


RESEARCH

Open Access



Handling unexpected inputs: incorporating source-wise out-of-distribution detection into SAR-optical data fusion for scene classification

Jakob Gawlikowski^{1,2}, Sudipan Saha³, Julia Niebling² and Xiao Xiang Zhu^{1*} 

*Correspondence:
xiaoxiang.zhu@tum.de

¹ Data Science in Earth
Observation, Technical University
Munich, Munich, Germany

² Institute of Data Science,
German Aerospace Center, Jena,
Germany

³ Yardi School of Artificial
Intelligence, Indian Institute
of Technology Delhi, New Delhi,
India

Abstract

The fusion of synthetic aperture radar (SAR) and optical satellite data is widely used for deep learning based scene classification. Counter-intuitively such neural networks are still sensitive to changes in single data sources, which can lead to unexpected behavior and a significant drop in performance when individual sensors fail or when clouds obscure the optical image. In this paper we incorporate source-wise out-of-distribution (OOD) detection into the fusion process at test time in order to not consider unuseful or even harmful information for the prediction. As a result, we propose a modified training procedure together with an adaptive fusion approach that weights the extracted information based on the source-wise in-distribution probabilities. We evaluate the proposed approach on the BigEarthNet multilabel scene classification data set and several additional OOD test cases as missing or damaged data, clouds, unknown classes, and coverage by snow and ice. The results show a significant improvement in robustness to different types of OOD data affecting only individual data sources. At the same time the approach maintains the classification performance of the baseline approaches compared. The code for the experiments of this paper is available on GitHub: https://github.com/JakobCode/OOD_DataFusion

Keywords: Data fusion, Out-of-distribution, Missing modality, Robustness, Remote sensing

1 Introduction

Out-of-distribution detection and uncertainty quantification are two important research topics in the machine learning community [1–3] and have also gained attention in the remote sensing community recently [4, 5]. In particular, out-of-distribution examples are caused by a shift in the data distribution between training and test time and are very common in remote sensing data. Such OOD examples can have unpredictable effects on the behaviour of a neural network as this neural network has never seen such shifted data before and was not trained on how to handle such data. Following, such data leads to epistemic (or model) uncertainty affecting the prediction [1, 2, 6].

Such distribution shifts can be caused for example by geographical differences, changing illumination, clouds or other causes. While shifts as regional distribution shifts can still lead to a valid prediction and are part of domain adaptation cross-domain learning works [7, 8], stronger distribution shifts often lead to very confident but false predictions [9, 10]. For stronger shifts as the appearance of unknown classes (which hence cannot be predicted correctly) specific OOD detection approaches maximize the aleatoric uncertainty expressed in the soft-max output [1].

In remote sensing, the fusion of optical data and Synthetic Aperture Radar (SAR) is a commonly used technique to benefit from complementary information caused by very different physical properties [11, 12]. Due to this complementarity the fusion in general improves the predictive performance compared to corresponding single-source approaches. Considering SAR-optical data fusion, potentially not only the effects of clouds in the optical images on a prediction could be softened, but also the effect of corrupted or unavailable data caused for example by a broken sensor or a disturbed data transmission as long as one data source is unaffected by the shift. While this is a commonly given motivation for the fusion of optical and SAR data, most approaches do only train with all data sources available in an undisturbed way and hence the network is not necessarily robust against changes in individual data sources [13–15]. This means, that in general common data fusion networks are trained on a joint data distribution over the optical and SAR data. Following, a strong shift on the individual distributions also strongly affects the joint distribution which the network learned as the in-distribution. This holds even though most bi-sensor fusion methods use a two-stream network to process two inputs independently [14, 16–18]. The independently processed input modalities are then fused at a fusion layer, followed by a combined part, which computes the network prediction based on the fused information [19, 20]. Theoretically, the distribution shifts in single data sources could be detected in the feature space of the individual branches and the fusion step can then be realized adaptively and based on the in-distribution probability.

Despite the relevance of the topic, there has been little research on distributional uncertainty assessment in the context of remote sensing data fusion.

In this work, we propose a training and testing procedure that not only works on the joint data distribution, but also on the individual data distributions. We test the proposed approach on a SAR-optical multi-label scene classification task and different out-of-distribution test cases affecting the SAR source as well as the optical data source.

As visualized in Fig. 1, the adaptive fusion predicts the in-distribution probability for each modality before realizing the fusion step. Following, we propagate the resulting estimated distributional uncertainty to the fused prediction which is given as a individually predicted probability for each class. Further, we introduce a training strategy for the network [21], where a classifier is trained to give a prediction for each kind of modality combination (i.e., in our case, SAR and optical input, masked-out SAR input, masked-out optical input, no input). As we consider a multi-label classification problem in this paper, the case with no inputs leads to the full weights given on the prior probability $p(y|0, 0) = 0.5$ for each class. Building up on the trained fusion networks, we train one OOD detector network for each modality to detect distribution shifts on the extracted features of the single modalities. This implies that the OOD detectors take the output

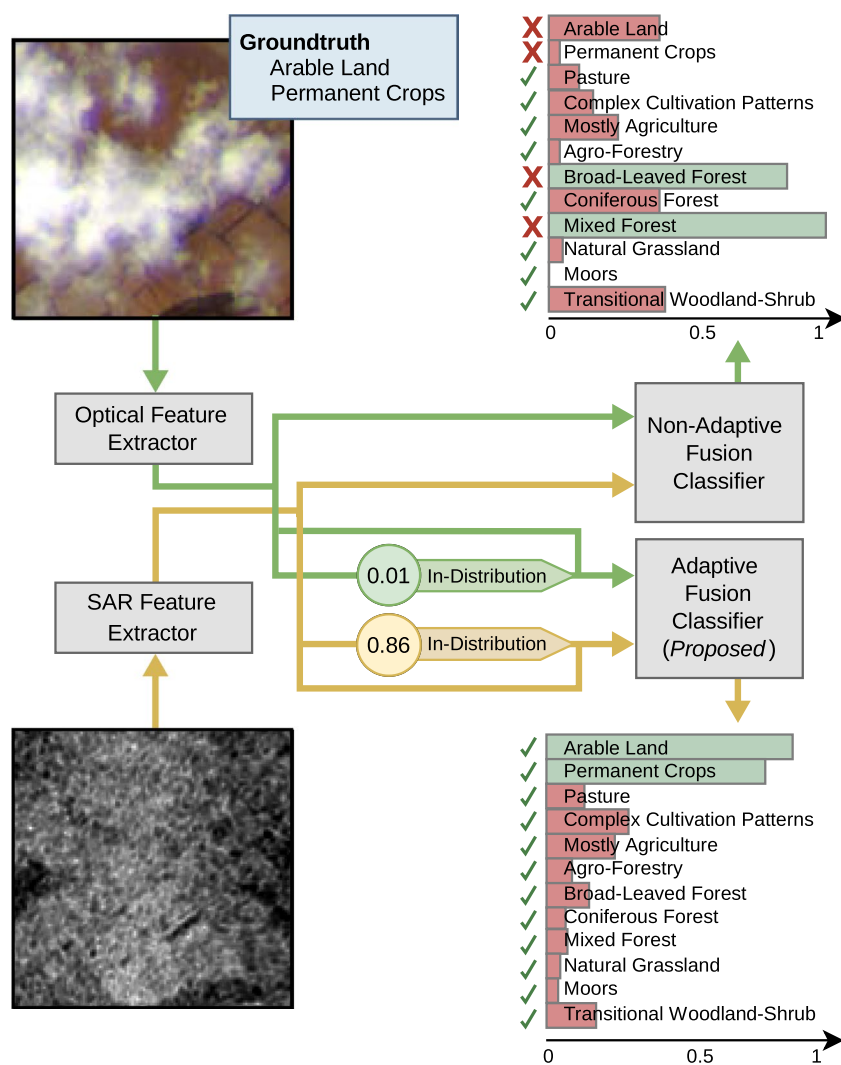


Fig. 1 A sketch of the adaptive data fusion as proposed in this work. The distributional uncertainty is determined by modality-wise out-of-distribution detectors and is considered within the adaptive fusion component. As a result, distribution shifts as for example caused by clouds have a less significant effect on the prediction. Even though the in-distribution probability for the SAR data is predicted as only 86%, the prediction is correct and significantly better than the non-adaptive model. Here, the non-adaptive model is given by a model trained equivalently as the baseline reported in [13]

of the single modality streams as input. We combine the components in such a way that the final approach is aware of the existence of uncertainty caused by distribution shifts affecting the individual sensors. The contributions of this work are as follows:

- 1 We introduce out-of-distribution detection in a data fusion setting on individual sources before the fusion step.
- 2 We propose a training and evaluation strategy for the adaptive fusion of data sources, based on the predicted in-distribution probabilities.
- 3 We show the improved robustness to distribution shifts in single sources and advanced OOD detection performances in comparison to baseline and state-of-the-art approaches.

