Domain adaptation in segmenting historical maps: A weakly supervised approach through spatial co-occurrence

Journal Article

Author(s):
Wu, Sidi; Schindler, Konrad; Heitzler, Magnus; Hurni, Lorenz

Publication date:
2023-03

Permanent link:
https://doi.org/10.3929/ethz-b-000599929

Rights / license:
Creative Commons Attribution 4.0 International

Originally published in:
ISPRS Journal of Photogrammetry and Remote Sensing 197, https://doi.org/10.1016/j.isprsjprs.2023.01.021

Funding acknowledgement:
188692 - HistoRiCH: Historical river change – Planning for the future by exploring the mapped past (SNF)

ISPRS Journal of Photogrammetry and Remote Sensing 197 (2023) 199–211

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Domain adaptation in segmenting historical maps: A weakly supervised approach through spatial co-occurrence

Sidi Wu [a],[*], Konrad Schindler [b], Magnus Heitzler [a], Lorenz Hurni [a]

[a] Institute of Cartography and Geoinformation, ETH Zurich, Switzerland
[b] Photogrammetry and Remote Sensing, ETH Zurich, Switzerland

ARTICLE INFO

Keywords:
Domain adaptation
Image segmentation
Historical maps
Change detection
Spatial co-occurrence

ABSTRACT

Historical maps depict past states of the Earth's surface and make it possible to trace the natural or anthropogenic evolution of geographic objects back through time. However, the state of the depicted reality is not the only source of change: maps of varying age can differ in terms of graphical design, and also in terms of storage conditions, physical ageing of pigments, and the scanning process for digitization. Consequently, a computer vision system learned from a specific (source) map series will often not generalize well to older or newer (target) maps, calling for *domain adaptation*. In the present paper we examine – to our knowledge for the first time – domain adaptation for segmenting historical maps. We argue that for geo-spatial data like maps, which are geo-localized by definition, the spatial co-occurrence of geographical objects provides a supervision signal for domain adaptation. Since only a subset of all mapped objects co-occur, and even those are not perfectly aligned due to both real topographic changes and variations in map generalization/production, they only provide *weak supervision* — still they can bring a substantial benefit over completely unsupervised domain adaptation methods. The core of our proposed method is a novel self-supervised co-occurrence network that detects co-occurring objects across maps (specifically, domains) with a novel loss function that allows for object changes and spatial misalignment. Experiments show that, for the task of segmenting hydrological objects such as rivers, lakes and wetlands, our system significantly outperforms two state-of-art baselines, even with limited supervision (e.g., 5%). The source code is publicly available at https://github.com/sian-wusidi/spatialcooccurrence.

1. Introduction

Historical maps are the only comprehensive, spatially explicit source of information about the Earth's surface before the invention of modern air- and space-borne Earth observation. They can be used to study past states of the Earth's surface (Bromberg and Bertness, 2005; Levin et al., 2010), and when combined with recent data to analyse the long-term evolution of geographic features (San-Antonio-Gómez et al., 2014; Burghardt et al., 2022; Tonolla et al., 2021; Picuno et al., 2019). Spatially detailed assessments of spatio-temporal developments, due to both anthropogenic or natural drivers, serve as a basis to for strategies to manage, preserve or restore the affected landscape (San-Antonio-Gómez et al., 2014; Walz, 2008; Hoyer and Chang, 2014). However, information from analog map sheets or raster scans is mostly retrieved by interactive, manual digitization. This is not only time-consuming, tedious and costly, but may also limit the geographical extent and the time window considered in a study. Deep encoder–decoder networks have greatly advanced image segmentation (Chen et al., 2018; Ronneberger et al., 2015), and thus also the automatic semantic analysis of historical maps (Uhl et al., 2020; Heitzler and Hurni, 2020; Wu et al., 2022b). However, maps produced at different times exhibit strong variations with respect to graphical style, scale, craftsmanship, drawing or printing quality, storage conditions of the analog sheets, and scanning process used to digitize them. It is therefore challenging to learn a single, generic segmentation model. Even maps created, stored and digitized with high standards vary considerably, especially in terms of the symbols used for foreground objects and the textures and colours of background elements, see the example in Fig. 1. A model trained on a specific map series typically does not generalize all that well to other map series. This is not surprising: it has repeatedly been pointed out that even subtle shifts between the training and test distributions of a neural network can swing its predictions and seriously harm its performance, e.g., Tzeng et al. (2017). Fortunately, despite differences in appearances, maps of the same type (e.g., topographic

[*] Corresponding author.
E-mail address: sidiwu@ethz.ch (S. Wu).

https://doi.org/10.1016/j.isprsjprs.2023.01.021
Received 1 October 2022; Received in revised form 13 January 2023; Accepted 25 January 2023
Available online 14 February 2023

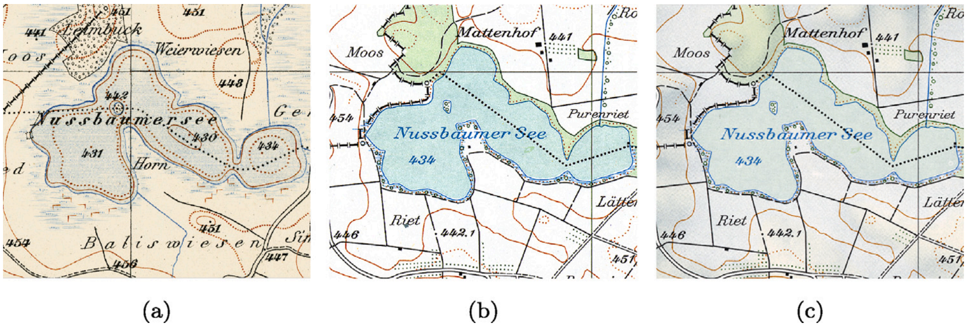


Fig. 1. Varying map designs in foreground objects (lakes/forests) between (a) and (b), and in background colours and textures between (a), (b) and (c). (a) (b) (c) are from *Siegfried Maps*, old *Swiss National Maps* (KOMB) and old *Swiss National Maps* (KREL), respectively.

maps) share the same underlying geography, as well as common principles of cartography. Intuitively, it appears possible to train a model on one "source" map series and then deploy it on another "target" series, if, during training, the model is exposed to the cartographic (dis-)similarities between the two series. Importantly, doing this does not require labels for the target domain.

Standard Domain Adaptation (DA) addresses shifts between the training (source) and test (target) domains by aligning them in a statistical sense. Depending on the availability of ground truth in the target domain, the alignment is carried out in supervised or unsupervised fashion. Different approaches have been developed that match the source and target distributions at the level of the input images (Li et al., 2019; Tasar et al., 2020), some intermediate feature space (Hong et al., 2018; Hoffman et al., 2016), or the predicted outputs (Tsai et al., 2018; Vu et al., 2019a). Our work aligns domain shifts in the output space, where we assume a certain level of consistency in structural properties like topology and spatial layout of geographical objects. E.g., a stream is unlikely to be a closed curve, or a lake should not be enclosed by a river. Rather than the direct prediction (Tsai et al., 2018), enforcing consistency in the prediction entropy (Vu et al., 2019a) has achieved state-of-art performance. On the one hand, according to Wu et al. (2022a), Vu et al. (2019a), entropy maps that capture normal object boundaries preserve meaningful structural information. On the other hand, target predictions without direct supervision tend to have higher entropy, which means lower confidence, than source predictions (Wu et al., 2022a). In total, aligning distributions of the prediction entropy can amplify the structural similarity between the source and target output and simultaneously reduce the entropy of the target output, so that the domain gap can be bridged. In our work, we apply adversarial entropy minimization (Vu et al., 2019a,b) for domain adaptation across different map designs.

Iqbal and Ali (2020) showed that, when segmenting built-up regions in aerial or satellite images, image-level labels as additional supervision improve DA. It is an interesting and intuitive finding that even a weak supervision signal helps to better bridge domain gaps. Yet it requires a sufficient quantity of image-level labels in the target domain, arguably a relatively contrived setting in the context of map digitization. On the contrary, we argue that a different form of weak supervision is almost always available for geo-spatial data, including not only maps but also satellite images, GPS trajectories, etc.: they are geo-referenced, so one can pair data from different domains after transforming them into the same coordinate reference system. Importantly, in many settings one *cannot* simply assume that data from two different domains are aligned perfectly within the required accuracy. If that were the case one could trivially solve the task by segmenting in the source domain and transferring the labels. Due to different acquisition techniques/times/conditions as well as different processing, geo-locating and generalizing systems, there are significant displacements. E.g., the location of the same road or river in different map sheets may differ by a lot more than its width, exemplified in Fig. 2, leading to a situation where it is present in both sheets but the two instances do not overlap in space; making it impossible to transfer labels via their geo-coordinates. In other words, the common geo-reference does not mean that objects *coincide* in different map sheets of the same location. But it does mean that most objects *co-occur* at roughly the same locations. As we will show, that weaker constraint is already helpful for DA. Note that co-occurrence is a property that applies to any two co-located maps, independent of ground truth labels. It can be seen as a form of self-supervision.

In our work, we investigate the DA problem in the context of segmentation in historic maps. Our focus is on hydrological objects. As explained above, the principle of our method is to use co-occurring geographical objects in the source domain as context information to adapt the segmentation engine to the target domain. However, different maps normally show the Earth's surface at different times. Due to natural as well as anthropogenic changes (e.g., rivers changing course, wetlands being drained, roads and buildings being constructed or torn down), not every object will have a matching instance in the source domain. To filter out such cases, we employ a co-occurrence detection network that discovers co-occurring objects between source and target data in a self-supervised way. The detection step includes a tolerance on the spatial location, to account for misalignment due to geo-localization errors, various distortions, and map generalization. Once co-occurrence has been established, the label prediction in the source domain, with no domain shift, can serve as additional evidence for the one in the target domain. We still follow the line of research on DA at the predicted output, specifically the output entropy, to give supervision in the entire output space. To achieve this, we adopt adversarial learning to minimize the discrepancy between entropy maps from the source and target domains. We assume this is vital to support regions of significant changes, where supervision of co-occurrence is unavailable.

