

Selection of unlabeled source domains for domain adaptation in remote sensing

Christian Geiß*, Alexander Rabuske, Patrick Aravena Pelizari, Stefan Bauer,
Hannes Taubenböck

German Remote Sensing Data Center (DFD), German Aerospace Center (DLR), 82234 Weßling-Oberpfaffenhofen, Germany

ARTICLE INFO

Index Terms:

Domain adaptation
Remote sensing
Multiple source domains
Similarity metrics
Regression
Built-up density and height

ABSTRACT

In the context of supervised learning techniques, it can be desirable to utilize existing prior knowledge from a source domain to estimate a target variable in a target domain by exploiting the concept of domain adaptation. This is done to alleviate the costly compilation of prior knowledge, i.e., training data. Here, our goal is to select a single source domain for domain adaptation from multiple potentially helpful but unlabeled source domains. The training data is solely obtained for a source domain if it was identified as being relevant for estimating the target variable in the corresponding target domain by a selection mechanism. From a methodological point of view, we propose unsupervised source selection by voting from (an ensemble of) similarity metrics that follow aligned marginal distributions regarding image features of source and target domains. Thereby, we also propose an unsupervised pruning heuristic to solely include robust similarity metrics in an ensemble voting scheme. We provide an evaluation of the methods by learning models from training data sets created with Level-of-Detail-1 building models and regress *built-up density* and *height* on Sentinel-2 satellite imagery. To evaluate the domain adaptation capability, we learn and apply models interchangeably for the four largest cities in Germany. Experimental results underline the capability of the methods to obtain more frequently higher accuracy levels with an improvement of up to 10% regarding the most robust selection mechanisms compared to random source-target domain selections.

1. Introduction

The extraction of thematic information from remote sensing imagery in an automated way is frequently addressed by resorting to supervised learning techniques due to their adaptive nature and accuracy properties. Such techniques determine assigning a discrete thematic label (in classification problems) or continuous value (in regression problems) based on a sufficient amount of properly encoded prior knowledge, i.e., training data. The training data is deployed to infer a rule (e.g., a decision function) that enables favorable generalizations capabilities for unseen instances [1,2].

Unfortunately, the compilation of prior knowledge is frequently very costly [3,4]. Consequently, it can be desirable to utilize existing prior knowledge from a *source domain* to estimate the target variable in a *target domain* by exploiting the concept of domain adaptation. Domain adaptation methods are frequently categorized according to the amount of available prior knowledge in the source and target domain.

Supervised domain adaptation methods can exploit labeled samples from both source and target domain. Thereby, labeled samples in the target domain are substantially less numerous than labeled samples available in the source domain (otherwise there would be no need to render the estimation problem as a domain adaption problem, i.e., solely using a supervised learning algorithm in the target domain would be appropriate). In contrast, unsupervised domain adaptation methods assume that two unlabeled domains need to be matched. If there is prior knowledge available in the source domain but not in the target domain, and if the joint distributions of source and target domain are different, but related enough to ensure that the source domain information can be of help for solving the prediction problem in the target domain, we can regard the situation as a semi-supervised domain adaptation problem [5].

Domain adaptation frequently includes two further subproblems, i.e., *covariate shift* and *sample selection bias*. The latter refers to changes in the nature of the objects observed regarding the source and target

* Corresponding author.

E-mail address: christian.geiss@dlr.de (C. Geiß).

domain, i.e., the samples of the source domain are non-representative for the (unlabeled) samples of the target domain. Covariate shift is a particular case of a sample selection bias. This kind of bias is induced by the independent variables, i.e., covariates. In the field of remote sensing, such a situation is frequently related to changes in the source and target domain imagery. Those changes are induced by varying data acquisition properties related to illumination and acquisition angle, among others [5]. Hence, by definition, it is crucial to avoid a negative model transfer from a source to a target domain, i.e., to ensure that the general assumption of transfer learning, namely that the domains have to be related, is not violated [6].

We address this aspect by a proper selection of the source domain the model should be learned from before transferring it, i.e., we aim to establish an efficient way for *domain adaptation when multiple potentially helpful but unlabeled source domains exist*. Hence, the goal is to select a source-target domain combination that allows for beneficial model estimates in the target domain, i.e., to enable a positive model transfer. Subsequent to selection, the training data is obtained for the selected source domain and a prediction model is learned thereof (Fig. 1). Previous works in the context of domain adaptation either dealt with a given labeled source domain or had access to multiple labeled source domains. Regarding the first family of methods, feature-based methods jointly map the distributions of source and target domain in a latent subspace for direct inference, e.g. Refs. [7,8], whereas instance-based methods estimate the instances in the original space by deploying semi-labeled samples of the target domain, e.g., Refs. [9,10]. Only recently, techniques in the context of remote sensing evolved, that are concerned with domain adaptation including multiple labeled source domains. Those works followed the idea of synthetization of sources, i.e., creating an optimal source domain from multiple suboptimal source domains: Lu et al. [11] consider the case where a single source domain does not contain all the categories of a target variable regarding the target domain. They use multiple complementary source domains to form the categories of the target domain and learn a multisource compensation network thereof. Likewise, Gong et al. [12] integrate multiple incomplete source domains and propose a separation mechanism so that only source domains with relevant manifestations of the target variable are considered for cross-domain alignment. Elshamli et al. [13] propose a CNN architecture that is learned and adaptively optimized across multiple source domains.

However, the uniqueness of this study is related to the selection of a

single source domain from multiple potentially helpful but unlabeled source domains and subsequent labeling only after selection. Our motivation is related to the circumstance that domains are particularly associated with distinct geographic locations in remote sensing. This frequently implies that the compilation of labels remains very costly since it requires, e.g., field campaigns in dedicated areas or the acquisition of additional geospatial data for an area to enable the compilation of labeled samples. Ideally, such labor-intensive efforts must solely be carried out for a source domain that was identified beforehand to likely enable a positive model transfer later on. From an application-oriented point of view, such a setting frequently occurs if the direct compilation of training data for a target domain is not possible or inefficient. The latter refers to situations where a target variable needs to be estimated over various domains, i.e., large geographical areas, whereby solely a small subset of them can be labeled due to limited resources.