The overall approach is not problem-agnostic and applicability to different tasks such as regression or classification can be considered. In this work, we apply and evaluate it on multi-label classification tasks. 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 rest of this paper is organized as follows. Related works are discussed in Sect. 2. We briefly discuss the proposed approach in Sect. 4. Experimental validation is detailed in Sect. 6 and discussed in Sect. 7. The paper is concluded in Sect. 8.

2 Related work

Despite its importance, not many works can be found in the literature that focus on the combination of robust data fusion approaches and strong distribution shifts leading to unusable data in single data sources. Considering the relevance to our work, we discuss multi-sensor and multi-modality data fusion in earth observation (Sec. 2.1), distribution shifts and out-of-distribution detection (Sec. 2.2) Julia, and the only few works that have worked towards combining strong distribution shifts and data fusion in the literature (Sec. 2.3).

2.1 Data fusion approaches for earth observation

With an increasing number of Earth observation data sources, data fusion has become a commonly used technique to benefit from the complementary, competitive or cooperative information [11, 22, 23]. While traditional methods are mainly based on statistics and strategies especially designed for the fusion process [24, 25], deep learning offers the possibility to stack the (pre-processed) data sources together and derive a corresponding fusion strategy in a data-driven manner [11]. In the field of Earth observation deep learning based data fusion approaches have delivered significant contributions over the last years [20, 26–29].

Many contributions focus on the design of fusion strategies. These contributions cover a variety of data sources as well as applications. For example in [30] social media data and satellite imagery are combined in order to determine building types. Others designed fusion strategy for disaster management based on satellite and social media data [31, 32] or on the spoken language and images [33]. Multi-sensor time series data has been applied for tasks as general change detection [20, 21], or crop classification [34–36]. Gao et al. [37] and Meraner et al. [38] make use of the fact that in contrast to optical data, SAR data is not affected by clouds and propose cloud removal approaches for optical images. In [39] a Bayesian modelling is used in order to improve spatial details and reduce spectral distortions within multispectral and hyperspectral based pan sharpening.

Another important task where data fusion is commonly used is the scene classification [13, 40–44]. For the scene classification on a pixel level Hong et al. introduce X-Modal-Net [28] which propagates labels and features between hyperspectral, multispectral and SAR data. In [45] an approach to align multi- and hyperspectral images and realize semi-supervised cross-modality learning for the task of pixel-wise scene classification. Other works approached the same task by second order attention based methods applied after the fusion process [46] and by including explicit domain knowledge [47].

Cross-modality learning aims to learn a joint feature space and exchange information between different data sources in order to improve the training process, as for example done in [28] or [48] for the multi-spectral and hyper-spectral image fusion. Especially in the field of multispectral and hyperspectral image fusion many approaches aim at learning a common feature representation for both sources, often in form of a dictionary [28, 45]. Further, it was shown that including (partially) cloudy images into the training process improves the classification performance for SAR-optical fusion approaches [49] or multi-temporal fusion [50]. Other works as [51] explicitly focus on evaluating different fusion strategies of multispectral and SAR data.

2.2 Distribution shifts

Data distribution shifts caused by geographic or illumination differences are highly present in the field of remote sensing [7, 52]. They are also discussed in some of the approaches discussed above. Other single sources approaches are for example the method in [53], where the authors use an augmented linear mixing model to address spectral variability of hyperspectral images, or the work of [54] where the features extracted from a convolutional and a graph neural network are fused. The problem of distribution shifts can also be tackled from the field of domain adaptation, where approaches are specifically designed to adapt to new distributions [7, 8]. Approaches in this field often focus on single data sources, like in [55] where multiple classifiers are combined to handle the distribution shift between different domains. But also data fusion approaches can be found as for example in [29] where a semi-supervised approach based on Sentinel-1 SAR and Sentinel-2 multispectral data is proposed for across-region generalization for built-up area mapping.

Such relatively small shifts in the data distribution can usually be tackled by elaborating domain-invariant features, which can be seen as learning the features which are not (or less) affected by the distribution shift. Following, methods approaching this problem in general do not consider possible (strong) distribution shifts, as for example, corrupted sensor measurements, a strong cloud coverage only at test time or unknown classes.

For a uni-modal setup, distribution shifts and out-of-distribution detection have been widely studied in the field of machine learning [1, 56, 57]. Considering the detection of out-of-distribution examples where the shift is such significant that a prediction is not possible anymore, the detection of unknown classes and input from different sensors have been evaluated based on a Dirichlet prior network [58]. In such scenarios and the single source case, the best a model can do is to express its uncertainty about the true prediction or give a prediction purely based on prior knowledge [57, 59].

2.3 Distribution shifts and data fusion

In contrast to this, multi-source approaches offer additional opportunities when only a subset of the available data sources is affected by distribution shifts. Such a scenario is for example given when working with optical and SAR data. While optical data has been shown to lead to a better performance in land cover classification tasks [12], it is also sensitive to illumination and clouds [20, 21]. While this fact is a very common motivation for the fusion of optical and SAR data [11, 12], these fusion methods are generally

designed to work with non-cloudy and well-illuminated data samples and with all data sources available.

In the field of deep learning based data fusion, only a few approaches tackle the problem of distribution shifts in single or multiple available data sources. In [60], at most one out of at least three modalities is assumed to experience a distribution shift caused by an adversarial attack, and a network was trained to identify which modality does not fit to the others. In [61] the authors propose EmbraceNet, focusing on missing modalities at test time and using a feature sampling based fusion strategy to be able to replace the features from one modality by the extracted features from another modality at test time. In [62], a meta learning procedure is introduced where missing input modalities are reconstructed from the available modalities. Based on this, missing modalities at training and test time can be simulated. For the latter two, the missing data can be interpreted as a type of distribution shift, but no actual detection of this shift is needed as it is for other distribution shifts. For the fusion of optical and SAR data, only the EmbraceNet approach seems promising since the generation of optical data from SAR images is a very challenging task [38] and the fusion procedure only consists of two modalities, i.e. no majority voting for excluding non-fitting modalities can be realized.

In the following, we propose an alternative training procedure and introduce additional OOD-detectors for the detection of distribution shifts at test time.

3 Uncertainty in neural networks

Consider a neural network $f_\theta : \mathbb{R}^I \rightarrow \mathbb{R}^C$, parameterized by parameters θ and corresponding outputs $\mu \in \mathbb{R}^C$. For a multi-label classification task with C classes and a training data set \mathcal{D} , the predictive uncertainty for an input sample $x^* \in \mathbb{R}^I$ having a corresponding label y is stated as $p(y|x^*, \mathcal{D})$. In practice, the transformation of the neural network logits μ into a probability distribution parameterization, by applying the soft-max function for classification tasks and the sigmoid function for multi-label classification tasks, aims to represent this uncertainty. But the predictive uncertainty is a composition of multiple types of uncertainties, stemming from different sources. Uncertainty quantification in neural networks is a vast field of research and multiple approaches have been proposed in order to quantify these uncertainties and give a proper predictive uncertainty. In general, uncertainties are split into two main types of uncertainty, namely data (or aleatoric) uncertainty and model (or epistemic) uncertainty [59]. The former one is caused by shortcomings in the data itself and hence cannot be reduced. The latter one is caused by shortcomings in the modeling and the training of the neural network. The epistemic uncertainty is reducible by adjusting the model and training procedure. Several works [56, 63] represent the model uncertainty in a prediction by explicitly representing it as a probability distribution over the model parameters θ :

$$p(y|x^*, \mathcal{D}) = \int p(y|x^*, \theta) \cdot p(\theta|\mathcal{D}) d\theta. \quad (1)$$

In this notation $p(\theta|\mathcal{D})$ can now be interpreted as model uncertainty and $p(y|x^*, \theta)$ as data uncertainty.