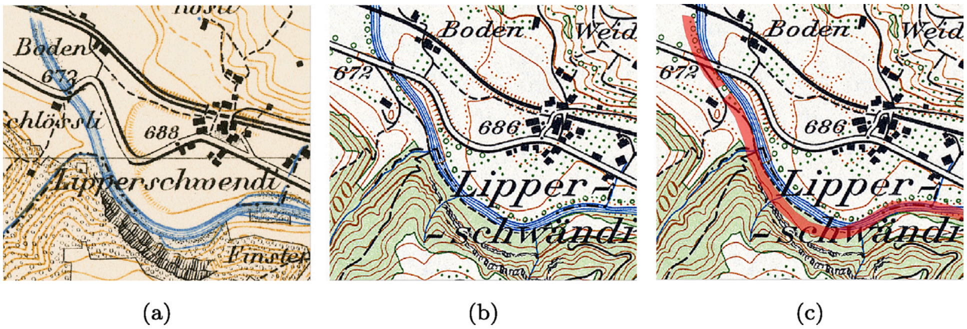
Our main contributions can be summarized as follows:

- We propose a novel weakly self-supervised DA pipeline that uses the co-occurrence of geographic objects as auxiliary supervision signal to adapt segmentation across different image domains, assisted by entropy-based adversarial learning. While the framework is general and potentially applicable to a range of geo-spatial imagery, our application scenario is the segmentation of historical maps from different series of different ages.
- To avoid erroneous supervision signals due to changes, we develop a dedicated network to detect co-occurrence, or its absence, between co-registered images in a self-supervised manner.
- For the co-occurrence detector, we introduce a loss function that allows for spatial misalignment between the source and target domains, relaxing the assumption of spatial co-incidence to a weaker notion of spatial co-occurrence.

2. Related work

2.1. Domain adaptation

Carefully supervised image segmentation models often fail to generalize to test data that follow even a slightly different distribution



(a)

(b)

(c)

Fig. 2. Displacements of the river Töss between a *Siegfried Map* sheet at 1882 (a) and an old *Swiss National Map* sheet at 1956 (b). Both sheets are geo-referenced in the same coordinate reference system. In (c) we overlay (b) with the ground-truth river from (a), shown in red. The displacement is significant. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

than the training data, a situation termed “domain shift”. Consequently, a dedicated line of research has emerged that investigates how a model trained on one (source) domain can be adapted to another (target) domain (Tzeng et al., 2015; Motiian et al., 2017; Zhuang et al., 2015; Tommasi et al., 2016; Tzeng et al., 2014; Ganin et al., 2016). The arguably most relevant case is Unsupervised Domain Adaptation (UDA), where the DA must be accomplished in the absence of labelled target data — if labelled examples from the target domain are available, the problem largely reduces to transfer learning which, with the rise of data-hungry deep networks, has become a standard procedure. A common practice is to pre-train the model with labelled source data and fine-tune it with labelled target data (Pires de Lima and Marfurt, 2019; Wang et al., 2018). UDA can be categorized into methods that bridge the domain gap at the level of input data, at the level of latent features, or at the output level (Toldo et al., 2020).

For input-level adaptation, generative networks (Goodfellow et al., 2020) are often used to translate images from the source domain to the target domain (Hoffman et al., 2018; Li et al., 2019; Tasar et al., 2020), assuming that a model trained on those translated images will be applicable to the target domain. The major drawback of these approaches is that the image-to-image translation is unaware of the labels, and not guaranteed to produce outputs with coherent and separable class-conditional distributions. Informally speaking, image-to-image translation is tuned to generate images that match the target domain in terms of global visual appearance, not in terms of local evidence for segmentation.

Feature-level adaptation aligns the distributions of latent activations within the network, such that they are representative of the target domain despite having been trained on the source domain. Chen et al. (2019) proposed the Progressive Feature Alignment Network (PFAN) that progressively learns to align latent features across domains. Hong et al. (2018), Hoffman et al. (2016) trained a discriminator (Goodfellow et al., 2020) to distinguish intermediate feature representations between source and target images and boost their similarity through adversarial training. However, unlike image classification, complex structures in the high-level feature space for semantic segmentation make it difficult to stably train the discriminator and may harm DA performance (Tsai et al., 2018; Yang and Soatto, 2020). To sidestep the complexity of the latent feature space, a further line of work seeks to cross the domain gap in the output space. While Tsai et al. (2018), Biasetton et al. (2019) train an adversarial mapping directly to the model predictions, Vu et al. (2019a,b) instead align the prediction entropies between the source and target domains, and achieve state-of-art performance. The underlying assumption of that approach is that predictions for the unseen target domain should have lower confidence and thus higher entropy, than those for the source domain — in earlier work (Wu et al., 2022a) we found this to indeed be the case for historical map segmentation. On the other hand, matching entropy distributions can also enforce structural similarity between source and target, since high entropy is often found along object edges (Vu et al., 2019a; Wu et al., 2022a).

2.2. Weakly supervised learning

Many practical learning tasks are hampered by the lack of huge, subject-specific training data sets with detailed labels. This has lead to an interest in methods that can learn from inexact, incomplete, or inaccurate labels that can be more easily obtained in large quantities. For instance, fine-grained pixel-level labels are tedious and expensive to annotate, and one could save effort by substituting them with coarse annotations, like a single label per image (Ahn et al., 2019; Li et al., 2018; Wei et al., 2016; Kolesnikov and Lampert, 2016; Wang et al., 2020; Chen et al., 2020), sparse, point-wise labels (Bearman et al., 2016; Wang et al., 2020; Yu et al., 2021), scribbles (Lin et al., 2016; Wu et al., 2018) or bounding boxes (Khoreva et al., 2017; Papandreou et al., 2015). The hope is that partial annotations can be propagated to unlabelled pixels that are similar in image or feature space (Lin et al., 2016; Kolesnikov and Lampert, 2016; Ahn et al., 2019). A complementary direction is to rely on labels that are available in large volumes, but noisy, e.g., when obtained by crowd-sourcing (Kaiser et al., 2017; Uzkent et al., 2019). The prevalent strategies to handle label noise are data cleansing and filtering (Malossini et al., 2006; Thongkam et al., 2008; Frenay and Verleysen, 2014), noise-tolerant learning techniques (Mnih and Hinton, 2012; Beigman and Beigman Klebanov, 2009; Patrini et al., 2017), and explicit modelling of the noise distribution (Lu et al., 2017; Xiao et al., 2015).

For geo-spatial data, a main source of label noise is inaccurate alignment, hence weak learning based on (partially) misaligned labels is related to our notion of co-occurrence. E.g., Kaiser et al. (2017) directly pair satellite images with public OpenStreetMap (OSM, http://www.openstreetmap.org/) data to train semantic segmentation models. It is worth mentioning that spatial occurrence has already been explored in the context of DA. Sakaridis et al. (2022) associated daytime street-view images, as the source domain, with night-time images as target domain, based on GNSS positions. The shared content was used to guide DA to the less favourable nighttime conditions.

The simple map-based matching does not account for changes in image content, and it was necessary to carefully separate and handle even the moderate amount of misalignment induced by moving objects like humans or vehicles. In our setting, misalignments are a lot more extreme, to the point where almost none of the pixels of corresponding map objects coincide. In maps, such offsets are frequent due to surveying or production bias, subjective map generalization, residual map distortion, or actual changes between different production times. They concern in particular small objects like houses and different point signatures, and narrow line objects like rivers, roads or iso-contours. The present work addresses this issue, by training a network to discover co-occurring, but not spatially coincident objects, and utilize them to guide adaptation only with matched/unchanged objects.

2.3. Historical map segmentation

Historical maps contain useful information about past states of the Earth's surface. One can trace the past evolution of natural and man-made objects in them, furthermore one can also combine them with modern Earth observation data to identify and understand long-term geospatial changes (Uhl et al., 2021). Given the large available volume of scanned historical maps, the past two decades have seen increased research efforts towards automatic map processing techniques (Bin and Cheong, 1998; Leyk, 2009; Chiang and Knoblock, 2009). Despite this interdisciplinary effort at the intersection of cartography, geo-information science and computer vision, many map digitization tasks still require human intervention, or custom tuning to every specific map style and object type (Wu et al., 2022b).

Similar to satellite remote sensing, historical map processing has embraced the deep learning era, in particular recent advances in computer vision (Chen et al., 2018; Ronneberger et al., 2015; He et al., 2017; Visin et al., 2016; Chen et al., 2016). Deep neural networks have been applied to extract buildings (Uhl et al., 2020; Heitzler and Hurni, 2020), roads (Ekim et al., 2021) and water bodies (Wu et al., 2022b) from maps. However, all those methods are restricted to a specific map series, without regard to generalization. Considering the diversity of maps from different epochs and cartographic traditions, it appears impractical to custom-tailor map processing to every single map series by annotating an extensive, representative amount of ground truth. To the best of our knowledge, our work is the first to investigate DA across different styles of historical maps.

3. Method

This section introduces our proposed domain adaptation pipeline for map segmentation. We begin with a definition of the problem and an overview of our proposed system, then we describe its components in more detail.

3.1. Problem definition and algorithm overview

Our task is an instance of UDA, where images in the source domain are labelled, whereas all images in the target domain are unlabelled. For our case we may assume that all map sheets have the same size (w.l.o.g., as it is common practice to resize or tile images for processing with neural networks). Instead of Let x_s denote a source image from a source set $\mathcal{X}_s \subset \mathbb{R}^{H\times W\times 3}$ with height H, width W and three RGB channels, and let y_s be the corresponding binary label image from an associated ground-truth set $\mathcal{Y}_s \subset \mathbb{R}^{H\times W\times C}$, with one channel for each of the C output classes to form the problem of multi-class binary segmentation. Moreover, we have unlabelled target images x_t from a target set $\mathcal{X}_t \subset \mathbb{R}^{H\times W\times 3}$, for which we want to make predictions $\hat{y}_t$, with the same dimensions as y_s. The source and target sets have size n_s and n_t, respectively. For some of the target images there is a source image that covers the same region of the Earths surface, forming $m \leq min(n_s, n_t)$ co-located pairs. We define $\text{cooc}_r = \frac{m}{min(n_s,n_t)}$ as the co-occurrence rate. These pairs depict many co-occurring objects with the same labels and approximately, but not exactly, the same coordinates. The co-located pairs provide weak supervision to the UDA process.

