



# Toward Targeted Change Detection with Heterogeneous Remote Sensing Images for Forest Mortality Mapping

Jørgen A. Agersborg, Luigi T. Luppino, Stian Normann Anfinsen & Jane Uhd Jepsen

To cite this article: Jørgen A. Agersborg, Luigi T. Luppino, Stian Normann Anfinsen & Jane Uhd Jepsen (2022): Toward Targeted Change Detection with Heterogeneous Remote Sensing Images for Forest Mortality Mapping, Canadian Journal of Remote Sensing, DOI: 10.1080/07038992.2022.2135497

To link to this article: <https://doi.org/10.1080/07038992.2022.2135497>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 20 Oct 2022.



Submit your article to this journal [↗](#)







View related articles [↗](#)



View Crossmark data [↗](#)

# Toward Targeted Change Detection with Heterogeneous Remote Sensing Images for Forest Mortality Mapping

## Vers une détection ciblée de changements à l'aide d'images de télédétection hétérogènes pour la cartographie de la mortalité sylvestre

Jørgen A. Agersborg<sup>a</sup> , Luigi T. Luppino<sup>a</sup> , Stian Normann Anfinsen<sup>a,b</sup> , and Jane Uhd Jepsen<sup>c</sup> 

<sup>a</sup>Department of Physics and Technology, UiT The Arctic University of Norway, PO Box 6050 Langnes, N-9037 Tromsø, Norway;

<sup>b</sup>NORCE Norwegian Research Centre, PO Box 6434, N-9294 Tromsø, Norway; <sup>c</sup>Norwegian Institute for Nature Research, PO Box 6606 Langnes, N-9296 Tromsø, Norway

### ABSTRACT

Several generic methods have recently been developed for change detection in heterogeneous remote sensing data, such as images from synthetic aperture radar (SAR) and multi-spectral radiometers. However, these are not well-suited to detect weak signatures of certain disturbances of ecological systems. To resolve this problem we propose a new approach based on image-to-image translation and one-class classification (OCC). We aim to map forest mortality caused by an outbreak of geometrid moths in a sparsely forested forest-tundra ecotone using multisource satellite images. The images preceding and following the event are collected by Landsat-5 and RADARSAT-2, respectively. Using a recent deep learning method for change-aware image translation, we compute difference images in both satellites' respective domains. These differences are stacked with the original pre- and post-event images and passed to an OCC trained on a small sample from the targeted change class. The classifier produces a credible map of the complex pattern of forest mortality.

### RÉSUMÉ

Plusieurs méthodes génériques de détection de changements à partir d'images satellites issues de sources hétérogènes (radar à synthèse d'ouverture, optique, etc.) ont été développées récemment. Cependant, celles-ci sont rarement adaptées à la détection des signatures spectrales peu distinctives de certaines perturbations des systèmes écologiques. Pour remédier à ce problème, nous proposons une nouvelle approche basée sur le transfert d'image et un algorithme de classification à une classe (OCC). Notre objectif est de cartographier l'effet d'une épidémie de papillons géométrides dans un écotone forêt-toundra peu boisé en utilisant des images satellites multisources. Les images précédant et suivant l'événement proviennent de Landsat-5 et RADARSAT-2, respectivement. En utilisant une méthode récente d'apprentissage profond de transfert d'images sensible aux changements, nous calculons les images de différence dans les domaines respectifs des deux satellites. Ces images de différences accompagnées des deux images originales sont alors traitées par un OCC qui peut être entraîné sur un petit échantillon de la classe des changements souhaitée. L'algorithme de classification produit une carte crédible de la complexité de la mortalité forestière.

### ARTICLE HISTORY

Received 25 March 2022


Accepted 15 September 2022

## Introduction

The forest-tundra ecotone, the sparsely forested transition zone between northern-boreal forest and low arctic tundra, is changing rapidly with a warming climate (CAFF 2013). In particular, changes in the distribution of woody vegetation cover through shrub encroachment, tree line advance, and altered pressure

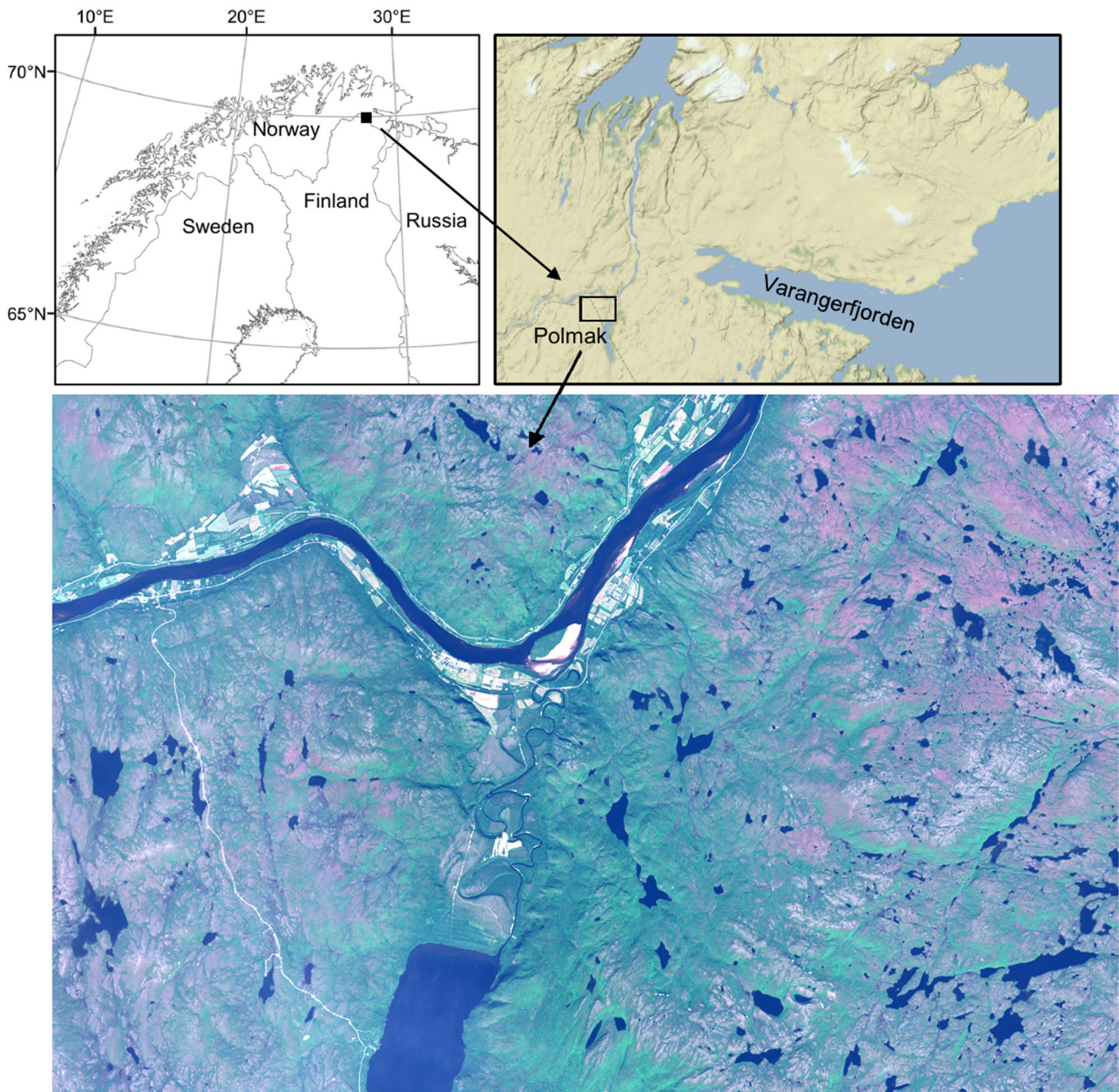
from browsers and forest pests, modify the structural and functional attributes of the forest-tundra ecotone with implications for biodiversity and regional climate feedback.

In Northern Norway, mountain birch (*Betula pubescens* var. *pumila*) forms the treeline ecotone toward the treeless Low Arctic tundra. In this region,

**CONTACT** Jørgen A. Agersborg  [jorgen.agersborg@uit.no](mailto:jorgen.agersborg@uit.no)

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Figure 1.** A true color Sentinel-2 image (July 26, 2017) of the Polmak study site, and maps showing its location on the Norwegian-Finnish border.

periodic outbreaks by forest defoliators are the most important natural disturbance factor and the only cause of large-scale forest die-off. In recent decades, such pest outbreaks have been intensified due to range expansions thought to be linked to more benign climatic conditions (Jepsen et al. 2008, 2011). Today, two species of geometrid moths; the autumnal moth (*Epirrita autumnata*) and the winter moth (*Operophtera brumata*) have overlapping population outbreaks at approximately decadal intervals, some of which cause regional scale defoliation and tree and shrub mortality in the forest-tundra ecotone. Outbreaks can thus lead to a reduction in forested areas as well as

cascading effects on other species (Biuw et al. 2014; Henden et al. 2020; Pedersen et al. 2021).

The Climate-ecological Observatory for Arctic Tundra (COAT) (Ims et al. 2013) is an adaptive, long-term monitoring program, aimed at documenting climate change impacts in Arctic tundra and tree-line ecosystems in Arctic Norway. One of the COATs monitoring sites, shown in Figure 1, is located near lake Polmak, partially on the Norwegian side and partially on the Finnish side of the border (28.0° E, 70.0° N). The chosen study site is subject to different reindeer herding regimes, where the area on the Finnish side of the border is grazed all year round (but mostly

during summer), while on the Norwegian side the region is mainly winter grazed (Biuw et al. 2014). The site's subarctic birch forest suffered a major outbreak by both autumnal moth and winter moth between 2006 and 2008, with effects that are still clearly visible in the form of high stem mortality (Biuw et al. 2014).

Remote sensing imagery is an important tool to observe and understand changes in the forest-tundra ecotone, both for large-scale monitoring and mapping on a local scale. In this work, we develop a method to find areas with forest mortality after the geometrid moth outbreak, based on satellite images and limited ground reference data. This is a challenging task for several reasons, where three significant factors stand out and guide our approach to solve the problem. Firstly, there are few remote sensing images available from our study site. This is due to the high cloud coverage at high latitudes of subarctic Fennoscandia, which limits the imaging opportunities of optical satellites. The available cloud-free optical images are relatively few and far between and consecutive images are often from different sensors. Synthetic aperture radar (SAR) is an active sensor largely uninfluenced by clouds, which can be utilized to monitor defoliation and deforestation (Bae et al. 2022; Perbet et al. 2019). However, the planned acquisition of SAR images of the Polmak study site did not start until after the outbreak. Detecting changes between images from different modalities (e.g., SAR and optical), and even between two images from different sensors of the same modality, is very challenging. If these challenges can be overcome, heterogeneous change detection would enable us to use all available historical data sources for long-term monitoring and increase the temporal resolution and responsiveness of the analysis. The second factor is that changes in the canopy state are difficult to detect in medium-resolution imagery. At this scale we do not observe the aggregated landscape level effect as in low-resolution satellite images, nor are the individual canopies visible as in high-resolution aerial photographs. For optical images, the loss of "greenness" or normalized difference vegetation index (NDVI) response caused by forest mortality can be offset by the understory vegetation that becomes increasingly visible as the canopy disappears. For SAR imagery, the change in scattering mechanisms may help detect forest mortality. However, depending on the forest density, these changes can be very subtle compared to other changes in the scene. Thus, when using unsupervised change detection methods, the presence of other man-made or natural changes can easily drown out less distinct

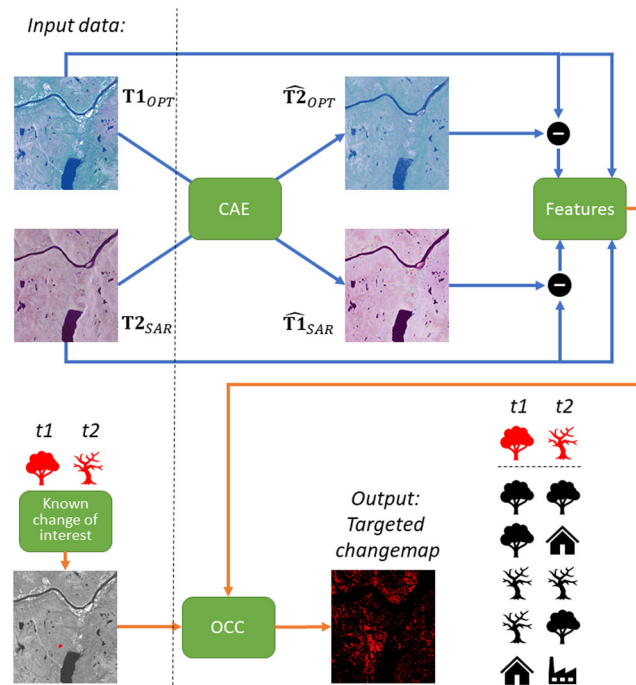
signs of forest mortality. Unsupervised methods tend to detect these strong change signatures and attenuate the weaker ones. Masking of pixels susceptible to such changes by manual inspection, creation of detailed databases of such areas, or pre-classification of images would add another layer of complexity to the task. The accuracy of the final detection result would also be very dependent on the ability of the masking operation to find all relevant areas, and the spatial resolution would be limited to that of the mask. Furthermore, it does not prevent other uninteresting changes in vegetation (i.e., not related to the canopy state) from appearing in an unsupervised result. We, therefore, create a targeted change detection algorithm to learn the change signature for forest mortality based on field data. The third factor is that learning these signatures with a supervised approach requires labeled data to train the classifier. While an ecological, field-based, monitoring of forest structure and forest regeneration in the Polmak study site was started in 2011, the scale of the field data is unsuitable to generate training labels for medium-resolution satellite images and does not cover the full extent of all ground cover classes contained in the images. To label data for all classes would be a tedious manual process, and we want to generate just enough training data for our application of detecting forest mortality since exhaustive classification is unnecessary for our application and could adversely affect the classification accuracy for the class of interest.