A special case of the model uncertainty is distributional uncertainty, which is caused by a shift between the training and the testing data distribution. Popular examples of distributional uncertainty are the occurrence of new classes unseen during the training or noise in the data. In the field of remote sensing, data distribution shifts have been evaluated for the cases of sensor change, spatial change, and unseen classes [58]. In the literature, several works seek to identify out-of-distribution samples by Bayesian or ensemble based approaches which quantify distributional uncertainty as a part of the epistemic uncertainty [56, 63]. As done in [1, 2] the distributional uncertainty can be extracted from the model uncertainty and is explicitly modeled as an additional distribution over the network's logits μ ,

$$p(y|x^*, \mathcal{D}) = \int p(y|\mu) \cdot p(\mu|x^*, \theta) \cdot p(\theta|\mathcal{D}) d\mu d\theta. \quad (2)$$

Here, $p(y|\mu)$ represents the data uncertainty, while $p(\mu|x^*, \theta)$ represents the distributional uncertainty, and $p(\theta|\mathcal{D})$ the model uncertainty. The distributional uncertainty depends on the network parameters only for the case that the network explicitly predicts the parameters of this distribution [1]. Approaches that quantify distributional uncertainty often target at learning a boundary around the in-distribution samples by including explicit out-of-distribution samples as a part of the training procedure [1–3].

4 Methodology

The goal of the proposed approach is to be able to quantify distributional uncertainty stemming from individual data sources and use this information to weight the input from the available sources. For this, the following components are needed and will be introduced in the following:

(1) Fusion neural network:

A deep learning structure that can be trained to give a prediction based on a SAR and an optical sample.

(2) Training strategy:

A training strategy that enables the network to give good predictions when individual data sources are excluded from the fusion step.

(3) Out-of-distribution detection:

Out-of-distribution detectors to evaluate the in-distribution probability for each data source.

(4) Fusion Strategy at Test Time:

A scheme to detect distributional uncertainty in individual data sources and also propagate it onto the final prediction.

The proposed training and testing procedure is also visualized in Fig. 2.

4.1 Underlying fusion neural network

Consider input samples of the form $x = (x_s, x_o)$. As a basic network structure we use a fusion network consisting of two input branches, $f_s(x_s) = \mu_s$ and $f_o(x_o) = \mu_o$, for the individual feature extraction from the SAR and optical data, respectively. The extracted features

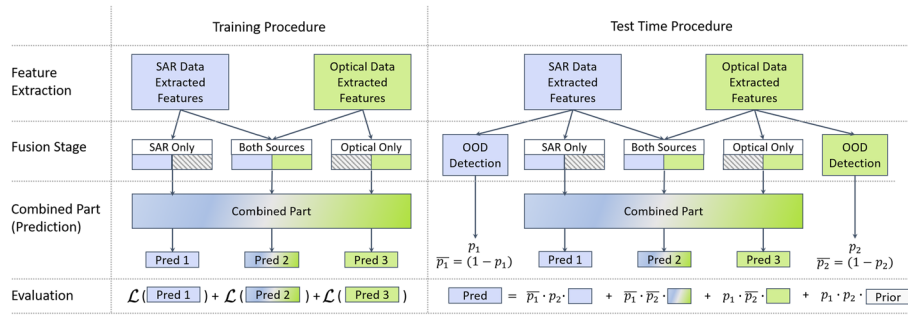


Fig. 2 A sketch of the considered network structure for the training (left) and the testing (right) time. At training time, the network receives the ground-truth and gives a prediction based on the fused features and the single modalities. At test time, modality-wise OOD detectors are used to propagate distributional uncertainty onto the fused prediction. The prior distribution represents the case where no information is available and is chosen as a probability of 0.5 for each class

are fused by a concatenation operation, $[\mu_s, \mu_o]$, and fed into a combined neural network part, $f_c([\mu_s, \mu_o]) = \hat{y}_f$, which outputs a prediction for the multi-label classification problem. In order to balance the extracted features, we scale the features to the range of $[-1, 1]$ by rescaling the sigmoid function.

At test time, we want to be able to exclude individual sources from the propagation through the combined part. Hence, a standard training strategy is not applicable but the network needs to be aware of missing features. A training strategy enabling the network to do so is presented in the next subsection.

4.2 Training strategy including source exclusion

In order to enable the network to be able to handle missing data sources, we perform multiple forward passes with masked-out features from excluded data sources. For this we introduce a masking operator, which returns all combinations of masked-out data sources, i.e.,

$$M : (\mu_s, \mu_o) \rightarrow \{[\mu_s, \mu_o], [\mu_s, 0], [0, \mu_o], [0, 0]\}, \quad (3)$$

where the masked-out features are replaced by zeros. The resulting predictions are given as $f_c([\mu_s, \mu_o]) = \hat{y}_f$, $f_c([\mu_s, 0]) = \hat{y}_s$ and $f_c([0, \mu_o]) = \hat{y}_o$.

For the evaluation performance we apply the binary cross-entropy loss, \mathcal{L}_{BCE} , the standard loss function for multi-label classification tasks. In order to also learn the missing data sources, we propose a loss function of the form

$$\begin{aligned} \mathcal{L}(y, \hat{y}_s, \hat{y}_o, \hat{y}_f) = & w_f \cdot \mathcal{L}_{\text{BCE}}(y, \hat{y}_f) \\ & + w_s \cdot \mathcal{L}_{\text{BCE}}(y, \hat{y}_s) \\ & + w_o \cdot \mathcal{L}_{\text{BCE}}(y, \hat{y}_o), \end{aligned} \quad (4)$$

with scalar weights $w_f, w_s, w_o \geq 0$. We found the approach to be robust to the choice of these hyper-parameters and ran our experiments with $w_f = 2$, $w_o = 1$, $w_s = 1$.

4.3 Out-of-distribution detection (OOD)

For the OOD detection on the individual data sources, we train a simple binary classification model consisting of two fully connected layers with 512 and 64 inner neurons, hyperbolic tangent activation and a dropout layer with dropout rate of 0.1. The OOD detectors for the SAR and the optical features as $g_s(\mu_s) \in [0, 1]$ and $g_o(\mu_o) \in [0, 1]$, representing the probability that the given features are not affected by a distribution shift. The considered OOD detectors are trained in a supervised way, this means, that out-of-distribution examples are needed in order to learn a boundary around the input data [2]. How we chose these out-of-distribution examples can be found in Sec. 5.

4.4 Adaptive OOD probability dependent fusion

In order to incorporate the out-of-distribution detection into the prediction process, we want to mask out the features based on the in-distribution probability. Thus, for an input sample $x = (x_s, x_o)$ and (branch and combined) network parameters $\theta = (\theta_b, \theta_c)$, let

$$p(y|x^*, \mathcal{D}) = \int p(y|\mu) \frac{p(\mu|\mu_s, \mu_o, \theta_c)}{p(\mu_s, \mu_o|x_s, x_o, \theta_b)} d\theta d\mu_s d\mu_o \quad (5)$$

be the predictive uncertainty and $p(\mu_s, \mu_o|x_s, x_o)$ the in-distribution probability. Following former works as [1, 2, 57], we use a deterministic point-estimation of the network parameters and hence drop the θ in the following notations. Further, we apply a strategy to use and propagate a source based on its in-distribution probability. This means, if the in-distribution probability is 80%, the sample is propagated in 80% out in 20%.

Under this setup, the feature probabilities for the SAR and the optical features that are fed into the combined part can be replaced by Bernoulli distributions of the form $p(\lambda_s|x_s, x_o)$ and $p(\lambda_o|x_s, x_o)$, where λ_s and $\lambda_o \in \{0, 1\}$ with 1 if the corresponding data source is in-distribution and 0 if it is out-of-distribution. This leads to a predictive uncertainty stated as

$$p(y|x^*, \mathcal{D}) = \int p(y|\lambda_s \cdot \mu_s, \lambda_o \cdot \mu_o) \cdot p(\lambda_s|x_s) p(\lambda_o|x_o) d\lambda_s d\lambda_o. \quad (6)$$

The formulation in (6) can be explicitly computed as

$$\begin{aligned} p(y|x^*, \mathcal{D}) &= \mathbb{E}_{\lambda_s, \lambda_o} [p(y|\lambda_s \cdot \mu_s, \lambda_o \cdot \mu_o)] \\ &= (1 - p_s) \cdot (1 - p_o) \cdot p(y|0, 0) \\ &\quad + p_s \cdot (1 - p_o) \cdot p(y|\mu_s, 0) \\ &\quad + (1 - p_s) \cdot p_o \cdot p(y|0, \mu_o) \\ &\quad + p_s \cdot p_o \cdot p(y|\mu_s, \mu_o), \end{aligned} \quad (7)$$

with $\lambda_s \sim \text{Ber}(p_s)$ and $\lambda_o \sim \text{Ber}(p_o)$. The formulation in (7) can be computed straight forward as a weighted sum of the combined network part,

$$\begin{aligned}
\hat{y} = & (1 - p_s) \cdot (1 - p_o) \cdot 0.5 \\
& + p_s \cdot (1 - p_o) \cdot \hat{y}_s \\
& + (1 - p_s) \cdot p_o \cdot \hat{y}_o \\
& + p_s \cdot p_o \cdot \hat{y}_f.
\end{aligned} \tag{8}$$

4.5 Computational complexity

The computational complexity and memory consumption is an important aspect for many applications. The only additional computational steps and memory consumption of our approach results from the multiple passes through the combined part of the network. Here the number of calculations and variables increases linearly with the number of combinations of masked data sources. For our concrete application example (i.e. two data sources) there are three possible combinations (no masked data sources, masked optical data, masked SAR data), so that three forward passes through the combined part are required for a prediction. It is important to note that in practice the combined part is often reduced to a few linear layers and that the multiple forward passes can be run in parallel as a batch.