A graphical depiction of the complete architecture is shown in Fig. 3. A shared encoder F and decoder G constitute the segmentation network $S = F \circ G$, which operates independently on the source and target domains. The two domains are linked via a co-occurrence detector O and a discriminator D. Since we aim at making predictions $\hat{y}_s = S(x_s)$ from source images and $\hat{y}_t = S(x_t)$ from target images consistent with each other, we train the discriminator D to distinguish whether a predicted segmentation map is from the source or target domain. The adversarial loss, in turn, forces the segmentation network S to generate predictions with similar distributions in the target domain and the source domain. Furthermore, for a co-located pair of source and target images, the encoded features $f_s = F(x_s)$ and $f_t = F(x_t)$ are concatenated and then passed to the co-occurrence detector O. F and O together form the weak supervision network $U = F \circ O$. For the segmentation of the source images, where ground truth for direct supervision is available, we employ the dice loss (Milletari et al., 2016), to account for the sparsity of hydrological objects in historical maps (Wu et al., 2022b):

$$\mathcal{L}_{\text{seg}_s} = 1 - \frac{2\sum_{i=1}^{H\times W\times C} \hat{y}_s{}^{(i)} y_s^{(i)}}{\sum_{i=1}^{H\times W\times C} \hat{y}_s{}^{(i)} + \sum_{i=1}^{H\times W\times C} y_s^{(i)}} \tag{1}$$

where $\hat{y}_s{}^{(i)}$ and $y_s^{(i)}$ are a class- and pixel-wise prediction score and its ground truth.

3.2. Co-occurrence detection

Given the feature maps f_s and f_t of the source and target images, the co-occurrence detector O outputs an attention map $\alpha_{st} = O(f_s, f_t)$ that has the same dimensions $H \times W \times C$ as the segmentation outputs. That map indicates unchanged regions between x_t and x_s where the prediction $\hat{y}_t = G(f_t)$ should locally match $\hat{y}_s = G(f_s)$ and y_s. The co-occurrence detection is learned in self-supervised fashion, by maximizing the similarity between the target prediction $\hat{y}_t \odot \alpha_{st}$ and the source label $y_s \odot \alpha_{st}$ masked by the attention map. In this way, the target labels will be pushed towards the source labels in regions where the map content is similar; whereas regions where the map content has actually changed will be assigned a low co-occurrence score, so as to avoid the associated penalty.

However, as explained above, we can only expect co-occurrence, not exact coincidence. Even for objects that have not changed in the world, the corresponding map elements in x_s and x_t will not exactly match, due to localization errors, distortions, and map generalization. To nevertheless detect the co-occurrence of such objects, we propose to include a distance tolerance in the loss function, to obtain a *Buffered Dice Coefficient (BDC)*:

$$BDC(\hat{y}, y) = \frac{2\left|\hat{y} \cap b_s(y)\right| |y|}{|y|^2 + |\hat{y}|^2} = \frac{2\sum_i \hat{y}^{(i)} b_s\left(y^{(i)}\right) \sum_i y^{(i)}}{\left(\sum_i y^{(i)}\right)^2 + \left(\sum_i \hat{y}^{(i)}\right)^2} \tag{2}$$

The BDC accounts for misalignment between $\hat{y}$ and the corresponding ground truth y simply by looking at a fixed-width buffer b_s around the ground truth object, a standard operation in GIS analysis (Bhatia et al., 2013). An efficient way to compute the buffer zone is to downsample the ground truth by a factor s with max pooling, and then upsample it back to the original resolution. See Fig. 4. This formulation can be differentiated and has the maximum value when $\hat{y} = y$ in the buffer $b(y)$. When $s = 1$, BDC functions as normal dice loss, as shown in Eq. (1).

We can use the masked BDC to weakly supervise the target images:

$$\mathcal{L}_{\text{seg}_t} = 1 - BDC\left(\hat{y}_t \odot \alpha_{st}, y_s \odot \alpha_{st}\right) \tag{3}$$

To prevent the co-occurrence detector from always predicting low value, we add a regularization term:

$$\mathcal{L}_{\text{reg}} = 1 - \frac{1}{H \times W} \sum_{j=1}^{H\times W} \sum_{c=1}^{C} \beta_c \alpha_{st}^{(j,c)} \tag{4}$$

where the β_c are class-specific weights.

3.3. Entropy-based adversarial learning

Through the co-occurrence detector, unchanged objects in the target image receive supervision from the source label. As additional supervision for all elements, including the changed ones, we adopt adversarial learning which is widely used for UDA. Following Vu et al. (2019a,b), we do not apply the discriminator D directly to the predicted class probabilities, but rather to their entropy:

$$E^{(h,w,c)} = -\hat{y}^{(h,w,c)} \cdot \log\left(\hat{y}^{(h,w,c)}\right) \tag{5}$$

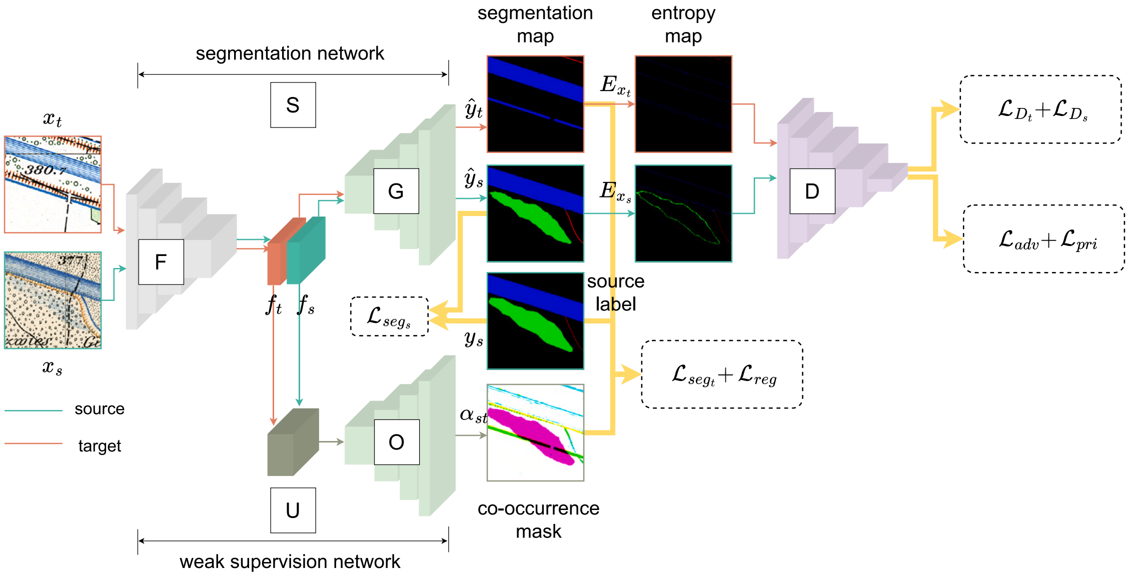


Fig. 3. Overview of our proposed algorithm. Features extracted by an encoder F from either source or target images are passed to the same decoder G for multi-class binary segmentation. Features from co-located source and target images are concatenated and passed to the co-occurrence detector O to generate co-occurrence masks. The source labels serve as weak supervision for the co-occurring/unchanged objects in the target images. An adversarial loss (discriminator) D on entropy maps of the two segmentations enforces consistency of the predictions.

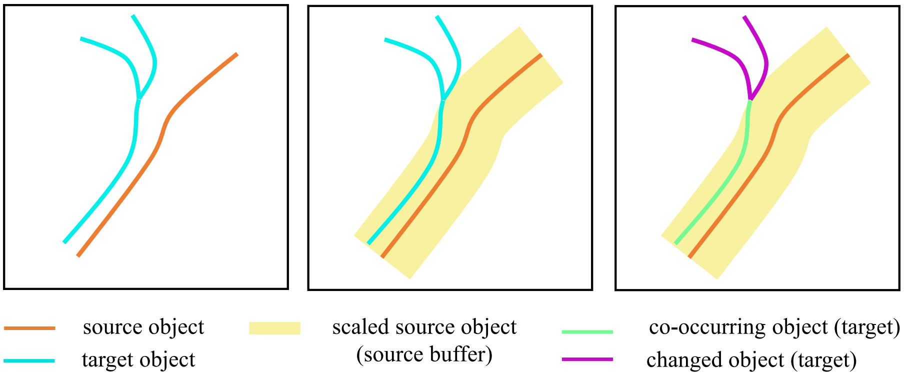


Fig. 4. Co-occurrence detection with BDC. To account for misalignment of map elements, a buffer is placed around the source element to find a potentially matching target element.

$E^{(h,w,c)}$ is the class- and pixel-wise entropy. The discriminator is a binary classifier that tries to distinguish entropy maps of source and target predictions, E_s and E_t; thereby encouraging the segmentation network for the target domain to predict class distributions that are similar to those in the source domain. Accordingly, its loss function is:

$$\mathcal{L}_{D_s} = \left\| D(E_s) - y_{source} \right\|^2$$
$$\mathcal{L}_{D_t} = \left\| D(E_t) - y_{target} \right\|^2 \tag{6}$$

where $y_{source} = 1$ and $y_{target} = 0$. Training the discriminator is alternated with training the segmentation network to fool the discriminator, via the loss term

$$\mathcal{L}_{\text{adv}} = \left\| D(E_t) - y_{source} \right\|^2 \tag{7}$$

This pushes the target predictions to have similar entropy distributions as the source predictions. As pointed out by Vu et al. (2019a), entropy-based adversarial learning favours easy classes and tends to over-fit to them. To alleviate that problem we follow the strategy recommended by Vu et al. (2019a) and guide the training with a class frequency prior P. We calculate the area (the number of pixels) per object class in the source images and normalize it over all the classes. Deviations of the predicted probabilities from that prior are penalized. To account for the variability of the frequency distribution between different batches, as well as frequency changes due to changed map elements, we relax the prior constraint with a factor $\mu \in [0, 1]$:

$$\mathcal{L}_{\text{pri}} = \sum_{c=1}^{C} \max\left(0, \mu \cdot \frac{\hat{y}_t^{(c)}}{\sum_c \hat{y}_t^{(c)}} - P(c)\right) \tag{8}$$

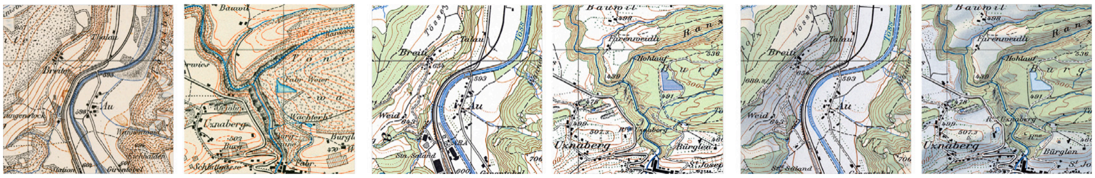
3.4. Optimization

Combining Eqs. (1), (3), (4), (7), (8) we derive the optimization problem for the segmentation network S, consisting of encoder F and the decoder G:

$$\min_{F,G} \frac{1}{n_s} \sum_{(\mathcal{X}_s,\mathcal{Y}_s)} \mathcal{L}_{\text{seg}_s} + \frac{\lambda_{\text{adv}}}{n_t} \sum_{\mathcal{X}_t} \mathcal{L}_{\text{adv}} + \frac{1}{n_t} \sum_{\mathcal{X}_t} \mathcal{L}_{\text{pri}}$$
$$\min_{F} \frac{1}{m} \sum_{(\mathcal{X}_s,\mathcal{Y}_s),\mathcal{X}_t} \mathcal{L}_{\text{seg}_t} + \frac{1}{m} \sum_{\mathcal{X}_s,\mathcal{X}_t} \mathcal{L}_{\text{reg}} \tag{9}$$

with the weight λ_{adv} for the adversarial loss. The parameters of the co-occurrence detector O are optimized by the combination of Eqs. (3) and (4):

$$\min_{O} \frac{1}{m} \sum_{(\mathcal{X}_s,\mathcal{Y}_s),\mathcal{X}_t} \mathcal{L}_{\text{seg}_t} + \frac{1}{m} \sum_{\mathcal{X}_s,\mathcal{X}_t} \mathcal{L}_{\text{reg}} \tag{10}$$



(a) Siegfried (b) KOMB (c) KREL

Fig. 5. Siegfried maps (labelled, source) and KOMB/KREL maps (unlabelled, target) used in our experiments. They are geo-referenced in the same coordinate reference system (CH1903).