We, therefore, select one-class classification (OCC) to delineate the targeted change in an approach with two main steps:

1. Change-aware image-to-image translation that allows direct comparison of heterogeneous pre- and post-event images through differencing.
2. OCC applied to a stack of difference images (from step 1) and original input images to detect defoliation.

We use the recently developed code-aligned autoencoders (CAE) algorithm (Luppino et al. 2022) to do the image-to-image translation. CAE performs unsupervised change detection. However, since it is based on obtaining the difference images, it can be used to translate images between domains. Furthermore, since it is designed with change detection in mind, the network learns to preserve the changes in the translations.

OCC is a semi-supervised learning approach that utilizes the available labels but does not require a big training set or access to labels for all classes. By



**Figure 2.** Illustration of our approach. The input optical and SAR images at times  $t_1$  and  $t_2$  are translated to the other domain using code-aligned autoencoders (CAE). The originals and the differences with the translated versions are stacked as feature vectors for every image pixel. A limited amount of training data in the form of known areas with forest mortality is then used to train a one-class classifier (OCC) to map the change for a large area in the presence of both unchanged areas and other changes.

learning the change signature of the phenomenon of interest, we can perform targeted change detection by solving a classification problem with limited and incomplete ground reference data. For OCC we select a flexible approach that utilizes all available ground reference data, i.e., also from outside the class of interest. Our approach is summarized in Figure 2.

The main contributions of this work are:

- We propose a method to detect a specific change in heterogeneous remote sensing images based on limited ground reference data and in presence of other changes.
- We adapt a deep learning method recently developed for unsupervised change detection to translate the images between domains while preserving the changes.
- We adapt a method developed to identify reliable negative samples in OCC for the text domain to work on image data.
- We analyze the effect of the number of labeled samples on benchmark datasets and show that relatively little training data is needed to achieve a good classification.
- We provide an ablation study for the various components of our method using benchmark datasets.
- Our approach allows a *post-hoc* study of change events, that is study areas that were not originally

subject to persistent monitoring, by using any available satellite imagery combination from before and after the event.

- Our method allows more responsive detection of changes in areas with high cloud cover.

## Theory and related work

In this section, we provide an overview of relevant theory for our application and work related to the detection of canopy damage caused by defoliating insects.

### Remote sensing of insect-induced canopy defoliation

Previous studies of canopy defoliation in the forest-tundra ecotone have mostly utilized low-resolution (250 m) MODIS data (Biuw et al. 2014; Jepsen et al. 2013, 2009; Olsson et al. 2016a, 2016b). This agrees with the findings in a review of remote sensing of forest degradation by Gao et al. (2020), which shows the prevalence of optical data in general and MODIS in particular. A literature review by Senf et al. (2017) found that studies of disturbances by broadleaved defoliators mainly used low or medium-resolution data, with Landsat being the most used sensor. These works typically use spectral indices and dense time series to detect the defoliation. Senf et al. (2017)

found that 82% of studies mapping insect disturbance of broadleaved forest used a single spectral index, typically NDVI. This is consistent with the observation by Hall et al. (2016) that image differencing of vegetation indices derived from spectral band ratios were most frequently used. However, problems with the low resolution of MODIS for mapping forest insect disturbance in fragmented Fennoscandian forest landscape were emphasized by Olsson et al. (2016a). Limitations of low spatial resolution sensors for detection of pest damage were also pointed out by Hall et al. (2016), due to the large number of different surface materials that can be contained in a pixel, only some of which are affected by the outbreak. To rely on a single spectral index makes results susceptible to changes from other sources than forest mortality or dependent on accurate forest masks to avoid this. The sole use of NDVI also ignores the information contained in other bands of the sensor.

When it comes to the use of SAR data, none of the studies of broadleaf defoliation listed by Senf et al. (2017) used SAR, nor did the works on remote sensing for assessing forest pest damage reviewed by Hall et al. (2016). A study by Mitchell et al. (2017) of approaches for monitoring forest degradation did include several SAR data applications. However, these were mainly for L- and X-band SAR data. Most of the summarized papers dealt with the scenario where entire trees (stems and branches) had been removed, for instance by fires or logging, or they used proxy indicators, such as the detection of forest roads, to monitor degradation (Mitchell et al. 2017). It was found that only a few studies had investigated the use of C-band data (Mitchell et al. 2017). The study of C-band SAR remains interesting, particularly because the Sentinel-1 satellites provide free data in that frequency band.

A study of insect-induced defoliation using C-band SAR was presented by Bae et al. (2022), which calculated the correlation between defoliation risk and smoothed time series of backscatter values averaged over five hectares (50,000 m<sup>2</sup>) plots. In a precursor to this work, we discriminated between live and dead canopy based on an accurate estimation of polarimetric covariance from a single, full-polarimetric C-band image (Agersborg et al. 2021).

### **Heterogeneous change detection**

Traditionally, change detection has been based on images from the same sensor and preferably with the same acquisition geometry before and after the change

event. This puts some limitations on the use of such methods. Firstly, the response time is at a minimum limited to the revisit time of that particular satellite or constellation. Secondly, the area of interest (AOI) might not be covered frequently by the sensor. Furthermore, the time period which can be studied is also limited to the active time period for that particular sensor. One way to alleviate these issues is to use heterogeneous change detection on imagery from two different sensors. This comes at the cost of solving a problem that is methodologically more challenging, especially in the unsupervised case, but is still our chosen solution given the practical constraints.

To enable reliable, long-term, persistent monitoring of forest mortality in our AOI, we need to use images from different sensors due to frequent cloud cover. SAR sensors have significantly higher imaging opportunities because of their near-weather-independent nature. For our AOI, we do not have SAR imagery before the geometrid moth outbreak, and we thus have to use Landsat data for the pre-event image. Furthermore, we do not have enough data to use time series for smoothing or monitoring gradual changes, as in the studies based on low-resolution MODIS NDVI (Biuw et al. 2014; Jepsen et al. 2009, 2013; Olsson et al. 2016a, 2016b) or the approach by (Bae et al. 2022). Hence, we must rely on bi-temporal change detection using pairs of images.

Heterogeneous change detection has received growing interest in the last years (Sun et al. 2021; Touati et al. 2020). Fully heterogeneous change detection should work in a range of settings, from the easiest where the images are acquired with the same sensor, but with different sensor parameters or under different environmental conditions, to the most challenging where images are obtained by sensors that use different physical measurement principles (e.g., SAR and optical) (Sun et al. 2021). The advances of deep learning have in recent years opened up several new directions for heterogeneous change detection. Image-to-image translation in particular offers the interesting prospect of comparing the images directly, once one or both have been translated to the opposite domain. Many traditional change detection methods involve a step for homogenizing the images, such as radiometric calibration, even for images from the same sensor (Coppin et al. 2004). Image-to-image translation can be seen as an evolution of this traditional preprocessing step; one (or both) images are “re-imagined” to what it would look like in the other domain.

## Image-to-image translation

Code-aligned autoencoders (CAE) is a recently developed method for general purpose change detection designed to work with heterogeneous remote sensing data (Luppino et al. 2022). The change detection is based on translating images  $\mathbf{U}$  and  $\mathbf{V}$  into  $\hat{\mathbf{V}}$  and  $\hat{\mathbf{U}}$ , respectively, such that the difference images  $\mathbf{U} - \hat{\mathbf{U}}$  and  $\mathbf{V} - \hat{\mathbf{V}}$  can be computed. The CAE algorithm must ensure that changed areas are not considered when learning the translation across domains to avoid falsely aligning the changed areas. CAE identifies changed areas in an unsupervised manner and reduces their impact on the learned image-to-image translation. The unsupervised nature of the algorithm means that we do not need training data. While CAE performs change detection directly, it will detect all changes between the two acquisition times, such as those caused by human activity, fluctuating water levels, seasonal vegetation changes, etc., and not just forest mortality. We, therefore, utilize it as an image-to-image translation method that does not require labeled training data, is change-aware, and is designed for heterogeneous remote sensing data.

## One-class classification

Canopy defoliation has a weak *change signature* compared to many other vegetation changes that can be discerned in medium-resolution imagery, meaning that the imposed change of the radiometric signal is much smaller than for disturbances, such as e.g., forest fires or clear cutting. Thus, it will not in general be detected by the CAE. An unsupervised change detection method will tend to highlight certain strong changes and ignore weaker ones. If we lower the threshold for detection, the forest mortality could be found, but it would be surrounded and accompanied by many unrelated changes in the final change map. We, therefore, use a semi-supervised approach to detect the change phenomenon of interest while ignoring all other changes.

Traditional supervised classification algorithms require that all classes that occur in the dataset are exhaustively labeled (Li et al. 2011). To obtain sufficient training data from all possible classes by manual labeling would be both time-consuming and costly (Li et al. 2011), and does not necessarily improve the classification accuracy with respect to our class of interest. While we cannot collect ground reference data for all classes, we still want to utilize the available data to train a change detection algorithm to find changes in a larger region extending beyond the study area. We instead use techniques from one-class classification (OCC) to learn the change signature from a very

limited number of labeled samples of ground reference data containing forest mortality information.

OCC is framed as a binary classification problem, where the class of interest is referred to as the positive class (or target class), with the label  $y = 1$ . In this setting, the negative class,  $y = 0$ , is either absent from the training data or the instances available “do not form a statistically representative sample” (Khan and Madden 2014). The negative class is usually a mixture of different classes (Li et al. 2020), defined as the complement of the positive class. The typical case for OCC is that the full dataset  $\mathbb{X}$  is divided into  $\mathbb{P}$ , the set of labeled positive samples, and an unlabeled set  $\mathbb{U}$  (also called mixed set) that consists of data from both the positive and the negative class. OCC seeks to build classifiers that work in such scenarios, which often naturally arise in real-world applications (Bekker and Davis 2020). In general, the results will not be as good as in true binary classification, where statistically representative samples from both the positive and negative classes are available for training.

In our case, the reference data is not randomly sampled, but selected systematically from a spatially limited area with a sampling bias that is somewhat related to landscape attributes. The lack of random sampling limits our choice of OCC algorithm to the so-called *two-step techniques* and discards the other possibilities mentioned in the taxonomy of Bekker and Davis (2020). Two-step techniques only require two quite general assumptions: smoothness and separability (Bekker and Davis 2020). Smoothness means that samples that are close are more likely to have the same label, while separability means that the two classes of interest can be separated (Bekker and Davis 2020). The steps are:

1. Given labeled positive samples  $\mathbb{P}$  from a dataset  $\mathbb{X}$ , reliable negative (RN) samples are identified from the remaining set of unlabeled samples  $\mathbb{U}$ .
2. A classifier is trained with the provided labeled positive samples  $\mathbb{P}$  and using the RN samples from Step 1 as the (initial) set of negative training data,  $\mathbb{N}$ .

Any (semi-)supervised classifier can be used in Step 2, as both positive and RN labels are available.