4.6 Adaptive and non-adaptive fusion

In the following experiments we evaluate the considered baselines against two different approaches based on our proposed procedure. The full approach with training and inference as described above, including the OOD detection, we refer to *Adaptive Fusion*. Further, we also evaluate the performance only based on the proposed training strategy without additionally using the OOD detectors at test time, i.e. assuming the inputs are in-distribution and only evaluating the output \hat{y}_f . For this we refer to *Non-Adaptive Fusion*.

5 Data

5.1 BigEarthNet-MM

For our experiments we use the BigEarthNet-MM data set, which is the multi-modal version of the BigEarthNet data set [13]. It consists of 125 co-registered Sentinel-1 and Sentinel-2 tiles distributed over 10 different countries all over Europe. The authors took care of a minimal temporal difference between the acquisition of the SAR and the optical data. The collected tiles are separated into 590,326 non-overlapping pictures, with atmospheric correction on the optical Sentinel-2 images. The patches are labeled for a multi-label classification task, i.e., each patch is annotated by a subset of in total 19 classes of land-cover.

In-distribution and OOD class splits

In order to train the proposed OOD detectors we consider the following split of the 19 classes into the following subsets:

- In-Distribution: Twelve classes containing different types of vegetation (i.e., class 3 to class 14).

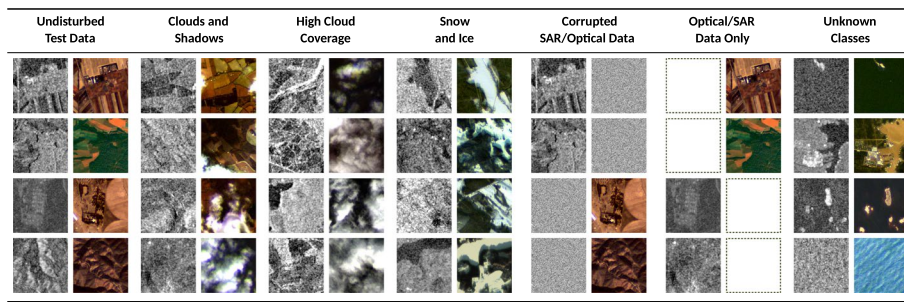


Fig. 3 Pairs of optical and SAR patches representing the considered test cases

- OOD Training Data: Two classes that contain only industrial classes (i.e., class 1 (Urban Fabric) and class 2 (Industrial)).
- OOD Testing Data: Five classes containing sand and water related classes (i.e., classes 15 to 19).

For the described splits, we excluded all patches that contain both in-distribution and OOD classes. This split leaves 149,820 in-distribution patches for training, 68,959 for validation and 70,750 for testing. For the OOD detection the split results in 4022 cloudy samples from which we selected 197 very cloudy samples in an additional subset, 23,054 samples containing snow and ice and 4,748 samples with only OOD training classes and 4,997 samples with only OOD testing classes.

Additional OOD examples

The BigEarthNet-MM data set explicitly excludes all samples that contain clouds in the optical data or where the land is covered by snow and ice. For testing our approach we make use of this and consider seven different types of OOD examples in single data sources. Example samples are visualized in Fig. 3.

- Samples with *clouds and shadows*:
4022 patches where the optical image is affected by cloud coverage in all different levels of intensities.
- Handpicked samples with high cloud coverage:
A subset of 197 samples which are fully covered by clouds.
- Samples with *Snow and Ice*:
23,054 samples where seasonal snow or ice occurs.
- Samples with *missing SAR modality*:
We simulate a missing SAR modality by using the original test data set (70750 samples) without the SAR modality.
- Simulated samples with *corrupted SAR modality*:
The SAR data of the test data set (70750 samples) is replaced by pixels sampled from a Gaussian distribution using the mean and standard deviation of the training data.
- Samples with *missing optical modality*:
We simulate a missing optical modality by using the original test data set (70750 samples) without the optical modality.
- Simulated samples with *corrupted optical modality*:

The optical data of the test data set (70750 samples) is replaced by pixels sampled from a Gaussian distribution using the mean and standard deviation of the training data.

The class distributions of the individual test cases can be found in the supplementary files contained in the code repository.

6 Experiments

6.1 Comparison of methods and training procedure

We compare the proposed fusion process to the EmbraceNet structure [61], simple baseline approaches with early and late fusion and individual modality approaches trained equivalently to the one proposed by the authors of the data set [13]. For our experiments we build the considered fusion architectures based on a ResNet18 [64]. The ResNets are structured in five convolutional blocks, followed by a linear classifier. For the baseline approach with late fusion and the EmbraceNet, the fusion takes place between the last convolutional block and the classifier part. For the baseline approach with early fusion, the data is concatenated before the first layer such that a normal ResNet is trained. The same holds for the optical-only and the SAR-only setup.

We train all approaches for 50 epochs and save the parameters for the best performances on the given validation split. We set the loss parameters to $w_o = 1$, $w_s = 1$ and $w_f = 2$. Following this, we train the OOD detectors for three epochs on the optical and the SAR branch. The OOD detectors are two layers with a tanh activation function and a scalar output representing the probability that the input is out-of-distribution. In order to get a better sensitivity to distribution shifts, we apply intensity augmentations on the OOD-data. For the optimization we use Adam with a learning rate of 0.0001.

For the evaluation, we consider the F1 and F2 scores for multi-label classification to measure the classification performance. We differentiate two different types of the F1 and F2 scores, namely the micro, macro and sample evaluation, referencing to weighting each class prediction equally, averaging the scores class-wise or averaging the scores sample-wise. For the F1 score these three metrics are defined as follows:

$$\mathbf{F1}(\mathbf{micro}) : \mathbf{F1}(\{\hat{y}_i, y_i\}_{i=1}^M) \quad (9)$$

$$\mathbf{F1}(\mathbf{macro}) : \frac{1}{C} \sum_{c=1}^C \mathbf{F1}(\{\hat{y}_{i,c}, y_{i,c}\}_{i=1}^M) \quad (10)$$

$$\mathbf{F1}(\mathbf{sample}) : \frac{1}{M} \sum_{i=1}^M \mathbf{F1}(\hat{y}_i, y_i), \quad (11)$$

where $\hat{y}_{i,c}$ is the prediction for class c based on the i -sample, $i = 1, \dots, M$, and $y_{i,c}$ is the corresponding ground-truth. For the F2 score the procedure follows the same principle. For missing modalities we assume the absence of the data to be recognized and adjust the propagation in the adaptive fusion and the EmbraceNet accordingly. For the corrupted data we assume that the noise is not recognized and the data is processed in a normal way. In order to evaluate the trustworthiness of the predictions, we state the calibration

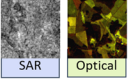
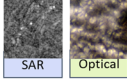
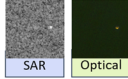
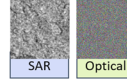



	In-Distribution	Cloudy Example	OOD Example	Noisy Optical Data
Input Data				
Labels	Arable Land, Mostly Agricultural, Coniferous Forest, Mixed Forest	Complex Cultivation Patterns, Mostly Agricultural, Transitional Woodland-Shrub	Water (out-of-distribution)	Arable Land, Coniferous Forest, Mixed Forest
In-Distribution Probability	97.92% 99.80%	98.46% 0.63%	1.64% 2.41%	98.91% 20.04%
Predictions	  	Broad-Leaved Forest	Arable Land	Arable Land, Mostly Agricultural, Mixed Forest
Adaptive Predictions	Arable Land, Mostly Agricultural, Coniferous Forest, Mixed Forest	Complex Cultivation Patterns, Mostly Agricultural	---	Arable Land, Coniferous Forest, Mixed Forest

Fig. 4 Example visualizations of different test cases: in-distribution samples, samples with the optical modality affected by clouds, out-of-distribution samples with unknown classes and a sample with corrupted optical data

error, given as the adaptive calibration error over 10 bins [65], and the mean entropy. Further, we indicate the separability of the different test sets from the original testing data set using the average under the receiver operating characteristic curve (AUROC) and the average under the precision recall curve (AUPR) evaluated with respect to the average entropy and the average confidence. For the proposed adaptive fusion approach we additionally state the AUROC and AUPR values based on the modality-wise OOD detectors.

