relationship", *IEEE Geoscience and Remote Sensing Magazine*, vol. 4, no. 4, pp. 6–23, 2016.

[2] Andrey Malinin and Mark Gales, "Predictive uncertainty estimation via prior networks", in *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, 2018, pp. 7047–7058.

[3] Jakob Gawlikowski, Sudipan Saha, Anna Kruspe, and Xiao Xiang Zhu, "An advanced dirichlet prior network for out-of-distribution detection in remote sensing", *IEEE Transactions on Geoscience and Remote Sensing*, 2022.

[4] Jakob Gawlikowski, Cedrique Rovile Njieutcheu Tassi, Mohsin Ali, Jongseok Lee, Matthias Humt, Jianxiang Feng, Anna Kruspe, Rudolph Triebel, Peter Jung, Ribana Roscher, et al., "A survey of uncertainty in deep neural networks", *arXiv preprint arXiv:2107.03342*, 2021.

[5] Jakob Gawlikowski, Michael Schmitt, Anna Kruspe, and Xiao Xiang Zhu, "On the fusion strategies of sentinel-1 and sentinel-2 data for local climate zone classification", in *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2020, pp. 2081–2084.

[6] Patrick Ebel, Sudipan Saha, and Xiao Xiang Zhu, "Fusing multi-modal data for supervised change detection", ISPRS, 2021.

[7] Sudipan Saha, Patrick Ebel, and Xiao Xiang Zhu, "Self-supervised multisensor change detection", *IEEE Transactions on Geoscience and Remote Sensing*, 2021.

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[9] Gencer Sumbul, Arne de Wall, Tristan Kreuziger, Filipe Marcelino, Hugo Costa, Pedro Benevides, Mário Caetano, Begüm Demir, and Volker Markl, "Bigearthnet-mm: A large scale multi-modal multi-label benchmark archive for remote sensing image classification and retrieval", *arXiv preprint arXiv:2105.07921*, 2021.

Authorized licensed use limited to: Deutsches Zentrum fuer Luft- und Raumfahrt. Downloaded on January 13,2023 at 16:32:23 UTC from IEEE Xplore. Restrictions apply.