## Methodology

### Feature selection

Feature engineering is an important part of machine learning, as selecting the right features and

normalization may have a big influence on the final classification result. For our bitemporal change detection problem, we originally have the co-registered pre-event ( $\mathbf{U}$ ) and post-event ( $\mathbf{V}$ ) images, which may be from different sensors and even different physical measurement principles. We want the feature vectors to be as descriptive as possible since we have a very limited amount of training data, and only from the change class of interest. A common change detection technique for homogeneous images is to subtract one image from the other to obtain the difference image,  $\mathbf{D} \in \mathbb{R}^{h \times w \times c}$ . The exact steps for finding the changes from the features of  $\mathbf{D}$  vary, but often involve a form of thresholding.

We seek to combine original image features and difference vectors for heterogeneous images without creating a specific method for weighting the contributions. Using CAE for image-to-image translation, we can obtain  $\hat{\mathbf{U}}$ , which is the post-event image translated to the domain of  $\mathbf{U}$ , and  $\hat{\mathbf{V}}$ , the pre-event image translated to the domain of  $\mathbf{V}$ . This allows us to compare the pre- and post-event images and utilize the difference images in our features:

$$\mathbf{D}_u = \hat{\mathbf{U}} - \mathbf{U} \quad (1)$$

$$\mathbf{D}_v = \mathbf{V} - \hat{\mathbf{V}} \quad (2)$$

If  $\mathbf{U}$  is an image with  $c_u$  channels and  $\mathbf{V}$  has  $c_v$  channels for each pixel position in the  $h \times w$  images, Equations (1) and (2) correspond to the element-wise differences:

$$\mathbf{d}_u = \hat{\mathbf{u}} - \mathbf{u} = [\hat{u}(1) - u(1), \dots, \hat{u}(c_u) - u(c_u)] , \quad (3)$$

$$\mathbf{d}_v = \mathbf{v} - \hat{\mathbf{v}} = [v(1) - \hat{v}(1), \dots, v(c_v) - \hat{v}(c_v)] , \quad (4)$$

where  $\mathbf{u} \in \mathbb{R}^{c_u \times 1}$  and  $\mathbf{v} \in \mathbb{R}^{c_v \times 1}$  are the multi-channel pixel vectors in the co-registered input images at a given position,  $\hat{\mathbf{u}} \in \mathbb{R}^{c_u \times 1}$  and  $\hat{\mathbf{v}} \in \mathbb{R}^{c_v \times 1}$  are the corresponding pixel vectors in the translated images, and the parentheses are used to index the channels of the image. For an optical image, the channels will be spectral bands, while for SAR they will typically contain polarimetric information.

By stacking the difference vectors from Equations (3) and (4) with the original multichannel input data, we form an array  $\mathbf{X} \in \mathbb{R}^{h \times w \times c}$  where at each of the  $h \times w$  pixel positions, the feature vector  $\mathbf{x}$  can be written as:

$$\mathbf{x} = [\mathbf{u}^T, \mathbf{d}_u^T, \mathbf{v}^T, \mathbf{d}_v^T]^T \in \mathbb{R}^{c \times 1} \quad (5)$$

where  $(\cdot)^T$  denotes the transpose operation, and the dimension of the feature vector is equal to the sum of the dimension for each component,  $c = c_u + c_u + c_v + c_v = 2c_u + 2c_v$ . The combined feature vector  $\mathbf{x}$  will

contain information about both the change, and the state before and after the event. Since we use labeled training data, we avoid hand-crafted features or dimensionality reduction methods, such as PCA, thus allowing the machine learning algorithm to learn which features are important for detecting the change of interest.

### Building the OCC

We select the two-step approach to OCC for our application since it is flexible and makes no assumptions about the random sampling of the positively labeled training data. A further benefit is that enables the use of ground reference data collected from other classes (i.e., not forest mortality) by adding it to the RN samples to augment the negative class. When selecting the methods for each step, we want to avoid those that have many parameters that require tuning. To further guide our choice, the OCC should work for a range of different numbers of positive samples. We also exclude methods that are highly specialized in the text domain.

#### Step 1

To obtain RNs in the first step, we use a Gaussian mixture model (GMM) updated once with the expectation maximization (EM) algorithm (Dempster et al. 1977). This is inspired by the first step in the Spy algorithm (Liu et al. 2002) and the well-established use of the naïve Bayes (NB) classifier solved with the EM algorithm as the second step (Bekker and Davis 2020). Our starting point is the same as for Liu et al. (2002): We want to train a Bayes classifier with the labeled positive data and to use all remaining (unlabeled) samples as negatives, before updating the classifier once using EM. The classifier uses Bayes' rule, where a data point  $\mathbf{x}$  should be considered a reliable negative if the probability that it belongs to the negative class ( $y=0$ ) is greater than the probability of belonging to the positive class ( $y=1$ ):

$$P(y=0|\mathbf{x}) > P(y=1|\mathbf{x}). \quad (6)$$

These probabilities are then reformulated using Bayes' rule and canceling the evidence ( $p(\mathbf{x})$ ) in the denominators, which gives the decision rule:

$$P(y=0)p(\mathbf{x}|y=0) > P(y=1)p(\mathbf{x}|y=1) \quad (7)$$

where  $P(y)$  is the prior probability of the positive or negative class and  $p(\mathbf{x}|y)$  is the likelihood.

The approach of NB is to make the so-called naïve assumption that all the features of  $\mathbf{x}$  are mutually independent when conditioned on the class  $y$ . Then



$p(\mathbf{x}|y)$  can be written as the product of the marginal univariate probability density functions (PDFs) of all features in  $\mathbf{x}$ . The use of NB by Liu et al. (2002) and other works stems from document classification, where the marginals are modeled as discrete probability mass functions which represent the probability of a given word occurring in a document of class  $y$ . These can be readily estimated by counting word occurrences while estimating  $p(\mathbf{x}|y)$  for the set of words in a document is intractable unless the vocabulary is small. In our case, the naïve assumption of mutual independence of features does not make it easier to calculate Equation (7), as that requires assumptions about the PDFs  $p(x_i|y)$  for each feature  $x_i \in \mathbf{x} = [x_0, x_1, \dots, x_{c-1}]$ . Instead, we choose to use a parametric model for the conditional probability density functions for the feature vectors for each class,  $p(\mathbf{x}|y)$ . Compared to using the naïve approach, this allows us to account for the correlation between the features of  $\mathbf{x}$ . Recall that the feature vectors consist of all channels from the original images as well as differences obtained with the translations, as given by Equation (5), so we must assume a correlation between features. Since the goal is to perform classification to obtain reliable negative samples which can be used to train a better, final, classifier in the second step, we do not attempt to optimize the selection of  $p(\mathbf{x}|y)$ . We instead argue that the Gaussian is a reasonable choice of PDF in this setting since its parameters are readily estimated by the mean and the sample covariance matrix, with the latter capturing the correlation between features. Thus, we model the marginal density for the positive class as  $p(\mathbf{x}|y=1) \sim \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ , and likewise for the negative class  $p(\mathbf{x}|y=0) \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ .

This is a two-component GMM, and we can use EM to provide an initial classification of the data to find RNs. The initial estimates for the mean and covariance of the first mixture component,  $\hat{\boldsymbol{\mu}}_1$  and  $\hat{\boldsymbol{\Sigma}}_1$ , are based on the labeled positive training set, while the estimates  $\hat{\boldsymbol{\mu}}_0$  and  $\hat{\boldsymbol{\Sigma}}_0$  are based on the unlabeled set, which contains both positive and negative samples. The standard sample mean and sample covariance matrix estimators are used. Using results from EM for GMM (see e.g., Bishop 2006) we can now find refined estimates for the parameters. The new estimates of the expectation for mixture component  $k$  is then:

$$\hat{\boldsymbol{\mu}}_k^{\text{new}} = \frac{1}{n_k} \sum_{j=0}^{N-1} \gamma(z_{jk}) \mathbf{x}_j \quad , \quad k = 0, 1 \quad (8)$$

where  $\gamma(z_{jk}) \in [0, 1]$  is called the responsibility and denotes the posterior probability of sample  $\mathbf{x}_j$

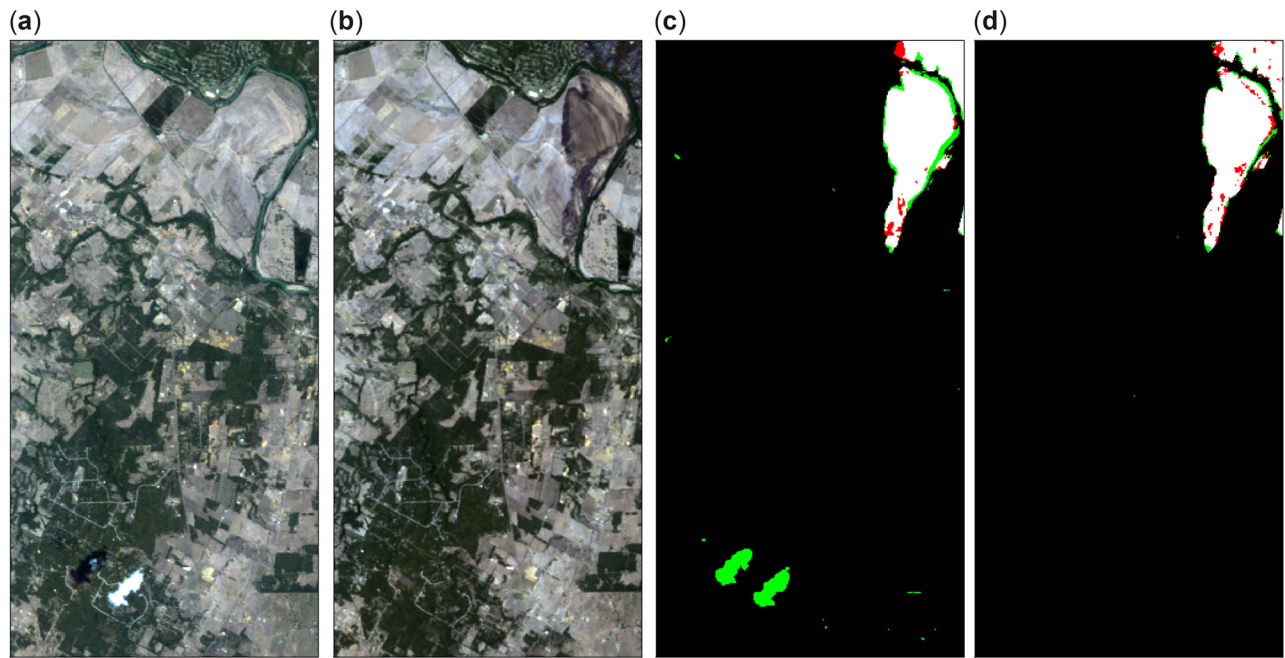
belonging to mixture component  $k$ , and  $n_k = \sum_{j=0}^{N-1} \gamma(z_{jk})$  is a normalization factor. The term indicates how much “responsibility” mixture component  $k$  has for explaining sample  $j$ . The  $\gamma(z_{jk})$  are calculated as the posterior probabilities of sample  $\mathbf{x}_j$  belonging to mixture component  $k$  given the parameters of the mixture components calculated in the previous (initial) iteration of the EM algorithm. The prior probabilities in Equation (7) are initialized as equal (uninformative)  $P(y=1) = P(y=0) = 0.5$ . Since we want mixture component  $k=1$  to model the positive class, we set  $\gamma(z_{j1}) = 1$  and  $\gamma(z_{j0}) = 0$  when  $\mathbf{x}_j$  is from the positive training set. The updated estimate for the covariance matrices are given in a similar manner as Equation (8) as

$$\hat{\boldsymbol{\Sigma}}_k^{\text{new}} = \frac{1}{n_k} \sum_{j=0}^{N-1} \gamma(z_{jk}) (\mathbf{x}_j - \hat{\boldsymbol{\mu}}_k)(\mathbf{x}_j - \hat{\boldsymbol{\mu}}_k)^T \quad , \quad k = 0, 1. \quad (9)$$

We note that the EM estimates are the maximum likelihood estimators (MLEs) for the expectation and covariance matrix weighted by  $\gamma(z_{jk})$ . These can be easily found by functions that can calculate weighted versions of the mean and sample covariance matrix, e.g., the *average* and *cov* functions from the Python `numpy` library. The reliable negatives are selected as the samples  $\mathbf{x}$  where the probability of belonging to the mixture component used to model the negative class is greater than that of the positive class according to Equation (7). The prior probabilities are calculated as the proportion of samples assigned to each class and the estimates in Equations (8) and (9) for the positive and negative classes are then used to evaluate the Gaussian PDFs. We used the *multivariate\_normal* function from the `scipy.stats` package. Intuitively, the mixture component for the negative class will be “wide” and have the largest value for most of the support, while the mixture component of the positive class will be “compact” and only have higher values close to where the positive training samples are located.