In Fig. 4 examples for the adaptive fusion under different types of test time data distributions are visualized.

6.2 Results

In the following, we first evaluate the influence of the fusion stage on the performance of the proposed adaptive fusion approach. Following this, we compare the proposed method against other approaches on multiple different data setups. Finally, we evaluate how well the different methods allow to differentiate between in-distribution and out-of-distribution data. The results are based on five repetitions. Besides the result figures in this paper, the results can be also found as tables in Additional File 1.

6.2.1 Evaluate best fusion stage for adaptive fusion

First, we compare the adaptive fusion approach with ResNet18 backbone and fusion strategies after each of the five convolutional blocks. The results in Fig. 5 show that later fusion stages show little difference between the different stages. One can see, that the standard deviation among the five repetitions is the largest for the corrupted single

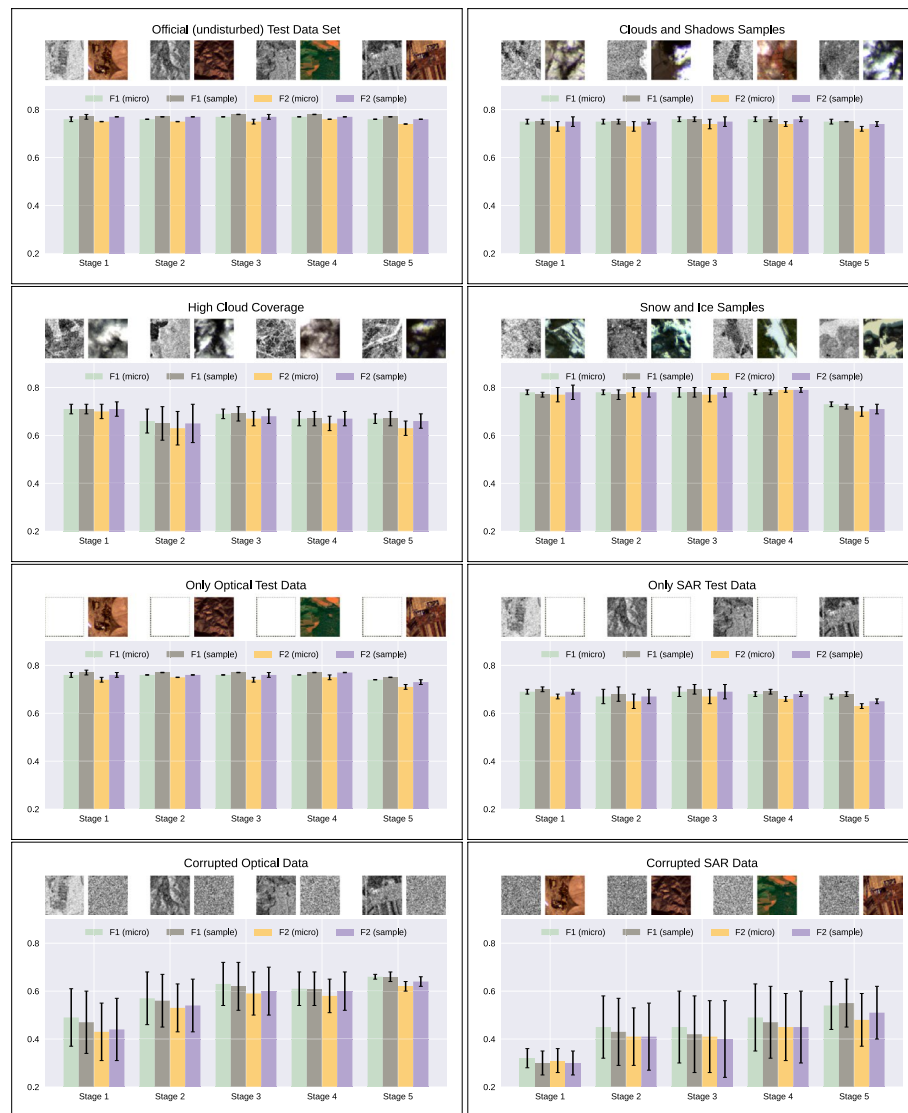


Fig. 5 Comparison of different stages of fusion for the proposed adaptive fusion approach. The results are given as mean and standard deviation over five runs. The different stages seem to be beneficial for different types of distributional uncertainty. Considering the high standard deviation on the results of some test settings, we decided to use the approach with fusion after the fourth block for the comparison against the baselines

modalities test cases. In the following, we use the fourth fusion stage, which performs slightly better than the other stages, as a representative for our adaptive fusion approach.

6.2.2 Comparison to other approaches

In Fig. 6 the performance of the proposed adaptive fusion approach (with fusion at stage four) is compared to the proposed non-adaptive strategy (also with fusion at stage four), the multi-modal baselines with early and late fusion, the uni-modal baselines and the EmbraceNet. For the original testing data set, the performance of the adaptive approach is slightly below the performance of the baseline models and the non-adaptive version,

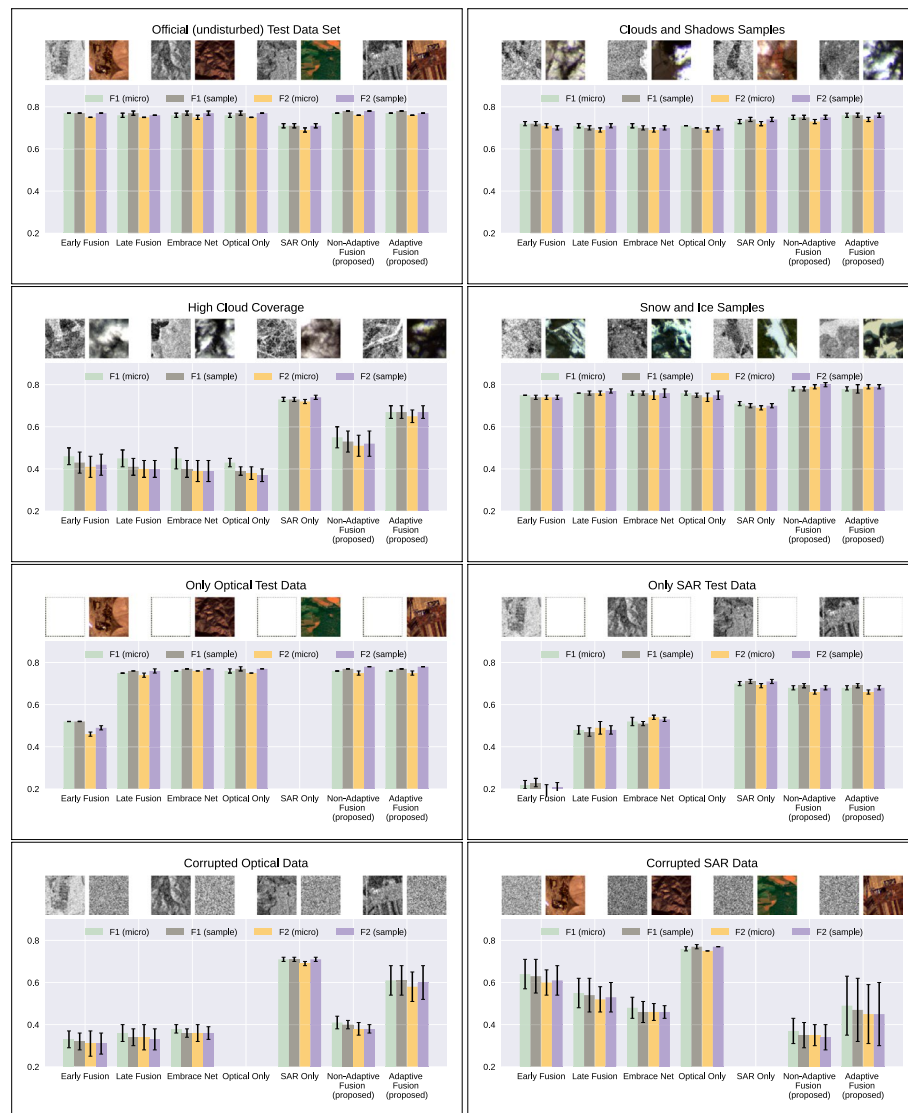


Fig. 6 Evaluation of the proposed method and the baseline approaches on different test settings. The results are given as mean and standard deviation over five runs. Based on the results in Figure 5 we use the fusion stage 4 for the proposed approaches. For the Optical Only and for the SAR Only approaches only the results with available data is listed, for the Baseline fusion approaches, the missing modalities are masked out with zeros

which performs best on the testing, the snow and ice data set the data set with all ranges of cloud coverage. The SAR only approach performs slightly worse on the original testing data and these two data sets, while the optical approach performs comparable well.