Table 1
Our training dataset. We conduct experiments with the combinations Siegfried-KOMB and Siegfried-KREL.

	Map	Scale	Time of production	Training samples
source, labelled	Siegfried	1:25 k	≈1880	10 560
target, unlabelled	KOMB	1:25 k	≈1950	10 560
target, unlabelled	KREL	1:25 k	≈1950	10 560

The discriminator D is trained by minimizing the corresponding losses for both the source and the target domain:

$$\min_{D} \frac{1}{n_s} \sum_{\mathcal{X}_s} \mathcal{L}_{D_s} + \frac{1}{n_t} \sum_{\mathcal{X}_t} \mathcal{L}_{D_t} \tag{11}$$

The optimization alternates between minimizing (9) and (10) together, and minimizing (11).

4. Experiments

4.1. Datasets

Our labelled dataset is a portion of the *Siegfried Maps* from around 1880, a Swiss national map series published between 1870 and 1949. Each map sheet we use is 7000 pixels wide and 4800 pixels high, at the scale of 1:25 k, with a spatial resolution of 1.25 m/pixel. We semi-automatically vectorize four types of hydrological objects: streams (lines), wetlands (polygons), rivers (polygons), and lakes (polygons). The vector objects of each class are separately rasterized into a binary map, and the four maps are stacked into four-channel raster images that form our ground-truth annotations for multi-class binary segmentation. Starting from the 1950ies, the *Swiss National Maps* superseded previous map series and became the official topographic base map of Switzerland. We make use of old *Swiss National Maps*, which were in production until 2008. At both scales of 1:25 k and 1:50 k, two types of maps were produced — normal topographic maps (KOMB) and topographic maps with relief shading (KREL), illustrated in Fig. 5. Early editions of both KOMB and KREL, from the 1950ies, at scale 1:25 k are used as unlabelled target datasets. Each map sheet of KOMB is 14000 pixels wide and 9600 pixels high, with a spatial resolution of 1.25 m/pixel, while each map sheet of KREL is 7000 pixels wide and 4800 pixels high with a resolution of 2.5 m/pixel. To sidestep differences in spatial resolution, we upsample KREL to 1.25 m/pixel. Maps from Siegfried, KOMB and KREL are geo-referenced in the same coordinate reference system — the local Swiss reference frame CH1903. For our experiments, we randomly sample tiles of 256 × 256 pixels from the map sheets. In total, we use 10560 labelled training tiles from 95 Siegfried map sheets and 10 560 unlabelled training tiles from 30 KOMB/KREL map sheets. An overview of the training dataset is shown in Table 1. 20 full sheets of KOMB and 20 full sheets of KREL serve as test data. The ratio between the number of positive samples (containing objects of interest) and negative samples (not containing objects of interest) is empirically fixed as 3:1.

4.2. Detailed network architectures

For the segmentation network $S = F \circ G$, we use a U-Net with an ASPP block to incorporate multi-scale contexts (Wu et al., 2022b). The encoder F is comprised of five encoding blocks and an ASPP block. The decoder G has five decoding blocks and a sigmoid layer for classification. Each encoding/decoding block consists of two consecutive rounds of convolution, batch normalization, and ELU activation, with a drop-out layer between the two rounds. Max pooling is used to down-sample activation maps at the end of each encoding block while transposed convolution is used to up-sample feature maps at the beginning of each decoding block. Features extracted from each encoding block are passed to the corresponding decoding block via skip connections.

For the weak supervision network U, we concatenate the bottleneck features of both source and target images, extracted with the same encoder F. The concatenated features are fed into the co-occurrence detector O, a decoder of the same structure as G, with a sigmoid layer at the end. Low-level features of source and target images encoded by F are also concatenated and fed into the corresponding decoder stage of O via skip connections.

The architecture of the discriminator is illustrated in Fig. 6. It has five encoding blocks. Blocks 2 to 5 are each comprised of a convolutional layer with stride 2, a batch normalization layer, a leaky-RELU layer, and a drop-out layer. In the first block, the convolutional stride is 1 to ensure all input pixels are accounted for, instead a max-pooling layer with stride 4 is added at the end to down-sample the activation map. The activations after the fifth encoding block are passed to a fully-connected classification layer. We empirically found that going deeper than five blocks tends to lose thin linear structures. We also noticed that a drop-out layer in every block stabilizes the discriminator training.

4.3. Implementation details

We implement our proposed approach using Keras (Chollet et al., 2015) with Tensorflow backend (Abadi et al., 2015), and run it on a single NVIDIA 2070Ti GPU with 11 GB memory. We use the Adam Optimizer (Kingma and Ba, 2015) with initial learning rate 0.0004 for the segmentation S and weak supervision U networks, and 0.0001 for the discriminator D and the adversarial training of the segmentation. The momentum of Adam is set to 0.9 and 0.99, respectively, and the weight decay is set to 1e−4. The hyperparameter λ_{adv} for adversarial learning is set to 0.01 and 0.001 for KOMB and KREL, respectively, when the co-occurrence rate cooc_r is fixed as 20%. We notice that when the co-occurrence rate increases/decreases, λ_{adv} should be increased/decreased accordingly to enhance/attenuate adversarial learning with more/less confidence. The class weights β_j in Eq. (4) are set as 0.2, 0.2, 0.4, 0.2, for streams, wetlands, rivers and lakes, respectively. The prior P is 0.02, 0.5, 0.08, 0.4 for the classes mentioned above, and μ_a ranges from 0 to 0.4 in our experiment for different classes under different configurations. To ensure the classifier training is not destabilized in its early stages, we only start adversarial learning after both the classifier and the discriminator achieve plausible accuracies ≥ 0.6.

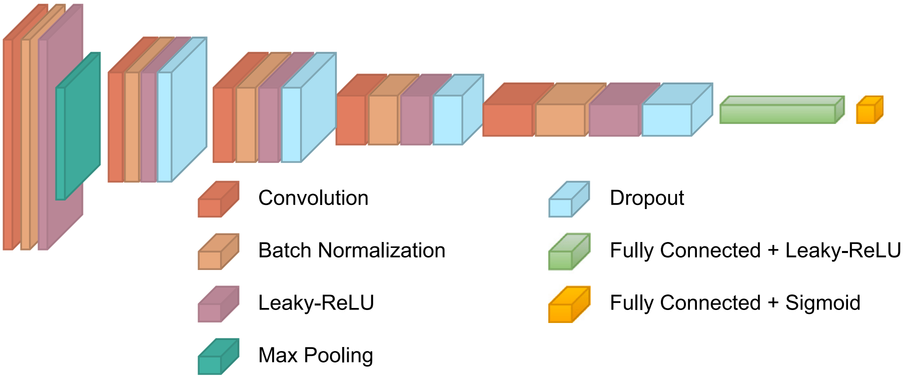


Fig. 6. The discriminator architecture.

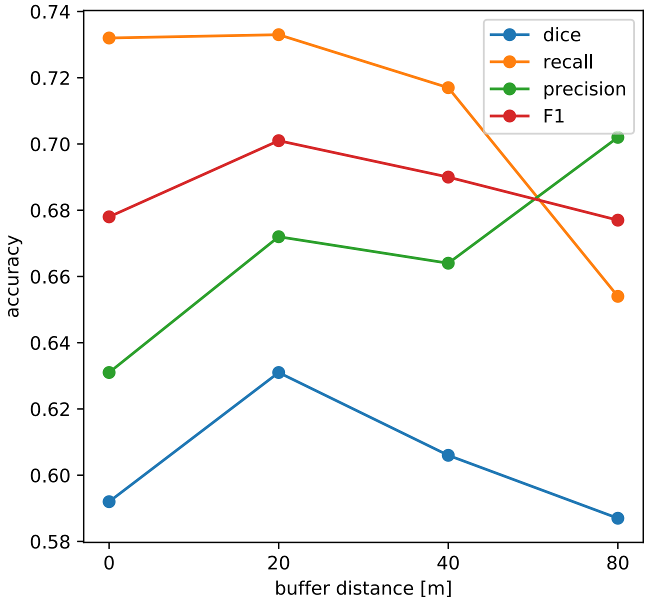


Fig. 7. Effect of the buffer distance for BDC on accuracy metrics. We fix the co-occurrence rate as 20% for this analysis.

5. Results

5.1. Impacts of the buffer distance

In the proposed loss function using BDC (2), the buffer distance plays an important role. We test the sensitivity of this parameter by varying the scaling factor s from 1 to 64 and observe the change of accuracy metrics. It equals to varying the buffer distance between 0 and 80 m, given that the spatial resolution of both our source and target data is 1.25 m/pixel. We only regard objects within up to 80 meters' spatial shifts as valid co-occurring pairs when the source and target data are carefully geo-referenced in the same coordinate reference system. When the buffer distance is set as 0, it is equal to using a normal dice loss function (Milletari et al., 2016). As we can see from Fig. 7, the overall accuracy increases when the buffer distance increases from 0 to 20 m, but decreases when the buffer distance gets larger. While precision is generally improved (fewer false positives) with a larger buffer distance, recall declines sharply (greater number of false negatives) when the buffer distance exceeds a certain limit, i.e., 40 m.