## Step 2

For the second step, any (semi-)supervised method could be used, as we now have training data for both classes with the labeled positives and RNs (Bekker and Davis 2020). We choose to base our second step on multi-layer perceptron (MLP) feed-forward neural networks. Neural networks are a good out-of-the-box method for problems, such as this, due to their high flexibility as function approximators. Unfortunately, there are no general guidelines for selecting the



**Figure 3.** The spectral bands corresponding to red, green, and blue (RGB) for (a) the pre-event and (b) the post-event image. The two rightmost plots show the confusion maps for (c) the unsupervised CAE result and (d) our OCC method, where white pixels represent true positive (TP) classifications, black true negative (TN), green false positive (FP), and red false negative (FN). The small clouds and their shadows in the Landsat 5 image are detected as changes compared to the EO-1 image in the unsupervised CAE change detection result, and appear pairwise as green areas (FP) in the confusion map.

number of hidden network layers or the number of neurons in each. Our initial data exploration revealed that there were some variations in the classification results depending on how the network architecture was selected. We, therefore, opted to combine five different MLPs in an ensemble model and use the majority vote to determine the class of each pixel. The architectures of the ensemble consisted of one MLP with a single hidden layer of 1000 neurons, two MLPs with two hidden layers, one with 100 and the other with 200 neurons in both layers, and two MLPs with three hidden layers, again with uniform layer size of either 100 or 200 neurons in all layers. Except for the architecture, all MLPs used the default parameters for the *MLPClassifier* function from the Python Scikit-learn `sklearn.neural_network` package (Pedregosa et al. 2011). The default parameter selection uses the rectified linear unit (ReLU) activation function and the Adam optimizer (Kingma and Ba 2014). The ensemble setup and MLP parameters are kept constant for all experiments. After training, the ensemble of MLPs is used on the full dataset to find the targeted change.

## Results

### *Illustrating targeted change detection*

To illustrate targeted change detection, we test our method on a dataset used in the change detection

literature. The Texas dataset consists of two  $1534 \times 808$  pixel multispectral optical images of Bastrop County, Texas, where a destructive wild-land-urban fire struck on September 4, 2011 (Volpi et al. 2015). The pre-event image is from Landsat 5 Thematic Mapper (TM) with 7 spectral channels and the post-event image is from Earth Observing-1 Advanced Land Imager (EO-1 ALI), with 10 spectral channels. The ground truth was provided by Volpi et al. (2015).

We apply our method with 1000 randomly drawn positive samples as the training data. This corresponds to 0.76% of the positive ground truth data (0.08% of the total image pixels). In Figure 3 we zoom in on an area containing both the targeted change (where the fire has occurred) and other changes (clouds) and show the original image data and the change detection result. In addition to the result from our approach, we show that of the unsupervised CAE change detection from Luppino et al. (2022). This illustrates how our semi-supervised OCC-based approach ignores changes not present in the labeled training set. The result in Figure 3c is reasonable since the clouds (and their shadows) represent an actual change between the two acquisition times. However, it is the effects of the forest fire which is interesting for us to map, and thus the cloud-related changes are marked as false positives in the ground truth data.

## Ablation study

We perform an ablation study to check the contributions of the various components of our approach. In this study, we remove or replace a component and assess the result when using the ablated procedure. The objective is to check if all components contribute positively to the change detection, or if a simpler method performs equally well or better. This is an important exercise in the field of machine learning due to the complexity of the methods involved. Since we wish to objectively measure the effects of the ablation, we must use benchmark datasets where the ground truth is available and we can evaluate the performance of our method numerically. As a part of this study, we also investigate the effect of the number of labeled positive training samples, and how many are needed for a stable result. We use three benchmark change detection datasets with heterogeneous pre- and post-event images for which ground truth is available. Though these datasets are not directly related to the task of mapping forest mortality, this study allows us to assert that our approach is valid for problems beyond our AOI. We present how we performed the study before discussing the result for one of the benchmark datasets in detail, and then briefly summarize the two others.

The labeled positive training data is created by randomly drawing several positive samples from the ground truth, and we vary the number used between 25 and 3000. For each positive sample set, we use it as training data for the two-step classifier and calculate the F1 score of the classification result. The F1 score is defined as the harmonic mean of precision and recall and is a popular metric for evaluating the performance of binary classifiers. Intuitively, the F1 score puts equal weight on the positive predictions being precise (few false detections) and on finding all positive samples. Due to its emphasis on both these aspects, the F1 score gives a better characterization of classifier performance than the traditional overall accuracy measurement, particularly when the classes are unbalanced. Expressed in terms of true (T) and false (F) classification of positives (P) and negatives (N), the F1 score is given as:

$$F1 = \frac{TP}{TP + 0.5(FP + FN)} \quad (10)$$

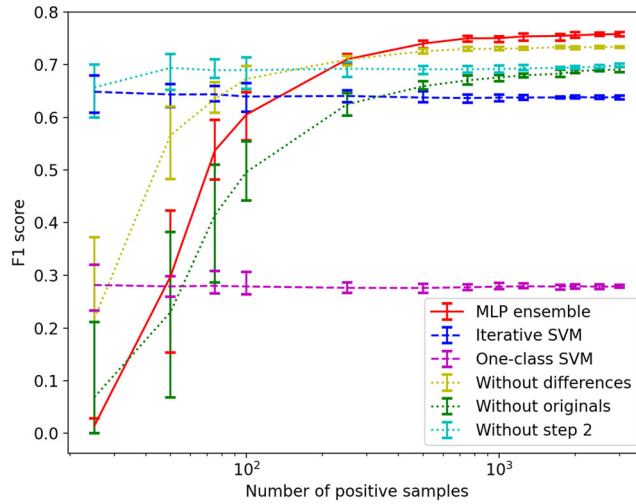
where we see that  $F1 \in [0, 1]$ . For each training set size,  $|\mathbb{P}|$ , we repeat the experiment 10 times with a different random sample and calculate the mean F1 score. We keep all samples from the preceding set as

the number of labeled positives is increased to keep the result consistent.

We compare our proposed approach to five different alternatives, three of which are straightforward ablations of our method; one where we drop the second step and two with a reduced feature vector. Compared to our proposed feature vector  $\mathbf{x}$  in Equation (5), the two feature-related ablations are: not including the differences  $\mathbf{d}_u$  and  $\mathbf{d}_v$ , and not including the original image pixel vectors  $\mathbf{u}$  and  $\mathbf{v}$ . The alternative difference vectors are then  $\mathbf{x}_{w/o \text{ differences}} = [\mathbf{u}^T, \mathbf{v}^T]^T$  and  $\mathbf{x}_{w/o \text{ originals}} = [\mathbf{d}_u^T, \mathbf{d}_v^T]^T$ . We also include the F1 score for the GMM with one EM update used to find the reliable negatives (RNs) in the first step. The difference between this result and our proposed method is the contribution of the second step to the final OCC result.

We also consider an alternative method in the second step, which uses the same RNs as our proposed approach. The method is called iterative support vector machine (SVM) (Bekker and Davis 2020) and is a common choice for step two. It successively trains SVM classifiers based on the labeled positive and RN samples available and then classifies the remaining unlabeled dataset. The samples which are classified as negative are added to the RN set and used to train a new SVM in the next iteration. After some convergence criterion is met, the final SVM trained is used to classify the entire dataset. Due to a large number of RNs found in the first step, we cannot use kernel SVM and must use the linear variant. The SVM penalty parameter (see e.g., Bishop 2006) is set by an automated search procedure, as recommended by Hsu et al. (2003), to select the best value based on the false negative rate (FNR). The iterative SVM training stops when the FNR exceeds 5%, as was first proposed by Li and Liu (2003). Both the MLP ensemble and the iterative SVM use the same RNs found in the first step by the GMM with one EM update.

Finally, we also include results for the popular one-class SVM (OCSVM) method (Khan and Madden 2014). While it is not an ablation of our approach, OCSVM is an alternative to the first-step method we use, and it is also a frequently used OCC method on its own. Furthermore, in contrast to the iterative SVM method, the training is only based on the relatively few labeled positive samples, which enables the use of kernel methods. We use the Scikit-learn implementation of OCSVM, *OneClassSVM* from `sklearn.svm`, with default parameter values, which uses an RBF



**Figure 4.** F1 scores for all the ablations and methods tested with the France dataset and for different positive training set sizes (logarithmic).

kernel with kernel size determined by the number of features and sample variance.

We plot the F1 scores as a function of the number of positive samples,  $|\mathbb{P}|$ , for a heterogeneous change detection dataset consisting of two optical RGB images acquired by Pleiades and WorldView 2 in May 2012 and July 2013, respectively. The ground truth is provided by Touati et al. (2020). The dataset shows construction work occurring in Toulouse, France, and the list of changes includes earthwork, concrete laying, building construction, and more. The areas that have been changed were a mixture of different landcover types in the pre-event image, including bare soil, urban, and vegetated. Figure 4 shows the average F1 score of the 10 runs plotted as a function of the increasing number of labeled positive training samples on a logarithmic x-axis. The error bars represent the 10th and 90th percentile F1 score. We see that the iterative SVM second step performs well when the number of labeled positive samples is small. However, the algorithm actually decreases the F1 score from the first step. The OCSVM has relatively consistent performance as a function of  $|\mathbb{P}|$ , but a considerably lower F1 score than the GMM with one EM update used in the first step. The OCSVM was originally designed for anomaly detection, a setting where there often is no unlabeled (or negative) data available. The poor performance in this setting compared to our GMM EM method illustrates the importance of utilizing the unlabeled data. The ablation of difference features performs better than the ablation of original features, and also better than with the full feature vector when  $|\mathbb{P}|$  is small. Our proposed approach has the best F1 score when  $|\mathbb{P}| \geq 250$ .

**Table 1.** F1 score for 1000 positive samples.

Dataset	W/o originals	W/o differences	W/o step 2	OCSVM	ISVM	MLP ensemble
France	0.676	0.730	0.692	0.279	0.638	0.750
Texas	0.804	0.940	0.869	0.635	0.965	0.951
Italy	0.729	0.664	0.581	0.503	0.607	0.782

In Table 1 we list the F1 scores for the different ablations when  $|\mathbb{P}| = 1000$ . In addition to the result for the France dataset, which corresponds to the F1 scores for one x-value in Figure 4, we include the result for two additional heterogeneous change detection datasets where ground truth information is available. One is the Texas dataset used for the example in Figure 3. The other dataset concerns mapping the extent of a lake overflow in Sardinia, Italy. Both images are obtained by Landsat-5, with the pre-event image obtained in September 1995 consisting of a single channel: the near-infrared (NIR) spectral band. The post-event image contains the RGB channels and is from July 1996. The ground truth provided by Touati et al. (2020).

Table 1 shows that both the originals and the differences are important to include in the feature vectors. The original images have the biggest contribution for the France and Texas datasets, while the difference vectors are the most important for Italy. For all datasets, the best result is achieved with the full feature vectors. The GMM with one EM update performs much better than OCSVM. For the second step, our MLP ensemble improves the F1 score significantly compared to the first step for all datasets. The ISVM results are more varied. It decreases the F1 score for the France dataset, as seen in Figure 4, but

achieves a slightly higher F1 score than the MLP ensemble for the Texas dataset.

We conclude that all components of our method contribute to the F1 score. Relatively few positive samples are needed for a stable classification result, which bodes well for our forest mortality classification. We also note that our method performs well for these datasets despite that the changed areas in the pre-event images of the benchmark datasets are more varied than for forest mortality mapping. Several different landcover types are affected by the construction, fire, and flooding, whereas for our application the changed areas are always live forest in the pre-event image.

### Creating the forest mortality map

There are few cloud-free optical images available from our AOI, which limits the selection of imagery that could be used to map the forest mortality that has occurred. We found one Landsat-5 (LS-5) Thematic Mapper (TM) image from 3 July 2005 reasonably close to the start of the geometrid moth outbreak (2006) which we use as the pre-event image.