Generally, we aim for a method that is independent of the learning algorithm. This is intended to account for the No-free-Lunch-Theorem. It states that there is no algorithm that can provide the highest accuracies in any domain at all times [14]. Algorithm-independent methods can be easily adapted to the problem at hand by employing suitable base learners and ensemble models. The latter were found to be particularly beneficial in terms of accuracy and reliability properties in the context of remote sensing [15,16]. Consequently, we designed an unsupervised source selection approach, which builds upon the robust quantification of the similarity between the source and target domains. We consider the main contributions of this work as follows:

- From a methodological point of view, we propose unsupervised source selection mechanisms by voting from similarity metrics that follow aligned marginal distributions regarding image features of source and target domains for remote sensing imagery. Thereby, we also establish an unsupervised pruning heuristic to solely consider robust similarity metrics in an ensemble voting scheme. Finally, labeled samples are solely obtained for a selected source domain and used to learn a prediction model which is subsequently applied to the target domain.
- We provide an evaluation of the method within an innovative application field. Recently, efforts were carried out to derive properties of the urban morphology for large areas based on regression techniques using Sentinel-2 imagery [15]. In this application context, we learn models from training data sets created with

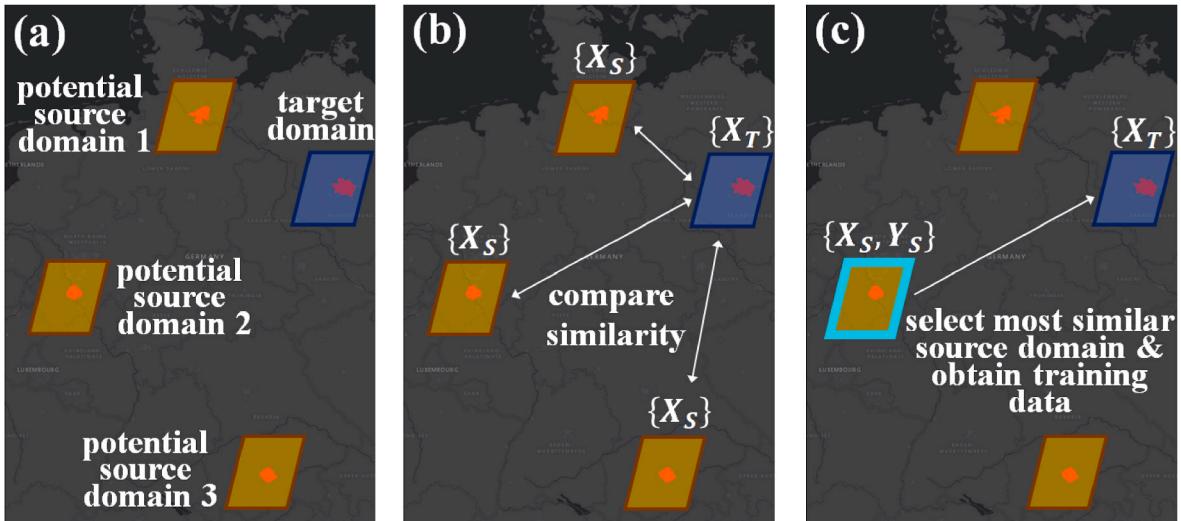


Fig. 1. Illustration of the considered unique problem setting of this work: a) multiple potentially helpful but unlabeled source domains exist; b) the source and target domains are compared individually regarding their similarity; c) the most similar source domain regarding the target domain is selected and solely for this source domain the (frequently very costly) training data is obtained. Subsequently, the selected source domain is deployed for learning the prediction model and transferring it to the target domain.

Level-of-Detail-1 (LoD-1) building models and regress *built-up density* and *height* for spatial processing units which correspond to urban neighborhood scales using features from multispectral Sentinel-2 imagery. To evaluate the domain adaptation capability of the methods, we learn and apply models interchangeably for the four largest cities in Germany.

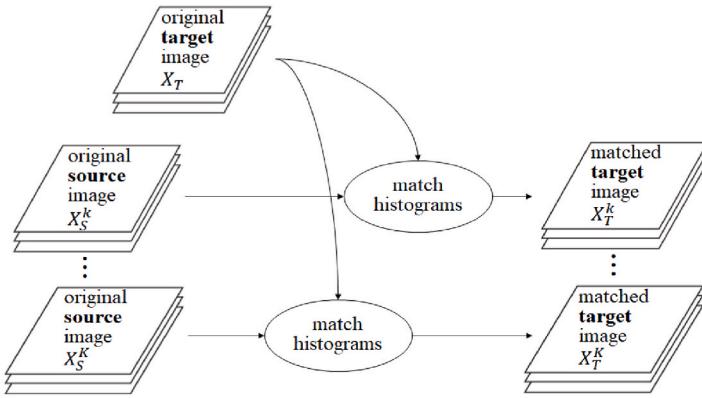
Section 2 details the source selection method. We describe the

experimental setup in section 3, report the results of experiments in section 4, and give concluding remarks in section 5.

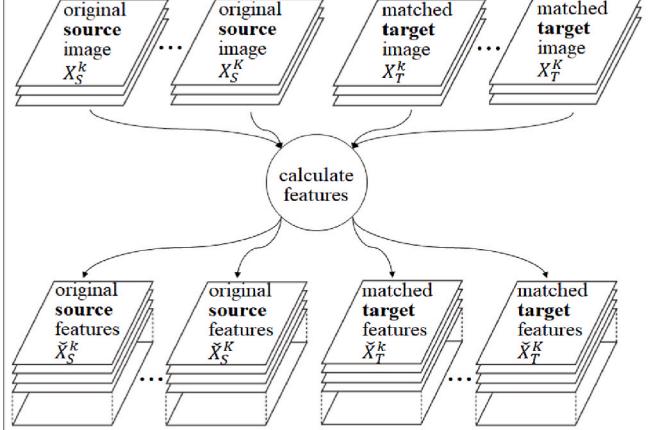
2. Proposed methodology

An overview of the approach with affiliated processing steps is provided in Fig. 2. The individual steps include matching the image data and computation of image features (sec. 2.1), comparative evaluation of