5.2. Ablation studies

We test the effectiveness of the discriminator D and the co-occurrence detector O. Without O, target images will receive all labels from their spatially-linked source images without considering regions of change. The quantitative and qualitative results are illustrated in Table 2 and Fig. 8, respectively. In Table 2 we see that O helps to improve the model performance significantly, especially for streams and wetlands, see also Fig. 8(a). It might be explained by large variations due to different generalization, delineation and painting processes, as well as actual changes like stream channelization and wetland destruction under a large temporal shift between source and target maps. The resulting large discrepancy between source labels and target images can harm the model performance if not being addressed. For completeness, we also test a version where the discriminator is applied to the latent feature representation rather than to the output. Table 2 shows that the feature space adaptation barely improves the performance. D in the output space boosts the segmentation of non-linear objects,i.e., wetlands, rivers and lakes, but has relatively limited improvement for linear objects, i.e., streams. In the qualitative results presented in Fig. 8 we see that adversarial learning corrects implausible patterns like wetland symbols misclassified as short streams in (c), and lake boundaries misclassified as rivers in (b).

5.3. Performance analysis

We compare the performances between our model and other state-of-art domain adaptation models. As the spatial co-occurrence rate cooc_r is crucial to our proposed method, we vary it and test its influence on the model accuracy. Two pairs of datasets are used in our experiment: Siegfried (source) – KOMB (target) and Siegfried (source) – KREL (target). As our proposed method is indifferent to specific neural network architectures, we retrain the state-of-art models (Tsai et al., 2018; Vu et al., 2019a) with the same architecture as ours, i.e., UNet-integrated ASPP (Wu et al., 2022b), instead of Deeplab-V2 (Chen et al., 2017) with ResNet-101 (He et al., 2016) or VGG-16 (Simonyan and Zisserman, 2015) to fairly compare the general concepts and workflows. We regard the vanilla segmentation model trained only on source images without any adaptation as our baseline. In Table 3 we see that the baseline model trained only on Siegfried maps yields poor results on both KOMB and KREL without adaptation. AdaptSeg (Tsai et al., 2018) and AdvEnt (Vu et al., 2019a) do not lead to a notable improvement — they generally reduce false negatives and improve the dice coefficient and the recall score, but increase false positives and worsen the precision. By comparison, our model, even with a small cooc_r (i.e., 5%), improves the accuracy significantly, especially for KREL. Interestingly, a larger cooc_r does not always enable better predictions. The accuracy starts declining when cooc_r increases from 20% and 10% onward for KOMB and KREL, respectively.

Qualitative comparisons are illustrated in Fig. 9 tested on KOMB (a–c) and KREL (d–f). Compared with the baseline model, AdvEnt and AdaptSeg mitigate the false positive segmentation of rivers misled by lakes and shaded relief from new designs in (b, d, e) and occasionally

Table 2
Ablation studies. We test the model performance without a discriminator D/co-occurrence detector O. Without O, target images will receive all labels from their spatially-linked source images without considering the regions of change. The co-occurence ratio cooc_r is fixed at 20%. As an additional ablation, we also apply the discriminator in the feature space rather than to the output.

Method	Average		Stream		Wetland		River		Lake	
	Dice	F1	Dice	F1	Dice	F1	Dice	F1	Dice	F1
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
w/o D, O	0.424	0.488	0.362	0.374	0.201	0.327	0.473	0.497	0.662	0.699
	0.466	0.512	0.314	0.461	0.311	0.346	0.407	0.641	0.833	0.602
w/o O	0.450	0.524	0.364	0.374	0.199	0.269	0.521	0.538	0.717	0.766
	0.503	0.547	0.280	0.564	0.422	0.197	0.417	0.757	**0.894**	0.670
w/o D	0.631	0.701	0.762	0.774	0.432	0.497	0.611	0.651	0.718	0.768
	0.672	0.733	0.701	0.865	0.707	0.383	0.501	**0.927**	0.778	0.759
Proposed	0.629	0.679	0.738	0.757	0.394	0.493	**0.685**	**0.711**	0.698	0.743
(feature)	0.607	0.771	0.670	0.872	0.399	**0.643**	**0.645**	0.793	0.713	0.776
Proposed	**0.713**	**0.762**	**0.772**	**0.781**	**0.582**	**0.631**	0.680	**0.711**	**0.816**	**0.843**
(output)	**0.745**	**0.780**	**0.702**	**0.880**	**0.828**	0.510	0.583	0.912	0.867	**0.821**

Table 3
Quantitative comparison for different methods with different levels of co-occurrence ratios cooc_r.

Method	Siegfried→ KOMB					Siegfried→ KREL				
	cooc_r	Dice	F1	Precision	Recall	cooc_r	Dice	F1	Precision	Recall
Baseline	–	0.395	0.527	0.540	0.514	–	0.186	0.347	0.299	0.412
	5%	0.533	0.606	0.595	0.617	5%	0.523	0.569	0.534	0.609
	10%	0.663	0.722	0.714	0.730	10%	**0.589**	**0.652**	**0.657**	0.648
Proposed	20%	**0.720**	**0.767**	0.741	**0.796**	20%	0.557	0.609	0.595	0.625
	50%	0.706	0.758	**0.752**	0.763	50%	0.565	0.623	0.608	0.639
	100%	0.697	0.743	0.696	**0.796**	100%	0.545	0.621	0.596	**0.649**
AdvEnt[a] (Vu et al., 2019a)	–	0.433	0.502	0.474	0.533	–	0.255	0.295	0.227	0.422
AdaptSeg[a] (Tsai et al., 2018)	–	0.406	0.494	0.457	0.538	–	0.230	0.290	0.259	0.331
Self-Training[a] (Xie et al., 2020)	–	0.458	0.578	0.572	0.585	–	0.178	0.369	0.352	0.387

[a] Denotes our retrained model. Since our algorithm is not tied to a specific network architecture, we reran (Tsai et al., 2018; Vu et al., 2019a; Xie et al., 2020) with the same architecture (ASPP-integrated UNet (Wu et al., 2022b)), for a meaningful comparison.

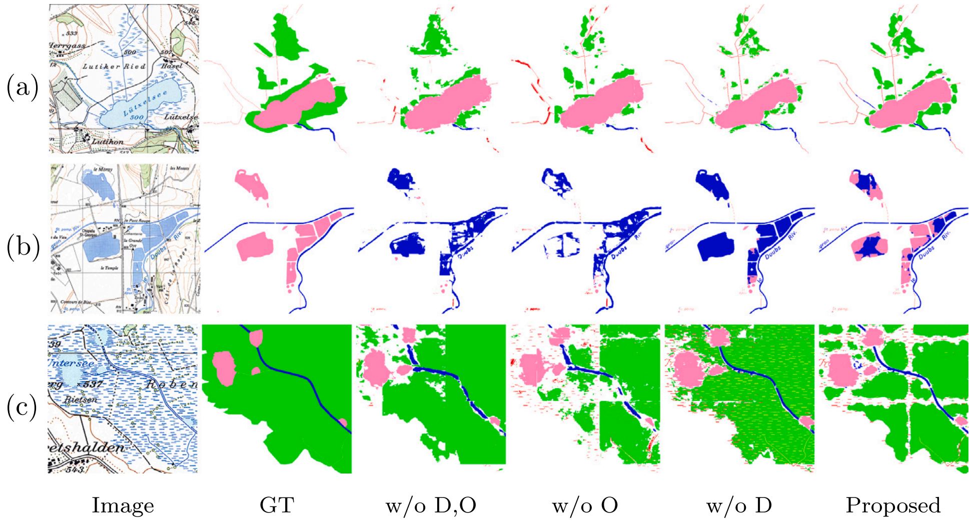


Fig. 8. Qualitative segmentation results with or without discriminator D (in the output space) and co-occurrence detector O. Streams, wetlands, rivers and lakes are represented in red, green, blue and pink, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

repair implausible topology in (a) for rivers represented as double lines that are wrongly classified as parallel streams. It also enhances the true positive segmentation of wetlands (b, c, d, e). However, on the other hand, they bring false positive predictions at arbitrary locations like wetlands and lakes in (a, b, d, e). This might be the bottleneck of these two methods to provide a unique solution that is exactly the expected output distribution without extra supervision, considering that multiple solutions can be learned from the diverse and complex output space. In addition, they inherit the limitation from the base segmentation model of false detection at out-of-distribution samples, i.e., when the map design changes completely, e.g., for lakes (b, f). By comparison, supervised by only 5% co-occurrence, the segmentation model can already generalize to the new map designs. Again, a higher cooc_r does not always lead to a better segmentation result.

We also compare our approach with self-training (Xie et al., 2020), where target predictions generated from the vanilla source segmentation model are used as pseudo-labels to retrain the target segmentation. Table 3 shows its limitation for DA and indicates again the importance of spatial co-occurrence.