For the post-event image, we use a fine-resolution quad-polarization RADARSAT-2 (RS-2) scene from July 25, 2017. We first performed radiometric calibration and terrain correction with the GETASSE30 digital elevation model (DEM) in the Sentinel Application Platform (SNAP), keeping the data in single-look complex (SLC) format with  $10.0\text{ m} \times 10.0\text{ m}$  spatial resolution. Then the polarimetric covariance matrices were estimated using the guided non-local means method (Agersborg et al. 2021). This method preserves SLC resolution and was shown to give estimates of the polarimetric features that better separate between live and dead canopy than alternative methods (Agersborg et al. 2021). The features relevant to canopy state classification were extracted from the covariance matrix  $C$ . These are the intensities in the HH, HV, and VV channels,  $C_{11}$ ,  $C_{22}$ , and  $C_{33}$ , respectively, and the cross-correlation between the complex scattering coefficients for HH and VV,  $C_{13} = |C_{13}|e^{j\angle C_{13}}$  (Agersborg et al. 2021).

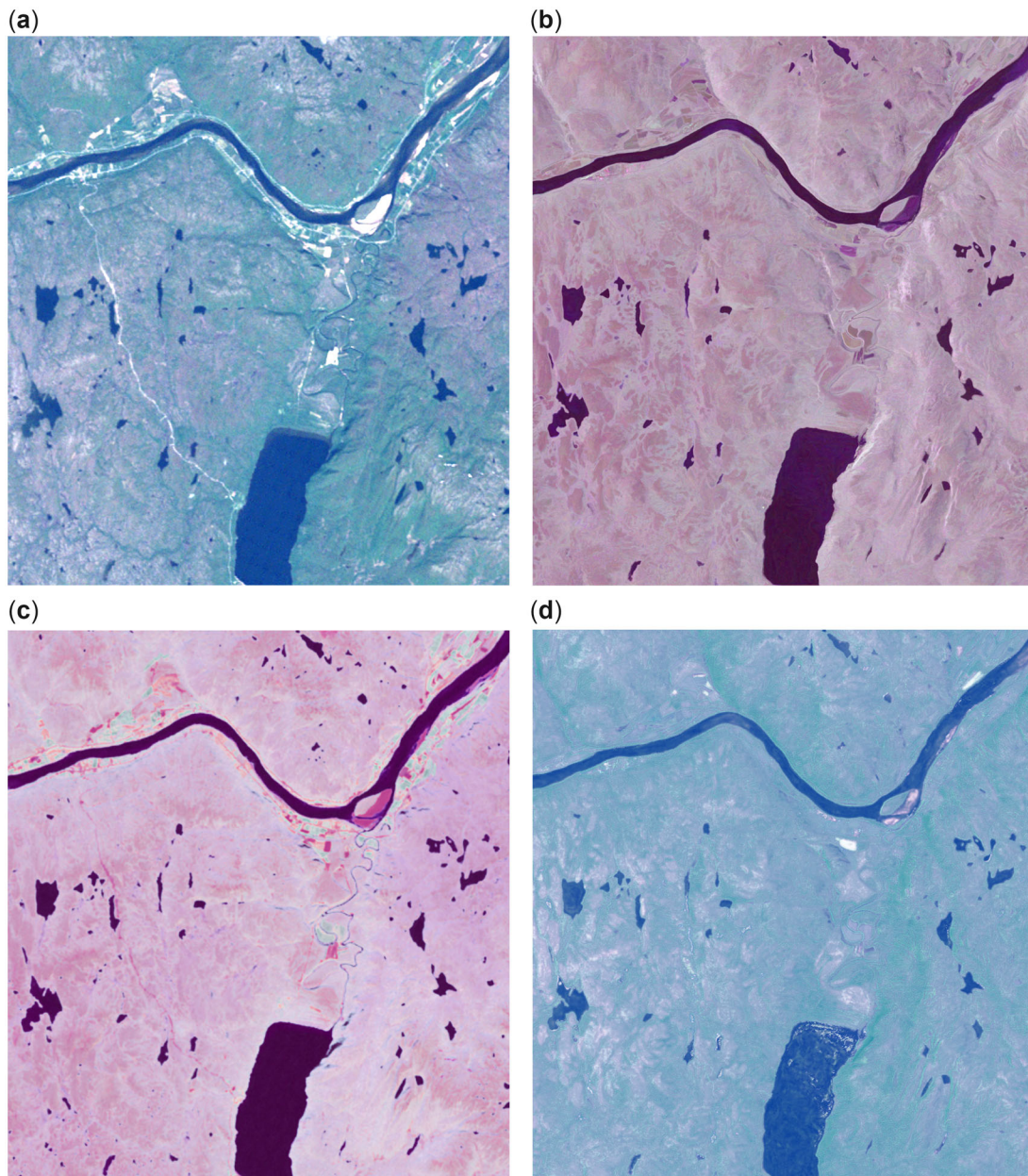
The Landsat-5 TM bands were upsampled from the original  $30.0\text{ m}$  ( $120.0\text{ m}$  for Band 6) spatial resolution to give a pixel size of  $10.0\text{ m} \times 10.0\text{ m}$  using bilinear interpolation resampling during the coregistration process with the SAR data. In this process, both the LS-5 and RS-2 images were geocoded to the UTM 35N projection and mapped on a common grid using QGIS before cropping and extraction of overlapping

images. This resulted in  $1399 \times 1278$  pixel images which were the input for the analysis.

The training data was created by experts carefully comparing high-resolution aerial photography from before (2005) and after the outbreak (2015), drawing polygons covering areas with forest mortality. We choose this approach for three reasons. Firstly, it was easier than selecting a greater number of smaller areas from all over the scene, especially considering that the aerial photographs only covered parts of it. The areas need to be relatively large as the original  $30.0\text{ m}$  resolution of the LS-5 data sets a lower limit for the polygon size. Secondly, by extracting from within larger homogeneous areas we minimize the effect of any misalignment between the ground reference data and the satellite imagery, which could cause the training data to contain pixels from the negative class. Thirdly, the manual creation of training data is tedious work, and we want to generate just enough training data for our OCC method to map the forest mortality for the entire scene. 15 polygons of roughly rectangular shapes with various sizes were created. Four of these intersect with transects studied in field work in 2017 (Agersborg et al. 2021). In total, the 15 polygons contained 1536 pixels, which given the excerpt size of  $1399 \times 1278$  constitutes 0.086% of all pixels.

Figure 5 shows the satellite images and the corresponding CAE translations. The red, green, and blue bands (Band 3, 2, and 1) of the Landsat-5 TM image are shown in the corresponding RGB channels. For the RADARSAT-2 scene, we use the intensities for the HH, HV, and VV polarizations as the red, green, and blue channels, respectively. The translations are shown below the corresponding original using the RGB channels as the other domain. Noteworthy features in the image include lake Polmak, in the center of the lower half of the image, and the Tana river, which runs approximately from east to west in the upper half of the image. There is also a dirt road in the leftmost part of the image going from the Tana river and south toward the western bank of lake Polmak, which is clearly visible in the optical imagery, but very hard to discern in the SAR data.

The CAE translations in Figure 5 are reasonably similar to the other image in the translated domain, but can easily be discerned from the original data. This is expected as it is not possible to exactly recreate the spectral information found in optical imagery from SAR data, and likewise, the polarimetric information about scattering characteristics cannot be replicated from the Landsat TM reflectance measurements. The translation from the optical to the SAR



**Figure 5.** Dataset pair and translation. (a) Landsat-5 RGB image. (b) RADARSAT-2 HH, HV, and VV intensities. (c) Translation of LS-5 image with same channels as (b). (d) Translation of RS-2 image with same channels as (a).

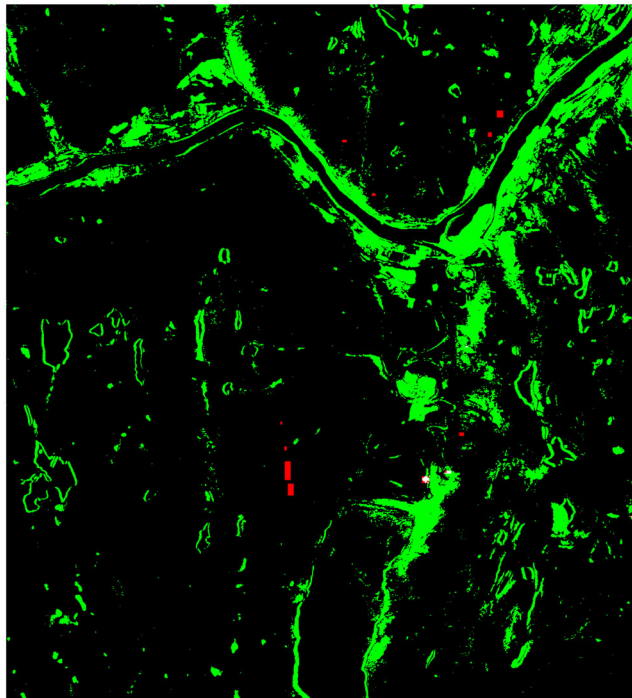
domain seems to match the original RS-2 image quite well and appears visually to be the better of the two translations, whereas the translation in 5(d) appears somewhat blurry with muted colors. There are some obvious translation artifacts, for instance, the bright pixels in lake Polmak in the translated RS-2 result.

#### **Unsupervised change detection with CAE**

To illustrate the changes detected by a non-targeted approach based on the translations in Figure 5, we show the CAE result. Change detection with CAE is based on thresholding the magnitude of the difference

vectors between the translations and the originals (Luppino et al. 2022). Figure 6 shows the confusion map for the CAE change detection based on the difference vectors obtained from the images shown in Figure 5.

Figure 6 also shows the 15 ground reference polygons with forest mortality created as training data for the OCC. Note that some of the polygons are too small to be discerned in Figure 6. Parts of the forest mortality polygons where CAE predicts no change are shown in red, while correct change predictions for these areas are white in Figure 6. Only 14% of the pixels with forest mortality are predicted as changed.



**Figure 6.** CAE confusion map, with predicted changes in green, predicted unchanged in black, correctly classified forest mortality from our limited ground reference dataset marked white, and red showing missed detections from the same dataset.

We notice that parts of the predicted changes resemble outlines of the water bodies in Figure 5, which indicates that CAE detects changes in water levels. There are also changes in the agricultural and settled land, primarily along the Tana river. The translation artifacts from lake Polmak in Figure 5d are also marked as changes. Compared to these phenomena, which result in high-magnitude difference vectors, the death of the canopy layer of the fragmented forest-tundra ecotone is a subtle change at this resolution level. To detect it we need to train a classifier to look for it specifically.

### **Target change detection with the proposed method**

Figure 7 shows the forest mortality map provided by our targeted change detection method. The feature vectors in Equation (5) are obtained from the images and translations shown in Figure 5. The two-step OCC is then trained on 1536 feature vectors from the 15 polygons with known forest mortality. The result indicates that large areas of forest have died following the outbreak. Particularly the western side of lake Polmak, on the Finnish side of the border, has been heavily afflicted. This is in line with the observations by Biuw et al. (2014) that the regeneration of mountain birch stands appeared to have been severely hampered by the year-round grazing on the Finnish side

of the border. Significant forest mortality is also detected north of the river. We note the fragmented nature of the forest mortality map, which is natural given that the sparse nature of the forest-tundra ecotone. Contrary to the CAE result, there are no detections along the agricultural and settled land around the Tana river.

We do not have another map of the effects of the outbreak to evaluate, so we cannot readily quantify the accuracy of the full forest mortality map. By use of NDVI measurements from before and after the outbreak, in the same period as the input imagery, and correlating with the result in Figure 7, we find a noticeable NDVI decrease in the areas of classified forest mortality. For these areas, shown as white in Figure 7, the NDVI decreased by 0.145 on average. In the other areas, the difference was virtually zero.

We also evaluate our result on a systematic grid of  $30\text{ m} \times 30\text{ m}$  cells aligned with transects examined during field work in 2017 (Agersborg et al. 2021). The cells were classified by an expert into four classes based on a study of aerial photographs. Three of the classes were related to forest mortality, as they describe the canopy state as “live” (33 cells), “dead” (44 cells), and “damaged” (30 cells) (Agersborg et al. 2021). All cells with a canopy state class were from the forest with a live canopy in the 2005 aerial image, with the canopy state of the cell in the 2015 image determining if it was classified as live, damaged,



**Figure 7.** Predicted forest mortality areas are shown in white using our approach.

or dead. The last class, denoted “other” (66 cells), is vegetation without canopy cover. For each grid cell, we find the pixels in the forest mortality map in Figure 7 that intersects the geographical coordinates defining the bounding box of that cell. Since the mortality map uses a  $10\text{ m} \times 10\text{ m}$  pixel spacing, each grid cell should correspond to  $3 \times 3 = 9$  pixels, though a few will be larger, as we choose to include partially covered pixels. Note that the grid is not aligned north-south and has not been co-registered to the satellite imagery. Hence, there may be some inaccuracies due to misalignment, especially since the size of the cells is the same as the resolution of the LS-5 pre-event image.