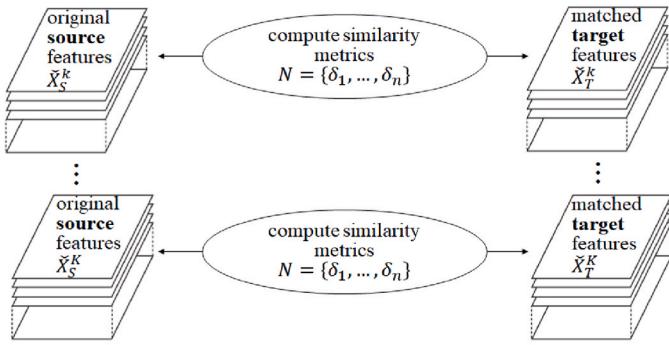
(a) unsupervised domain alignment: histogram matching



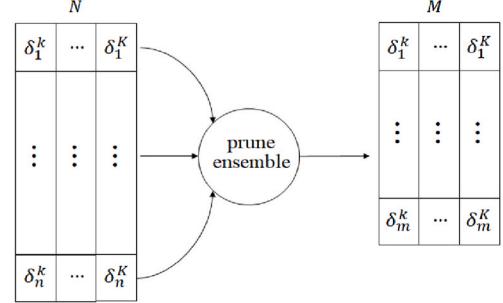
(b) feature calculation



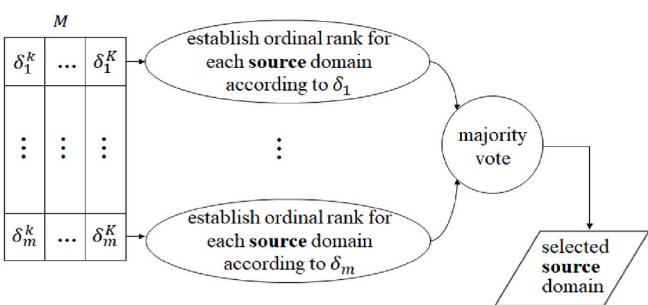
(c) computation of similarity metrics



(d) unsupervised ensemble pruning



(e) ranking strategy



(f) learning and application of regression model

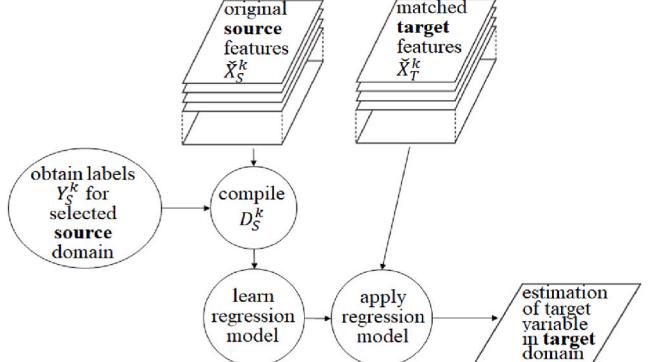


Fig. 2. Overview of the individual processing steps; (a) a histogram matching procedure aligns the shapes of the histograms of the spectral bands of the satellite imagery which covers a target domain to the spectral bands of the satellite imagery which cover the potential source domains individually; (b) the matched data is deployed for computing an exhaustive set of spectral-spatial features; (c) the features are deployed to compute a variety of similarity metrics; (d) an ensemble pruning strategy aims for identifying a robust subset of similarity metrics; (e) the selected similarity metrics are used to evaluate all possible source-target domain combinations and identify the combination which is most similar; the identified combinations from the similarity metrics are combined via a majority voting strategy; (f) labeled samples are obtained for a selected source domain and a regression model is learned from the corresponding source image features and applied to estimate the target variable in the target domain.

source-target domain combinations by computing a variety of similarity metrics, ensemble pruning for identifying a set of robust similarity metrics (sec. 2.2), implementation of a voting strategy to select a favorable source-target domain combination and learning a regression model thereof (sec. 2.3).

2.1. Histogram matching and feature computation

To enable domain adaptation under covariate shift, we implement a histogram matching procedure of the imagery covering the source and target domains first. At this point, we consider image data from K potentially helpful but unlabeled source domains $\{S^k\}_{k=1}^K$, where $S^k = X_S^k = \{\mathbf{x}_{S_i}^k\}_{i=1}^{n_S^k}$ represents the set of n_S^k unlabeled source samples of the k th domain and image data from the target domain where $X_T = \{\mathbf{x}_T\}_{j=1}^{n_T}$ represents the set of n_T unlabeled target samples, with samples $\mathbf{x}_{S_i}^k, \mathbf{x}_T \in \mathbb{R}^d \forall i, j$. Histogram matching is a method for adapting data distributions. To that purpose, the representations of the original data are matched independently of the subsequent processing model. The latter aspect is relevant since our goal is to establish a method, that can handle multiple independently learned models. Histogram matching implements a relative normalization. Thus, the method provides similarly distributed digital numbers rather than physical units as an output [5]. A nonlinear transformation φ is deployed to align the shapes of the cumulative histograms of the spectral bands of the image of the target domain to the spectral bands of the image of the source domain [17]. This procedure is carried out for the considered possible source-target domain combinations individually to obtain a set of matched target domain distributions concerning K source domains, i.e., $X_T \rightarrow \varphi(X_T) = X_T^k$ (Fig. 2a). Generally, we deploy histogram matching as a method for domain alignment here, since it showed viable properties for domain adaptation problems in the past [5]. However, given the modular nature of the proposed method, also other unsupervised domain alignment methods could be considered.