Compared to the original test data set, the performance on data that is shifted by including clouds and shadows in the optical version, results in a slightly worse prediction performance in all approaches except the SAR only approach, where the performance even improved compared to the clear test data set. The performance drop is significantly larger for the EmbraceNet, the early and late fusion baselines and the

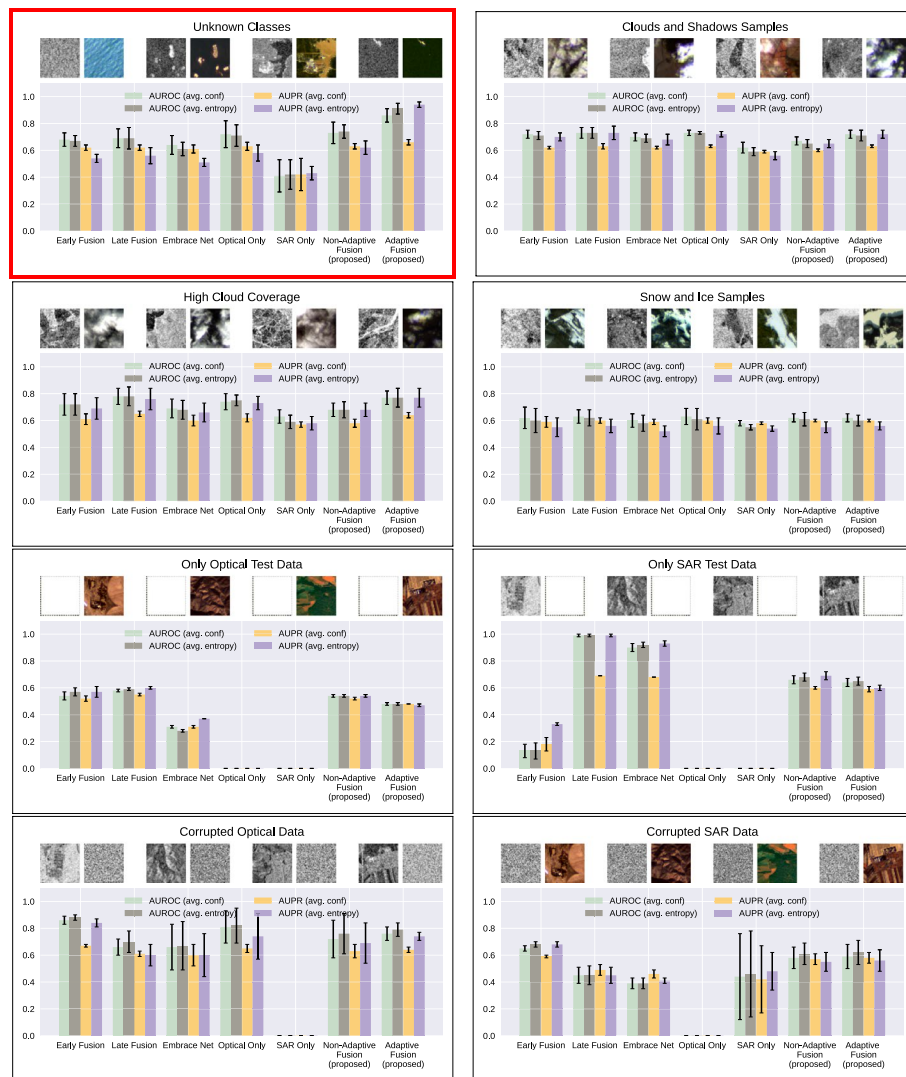


Fig. 7 Separation of in-distribution samples and OOD samples. The given values only indicate how well the considered approaches and metrics separate the given data sets from in-distribution test samples. For the cloudy data set, for example, many samples contain almost no clouds and hence the sample might not be interpreted as a clear OOD example. For the Optical Only and for the SAR Only approaches only the results with available data is listed, for the Baseline fusion approaches, the missing modalities are masked out with zeros. The thicker red box marks the case where a prediction is not possible and only the separation in in- and out-of-distribution is achievable

optical only approaches. For the high cloud coverage test case again all approaches but the SAR only approach result in worse classification performance. The proposed adaptive fusion shows the smallest drop while all other approaches except the SAR-only approach shows a significantly larger drop in the predictive performance. For the snow and ice test case, a slight decrease of the classification performance is observed for the early fusion baseline only.

For the missing modality and the data corruption experiments, the adaptive approach is ranked between the compared multi-modal baselines, which perform worse, and the

single modality approaches, which perform slightly better if the corresponding modality is available and not affected by the data corruption.

6.2.3 Out-of-distribution detection performance

The separability of in-distribution data and (possibly) OOD data is shown in Fig. 7 as the Average under the Curve Receiver Operating Characteristic (AUROC) and precision recall (AUPR) based on the average entropy and the average confidence in the network predictions. In comparison to this, the performance of the explicitly trained OOD detectors is shown in Table 1. In Fig. 7 one can see for the case of unknown classes, that the trained OOD detectors out-perform the non-supervised approaches significantly. In Table 1 one can also see, that the optical OOD detector gives a good separability especially on the high cloud coverage data set and the corrupted optical examples. For the SAR OOD detector, only for the unknown classes and the corrupted SAR data the OOD detector shows a tendency towards separating the in-distribution and OOD samples.

7 Discussion

The presented results underline that our proposed training strategy together with the quantification of distributional uncertainty on single modalities can improve the performance especially when one modality experiences a significant change in the data distribution. While the compared approaches experience a significant drop in the performance, the proposed approach suffers much less from the distribution shift. Interestingly, corrupted SAR modalities have a much smaller effect on the performance of the multi-modal baselines and on the EmbraceNet than corrupted optical data. This indicates, that these approaches focus more on the optical data than on the SAR data, what is a clear drawback when missing optical data can appear or the optical data is corrupted. At the same time, the quantification of the distributional uncertainty within

Table 1 Evaluation of the distribution separation performance of the OOD detectors which are used for the adaptive fusion approach

Data Set	Metric	Approach	
		OOD detector optical	OOD detector SAR
Clouds and Shadows	AUROC	0.73 ± 0.06	0.36 ± 0.05
	AUPR	0.80 ± 0.04	0.43 ± 0.03
Very Cloudy	AUROC	$\uparrow 0.97 \pm 0.01$	0.20 ± 0.04
	AUPR	$\uparrow 0.98 \pm 0.01$	0.35 ± 0.01
Snow and Ice	AUROC	0.59 ± 0.11	0.38 ± 0.03
	AUPR	0.61 ± 0.08	0.42 ± 0.01
Unknown Classes	AUROC	$\uparrow 1.00 \pm 0.00$	0.95 ± 0.02
	AUPR	$\uparrow 1.00 \pm 0.00$	0.95 ± 0.02
Optical Corrupted	AUROC	$\uparrow 0.97 \pm 0.02$	–
	AUPR	$\uparrow 0.93 \pm 0.05$	–
SAR Corrupted	AUROC	–	$\uparrow 0.70 \pm 0.19$
	AUPR	–	$\uparrow 0.60 \pm 0.15$

The table shows the separation performance of in-distribution samples from the test set and OOD samples for different test cases. For cases where the separation of in-distribution and OOD examples is clear, the direction of improvement is indicated by an arrow

the adaptive fusion approach, leads to a better representation of the predictive uncertainty by weighting the OOD examples less. This is not only useful in the cases where single modalities are corrupted (see for example the handpicked examples with high cloud coverage), but also for the case where samples only contain unknown classes. 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 the trained OOD detectors show the best separation between the test set and the considered OOD test sets, the most potential for possible improvements still lies in the OOD detectors, which are very simple for this example. More advanced approaches have the clear potential to further boost the performance of the proposed adaptive fusion. This can be also seen since 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 data sets and is negligible at this point.

8 Conclusion

In this work we presented an approach for the training and inference of an optical-SAR data fusion neural network which takes the occurrence of out-of-distribution examples in individual data sources at test time. We introduced an advanced training strategy and applied OOD detectors on individual data sources and propagate the distributional uncertainty of the individual data sources onto the fused prediction. The proposed method has shown great potential in making data fusion approaches more robust to distribution changes. Compared to the baseline approaches, the proposed approach significantly improved the predictive performance on OOD examples while keeping comparable performance on in-distribution examples.