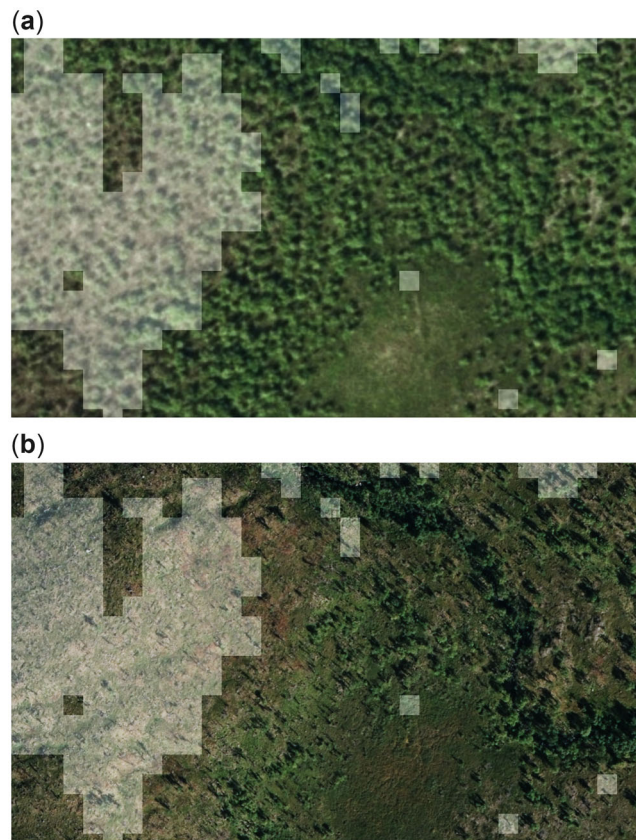
For the classes “live” and “other,” we observe a low false alarm rate with  $\widehat{\text{TN}}_{\text{live grid}} = 98.7\%$  and  $\widehat{\text{TN}}_{\text{other grid}} = 96.9\%$  of the pixels correctly classified as no forest mortality. As part of the work to create the polygons with forest mortality, 15 polygons of live forest were also extracted. These were similar in shape (rectangular), size and grouping, and located relatively close to areas with forest mortality. In total, the 15 polygons with live forest consist of 2070 pixels. For the live polygons, our approach performs well with  $\widehat{\text{TN}}_{\text{live poly}} = 97.2\%$ , which is consistent with the result for the “live” class for the grid cells. Note that the both the “live” and the “other” classes are subsets of the negative superclass and are not necessarily accurate estimates for the true negative rate for the complete negative class. Comparing the forest

mortality map in Figure 7 with the optical image in Figure 5a and the unsupervised CAE change detection results in Figure 6, indicates that we avoid classifying changes in agricultural and settled areas as forest mortality, which is important for the true negative rate.

For the “dead” grid cells we observe a true positive rate of  $\widehat{\text{TP}}_{\text{grid}} = 66.4\%$ . Misalignment could be a contributing factor to the missed detections as 37 of 44 (84.1%) of the cells do contain pixels classified as forest mortality. The number of missed detections along with the low false alarm rate could also be an indication that our method is conservative in predicting forest mortality, though this was not seen for the training data where the true positive rate was 96.8%. While accuracy scores for the training data always are of questionable validity, this indicates that our method does not have to “sacrifice” much accuracy on the positive training set to increase accuracy on the RNs during training.

It is not obvious if the grid cells with damaged canopy state should be considered as the positive or negative class for evaluation. This state contains a mixture of dead stems and trees where parts of the canopy have re-sprouted. 28.1% of the pixels in the grid cells of this category were classified as forest mortality. Arguments could be made for considering the damaged state as a part of the positive class since there is a clear decrease in canopy cover. However, for operational monitoring, it could be more important to focus on areas where the forest has died





**Figure 8.** Forest mortality map as an overlay to aerial photographs. (a) Before the outbreak (2005). (b) After the outbreak (2015).

completely, as it can be assumed that most of the forest will suffer canopy damage in a major outbreak, and our targeted change detection approach should be able to map only the former. Nonetheless, that only 28.1% of the pixels in the damaged class were labeled as forest mortality may be another indication that our prediction of forest mortality is on the conservative side.

To evaluate how well the forest mortality map corresponds to the high-resolution aerial photographs, we show the map as an overlay to the images from 2005 and 2015. The mortality map is shown as a bright see-through layer in Figure 8. The image from 2005 is shown at the top of Figure 8a with 2015 in Figure 8b. The latter image has a higher resolution of  $0.25\text{ m} \times 0.25\text{ m}$  compared to  $0.50\text{ m} \times 0.50\text{ m}$  for the 2005 image, and appears to be more easily interpretable. The “pixelated” appearance of the mortality map is due to the  $10\text{ m} \times 10\text{ m}$  pixel spacing. Most of the areas with forest mortality have been accurately detected, although the mapped area in the left half of Figure 8 could have included a larger area. There also appear to be some false alarms, with the three single  $10\text{ m} \times 10\text{ m}$  areas of the mortality map in the lower right half of the images in Figure 8 appear to have a live canopy in the 2015 image. Apart from that, our

method appears to avoid misclassifying other land-cover types as forest mortality.

We also note the challenge of accurately mapping forest mortality due to the sparse nature of the forest-tundra ecotone and the entangled pattern of live and dead trees. Since the resulting map appears quite good, setting the pixel spacing of the mortality map to  $10\text{ m} \times 10\text{ m}$  to match the RS-2 resolution, seems warranted. However, we should keep in mind that the resolution of LS-5 used as the pre-event image is  $30\text{ m} \times 30\text{ m}$ , which will affect the accuracy of the mortality map. An example of the mortality map failing to delineate thin separations between classes is shown in Figure 9. Here there is a separation between areas where the trees have died that are included in the mortality map. Similarly to the result in Figure 8, the map shown in Figure 9 misses parts of the forest that has died which borders on the areas correctly mapped as dead. It also appears that there is an area that is falsely classified as forest mortality in the bottom right part of Figure 9.

The results in Figures 8 and 9, and the numerical evaluations on the grid cells and live polygons, indicate that the mortality map is conservative. One way of easily increasing the number of pixels predicted as forest mortality is to adjust the ensemble voting done



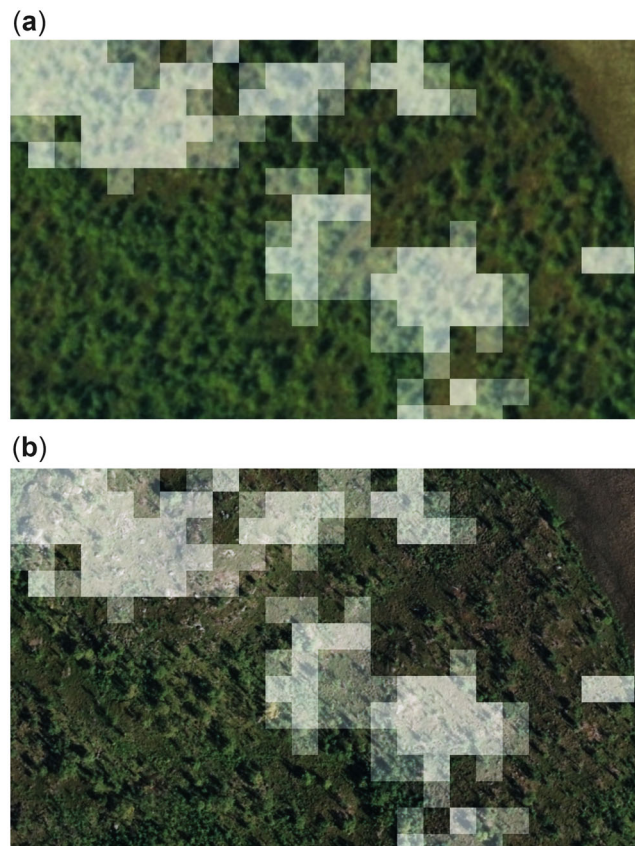
**Figure 9.** Forest mortality map as an overlay to aerial photographs. (a) Before the outbreak (2005). (b) After the outbreak (2015).

in the second step of the OCC method, where five MLPs classify the input data. Reducing the number of MLPs required to predict the positive class would therefore increase the prediction of forest mortality. This is equivalent to considering the output of the ensemble as the average number of predictors which predict the positive class as a number between 0 and 1 that is thresholded to give the final binary output map. Then, reducing the votes necessary is the same as reducing this threshold  $t$ . By reducing the number of votes required from three (majority) to two we obtain a higher true positive rate of  $\widehat{TP}_{\text{grid}} = 76.6\%$ . In terms of threshold, this corresponds to lowering the threshold from  $t=0.5$  to  $t=0.3$ . However, there is a decrease in the true negative rate, with  $\widehat{TN}_{\text{live poly}} = 94.6\%$ ,  $\widehat{TN}_{\text{live grid}} = 97.5\%$ , and  $\widehat{TN}_{\text{other grid}} = 94.6\%$ . There was also an increase in the number of pixels from the “damaged” state classified as forest mortality to 33.7%. When evaluating this result on the high-resolution aerial photographs, we see that more forest mortality is correctly detected, but at the cost of a higher number of false alarms. An example of this is shown in Figure 10, where the result of our method with  $t=0.5$  to  $t=0.3$  are both overlain the photographs from before and after the outbreak. The brightest overlay areas are predicted as forest

mortality with both thresholds, while the more see-through overlay corresponds to the prediction from  $t=0.3$  alone. For the area in the upper right corner, the  $t=0.3$  appears better, while the difference in prediction at the bottom right of Figure 10 appears to be mostly false alarms.

### Conclusions and future work

In this work, we have presented a method for detecting forest mortality from a pair of medium-resolution heterogeneous remote sensing images to map the effect of a geometrid moth outbreak. This is a challenging problem, and we were unable to achieve it using the unsupervised CAE results alone since the phenomenon of interest has a weak signature compared to other changes. By utilizing the CAE for change-aware image-to-image translation, we obtained multitemporal difference vectors despite the heterogeneity of the input images. When combined with the original image features, a semi-supervised one-class classifier can learn to map the changes of interest from a very limited set of training data consisting of  $<0.1\%$  of the image pixels of these extended feature vectors. The ablation study shows that all the different components of our targeted change detection approach contribute to the final



**Figure 10.** Two forest mortality maps corresponding to different thresholds as overlays to images before and after the outbreak. The brightest overlay shows areas classified as forest mortality with both thresholds, while the more transparent overlay are classifications only made with  $t=0.3$ . (a) Image from 2005. (b) Image from 2015.

output, and we achieve good results for benchmark datasets, despite not being intended as a general change detection method.

The evaluation for our AOI indicates that we achieve a low false alarm rate, but that the predicted forest mortality map may be a bit conservative. It is possible to increase the true positive rate at the expense of more false alarms by adjusting a threshold in the ensemble voting done in the second step of the OCC method. However, we prefer a low false alarm rate to a higher number of detections in the tradeoff, for instance in case the map should be used to determine new areas for field work to study the effect of the outbreak. Future work should seek to assess performance on datasets with complete ground truth data available. This should preferably be done on datasets suitable for targeted change detection, where changes unrelated to the phenomenon of interest are included in the negative class.

Our approach expands the potential for detecting the extent of changes that we know have occurred at one or more locations by using whatever satellite imagery available from before and after the event as long as it can be co-registered. This allows us to map a phenomenon of interest over large areas. It does not require a

dense time series of data, in the same manner as NDVI-based approaches, which can be problematic for our AOI given the high cloud cover percentage. The modular nature of our approach means that components can be replaced if the particular dataset warrants it. This could also be investigated in future work, and our ablation study hints at some interesting directions.

### Disclosure statement

No potential conflict of interest was reported by the author(s).

### Funding

This work was supported in by the COAT Tools project ([www.coat.no/en/Education/COAT-Tools](http://www.coat.no/en/Education/COAT-Tools)) by an Interdisciplinary Project Grant from the Developing the High North Strategic Programme of UiT The Arctic University of Norway, and by the Research Council of Norway under Grant 301922.