Subsequently, features are computed from the matched imagery (Fig. 2b). To estimate the target variables, we integrate into this work multispectral Sentinel-2 imagery. In particular, we deploy the R-G-B-NIR bands of Sentinel-2, which feature a pixel spacing of 10 m [18]. Previous works demonstrated already the capacity of multispectral Sentinel-2 imagery to describe the built environment in a discriminative way [15]. We computed an exhaustive set of spectral-spatial features, i.e., spectral features derived from the R-G-B-NIR bands, features that are based on the concept of mathematical morphology [19], and texture measures using the gray-level co-occurrence matrix [20]. We deploy features as described in detail in Geiß et al. [15]. Given the different feature categories and various window sizes used for the aforementioned spatial features, each characterized source and target domain \check{X}_S^k and \check{X}_T^k carries a 413-dimensional feature vector. Thus, it can be noted that we consider a homogeneous domain adaptation setting here since all domains share the same features [13].

2.2. Computation of similarity metrics and unsupervised ensemble pruning

To address problems related to sample selection bias, we compute similarity metrics regarding a characterized source domain – aligned target domain combination, i.e., \check{X}_S^k and \check{X}_T^k , and select a proper combination thereof. The empirical estimate of the similarity of a source and a target domain δ is computed with eight different metrics in this study (Table 1; Fig. 2c).

We aimed to consider metrics that feature a low level of sensitivity regarding, e.g., outliers, the scale of measurement, or data dimensionality [21]. Besides evaluating the suitability of a single metric to select a proper source-target domain combination, we also follow the idea to

Table 1

Deployed metrics with formulae; arrows ($\downarrow\uparrow$) indicate whether a high or low numerical value, respectively, indicates a high agreement between the instances (rectangular grid cells describing the built environment in this study) of a source and a target domain.

metric	Description	agreement	no.
Bray-Curtis	$\delta = 1 - \frac{\sum_{i=1}^n \check{x}_{S_i} - \check{x}_{T_i} }{\sum_{i=1}^n (\check{x}_{S_i} + \check{x}_{T_i})}$	\downarrow	(1)
Chord	$\delta = \left(2 - 2 \frac{\sum_{i=1}^n \check{x}_{S_i} \cdot \check{x}_{T_i}}{\ \check{x}_{S_i}\ _2 \ \check{x}_{T_i}\ _2} \right)^{\frac{1}{2}}$, where $\ \check{x}_S\ _2$ is the L^2 -norm $\ \check{x}_S\ _2 = \sqrt{\sum_{i=1}^n \check{x}_{S_i}^2}$	\downarrow	(2)
Hellinger	$\delta = \sqrt{\sum_{i=1}^n \left(\sqrt{\frac{\check{x}_{S_i}}{\sum_{i=1}^n \check{x}_{S_i}}} - \sqrt{\frac{\check{x}_{T_i}}{\sum_{i=1}^n \check{x}_{T_i}}} \right)^2}$	\downarrow	(3)
Whittaker β	$\delta = \frac{\ \frac{\check{x}_{S_i}}{\sum_{i=1}^n \check{x}_{S_i}} - \frac{\check{x}_{T_i}}{\sum_{i=1}^n \check{x}_{T_i}}\ _2}{\sum_{i=1}^n \frac{\check{x}_{S_i}}{\sum_{i=1}^n \check{x}_{S_i}}}$	\downarrow	(4)
cosine	$\delta = \frac{\sum_{i=1}^n \check{x}_{S_i} \cdot \check{x}_{T_i}}{\ \check{x}_S\ _2 \ \check{x}_T\ _2}$, where $\ \check{x}_T\ _2$ is the Euclidean norm of vector $\check{x}_T = (\check{x}_{T_1}, \check{x}_{T_2}, \dots, \check{x}_{T_n})$, i.e., $\ \check{x}_T\ _2 = \sqrt{\check{x}_{T_1}^2 + \check{x}_{T_2}^2 + \dots + \check{x}_{T_n}^2}$	\uparrow	(5)
χ^2	$\delta = \sum_{i=1}^n \frac{(\check{x}_{S_i} - \check{x}_{T_i})^2}{\check{x}_{S_i}}$	\uparrow	(6)
Sørensen-Dice	$\delta = \frac{2 \check{X}_S \cap \check{X}_T }{ \check{X}_S + \check{X}_T }$, where $ \check{X}_S $ and $ \check{X}_T $ are the cardinalities of the two sets, i.e., the number of elements in each set.	\uparrow	(7)
Pearson	$\delta = \frac{\sum_{i=1}^n (\check{x}_{S_i} - \mu_S)(\check{x}_{T_i} - \mu_T)}{\sqrt{\sum_{i=1}^n (\check{x}_{S_i} - \mu_S)^2(\check{x}_{T_i} - \mu_T)^2}}$, where μ_S and μ_T are the mean values regarding source and target domain, respectively.	\uparrow	(8)

establish an ensemble of metrics as a selection mechanism. In our work, an ensemble strategy foresees the combination of votes for certain source-target domain combinations from multiple metrics via a decision fusion strategy to obtain the final selection. This strategy is implemented to alleviate the risk of eventually picking an inadequate single metric. From a conceptual point of view, in contrast to, e.g., supervised classification problems, we do not have access here to knowledge on actual prediction accuracy which could facilitate the selection of a suitable metric. Instead, we need to select a proper source-target domain combination in an unsupervised way.

To compile a robust ensemble, we establish a heuristically-guided ensemble pruning strategy. The goal is to select a subset of metrics $M = \{\delta_1, \dots, \delta_m\}$ from the full set of metrics $N = \{\delta_1, \dots, \delta_n\}$, which enables a robust source-target domain selection later on (Fig. 2d). We exploit the notion of a small margin distance of empirical estimates obtained with a certain similarity metric δ_i regarding all possible source-target domain combinations $k = 1, \dots, K$. To this purpose, we compute the percentage deviation \hat{p} of the numerical value which indicates the highest agreement between a source and target domain a_{δ_i} and the mean value of all estimates μ_{δ_i} , i.e., $\hat{p} = \frac{|a_{\delta_i} - \mu_{\delta_i}|}{\mu_{\delta_i}} \cdot 100$. This is motivated by the circumstance that metrics that can properly describe the similarity of source-target domain combinations feature a rather small range of numerical values. Subsequently, we compile M based on the smallest numerical values of \hat{p} , while simultaneously selecting interchangeably metrics that quantify the distance (eq. 1-4) and similarity (eq. 5-8) (Table 1) of source-target domain combinations to account for a multi-view perspective, i.e., to impose the consideration of complementary subspaces with a particular physical meaning and statistical properties [22]. Thereby, the number of metrics to be included in the pruned ensemble

can be set according to a stopping criterion.