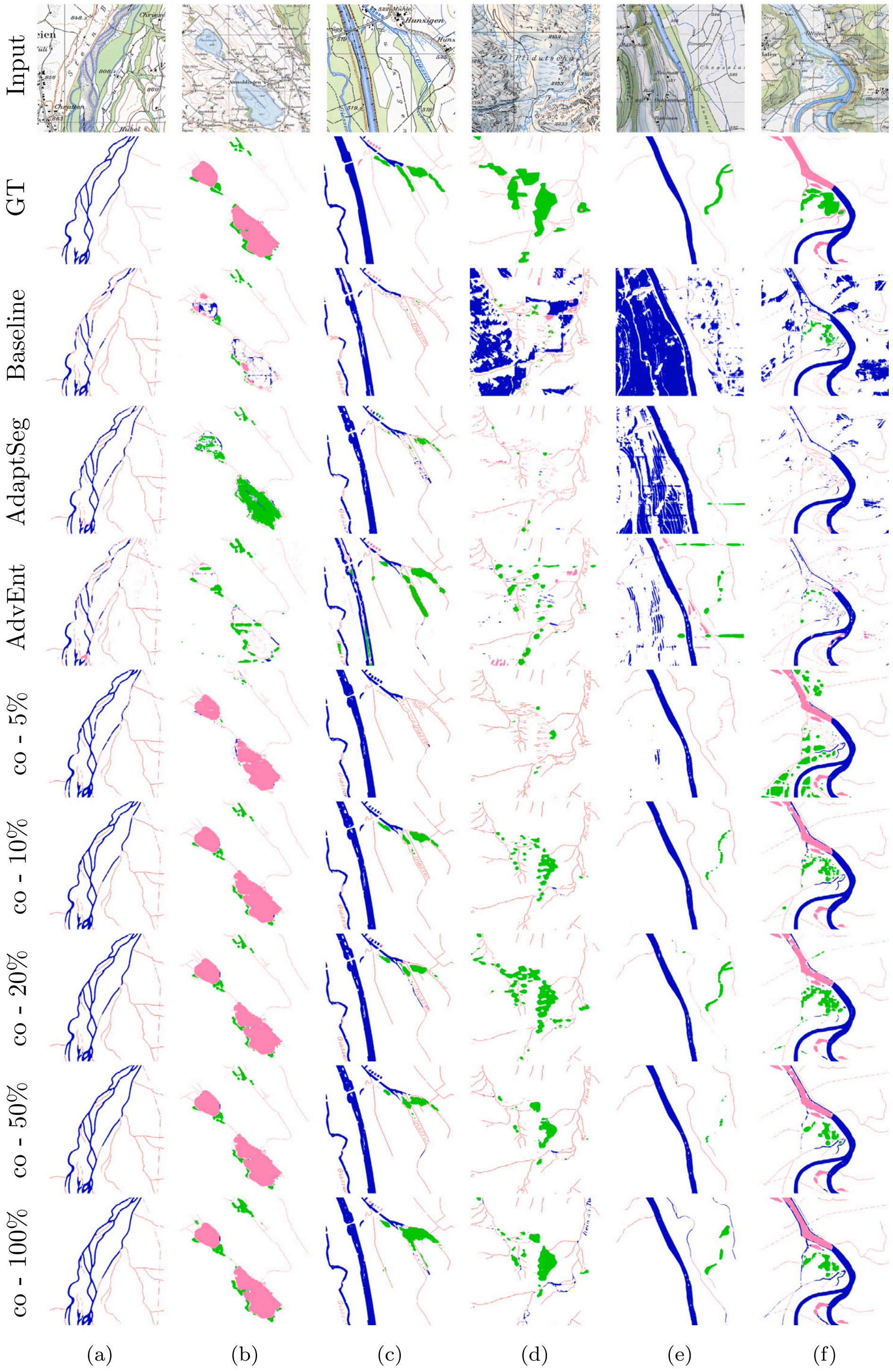


Fig. 9. Qualitative comparisons of segmentation adapted from Siegfried to KOMB (a–c) and KREL (d–f), respectively. We compare our model of different co-occurrence rates with baseline (vanilla source segmentation without adaptation), AdvEnt (Vu et al., 2019a) and AdaptSeg (Tsai et al., 2018).

6. Discussion

6.1. Co-occurrence detection

In this section, we evaluate our co-occurrence detection model and discuss its impact on DA. Specifically, we compare the predicted changes (the reverse of the predicted co-occurrence mask) with the corresponding ground truth. Table 4 illustrates the co-occurrence detection results under different configurations, (i.e., with different buffer distances, cooc_r, and with or without adversarial training). B00–B80 are experiments with a varying buffer distance from 0 to 80 m without adversarial learning for KOMB. KB-5%–KB-100% and KL-5%–KL-100% are experiments with a varying cooc_r from 5% to 100% and a fixed buffer distance of 20 m with adversarial learning for KOMB and KREL,

Table 4

Co-occurrence detection results. We compare the predicted changes (1 - co-occurrence) under different configurations with the corresponding ground truth. B00–B80 are experiments (without D) with different buffer distances and a fixed cooc_r of 20% for KOMB. KB- and KL- are experiments (with D) with different cooc_r and a fixed buffer distance of 20 for KOMB and KREL, respectively.

	B00	B20	B40	B80	KB-5%	KB-10%	KB-20%	KB-50%	KB-100%	KL-5%	KL-10%	KL-20%	KL-50%	KL-100%
Dice	0.311	0.397	0.363	0.192	0.281	0.405	**0.453**	0.440	0.363	0.286	0.321	0.258	0.304	0.336
F1	0.353	0.468	0.443	0.282	0.311	0.437	**0.486**	0.478	0.394	0.311	0.350	0.287	0.343	0.373
Precision	0.244	0.336	0.314	0.249	0.211	0.323	**0.345**	0.337	0.289	0.215	0.243	0.198	0.226	0.267
Recall	0.642	0.769	0.753	0.326	0.596	0.676	**0.822**	0.820	0.620	0.562	0.624	0.524	0.708	0.617

respectively. From B00–B80 we see that when the buffer size increases, the accuracy of co-occurrence detection increases first but then drops sharply, which is also illustrated in Fig. 10 (a–d), especially for streams. Comparing B20 with KB-20%, we find that adding adversarial learning has enhanced the detection, improving the recall score particularly. Our model generally performs better on KOMB than KREL for co-occurrence detection. This is reasonable to us since KREL has a larger difference from Siegfried than KOMB regarding designs of both foreground objects and background textures. We notice that our model has significantly higher recall than precision despite the experiment configuration, indicating that it tends to overestimate changes, or in other words, underestimate co-occurrence. One possible explanation is that our model is prone to false guidance and thus tends to underestimate the co-occurrence rather than receive more, but possibly inaccurate supervision. Adding more co-occurrence examples (higher cooc_r) does not improve the detection necessarily. Nevertheless, in Fig. 10 we see that the false positive predictions of wetland changes (e–f) do not impact the target segmentation as much as the false negative predictions of changes in (a, d, e).

To further investigate the impacts of co-occurrence detection on the DA performance, we plot the correlation maps between the accuracy of target segmentation and that of co-occurrence detection under all experiment configurations, as illustrated in Fig. 11. The Pearson correlation coefficient r is calculated for each class and the overall accuracy. The more the absolute value approaches 1, the higher the correlation is. We can find that the overall accuracy of target segmentation highly correlates with the accuracy of co-occurrence detection. However, this high correlation is not always true for individual object class: while a similar correlation can be observed for polygon objects like wetlands and lakes, this is not obvious for linear objects like streams and rivers. This is probably because the design of linear objects varies less between the source and target images than that of polygon objects, e.g., stroke and density of wetland symbols, or colour and texture of lakes. Therefore, more co-occurrence supervision can bring a more significant benefit to the latter.

6.2. Limitations of DA in Output Space

As demonstrated from previous results, simply enforcing the cross-domain structural similarity through adversarial learning in the output space can only correct the topology/geometry of the prediction to a fairly limited degree without extra supervision. This might be explained by the complexity of geospatial patterns when compared with street-view images (Vu et al., 2019a; Tsai et al., 2018) and satellite images of settlements (Iqbal and Ali, 2020; Peng et al., 2021), where purely output space adaptation can already achieve remarkable results. Adding supervision from co-occurring labels improves the segmentation significantly. Nevertheless, the co-occurrence supervision per se can bring false positive segmentation when the model tends to overestimate the changes at (1) wetland symbols as streams (2) power lines (3) text labels, as presented in Fig. 12 (a, c). While adversarial learning helps to reduce false positive segmentation for (1), as depicted in (a), it occasionally cuts off true intersections of streams and wetlands. In (b) we see that it also increases false positive segmentation of streams at lake boundaries, possibly overfitting the topology where streams flow in/out of lakes. More topology constraints can be explored in the future to enhance the awareness of the discriminator regarding topology-correctness. (c) shows another limitation of our DA method to reduce false segmentation of power lines and blue labels. One possible cause is the small proportion of the objects occupying the whole map sheet and the resulting scarcity of training samples. Schemes like active learning can be potentially applied to increment samples of these objects.

7. Conclusion

Our work has investigated the DA problem in the context of segmenting hydrological features from historical maps. We have followed a standard DA approach using adversarial training at the prediction output, specifically the output entropy, to enforce the structural consistency between the source and target output. However, a unique solution that matches the exact output distribution cannot be found merely through adversarial learning without extra supervision, considering diverse and complex geographical patterns when compared with street-view images in other applications. To overcome this bottleneck, we have made use of geographical objects that co-occur in the source and target domains. Even if they do no perfectly co-incide, such objects provide weak supervision to better bridge the domain gap. Specifically, we have proposed a novel co-occurrence detection network to detect unchanged objects between co-registered images in a self-supervised way with a novel loss function that gives a distance tolerance to relax exact spatial co-incidence to a weaker notion of spatial co-occurrence, borrowing the concept of a “buffer” from GIS. Empirically, the new loss function improves model performance within a certain distance tolerance. Obviously, the proposed approach is tailored to the specific setting where co-registered instances from the source and target domains exist and a significant amount of co-occurring objects are included. If that assumption is met, as in many geo-spatial analysis tasks, co-occurrence provides a strong cue for DA. In our experiments, we found that it supersedes generic DA methods even with quite limited co-occurrence supervision. Expectedly, the overall segmentation accuracy is strongly correlated with the performance of co-occurrence detection. However, a higher rate of co-occurring examples does not necessarily lead to better co-occurrence detection and better DA for target segmentation. In fact, our detector tends to underestimate the co-occurrence. On the one hand, it tries to avoid guiding target segmentation with false co-occurrence. On the other hand, the limited supervision resulted from underestimated co-occurrence also bounds the accuracy of target segmentation. While co-occurrence supervision already improves the performance of DA significantly, the topology and geometry of predicted outputs can be further corrected through adversarial learning. Nevertheless, we find that the discriminator can only gain limited knowledge in topology-correctness, which might be improved in the future through explicit constraints. In addition, active learning can be potentially leveraged to handle the scarcity of training

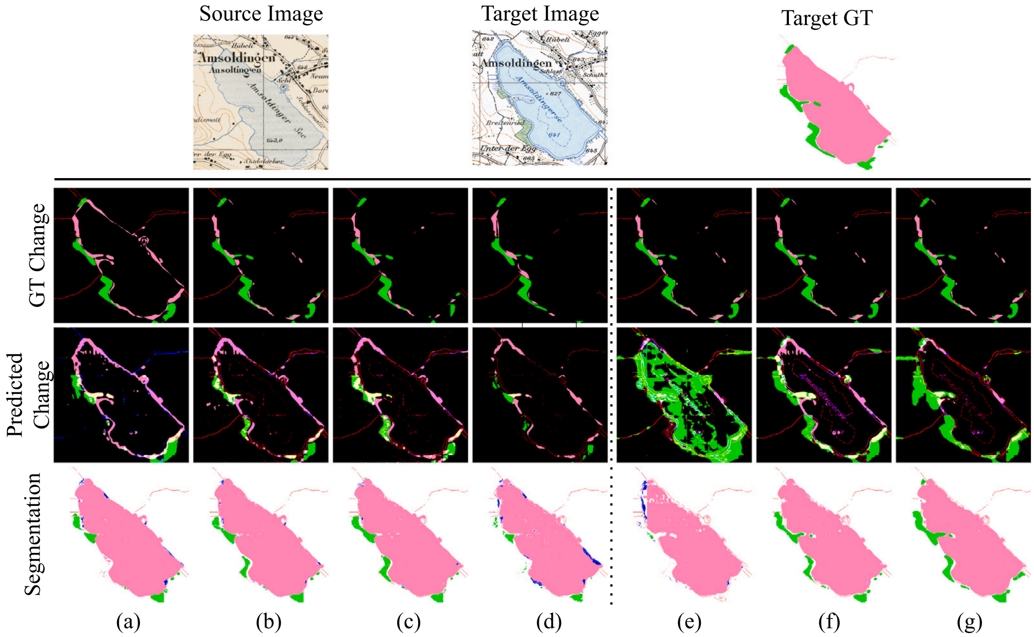


Fig. 10. Co-occurrence detection. Given the same input pair, (a–d) are the detection results with buffer distances of 0, 20, 40, 80 m and a fixed co-occurrence rate of 20%; (e–g) are the detection results with varying co-occurrence rates at 5%, 20%, 100% and a fixed buffer distance of 20 m. We revert the co-occurrence mask to highlight changes.