### ORCID

Jørgen A. Agersborg  <http://orcid.org/0000-0002-1526-6659>

Luigi T. Luppino  <http://orcid.org/0000-0002-1288-817X>  
 Stian Normann Anfinsen  <http://orcid.org/0000-0002-3758-4295>  
 Jane Uhd Jepsen  <http://orcid.org/0000-0003-1517-1569>

## References

- Agersborg, J.A., Anfinsen, S.N., and Jepsen, J.U. 2021. "Guided nonlocal means estimation of polarimetric covariance for canopy state classification." *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 60(No. 5208417): pp. 1–17. doi:10.1109/TGRS.2021.3090831.
- Bae, S., Müller, J., Förster, B., Hilmers, T., Hochrein, S., Jacobs, M., Leroy, B.M., Pretzsch, H., Weisser, W.W., and Mitesser, O. 2022. "Tracking the temporal dynamics of insect defoliation by high-resolution radar satellite data." *Methods in Ecology and Evolution*, Vol. 13(No. 1): pp. 121–132. doi:10.1111/2041-210X.13726.
- Bekker, J., and Davis, J. 2020. "Learning from positive and unlabeled data: A survey." *Machine Learning*, Vol. 109(No. 4): pp. 719–760. doi:10.1007/s10994-020-05877-5.
- Bishop, C. 2006. *Pattern Recognition and Machine Learning*. New York, NY: Springer. ISBN: 0387310738.
- Biuw, M., Jepsen, J.U., Cohen, J., Ahonen, S.H., Tejesvi, M., Aikio, S., Wäli, P.R., et al. 2014. "Long-term impacts of contrasting management of large ungulates in the Arctic tundra-forest ecotone: Ecosystem structure and climate feedback." *Ecosystems*, Vol. 17(No. 5): pp. 890–905. doi:10.1007/s10021-014-9767-3.
- CAFF. 2013. *Arctic Biodiversity Assessment 2013*. Akureyri, Iceland: Conservation of Arctic Flora/Fauna.
- Camps-Valls, G., Gómez-Chova, L., Muñoz-Marí, J., Luis Rojo-Álvarez, J., and Martínez-Ramón, M. 2008. "Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection." *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 46(No. 6): pp. 1822–1835. doi:10.1109/TGRS.2008.916201.
- Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B., and Lambin, E. 2004. "Digital change detection methods in ecosystem monitoring: A review." *International Journal of Remote Sensing*, Vol. 25(No. 9): pp. 1565–1596. doi:10.1080/0143116031000101675.
- Dempster, A.P., Laird, N.M., and Rubin, D.B. 1977. "Maximum likelihood from incomplete data via the EM algorithm." *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 39(No. 1): pp. 1–22. doi:10.1111/j.2517-6161.1977.tb01600.x.
- Elkan, C., and Noto, K. 2008. "Learning classifiers from only positive and unlabelled data." Proceedings of 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 213–220.
- Gao, Y., Skutsch, M., Paneque-G'Alvez, J., and Ghilardi, A. 2020. "Remote sensing of forest degradation: A review." *Environmental Research Letters*, Vol. 15(No. 10): pp. 103001. doi:10.1088/1748-9326/abaad7.
- Hall, R.J., Castilla, G., White, J.C., Cooke, B.J., and Skakun, R.S. 2016. "Remote sensing of forest pest damage: A review and lessons learned from a Canadian perspective." *The Canadian Entomologist*, Vol. 148(No. S1): pp. S296–S356. doi:10.4039/tce.2016.11.
- Henden, J.-A., Ims, R.A., Yoccoz, N.G., Asbjørnsen, E.J., Stien, A., Mellard, J.P., Tveraa, T., Marolla, F., and Jepsen, J.U. 2020. "End-user involvement to improve predictions and management of populations with complex dynamics and multiple drivers." *Ecological Applications*, Vol. 30(No. 6): pp. e02120. doi:10.1002/eap.2120.
- Hsu, C.-W., Chang, C.-C., and Lin, C.-J. 2003. *A Practical Guide to Support Vector Classification*.
- Ims, R.A., Uhd Jepsen, J., Stien, A., and Yoccoz, N.G. 2013. *Science Plan for COAT: Climate-Ecological Observatory for Arctic Tundra*. Fram Centre Report Series 1. Tromsø: Fram Centre.
- Jepsen, J.U., Kapari, L., Hagen, S.B., Schott, T., Vindstad, O.P.L., Nilssen, A.C., and Ims, R.A. 2011. "Rapid northwards expansion of a forest insect pest attributed to spring phenology matching with sub-Arctic birch." *Global Change Biology*, Vol. 17(No. 6): pp. 2071–2083. doi:10.1111/j.1365-2486.2010.02370.x.
- Jepsen, J.U., Biuw, M., Ims, R.A., Kapari, L., Schott, T., Vindstad, O.P.L., and Hagen, S.B. 2013. "Ecosystem impacts of a range expanding forest defoliator at the forest-tundra ecotone." *Ecosystems*, Vol. 16(No. 4): pp. 561–575. doi:10.1007/s10021-012-9629-9.
- Jepsen, J.U., Hagen, S.B., Høgda, K.A., Ims, R.A., Karlsen, S.R., Tømmervik, H., and Yoccoz, N.G. 2009. "Monitoring the spatio-temporal dynamics of geometrid moth outbreaks in birch forest using MODIS-NDVI data." *Remote Sensing of Environment*, Vol. 113(No. 9): pp. 1939–1947. doi:10.1016/j.rse.2009.05.006.
- Jepsen, J.U., Hagen, S.B., Ims, R.A., and Yoccoz, N.G. 2008. "Climate change and outbreaks of the geometrids *Operophtera brumata* and *Epirrita autumnata* in subarctic birch forest: evidence of a recent outbreak range expansion." *The Journal of Animal Ecology*, Vol. 77(No. 2): pp. 257–264. doi:10.1111/j.1365-2656.2007.01339.x.
- Jian, P., Chen, K., and Cheng, W. 2022. "GAN-based one-class classification for remote-sensing image change detection." *IEEE Geoscience and Remote Sensing Letters*, Vol. 19: pp. 1–5. doi:10.1109/LGRS.2021.3066435.
- Khan, S.S., and Madden, M.G. 2014. "One-class classification: Taxonomy of study and review of techniques." *The Knowledge Engineering Review*, Vol. 29(No. 3): pp. 345–374. doi:10.1017/S026988891300043X.
- Kingma, D.P., and Ba, J. 2014. "Adam: A method for stochastic optimization." *arXiv preprint arXiv:1412.6980*.
- Li, W., Guo, Q., and Elkan, C. 2011. "A positive and unlabeled learning algorithm for one-class classification of remote-sensing data." *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 49(No. 2): pp. 717–725. doi:10.1109/TGRS.2010.2058578.
- Li, W., Guo, Q., and Elkan, C. 2020. "One-class remote sensing classification from positive and unlabeled background data." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 14: pp. 730–746.
- Li, X., and Liu, B. 2003. "Learning to classify texts using positive and unlabeled data." *IJCAI*, Vol. 3, pp. 587–592. Citeseer.

- Liu, B., Sun Lee, W., Yu, P.S., and Li, X. 2002. "Partially supervised classification of text documents." *ICML*, Vol. 2, pp. 387–394.
- Li, P., and Xu, H. 2010. "Land-cover change detection using one-class support vector machine." *Photogrammetric Engineering & Remote Sensing*, Vol. 76(No. 3): pp. 255–263. doi:10.14358/PERS.76.3.255.
- Liu, Bing, Dai, Yang., Li, Xiaoli, Sun Lee, Wee., and Yu, PhilipS. 2003. "Building text classifiers using positive and unlabeled examples." Third IEEE International Conference on Data Mining, 179–186. IEEE.
- Luppino, L.T. 2020. *Unsupervised change detection in heterogeneous remote sensing imagery*. PhD diss. UiT The Arctic University of Norway, Department of Physics and Technology.
- Luppino, L.T., Hansen, M.A., Kampffmeyer, M., Bianchi, F.M., Moser, G., Jenssen, R., and Anfinsen, S.N. 2022. "Code-aligned autoencoders for unsupervised change detection in multimodal remote sensing images." *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–13. doi:10.1109/TNNLS.2022.3172183.
- Mitchell, A.L., Rosenqvist, A., and Mora, B. 2017. "Current remote sensing approaches to monitoring forest degradation in support of countries measurement, reporting and verification (MRV) systems for REDD+." *Carbon Balance and Management*, Vol. 12(No. 1): pp. 1–22. doi:10.1186/s13021-017-0078-9.
- Múnoz-Marí, J., Bovolo, F., Gómez-Chova, L., Bruzzone, L., and Camp-Valls, G. 2010. "Semisupervised one-class support vector machines for classification of remote sensing data." *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 48(No. 8): pp. 3188–3197. doi:10.1109/TGRS.2010.2045764.
- Olsson, P.-O., Kantola, T., Lyytikäinen-Saarenmaa, P., Jönsson, A., and Eklundh, L. 2016a. "Development of a method for monitoring of insect induced forest defoliation-limitation of MODIS data in Fennoscandian forest landscapes." *Silva Fennica*, Vol. 50(No. 2): pp. 1–22. doi:10.14214/sf.1495.
- Olsson, P.-O., Lindström, J., and Eklundh, L. 2016b. "Near real-time monitoring of insect induced defoliation in sub-alpine birch forests with MODIS derived NDVI." *Remote Sensing of Environment*, Vol. 181: pp. 42–53. doi:10.1016/j.rse.2016.03.040.
- Pedersen, Å.Ø., Jepsen, J.U., Paulsen, I.M.G., Fuglei, E., Mosbacher, J.B., Ravolainen, V., and Yoccoz, N.G. 2021. *Norwegian Arctic tundra: a panel-based assessment of ecosystem condition*. Technical report. Norsk Polarinstitut.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., et al. 2011. "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research*, Vol. 12: pp. 2825–2830.
- Perbet, P., Fortin, M., Ville, A., and B'Elend, M. 2019. "Near real-time deforestation detection in Malaysia and Indonesia using change vector analysis with three sensors." *International Journal of Remote Sensing*, Vol. 40(No. 19): pp. 7439–7458. doi:10.1080/01431161.2019.1579390.
- Ran, Q., Zhang, M., Li, W., and Du, Q. 2016. "Change detection with one-class sparse representation classifier." *Journal of Applied Remote Sensing*, Vol. 10(No. 4): pp. 042006. doi:10.1117/1.JRS.10.042006.
- Ran, Q., Li, W., and Du, Q. 2018. "Kernel one-class weighted sparse representation classification for change detection." *Remote Sensing Letters*, Vol. 9(No. 6): pp. 597–606. doi:10.1080/2150704X.2018.1452063.
- Senf, C., Seidl, R., and Hostert, P. 2017. "Remote sensing of forest insect disturbances: Current state and future directions." *International Journal of Applied Earth Observation and Geoinformation*, Vol. 60: pp. 49–60. doi:10.1016/j.jag.2017.04.004.
- Sun, Y., Lei, L., Li, X., Sun, H., and Kuang, G. 2021. "Nonlocal patch similarity based heterogeneous remote sensing change detection." *Pattern Recognition*, Vol. 109: pp. 107598. doi:10.1016/j.patcog.2020.107598.
- Touati, R., Mignotte, M., and Dahmane, M. 2020. "Multimodal change detection in remote sensing images using an unsupervised pixel pairwise-based Markov random field model." *IEEE Transactions on Image Processing*, Vol. 29: pp. 757–767. doi:10.1109/TIP.2019.2933747.
- Volpi, M., Camps-Valls, G., and Tuia, D. 2015. "Spectral alignment of multitemporal cross-sensor images with automated kernel canonical correlation analysis." *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 107: pp. 50–63. doi:10.1016/j.isprsjprs.2015.02.005.
- Ye, S., Chen, D., and Yu, J. 2016. "A targeted change-detection procedure by combining change vector analysis and post-classification approach." *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 114: pp. 115–124. doi:10.1016/j.isprsjprs.2016.01.018.
- Yu, H. 2005. "Single-class classification with mapping convergence." *Machine Learning*, Vol. 61(No. 1–3): pp. 49–69. doi:10.1007/s10994-005-1122-7.