2.3. Ranking strategy and learning of regression model

To select a proper source-target domain combination, we translate the numerical values as obtained with the metrics in ordinal ranks, which reproduce the levels of agreement of source-target domain combinations. In the context of an ensemble strategy, i.e., creating a committee of experts, the ranks from multiple metrics are integrated via majority voting. The source domain which is ranked as having the highest level of agreement according to the majority voting is finally selected (Fig. 2e). The full procedure to select a source domain with a pruned ensemble of similarity metrics is also documented in the pseudocode of Algorithm 1. Lastly, we obtain the manifestations of the target variable $Y_S^k = \{Y_{S_i}^k\}_{i=1}^{n_S^k}$ for the selected source domain. Subsequently, we compile the training set $D_S^k = \{X_S^k, Y_S^k\}$ which is fed to the learning algorithm. The learned model is deployed to estimate the target variable on \tilde{X}_T^k (Fig. 2f).

Algorithm 1. Unlabeled source selection with a pruned ensemble of similarity metrics

Input: K unlabeled source domains $\{S^k\}_{k=1}^K$, unlabeled target domain X_T
Output: selected source domain
1: Align the image of the target domain to the images of the source domains with φ for all source-target combinations individually
2: Compute features for all source domains and aligned target domains
3: Compute N for each characterized source domain-aligned target domain combination
4: **do**
5: Select interchangeably metrics that quantify the distance (eq. 1-4) and similarity (eq. 5-8) based on the smallest numerical values of \hat{p}
6: **while** stopping criterion is not met
7: **return** M
8: Establish ordinal rank for each source domain according to similarity metrics contained in M
9: Integrate ranks for each source domain via majority voting
10: Select source domain with lowest rank

3. Experimental setup

We evaluate the methods in the context of predicting *built-up density* and *height* using features from Sentinel-2 multispectral satellite imagery. The imagery was atmospherically corrected according to level 2A [23]. The acquisition timeframe was constrained to autumn and winter of the years 2015–2016 to minimize the influence of photosynthetically active vegetation on the *built-up density* and *height* estimates. Moreover, the acquisition timeframe offers the possibility to exploit shadow information, that encodes further helpful information regarding the target variables. Results are calculated for the settlement areas of the four largest German cities (Fig. 3a). The covered areas of the domains comprise 513 km² for Berlin, 577 km² for Hamburg, 535 km² for Munich, and 508 km² for Cologne, respectively.

Fig. 3b briefly visualizes the generation of the target variables to be estimated. We integrate the so-called Global Urban Footprint (GUF) layer that distinguishes “built-up” from “non-built-up” in a binary way. The data set covers the whole globe with a high spatial resolution of 12 m and a classification accuracy beyond 85% [24]. Generally, the spatial resolution characteristics of the Sentinel-2 imagery restrict analyses on the individual building level. This is due to the circumstance that the pixel spacing of 10 m of the R-G-B-NIR bands of Sentinel-2 can exceed the extent of individual buildings. Consequently, analyses are carried out for spatial entities with a higher level of aggregation, i.e., we create rectangular grid cells to compute *built-up density* and *height* thereof. In our setup, we establish rectangular grid cells with linearly progressing side lengths of 200, 500, and 800 m, respectively. For the actual computation of the target variables, we integrated LoD-1 building geometries (LoD-1 resolution foresees the representation of buildings as extruded footprints [25]) and corresponding height measurements from cadastral information: *Built-up density* per grid cell is computed by the ratio between the area of elevated pixels according to LoD-1 footprints and the whole settlement area as indicated by the GUF data set; *built-up*

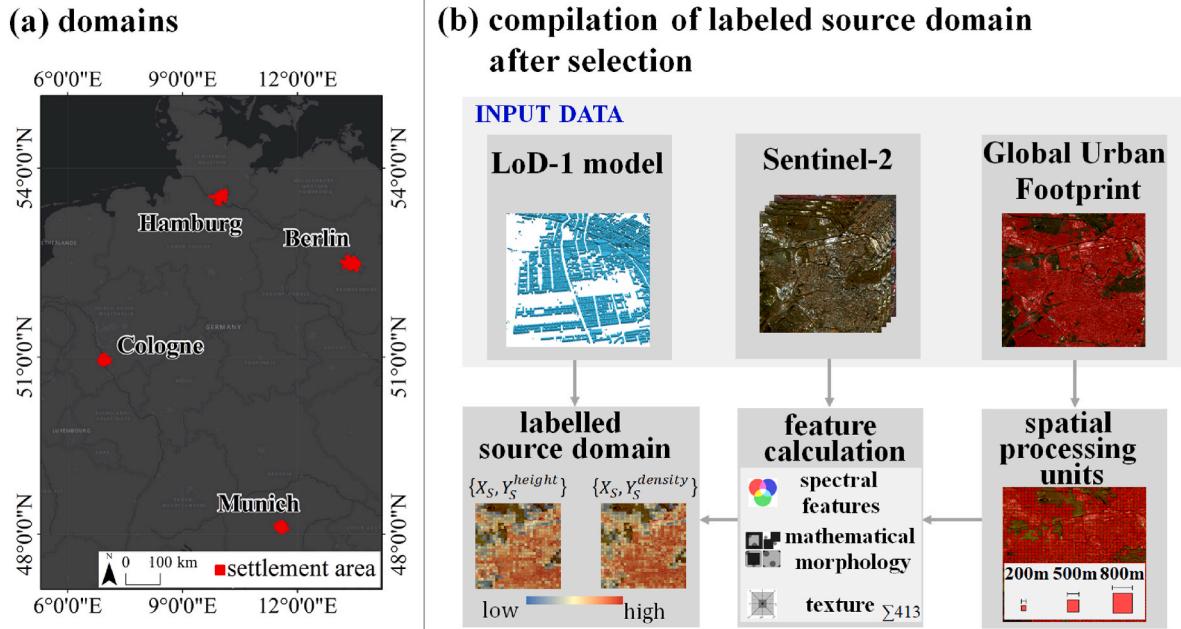


Fig. 3. Overview of the experimental setup; (a) study area with the four largest cities in Germany which correspond to the different domains; (b) procedure for establishing a fully characterized and labeled source domain after it was selected regarding the two target variables: i) we integrate the so-called Global Urban Footprint (GUF) layer that distinguishes “built-up” from “non-built-up” and establish rectangular grid cells thereof; ii) multispectral Sentinel-2 imagery is deployed to compute a set of spectral features, features that are based on the concept of mathematical morphology, and texture measures; iii) for the computation of the target variables, we integrated LoD-1 building models and affiliated height measurements from cadastral sources.

height per grid cell is computed by extracting the median height value according to the LoD-1-based height values within a grid cell [26].