9 Outlook

While we used simple OOD detectors and a multi-label classification setup, we are planning to investigate more complex out-of-distribution detection approaches for simpler classification tasks. Especially the usage of unsupervised OOD detection approaches (i.e., approaches that do not need OOD examples for the training) has a large potential to make the data fusion more applicable in real life scenarios. OOD detectors explicitly designed for specific modalities will also play an important key role in our up-coming work, as the distribution shift has been shown to be significantly harder to detect in the SAR data. Further, the evaluation of contrastive information from the different modalities and different fusion strategies, as for example weighted sums, and an evaluation of the applicability of including other sources of uncertainties (e.g., aleatoric uncertainty) during the fusion process, will be part of future research.

Abbreviations

AUPR	Average under precision recall curve
AUROC	Average under receiver operating curve
OOD	Out-of-distribution
PR	Precision recall
ROC	Receiver operating curve
SAR	Synthetic aperture radar

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13634-023-01008-z>.

Additional file 1. The file contains the experimental results presented in table form and visualizations of the data distributions for the used training and testing data sets.

Acknowledgements

Not applicable.

Author contributions

JG designed and implemented the methods and prepared the manuscript. SS supported in designing the methodology and the preparation of the manuscript. JN and XZ contributed with feedback on the methodology and the manuscript and provided the funding. All authors read and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. This research was funded by the German Aerospace Center (DLR) and the international AI4EO FutureLab. The work of X. Zhu is jointly supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. [ERC-2016-StG-714087], Acronym: **So2Sat**), by the Helmholtz Association through the Framework of Helmholtz AI (Grant number: ZT-I-PF-5-01) - Local Unit "Munich Unit @Aeronautics, Space and Transport (MASTr)" and Helmholtz Excellent Professorship "Data Science in Earth Observation - Big Data Fusion for Urban Research" (Grant number: W2-W3-100), by the German Federal Ministry of Education and Research (BMBF) in the framework of the international future AI lab "AI4EO - Artificial Intelligence for Earth Observation: Reasoning, Uncertainties, Ethics and Beyond" (Grant number: 01DD20001) and by German Federal Ministry for Economic Affairs and Climate Action in the framework of the "national center of excellence ML4Earth" (Grant number: 50EE2201C).

Availability of data and materials

The data sets generated and/or analysed during the current study are available on <https://bigearth.net/>. The code to pre-process the data and train the networks is available as a GitHub repository: https://github.com/JakobCode/OOD_DataFusion.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 4 January 2023 Accepted: 28 March 2023

Published online: 01 May 2023

References

1. A. Malinin, M. Gales, Predictive uncertainty estimation via prior networks, in Proceedings of the 32nd International Conference on Neural Information Processing Systems, pp. 7047–7058 (2018)
2. J. Nandy, W. Hsu, M. L. Lee, Towards maximizing the representation gap between in-domain & out-of-distribution examples, in H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, H. Lin, (eds.) Advances in Neural Information Processing Systems, pp. 9239–9250 (2020)
3. A. Malinin, M. Gales, Reverse kl-divergence training of prior networks: Improved uncertainty and adversarial robustness, in H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, R. Garnett, (eds.) Advances in Neural Information Processing Systems, pp. 14547–14558 (2019)
4. M. Kampffmeyer, A.-B. Salberg, R. Jenssen, Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 1–9 (2016)
5. A.M. Sayer, Y. Govaerts, P. Kolmonen, A. Lipponen, M. Luffarelli, T. Mielonen, F. Patadia, T. Popp, A.C. Povey, K. Stebel et al., A review and framework for the evaluation of pixel-level uncertainty estimates in satellite aerosol remote sensing. *Atmos. Measur. Tech.* **13**(2), 373–404 (2020)
6. A. Kendall, Y. Gal, What uncertainties do we need in Bayesian deep learning for computer vision? in Advances in Neural Information Processing Systems, pp. 5574–5584 (2017)
7. D. Tuia, C. Persello, L. Bruzzone, Domain adaptation for the classification of remote sensing data: an overview of recent advances. *IEEE Geosci. Remote Sens. Mag.* **4**(2), 41–57 (2016)
8. D. Tuia, C. Persello, L. Bruzzone, Recent advances in domain adaptation for the classification of remote sensing data. *arXiv preprint arXiv:2104.07778* (2021)

9. M. Hein, M. Andriushchenko, J. Bitterwolf, Why relu networks yield high-confidence predictions far away from the training data and how to mitigate the problem, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 41–50 (2019)
10. J. Gawlikowski, P. Ebel, M. Schmitt, X.X. Zhu, Explaining the effects of clouds on remote sensing scene classification, in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (2022)
11. J. Li, D. Hong, L. Gao, J. Yao, K. Zheng, B. Zhang, J. Chanussot, Deep learning in multimodal remote sensing data fusion: a comprehensive review. *Int. J. Appl. Earth Obs. Geoinf.* **112**, 102926 (2022)
12. S. Mahyoub, A. Fadil, E. Mansour, H. Rhinane, F. Al-Nahmi, Fusing of optical and synthetic aperture radar (sar) remote sensing data: a systematic literature review (slr). *Int. Arch. Photogram. Remote Sens. Spatial Inf. Sci.* **42**, 127–138 (2019)
13. G. Sumbul, A. de Wall, T. Kreuziger, F. Marcelino, H. Costa, P. Benevides, M. Caetano, B. Demir, V. 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)
14. B. Adriano, N. Yokoya, J. Xia, H. Miura, W. Liu, M. Matsuoka, S. Koshimura, Learning from multimodal and multitemporal earth observation data for building damage mapping. *ISPRS J. Photogramm. Remote Sens.* **175**, 132–143 (2021)
15. J. Shermeyer, D. Hogan, J. Brown, A. Van Etten, N. Weir, F. Pacifici, R. Hansch, A. Bastidas, S. Soenen, T. Bacastow, R. Lewis, Spacenet 6: Multi-sensor all weather mapping dataset, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops* (2020)
16. R. Hang, Z. Li, P. Ghamisi, D. Hong, G. Xia, Q. Liu, Classification of hyperspectral and lidar data using coupled cnns. *IEEE Trans. Geosci. Remote Sens.* **58**(7), 4939–4950 (2020)
17. V. Vielzeuf, A. Lechervy, S. Pateux, F. Jurie, Centralnet: a multilayer approach for multimodal fusion, in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pp. 0–0 (2018)
18. Y. Wang, W. Huang, F. Sun, T. Xu, Y. Rong, J. Huang, Deep multimodal fusion by channel exchanging. *Adv. Neural. Inf. Process. Syst.* **33**, 4835–4845 (2020)
19. S. Cui, A. Ma, L. Zhang, M. Xu, Y. Zhong, Map-net: Sar and optical image matching via image-based convolutional network with attention mechanism and spatial pyramid aggregated pooling. *IEEE Trans. Geosci. Remote Sens.* **60**, 1–13 (2021)
20. P. Ebel, S. Saha, X.X. Zhu, Fusing multi-modal data for supervised change detection. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **43**, 243–249 (2021)
21. S. Saha, P. Ebel, X.X. Zhu, Self-supervised multisensor change detection. *IEEE Trans. Geosci. Remote Sens.* **60**, 1–10 (2021)
22. J. Zhang, Multi-source remote sensing data fusion: status and trends. *Int. J. Image Data Fusion* **1**(1), 5–24 (2010)
23. D. Schulze-Brüninghoff, M. Wachendorf, T. Astor, Remote sensing data fusion as a tool for biomass prediction in extensive grasslands invaded by *I. polyphyllus*. *Remote Sens. Ecol. Conserv.* **7**(2), 198–213 (2021)
24. H. Nguyen, N. Cressie, A. Braverman, Spatial statistical data fusion for remote sensing applications. *J. Am. Stat. Assoc.* **107**(499), 1004–1018 (2012)
25. L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, M. Selva, Multispectral and panchromatic data fusion assessment without reference. *Photogram. Eng. Remote Sens.* **74**(2), 193–200 (2008)
26. W. Han, J. Li, S. Wang, X. Zhang, Y. Dong, R. Fan, X. Zhang, L. Wang, Geological remote sensing interpretation using deep learning feature and an adaptive multi-source data fusion network, in *IEEE Transactions on Geoscience and Remote Sensing* (2022)
27. Y. Han, Y. Liu, Z. Hong, Y. Zhang, S. Yang, J. Wang, Sea ice image classification based on heterogeneous data fusion and deep learning. *Remote Sens.* **13**(4), 592 (2021)
28. D. Hong, N. Yokoya, G.-S. Xia, J. Chanussot, X.X. Zhu, X-modalnet: a semi-supervised deep cross-modal network for classification of remote sensing data. *ISPRS J. Photogramm. Remote Sens.* **167**, 12–23 (2020)
29. S. Hafner, Y. Ban, A. Nascetti, Unsupervised domain adaptation for global urban extraction using sentinel-1 sar and sentinel-2 msi data. *Remote Sens. Environ.* **280**, 113192 (2022)
30. M. Häberle, E.J. Hoffmann, X.X. Zhu, Can linguistic features extracted from geo-referenced tweets help building function classification in remote sensing? *ISPRS J. Photogramm. Remote Sens.* **188**, 255–268 (2022)
31. N. Algiriyage, R. Prasanna, K. Stock, E.E. Doyle, D. Johnston, Multi-source multimodal data and deep learning for disaster response: a systematic review. *SN Comput. Sci.* **3**(1), 1–29 (2022)
32. Z. Ahmad, R. Jindal, N. Mukuntha, A. Ekbal, P. Bhattacharyya, Multi-modality helps in crisis management: an attention-based deep learning approach of leveraging text for image classification. *Expert Syst. Appl.* **195**, 116626 (2022)
33. G. Mao, Y. Yuan, L. Xiaoqiang, Deep cross-modal retrieval for remote sensing image and audio, in *2018 10th IAPR Workshop on Pattern Recognition in Remote Sensing (PRRS)*, pp. 1–7 (2018). <https://doi.org/10.1109/PRRS.2018.8486338>
34. A. Farooq, X. Jia, J. Hu, J. Zhou, Transferable convolutional neural network for weed mapping with multisensor imagery. *IEEE Trans. Geosci. Remote Sens.* **60**, 1–16 (2021)
35. M. Liu, B. Fu, D. Fan, P. Zuo, S. Xie, H. He, L. Liu, L. Huang, E. Gao, M. Zhao, Study on transfer learning ability for classifying marsh vegetation with multi-sensor images using deeplabv3+ and hrnet deep learning algorithms. *Int. J. Appl. Earth Obs. Geoinf.* **103**, 102531 (2021)
36. Z. Li, G. Chen, T. Zhang, A cnn-transformer hybrid approach for crop classification using multitemporal multisensor images. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **13**, 847–858 (2020)
37. J. Gao, Q. Yuan, J. Li, H. Zhang, X. Su, Cloud removal with fusion of high resolution optical and sar images using generative adversarial networks. *Remote Sens.* **12**(1), 191 (2020)
38. P. Ebel, A. Meraner, M. Schmitt, X.X. Zhu, Multisensor data fusion for cloud removal in global and all-season sentinel-2 imagery. *IEEE Trans. Geosci. Remote Sens.* **59**(7), 5866–5878 (2020)
39. P. Guo, P. Zhuang, Y. Guo, Bayesian pan-sharpening with multiorder gradient-based deep network constraints. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **13**, 950–962 (2020)