## Appendix

### Code-aligned autoencoders

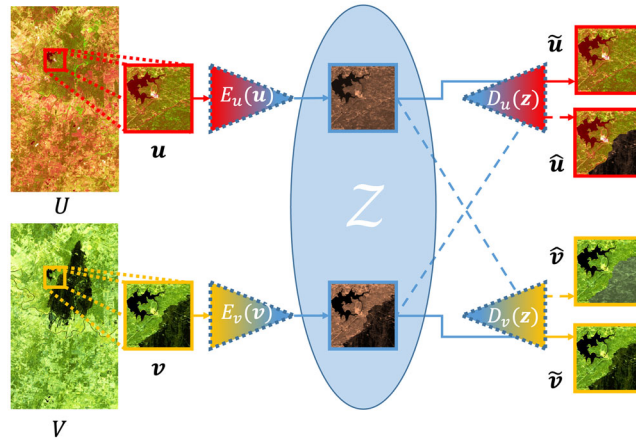
As the name implies, the CAE algorithm uses an autoencoder architecture that learns a pair of convolutional neural networks, the encoder, and the decoder, for each of the images. The domain-specific encoders are trained to encode their respective input images into a code representation, while the decoders are trained to reconstruct the input images with high fidelity from these codes. That is, if we denote the encoder associated with the pre- and post-event images as  $E_u$  and  $E_v$ , respectively, these can be used to obtain encoded representations of  $\mathbf{U}$  and  $\mathbf{V}$  as  $E_u(\mathbf{U}) = Z_u$  and  $E_v(\mathbf{V}) = Z_v$ . The corresponding decoders,  $D_u$  and  $D_v$ , then reconstruct the encoded input as:

$$D_u(Z_u) = D_u(E_u(\mathbf{U})) = \tilde{\mathbf{U}} \approx \mathbf{U} \quad (11)$$

$$D_v(Z_v) = D_v(E_v(\mathbf{V})) = \tilde{\mathbf{V}} \approx \mathbf{V} \quad (12)$$

where the reconstructed original pre- and post-event images,  $\tilde{\mathbf{U}}$  and  $\tilde{\mathbf{V}}$ , should be approximately equal to the corresponding original images. This objective is formulated as a loss function, the reconstruction loss, which is an inherent part of all autoencoders. For CAE the reconstruction loss is one of four terms in the total loss function used to train the full network.

In general, the code layer representations of two separately trained autoencoders are not similar. By aligning the



**Figure A1.** Illustration of code-aligned autoencoder network showing the translation of patches  $\mathbf{u}$  from  $\mathbf{U}$  and  $\mathbf{v}$  from  $\mathbf{V}$ .

code spaces one can obtain a translation between domains by using the decoder from one domain on the coded representation of the other domain. That is

$$D_u(Z_v) = D_u(E_v(\mathbf{V})) = \hat{\mathbf{U}} \quad (13)$$

$$D_v(Z_u) = D_v(E_u(\mathbf{U})) = \hat{\mathbf{V}} \quad (14)$$

where the encoders and decoders are the same as for Equations (11) and (12), and  $\hat{\mathbf{U}}$  and  $\hat{\mathbf{V}}$  are the pre- and post-event images translated to the other domain. Figure A1, adapted from Luppino et al. (2022), illustrates the network, showing the result of encoding and decoding a pair of coregistered image patches from the Texas dataset.

CAE enforces alignment of the code layers of the two autoencoders by adding a loss term that ensures their alignment in both distribution and location of land covers (Luppino et al. 2022). This code correlation loss is a novel feature of CAE and enforces that  $Z_u$  should be similar to  $Z_v$  (Luppino et al. 2022). The similarity is based on a cross-modal distance between training patches in the input domains. This allows pixels that have changed to be distinguished from those that have not, and the loss term seeks to preserve these relationships in the code layer. Contrary to the other loss functions, which are used to train the both encoders and decoders, the code correlation loss is only used for the encoders.

A cycle-consistency loss enforces that data translated from one domain to the other, and then back again, should be identical to the input. In a sense, it is similar to the reconstruction loss, except that the cycle-consistency loss involves all encoders and decoders of the network.

The final loss term requires that the translations  $\hat{\mathbf{U}}$  and  $\hat{\mathbf{V}}$  in Equations (13) and (14) should be similar to the data in the original domain  $\mathbf{U}$  and  $\mathbf{V}$ , except for pixels where changes have occurred. If there is a significant chance that a change has occurred for a particular pixel, its contribution to this loss term is strongly suppressed, whereas pixels of  $\hat{\mathbf{U}}$  ( $\hat{\mathbf{V}}$ ) from likely unchanged areas should be close to  $\mathbf{U}$  ( $\mathbf{V}$ ). To distinguish between changed and unchanged areas, this loss term includes a weighting factor updated iteratively during training. This is based on preliminary change detection results obtained with the image translations at the current stage of training. Note that while the final change-

detection result of CAE is a binary change map, the weighting factor is a continuous variable between zero and one where lower values indicate where there is a high probability of change.

We have made some adaptations to the CAE related to network training, with the aim of improving the visual quality and detail preservation of the translations, as briefly summarized:

- The patch size used for training is reduced from  $100 \times 100$  pixels to  $20 \times 20$ , so the code correlation loss is calculated for all pixel pairs in the input patches. In the original implementation, the cross-modal distance between training patches was based on a  $20 \times 20$  pixels excerpt from the center of the full  $100 \times 100$  training patch due to memory constraints (Luppino 2020). By reducing its size, the full training patch is used for the code correlation loss to better align the two domains.
- The number of patches per training batch is increased from 10 to 20 and the number of batches per epoch from 10 to 600 to compensate for the lower number of pixels seen during training due to the reduced patch size. A  $100 \times 100$  patch contains 25 times as many pixels as a  $20 \times 20$  one, and it is, therefore, necessary to increase the batch size and the number of epochs compared to Luppino (2020).
- The preliminary evaluation of the difference image is changed from using the reconstructed versions ( $\hat{\mathbf{U}}$  and  $\hat{\mathbf{V}}$ ) to using the originals in the weighted translation loss to better preserve details in the translations.

In our experience, these modifications improve the visual quality of the translations, and through more meaningful difference images also the accuracy of the final classification.

### One-class classification

OCC is framed as a binary classification problem, where the positive class of interest has label  $y=1$  and the negative class,  $y=0$ , is usually defined as the complement of the positive class (Li et al. 2020). The full dataset  $\mathbb{X}$  is typically divided into  $\mathbb{P}$ , the set of labeled positive samples, and an

unlabeled set  $\cup$  (also called mixed set) that consists of data from both the positive and the negative class.

Text classification and document retrieval are applications where OCC has seen much use and several OCC methods that have been developed are customized for the text domain. In the taxonomy of OCC techniques proposed by Khan and Madden (2014), the applications were divided into two categories: “text/document classification” and “other applications.” The terms single-class classification (SCC) (Yu 2005), partially supervised classification (Liu et al. 2002), and others (see Khan and Madden 2014 for a brief summary) have been used for one-class classification problems. OCC is also related to positive and unlabeled learning (PUL), and the distinction between the two is somewhat blurry. Just as in the OCC setting, PUL assumes that a labeled positive training set and an unlabeled set that contains mixed samples from both the positive and negative classes are available. Many PUL methods are based on estimating dataset attributes, such as the labeling frequency and class prior probability of the positive class, and thus need to make assumptions about how the labeled positive samples were generated (Bekker and Davis 2020). Some methods assume that the labeled positive samples were selected completely at random (SCAR). PUL also considers two different settings: the single-training-set scenario, where the positive and unlabeled samples are from the same dataset, and the case-control scenario, where they are from different datasets. We choose to draw the distinction between OCC and PUL by saying that the latter makes assumptions about how the positive samples were labeled, or that it is designed to work in the case-control scenario or both. Further, unlike PUL, OCC also opens up for the possibility that some labeled negative data may be available, although not well enough sampled or statistically representative enough to build a traditional binary classifier. Note that the PUL acronym has also been used by Elkan and Noto (2008) about a learning algorithm using the SCAR assumption and a specially held-out validation set to estimate the probability of a positive sample being labeled.

The SCAR assumption does not hold for our application with the available ground reference data from our AOI. The weaker “selected at random” condition, which assumes that labeled examples “are a biased sample from the positive distribution, where the bias completely depends on the attributes” (Bekker and Davis 2020), also does not hold. Our ground reference data is selected systematically from a limited area, and the sample selection bias is related to the attributes of the feature vectors due to some spatial correlation in the type of background vegetation, soil conditions, tree densities, mortality rates, etc. However, this bias is not completely dependent on the attributes of the feature vectors. Therefore, according to our definition above, we use the term OCC and not PUL in this work, although we acknowledge that the distinction between the terms is not well-established.

There are many different OCC methods to choose from, and the choice should naturally depend on the application. The OCC taxonomy by Khan and Madden (2014) divides the methodology into “one-class support vector machine” (OCSVM) and “non-OCSVM.” However, this taxonomy (Khan and Madden 2014) is focused on the text domain and only mentions in passing other application domains

without summarizing which OCC methods they use. A more useful taxonomy, at least when it comes to guiding our choice of methods, is provided in the review article by Bekker and Davis (2020), which lists three different categories of PUL methods: two-step techniques, biased learning, and class prior incorporation. The latter two invoke the SCAR assumption and are thus not applicable in our case. We, therefore, end up using the two-step technique, which was previously reviewed.

OCC has been used to extract a particular land cover type from remote sensing images. It has also been applied for targeted change detection. Camps-Valls et al. (2008) tested a one-class kernel support vector domain descriptor (SVDD) method for heterogeneous change detection of transitions between urban and non-urban landcover on an optical-SAR image pair. SVDD is similar to one-class SVM (OCSVM) but uses a hypersphere instead of a hyperplane for separation (Múnoz-Marí et al. 2010). It can be shown that with normalized data and isotropic kernels, the two methods give the same results (Múnoz-Marí et al. 2010). The SVDD compared favorably to a simple single hidden layer multilayer perceptron (MLP) and other SVM classifiers trained on both changed and unchanged pixels (Camps-Valls et al. 2008).

Li and Xu (2010) used OCSVM with a radial basis function (RBF) kernel to detect changes in two bitemporal pairs with features derived from Landsat-5 Thematic Mapper (TM) images. In the more complicated experiment of urban expansion, OCSVM was used separately for each of the different change types (water-urban, soil-urban, vegetation-urban) (Li and Xu 2010), thus requiring three separate classification procedures and three different labeled training datasets. OCSVM compared favorably to post-classification comparison (PCC) in tests performed on balanced samples from both datasets.

Two modified SVM methods were tested on four remote sensing problems by Múnoz-Marí et al. (2010), one of which was a homogeneous change detection problem. The two methods considered were variations of OCSVM and biased SVM (Liu et al. 2003) adjusted to better utilize the unlabeled data. OCSVM was modified to include unlabeled data to adjust the SVM kernel, while the modification of biased SVM was an adjustment to the cost function (Múnoz-Marí et al. 2010). However, only a small subset of the unlabeled samples was used due to the computational cost of inverting the kernel matrix that increases exponentially with the number of samples (Múnoz-Marí et al. 2010).

An OCC method based on a sparse representation of the features was tested for bitemporal change detection of a flood in a homogeneous multispectral dataset and compared with an RBF-kernel OCSVM by Ran et al. (2016). Ran et al. (2018) reported results using kernelized versions of the sparse representation classifiers.

Ye et al. (2016) presented an OCC-based method for targeted change detection combining SVDD with results from change vector analysis (CVA) and PCC. However, due to difficulties in finding tight boundaries for the target cluster in the feature space, the changed class was subdivided into several subclasses depending on the spectral signature, each dependent on a balanced training dataset (Ye et al. 2016). The approach was tested on homogeneous image pairs, and

good results were reported on selected balanced training and test datasets.

Jian et al. (2022) used generative adversarial networks (GANs) to perform OCC change detection. It was based on “spatial-spectral” features extracted from a stack of three homogeneous RGB remote sensing images, where the change occurred in the latest image of the stack. Contrary to the other methods discussed, the training dataset was created from unchanged data from the first two images of the stack. Two GANs generate samples from a “change data” distribution, and a discriminator was trained to separate unchanged from generated change samples (Jian et al.

2022). This discriminator was then finally used to detect changes in the spatial-spectral features generated from the second and third images in the stack where the change occurred. The method performed comparably to deep learning-based unsupervised change detection methods and other OCC methods when these, contrary to the normal setting of labeled positive data, were trained with data from the unchanged negative class. It should be noted that while the concept of using only unchanged data for OCC-based change detection is intriguing, it requires additional unchanged images and is unable to perform targeted change detection.