We used both Random Forest Regression (RFR) and Support Vector Regression (SVR) as learning algorithms for estimating the two target variables. For the RFR models [27] hyperparameters tuning was carried out as follows: $n_{tree} = 500$ and $m_{try} = 1, 2, \dots, 51$. Regarding the SVR models [28], we used Gaussian radial basis function kernels: $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma^2)$. The regularization parameter C , the tolerance value ϵ , and the kernel parameter σ were optimized according to $\sigma = \{0.01, 0.02, \dots, 0.1\}$, $C = \{5, 6, \dots, 15\}$, and $\epsilon = \{0.05, 0.06, \dots, 0.15\}$, respectively. In the experiments, we learn and optimize a model based on cross-validation in the source domain and compute generalization capabilities from unseen labeled samples in the target domain. Corresponding numbers of available labeled samples, i.e., labeled grid cells, are provided in Table 2.

Model selection was carried out using the root-mean-square error (RMSE). In the context of the ensemble pruning strategy, we set the number of metrics to be selected to four to establish a tradeoff between the computational burden and enabling an extensive description of a source-target domain combination. Beyond, we also created an ensemble that includes all metrics and ensembles based on the random selection of four metrics (20 random selections) for benchmarking our proposed ensemble scheme. Overall, we evaluate 11 selection mechanisms, whereby 8 correspond to the deployment of the considered similarity metrics, and 3 represent ensemble schemes. Results were also tested regarding statistical significance using a paired *t*-test with a significance cutoff of $p = 0.05$.

4. Experimental results and discussion

To investigate whether the proposed domain adaptation techniques are generally beneficial in this empirical setting, Fig. 4 provides mean absolute errors (MAEs) for estimated *built-up density* and *height* regarding four different situations: i) non-domain adaptation setting, i.e., supervised learning and applying a model within a domain (i.e., upper benchmark); ii) proposed domain adaptation strategy that foresees both a domain alignment and selection of a proper source domain (mean values across all 11 selection mechanisms are plotted); iii) domain adaptation strategy that foresees solely a domain alignment without selection, i.e., all source domains are deployed interchangeably and accuracy estimates are averaged; iv) no domain adaptation strategy was implemented, i.e., a model from a source domain is learned and applied to the target domain without adaptation (i.e., lower benchmark). Thereby, results from 24 realizations per target variable are plotted (composed of four cities, three sizes of the spatial processing units, and two regressors).

First, it can be noted that the unadapted models, despite featuring the largest deviations in this setup, yield accuracy levels that indicate that the models did not fail when applied to a new domain. This can be attributed to the exhaustive set of spectral-spatial features which imposes invariance in the models. However, when actually aligning source and target domains, a substantial decrease of MAE can be achieved. Compared to the domain adaptation strategy which solely foresees to align domains, our proposed additional mechanisms to select a proper source domain for an aligned target domain enable on average a further

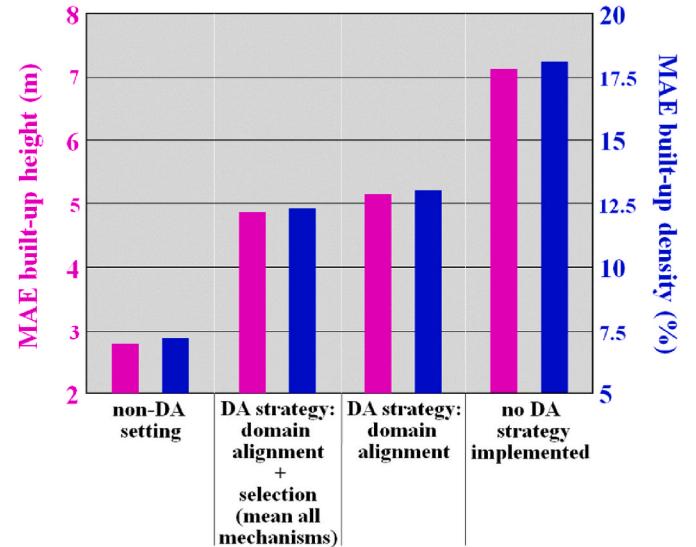


Fig. 4. MAE values for *built-up height* and *density* obtained when rendering the prediction problem as a non-domain adaptation (DA) setting (i.e., learning a model within a domain; this correspond to a regular supervised learning task and represents an upper benchmark here), when deploying the proposed DA strategy that foresees both a domain alignment and selection of a proper source domain (mean values across all 11 selection mechanisms are plotted), using a DA strategy that foresees solely a domain alignment, and without a dedicated DA strategy (i.e., learning a model over a source domain and apply it over a target domain without prior domain alignment and selection of a suitable source domain; this represents a lower benchmark here).

increase of 4.7 and 4.9 percentage points (p.p.) regarding *built-up density* and *height*, respectively. Expectably, MAEs are by far the lowest when learning and applying a model within a domain. Thus, the difference in accuracy compared to the domain adaptation settings can be regarded as an error cost, that is related to the acquisition of training data for a target domain.

However, to enable a more detailed perspective which source selection mechanisms are most beneficial, more differentiated results are revealed in Fig. 5. It shows the median percentage differences in terms of MAE as obtained with source-target domain selections with the (ensemble) similarity metrics compared to random source-target domain selections computed from the aligned data across all configurations. Thereby, we also depict empirical MAEs when always the worst source domain for a certain target domain would be selected (lower bound) and when always the best source domain for a certain target domain would be selected (upper bound), respectively.