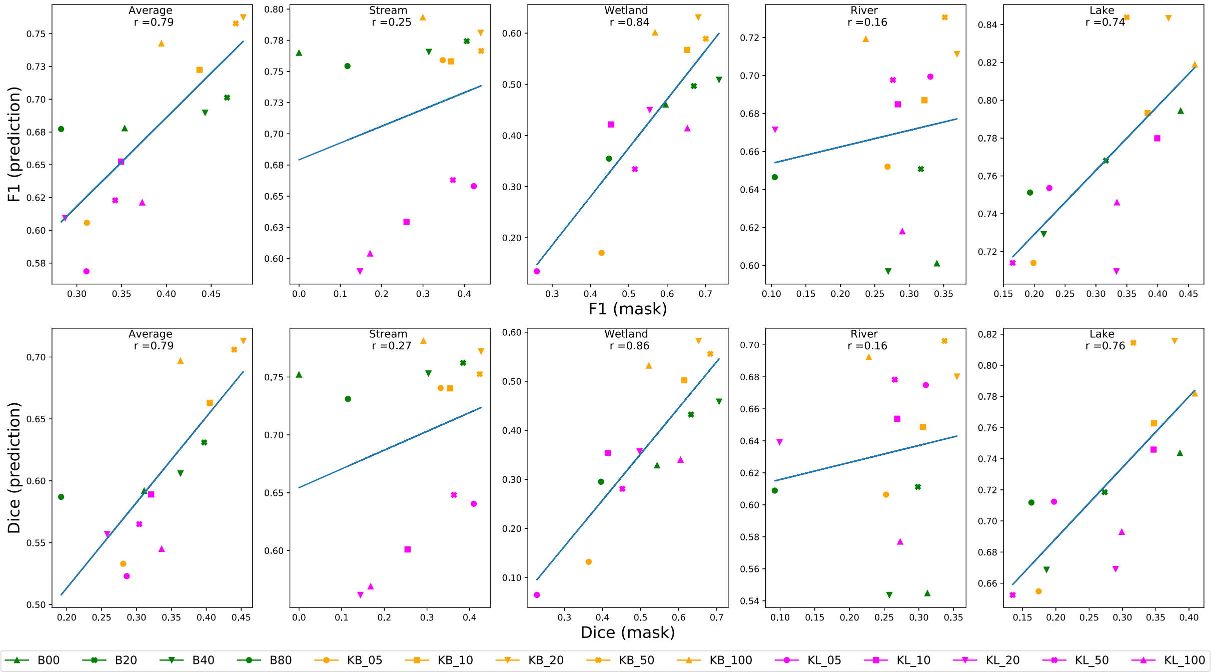


Fig. 11. Correlation between co-occurrence detection and DA. We present F1 and Dice of both co-occurrence detection (x-axis) and target segmentation (y-axis) under all different configurations. The Pearson correlation coefficient r is calculated for each class and the average accuracy. The more the absolute value approaches 1, the higher the correlation is.

samples for labels and power lines and thus to provide the adversarial learning with sufficient examples of these implausible patterns. Despite the fact that our application is the segmentation of historical maps, the proposed framework is general enough to be applied to other geo-spatial data (e.g., satellite imagery) and potentially extended to other relevant tasks like unsupervised change detection. Our described scheme can potentially serve as a tool to detect and track changes in various geo-spatial contexts, where domain gaps occur across time or between different sensors and mapping methods, but can be bridged with the help of spatial co-occurrence.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported and funded by the Swiss National Science Foundation (SNSF), as the part of the project “HistoRiCH: Historical river change - Planning for the future by exploring the mapped past” [No. 188692].

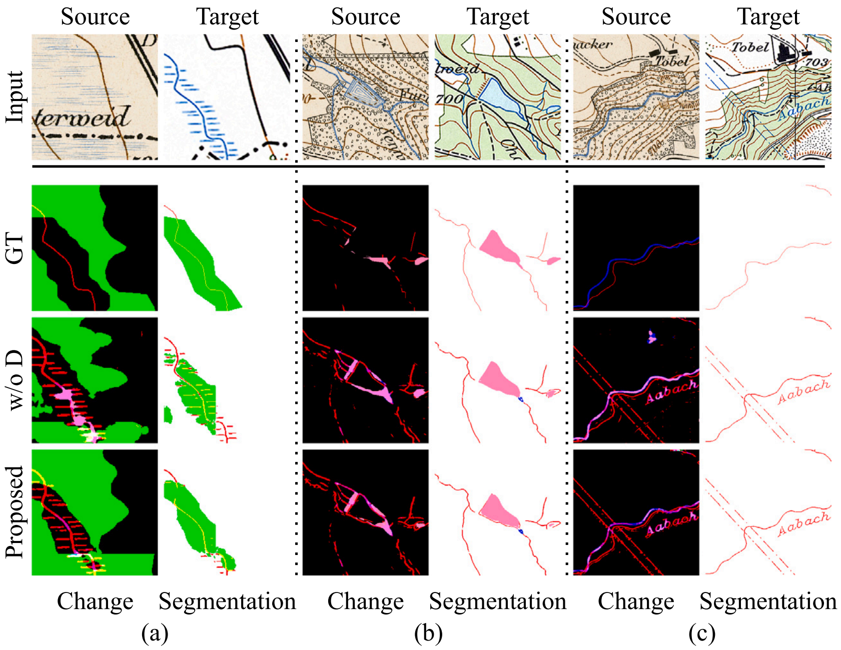


Fig. 12. Limitations of DA in output space. Three cases are presented for wetlands and streams (a), lakes (b) as well as labels and power lines (c). w/o D and Proposed are the configurations without and with adversarial learning, respectively.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2015. TensorFlow: Large-scale machine learning on heterogeneous systems. URL https://www.tensorflow.org/, Software available from tensorflow.org.

Ahn, J., Cho, S., Kwak, S., 2019. Weakly supervised learning of instance segmentation with inter-pixel relations. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 2209–2218.

Bearman, A., Russakovsky, O., Ferrari, V., Fei-Fei, L., 2016. What's the point: Semantic segmentation with point supervision. In: European Conference on Computer Vision. pp. 549–565.

Beigman, E., Beigman Klebanov, B., 2009. Learning with annotation noise. In: Annual ACL Meeting / International AFNLP Joint Conference on Natural Language Processing. pp. 280–287.

Bhatia, S., Vira, V., Choksi, D., Venkatachalam, P., 2013. An algorithm for generating geometric buffers for vector feature layers. Geo-Spatial Inform. Sci. 16 (2), 130–138.

Biasetton, M., Michieli, U., Agresti, G., Zanuttigh, P., 2019. Unsupervised domain adaptation for semantic segmentation of urban scenes. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 1211–1220.

Bin, D., Cheong, W.K., 1998. A system for automatic extraction of road network from maps. In: International Joint Symposia on Intelligence and Systems. pp. 359–366.

Bromberg, K.D., Bertness, M.D., 2005. Reconstructing New England salt marsh losses using historical maps. Estuaries 28 (6), 823–832. http://dx.doi.org/10.1007/BF02696012.

Burghardt, K., Uhl, J.H., Lerman, K., Leyk, S., 2022. Road network evolution in the urban and rural United States since 1900. Comput. Environ. Urban Syst. 95, 101803. http://dx.doi.org/10.1016/j.compenvurbsys.2022.101803.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2017. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans. Pattern Anal. Mach. Intell. 40 (4), 834–848.

Chen, L., Wu, W., Fu, C., Han, X., Zhang, Y., 2020. Weakly supervised semantic segmentation with boundary exploration. In: European Conference on Computer Vision. pp. 347–362.

Chen, C., Xie, W., Huang, W., Rong, Y., Ding, X., Huang, Y., Xu, T., Huang, J., 2019. Progressive feature alignment for unsupervised domain adaptation. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 627–636.

Chen, L.-C., Yang, Y., Wang, J., Xu, W., Yuille, A.L., 2016. Attention to scale: Scale-aware semantic image segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 3640–3649.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: European Conference on Computer Vision. pp. 833–851.

Chiang, Y.Y., Knoblock, C.A., 2009. Extracting road vector data from raster maps. In: IAPR International Workshop on Graphics Recognition. pp. 93–105.

Chollet, F., et al., 2015. Keras. GitHub, URL https://github.com/fchollet/keras.

Ekim, B., Sertel, E., Kabadayı, M.E., 2021. Automatic road extraction from historical maps using deep learning techniques: A regional case study of Turkey in a German World War II map. ISPRS Int. J. Geo-Inf. 10 (8).

Frenay, B., Verleysen, M., 2014. Classification in the presence of label noise: A survey. IEEE Trans. Neural Netw. Learn. Syst. 25 (5), 845–869.

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V., 2016. Domain-adversarial training of neural networks. J. Mach. Learn. Res. 17 (1), 1–35.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2020. Generative adversarial networks. Commun. ACM 63 (11), 139–144.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask R-CNN. In: IEEE International Conference on Computer Vision. pp. 2961–2969.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

Heitzler, M., Hurni, L., 2020. Cartographic reconstruction of building footprints from historical maps: A study on the Swiss Siegfried map. Trans. GIS 24 (2), 442–461.

Hoffman, J., Tzeng, E., Park, T., Zhu, J.-Y., Isola, P., Saenko, K., Efros, A., Darrell, T., 2018. Cycada: Cycle-consistent adversarial domain adaptation. In: International Conference on Machine Learning. pp. 1989–1998.