40. S. Talukdar, P. Singha, S. Mahato, S. Pal, Y.-A. Liou, A. Rahman, Land-use land-cover classification by machine learning classifiers for satellite observations-a review. *Remote Sens.* **12**(7), 1135 (2020)
41. T. Hoese, C. Kuenzer, Object detection and image segmentation with deep learning on earth observation data: a review-part I: Evolution and recent trends. *Remote Sens.* **12**(10), 1667 (2020)
42. B. Chen, B. Huang, B. Xu, Multi-source remotely sensed data fusion for improving land cover classification. *ISPRS J. Photogramm. Remote. Sens.* **124**, 27–39 (2017)
43. Y. Xu, B. Du, L. Zhang, D. Cerra, M. Pato, E. Carmona, S. Prasad, N. Yokoya, R. Hänsch, B. Le Saux, Advanced multi-sensor optical remote sensing for urban land use and land cover classification: outcome of the 2018 IEEE grss data fusion contest. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **12**(6), 1709–1724 (2019)
44. N. Joshi, M. Baumann, A. Ehammer, R. Fensholt, K. Grogan, P. Hostert, M.R. Jepsen, T. Kuemmerle, P. Meyfroidt, E.T. Mitchard et al., A review of the application of optical and radar remote sensing data fusion to land use mapping and monitoring. *Remote Sens.* **8**(1), 70 (2016)
45. D. Hong, N. Yokoya, N. Ge, J. Chanussot, X.X. Zhu, Learnable manifold alignment (lema): a semi-supervised cross-modality learning framework for land cover and land use classification. *ISPRS J. Photogramm. Remote. Sens.* **147**, 193–205 (2019). <https://doi.org/10.1016/j.isprsjprs.2018.10.006>
46. X. Li, L. Lei, Y. Sun, M. Li, G. Kuang, Multimodal bilinear fusion network with second-order attention-based channel selection for land cover classification. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **13**, 1011–1026 (2020)
47. Y. Li, Y. Zhou, Y. Zhang, L. Zhong, J. Wang, J. Chen, Dkdfn: domain knowledge-guided deep collaborative fusion network for multimodal unimodal remote sensing land cover classification. *ISPRS J. Photogramm. Remote. Sens.* **186**, 170–189 (2022)
48. D. Hong, L. Gao, N. Yokoya, J. Yao, J. Chanussot, Q. Du, B. Zhang, More diverse means better: multimodal deep learning meets remote-sensing imagery classification. *IEEE Trans. Geosci. Remote Sens.* **59**(5), 4340–4354 (2021). <https://doi.org/10.1109/TGRS.2020.3016820>
49. R. Zhang, X. Tang, S. You, K. Duan, H. Xiang, H. Luo, A novel feature-level fusion framework using optical and sar remote sensing images for land use/land cover (lulc) classification in cloudy mountainous area. *Appl. Sci.* **10**(8), 2928 (2020)
50. M. Rußwurm, M. Körner, Convolutional lstms for cloud-robust segmentation of remote sensing imagery, in *Proceedings of the Conference on Neural Information Processing Systems Workshops (NeurIPSW)* (2018)
51. H. Zhang, R. Xu, Exploring the optimal integration levels between sar and optical data for better urban land cover mapping in the pearl river delta. *Int. J. Appl. Earth Obs. Geoinf.* **64**, 87–95 (2018)
52. C. Luo, L. Ma, Manifold regularized distribution adaptation for classification of remote sensing images. *IEEE Access* **6**, 4697–4708 (2018)
53. D. Hong, N. Yokoya, J. Chanussot, X.X. Zhu, An augmented linear mixing model to address spectral variability for hyperspectral unmixing. *IEEE Trans. Image Process.* **28**(4), 1923–1938 (2019). <https://doi.org/10.1109/TIP.2018.2878958>
54. D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, J. Chanussot, Graph convolutional networks for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **59**(7), 5966–5978 (2021). <https://doi.org/10.1109/TGRS.2020.3015157>
55. H. Wei, L. Ma, Y. Liu, Q. Du, Combining multiple classifiers for domain adaptation of remote sensing image classification. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **14**, 1832–1847 (2021)
56. B. Lakshminarayanan, A. Pritzel, C. Blundell, Simple and scalable predictive uncertainty estimation using deep ensembles, in *Advances in Neural Information Processing Systems*, pp. 6402–6413 (2017)
57. J. Bitterwolf, A. Meinke, M. Hein, Certifiably adversarially robust detection of out-of-distribution data. *Adv. Neural. Inf. Process. Syst.* **33**, 16085–16095 (2020)
58. J. Gawlikowski, S. Saha, A. Kruspe, X.X. Zhu, An advanced dirichlet prior network for out-of-distribution detection in remote sensing, in *IEEE Transactions on Geoscience and Remote Sensing* (2022)
59. E. Hüllermeier, W. Waegeman, Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. *Mach. Learn.* **110**, 457–506 (2021)
60. K. Yang, W.-Y. Lin, M. Barman, F. Condessa, Z. Kolter, Defending multimodal fusion models against single-source adversaries, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3340–3349 (2021)
61. J.-H. Choi, J.-S. Lee, Embracenet: a robust deep learning architecture for multimodal classification. *Inf. Fusion* **51**, 259–270 (2019)
62. M. Ma, J. Ren, L. Zhao, S. Tulyakov, C. Wu, X. Peng, Smil: Multimodal learning with severely missing modality, in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, pp. 2302–2310 (2021)
63. Y. Gal, Z. Ghahramani, Bayesian convolutional neural networks with bernoulli approximate variational inference. *arXiv preprint arXiv:1506.02158* (2015)
64. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778 (2016)
65. C. Guo, G. Pleiss, Y. Sun, Weinberger, K.Q.: On calibration of modern neural networks, in *International Conference on Machine Learning*, pp. 1321–1330 (2017). PMLR

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.