First, it can be noted that all values of all considered (ensemble) similarity metric-based selections are positive, i.e., the majority of models could benefit from the guided source-target domain selection based on (ensemble) similarity metrics (Fig. 5a). Regarding the individual similarity metrics, the Pearson correlation coefficient proved to be the most useful similarity metric here, since an improvement of 9.9 p.p. could be achieved. This measure turned out to be useful in the context of, e.g., unsupervised classification of high-dimensional data sets also [21].

However, our proposed pruned ensemble enabled the 2nd highest overall improvement (9.4 p.p.). Its implementation was particularly motivated to alleviate the risk of eventually picking an inadequate single metric. In this manner, our pruned ensemble of four also enabled more favorable improvements compared to an ensemble that includes all metrics (7.1 p.p.) and random selections of four metrics (6.0 p.p.). Besides, also *t*-test results confirm the significant improvement of source selection based on Hellinger, Pearson, and the ensemble compiled with our proposed heuristically-guided ensemble pruning strategy. *T*-statistics of 2.33, 5.41, and 2.4, respectively, indicate that the null hypothesis

Table 2

Number of labeled rectangular grid cells for each domain used in the regression experiment.

domain	Number of grid cells for model learning/validation		
	200 m	500 m	800 m
Berlin	10,450	2,051	859
Hamburg	9,982	2,307	1,054
Munich	9,250	2,138	953
Cologne	9,131	2,030	895

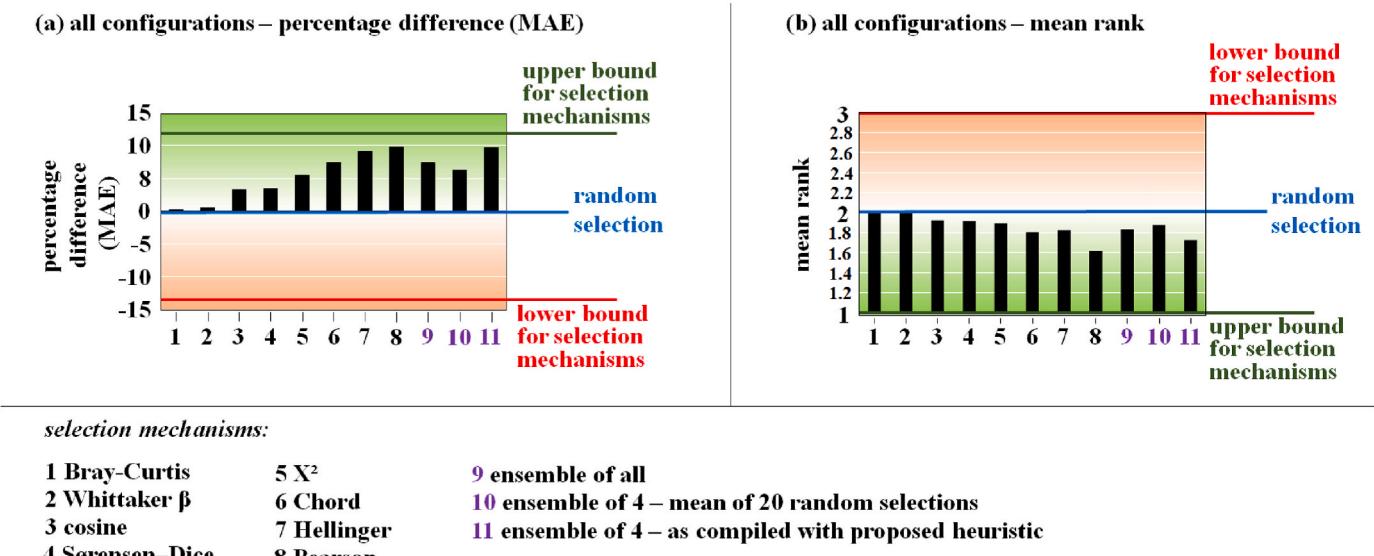


Fig. 5. (a) Median percentage difference in terms of MAE as obtained with source-target domain selections with (ensemble) similarity metrics compared to random source-target domain selections considering all configurations; results are obtained from 48 realizations (composed of four cities, three sizes of the spatial processing units, two regressors, and two target variables) except for the randomly compiled ensemble of four which was assembled and applied 20 times (i.e., the numerical value regarding 960 realizations is plotted); (b) corresponding mean ranks of the ordinal voting schemes.

(i.e., the selection mechanisms feature the same accuracy as random selection) can be rejected since the corresponding critical t -value of 1.68 is exceeded.

Overall, the experimental results underline the capability of the proposed methods to obtain substantially higher accuracy levels more frequently compared to random source-target domain selections. Those accuracy patterns are also unambiguously reflected in the achieved average ranks of the selection mechanisms which are shown in Fig. 5b. For instance, Hellinger, Pearson, and the ensemble compiled with our proposed heuristically-guided ensemble pruning strategy also feature the lowest average ranks.

5. Conclusions

In this work, we established a way for domain adaptation by source selection when multiple potentially helpful but unlabeled source domains exist. In contrast to previous works, we aimed to select an unlabeled source domain for an unlabeled target domain and obtain labeled samples for the source domain solely after it was selected successfully. The main findings comprise:

- Domain alignment based on histogram matching turned out to allow viable estimates of the target variables in the regression experiment, which aimed to estimate properties of the urban morphology, i.e., *built-up density* and *height* computed from Sentinel-2 imagery.
- The use of similarity metrics as applied to aligned marginal distributions of various image features proved useful to select proper source-target domain combinations, which allowed obtaining frequently higher accuracy levels compared to random source-target domain selections.
- We evaluated different similarity metrics, whereby Hellinger and Pearson turned out to be the most favorable individual metrics in the considered empirical setting.
- Our proposed heuristically-guided ensemble pruning strategy, which was particularly motivated to alleviate the risk of eventually picking an inadequate single metric in this unsupervised selection problem, allowed the 2nd best overall estimations.