Hoffman, J., Wang, D., Yu, F., Darrell, T., 2016. FCNs in the wild: Pixel-level adversarial and constraint-based adaptation. arXiv:1612.02649.

Hong, W., Wang, Z., Yang, M., Yuan, J., 2018. Conditional generative adversarial network for structured domain adaptation. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1335–1344.

Hoyer, R., Chang, H., 2014. Assessment of freshwater ecosystem services in the Tualatin and Yamhill basins under climate change and urbanization. Appl. Geogr. 53, 402–416. http://dx.doi.org/10.1016/j.apgeog.2014.06.023.

Iqbal, J., Ali, M., 2020. Weakly-supervised domain adaptation for built-up region segmentation in aerial and satellite imagery. ISPRS J. Photogramm. Remote Sens. 167, 263–275.

Kaiser, P., Wegner, J.D., Lucchi, A., Jaggi, M., Hofmann, T., Schindler, K., 2017. Learning aerial image segmentation from online maps. IEEE Trans. Geosci. Remote Sens. 55 (11), 6054–6068.

Khoreva, A., Benenson, R., Hosang, J., Hein, M., Schiele, B., 2017. Simple does it: Weakly supervised instance and semantic segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 876–885.

Kingma, D.P., Ba, J., 2015. Adam: A method for stochastic optimization. In: International Conference on Learning Representations.

Kolesnikov, A., Lampert, C.H., 2016. Seed, expand and constrain: Three principles for weakly-supervised image segmentation. In: European Conference on Computer Vision. pp. 695–711.

Levin, N., Kark, R., Galilee, E., 2010. Maps and the settlement of southern Palestine, 1799–1948: An historical/GIS analysis. J. Historical Geogr. 36 (1), 1–18. http://dx.doi.org/10.1016/j.jhg.2009.04.001.

Leyk, S., 2009. Segmentation of colour layers in historical maps based on hierarchical colour sampling. In: International Workshop on Graphics Recognition. pp. 231–241.

Li, K., Wu, Z., Peng, K.-C., Ernst, J., Fu, Y., 2018. Tell me where to look: Guided attention inference network. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 9215–9223.

Li, Y., Yuan, L., Vasconcelos, N., 2019. Bidirectional learning for domain adaptation of semantic segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 6936–6945.

Lin, D., Dai, J., Jia, J., He, K., Sun, J., 2016. ScribbleSup: Scribble-supervised convolutional networks for semantic segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 3159–3167.

Lu, Z., Fu, Z., Xiang, T., Han, P., Wang, L., Gao, X., 2017. Learning from weak and noisy labels for semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39 (3), 486–500.

Malossini, A., Blanzieri, E., Ng, R.T., 2006. Detecting potential labeling errors in microarrays by data perturbation. Bioinformatics 22 (17), 2114–2121.

Milletari, F., Navab, N., Ahmadi, S.A., 2016. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In: International Conference on 3D Vision. pp. 565–571.

Mnih, V., Hinton, G., 2012. Learning to label aerial images from noisy data. In: International Conference on Machine Learning. pp. 203–210.

Motiian, S., Piccirilli, M., Adjeroh, D.A., Doretto, G., 2017. Unified deep supervised domain adaptation and generalization. In: IEEE International Conference on Computer Vision. pp. 5715–5725.

Papandreou, G., Chen, L.-C., Murphy, K.P., Yuille, A.L., 2015. Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation. In: IEEE International Conference on Computer Vision. pp. 1742–1750.

Patrini, G., Rozza, A., Krishna Menon, A., Nock, R., Qu, L., 2017. Making deep neural networks robust to label noise: A loss correction approach. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1944–1952.

Peng, D., Bruzzone, L., Zhang, Y., Guan, H., Ding, H., Huang, X., 2021. SemiCDNet: A semisupervised convolutional neural network for change detection in high resolution remote-sensing images. IEEE Trans. Geosci. Remote Sens. 59 (7), 5891–5906.

Picuno, P., Cillis, G., Statuto, D., 2019. Investigating the time evolution of a rural landscape: How historical maps may provide environmental information when processed using a GIS. Ecol. Eng. 139, 105580. http://dx.doi.org/10.1016/j.ecoleng.2019.08.010.

Pires de Lima, R., Marfurt, K., 2019. Convolutional neural network for remote-sensing scene classification: Transfer learning analysis. Remote Sens. 12 (1), 86.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention. pp. 234–241.

Sakaridis, C., Dai, D., Gool, L.V., 2022. Map-guided curriculum domain adaptation and uncertainty-Aware evaluation for semantic nighttime image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 44, 3139–3153.

San-Antonio-Gómez, C., Velilla, C., Manzano-Agugliaro, F., 2014. Urban and landscape changes through historical maps: The Real Sitio of Aranjuez (1775–2005), a case study. Comput. Environ. Urban Syst. 44, 47–58. http://dx.doi.org/10.1016/j.compenvurbsys.2013.12.001.

Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition. In: International Conference on Learning Representations.

Tasar, O., Happy, S., Tarabalka, Y., Alliez, P., 2020. ColorMapGAN: Unsupervised domain adaptation for semantic segmentation using color mapping generative adversarial networks. IEEE Trans. Geosci. Remote Sens. 58 (10), 7178–7193.

Thongkam, J., Xu, G., Zhang, Y., Huang, F., 2008. Support vector machine for outlier detection in breast cancer survivability prediction. In: Advanced Web and Network Technologies, and Applications.

Toldo, M., Maracani, A., Michieli, U., Zanuttigh, P., 2020. Unsupervised domain adaptation in semantic segmentation: A review. Technologies 8 (2), 35.

Tommasi, T., Lanzi, M., Russo, P., Caputo, B., 2016. Learning the roots of visual domain shift. In: European Conference on Computer Vision. pp. 475–482.

Tonolla, D., Geilhausen, M., Doering, M., 2021. Seven decades of hydrogeomorphological changes in a near-natural (Sense River) and a hydropower-regulated (Sarine River) pre-Alpine river floodplain in Western Switzerland. Earth Surf. Process. Landf. 46 (1), 252–266. http://dx.doi.org/10.1002/esp.5017.

Tsai, Y.-H., Hung, W.-C., Schulter, S., Sohn, K., Yang, M.-H., Chandraker, M., 2018. Learning to adapt structured output space for semantic segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 7472–7481.

Tzeng, E., Hoffman, J., Darrell, T., Saenko, K., 2015. Simultaneous deep transfer across domains and tasks. In: IEEE International Conference on Computer Vision. pp. 4068–4076.

Tzeng, E., Hoffman, J., Saenko, K., Darrell, T., 2017. Adversarial discriminative domain adaptation. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 7167–7176.

Tzeng, E., Hoffman, J., Zhang, N., Saenko, K., Darrell, T., 2014. Deep domain confusion: Maximizing for domain invariance. arXiv:1412.3474.

Uhl, J.H., Leyk, S., Chiang, Y.Y., Duan, W., Knoblock, C.A., 2020. Automated extraction of human settlement patterns from historical topographic map series using weakly supervised convolutional neural networks. IEEE Access 8, 6978–6996.

Uhl, J.H., Leyk, S., Li, Z., Duan, W., Shbita, B., Chiang, Y.-Y., Knoblock, C.A., 2021. Combining remote-sensing-derived data and historical maps for long-term back-casting of urban extents. Remote Sens. 13 (18), 3672.

Uzkent, B., Sheehan, E., Meng, C., Tang, Z., Burke, M., Lobell, D., Ermon, S., 2019. Learning to interpret satellite images using Wikipedia. In: International Joint Conference on Artificial Intelligence. pp. 3620–3626.

Visin, F., Ciccone, M., Romero, A., Kastner, K., Cho, K., Bengio, Y., Matteucci, M., Courville, A., 2016. ReSeg: A recurrent neural network-based model for semantic segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 41–48.

Vu, T.-H., Jain, H., Bucher, M., Cord, M., Pérez, P., 2019a. ADVENT: Adversarial entropy minimization for domain adaptation in semantic segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 2517–2526.

Vu, T.-H., Jain, H., Bucher, M., Cord, M., Pérez, P., 2019b. Dada: Depth-aware domain adaptation in semantic segmentation. In: IEEE International Conference on Computer Vision. pp. 7364–7373.

Walz, U., 2008. Monitoring of landscape change and functions in Saxony (Eastern Germany) – Methods and indicators. Ecol. Indic. 8 (6), 807–817. http://dx.doi.org/10.1016/j.ecolind.2007.09.006, Ecological indicators at multiple scales.

Wang, S., Chen, W., Xie, S.M., Azzari, G., Lobell, D.B., 2020. Weakly supervised deep learning for segmentation of remote sensing imagery. Remote Sens. 12 (2), 207.

Wang, A.X., Tran, C., Desai, N., Lobell, D., Ermon, S., 2018. Deep transfer learning for crop yield prediction with remote sensing data. In: Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies. pp. 1–5.

Wei, Y., Liang, X., Chen, Y., Shen, X., Cheng, M.-M., Feng, J., Zhao, Y., Yan, S., 2016. Stc: A simple to complex framework for weakly-supervised semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39 (11), 2314–2320.

Wu, S., Heitzler, M., Hurni, L., 2022a. A closer look at segmentation uncertainty of scanned historical maps. Int. Arch. Photogramm. Remote Sens. Spatial Inform. Sci. 43, 189–194.

Wu, S., Heitzler, M., Hurni, L., 2022b. Leveraging uncertainty estimation and spatial pyramid pooling for extracting hydrological features from scanned historical topographic maps. GIScі. Remote Sens. 59 (1), 200–214.

Wu, W., Qi, H., Rong, Z., Liu, L., Su, H., 2018. Scribble-supervised segmentation of aerial building footprints using adversarial learning. IEEE Access 6, 58898–58911.

Xiao, T., Xia, T., Yang, Y., Huang, C., Wang, X., 2015. Learning from massive noisy labeled data for image classification. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 2691–2699.

Xie, Q., Luong, M.-T., Hovy, E., Le, Q.V., 2020. Self-training with noisy student improves ImageNet classification. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10687–10698.

Yang, Y., Soatto, S., 2020. Fda: Fourier domain adaptation for semantic segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 4085–4095.

Yu, K., Frank, H., Wilson, D., 2021. Points2Polygons: Context-based segmentation from weak labels using adversarial networks. arXiv:2106.02804.

Zhuang, F., Cheng, X., Luo, P., Pan, S.J., He, Q., 2015. Supervised representation learning: Transfer learning with deep autoencoders. In: International Joint Conference on Artificial Intelligence.