Overall, the beneficial accuracy properties of the proposed methods

encourage further deployment for remote sensing data analysis. Given the modular nature of the method, we are interested in adapting the method for multiple image modalities, e.g., integrating SAR measurements with corresponding feature set. Further, we aim to integrate this domain adaptation approach, which is independent of the learning algorithm, in supervised multitask ensemble learning models [15,29].

Credit author statement

Christian Geiß: Conceptualization, Methodology, Software, Visualization, Writing – original draft. Alexander Rabuske: Conceptualization, Methodology, Software. Patrick Aravena Pelizari: Conceptualization, Methodology, Writing – review & editing. Stefan Bauer: Conceptualization, Methodology, Writing – review & editing. Hannes Taubenböck: Conceptualization, Methodology, Writing – review & editing.

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.

Acknowledgments

This research was funded in part by the German Federal Ministry of Education and Research (BMBF) under grant no. 03G0876 (project RIESGOS).

References

- [1] Geiß C, Aravena Pelizari P, Blickensdörfer L, Taubenböck H. Virtual Support Vector Machines with self-learning strategy for classification of multispectral remote sensing imagery. *ISPRS J Photogrammetry Remote Sens* 2019;151:42–58.
- [2] Xu J, Lu K, Wang H. Attention fusion network for multi-spectral semantic segmentation. *Pattern Recogn Lett* 2021;146:179–84.
- [3] Persello C, Boulias A, Dalponte M, Gobakken T, Naesset E, Schölkopf B. Cost-Sensitive active learning with lookahead: optimizing field surveys for remote sensing data classification. *IEEE TGRS* 2014;52(10):6652–64.
- [4] Geiß C, Thoma M, Taubenböck H. Cost-sensitive multitask active learning for characterization of urban environments with remote sensing. *IEEE GRSL* 2018;15(6):922–6.

- [5] Tuia D, Persello C, Bruzzone L. Domain adaptation for the classification of remote sensing data: an overview of recent advances. *IEEE Geosci Remote Sensing Magazine* 2016;4(2):41–57.
- [6] Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng* 2010;22(10):1345–59.
- [7] Matasci G, Volpi M, Kanevski M, Bruzzone L, Tuia D. Semisupervised transfer component analysis for domain adaptation in remote sensing image classification. *IEEE TGRS* 2015;53(7):3550–64.
- [8] J. Peng, W. Sun, L. Ma, Q du, discriminative transfer joint matching for domain adaptation in hyperspectral image classification, *IEEE GRSL* 16 (6) 972–976.
- [9] Bruzzone L, Marconcini M. Domain adaptation problems: a DASVM classification technique and a circular validation strategy. *IEEE Trans Pattern Anal Mach Intell* 2010;32(5):770–87.
- [10] Zhu L, Ma L. Class centroid alignment based domain adaptation for classification of remote sensing images. *Pattern Recogn Lett* 2016;83(2):124–32.
- [11] Lu X, Gong T, Zheng X. Multisource compensation network for remote sensing cross-domain scene classification. *IEEE TGRS* 2020;58(4):2504–15.
- [12] Gong T, Zheng X, Lu X. Cross-domain scene classification by integrating multiple incomplete sources. *IEEE TGRS* 2021;59(12):10035–46.
- [13] Elshamli A, Graham WT, Shawki Areibi S. Multisource domain adaptation for remote sensing using deep neural networks. *IEEE TGRS* 2020;58(5):3328–40.
- [14] Wolpert H. Stacked generalization. *Neural Network* 1992;5(2):241–59.
- [15] Geiß C, Schrade H, Aravena Pelizari P, Taubenböck H. Multistrategy ensemble regression for mapping of built-up height and density with sentinel-2 data. *ISPRS J Photogrammetry Remote Sens* 2020;170:57–71.
- [16] Feilhauer H, Asner PG, R Martin E. Multi-method ensemble selection of spectral bands related to leaf biochemistry. *Remote Sens Environ* 2015;164:57–65.
- [17] Gonzalez RC, Woods RE. Digital image processing. second ed. Upper-Saddle River, NJ, USA: Prentice-Hall; 2002.
- [18] Drusch M, Del Bello U, Carlier S, Colin O, Fernandez V, Gascon F, Hoersch B, Isola C, Laberinti P, Martimort P, Meygret A, Spoto F, Sy O, Marchese F, Bargellini P. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sens Environ* 2012;120:25–36.
- [19] Soille P. Morphological image analysis: principles and applications. Germany: Springer; 2004.
- [20] Haralick RM. Statistical and structural approaches to texture. *Proc IEEE* 1979;67(5):786–804.
- [21] Shirkhorshidi AS, Aghabozorgi S, Wah TY. A comparison study on similarity and dissimilarity measures in clustering continuous data. *PLoS One* 2015;10(12):e0144059.
- [22] Aravena Pelizari P, Spröhnle K, Geiß C, Schoepfer E, Plank S, Taubenböck H. Multi-sensor feature fusion for very high spatial resolution built-up area extraction in temporary settlements. *Remote Sens Environ* 2018;209:793–807.
- [23] European Space Agency (ESA). – Sen2Cor, <http://step.esa.int/main/third-party-plugins-2/sen2cor/>. [Accessed 8 July 2021].
- [24] Klotz M, Kemper T, Geiß C, Esch T, Taubenböck H. How good is the map? A multi-scale cross-comparison framework for global settlement layers: evidence from Central Europe. *Remote Sens Environ* 2016;178:191–212.
- [25] Luebke D, Watson B, Cohen JD, Reddy M, Varshney A. Level of detail for 3D graphics. Elsevier Science Inc; 2002.
- [26] Geiß C, Leichtle T, Wurm M, Aravena Pelizari P, Standfuß I, Xiao Xiang X, So E, Siedentop S, Esch T, Taubenböck H. Large-area characterization of urban morphology – mapping of built-up height and density using TanDEM-X and sentinel-2 data. *IEEE JSTARS* 2019;12(8):2912–27.
- [27] Liaw A, Wiener M. Classification and regression by random forest. *R News: Newsletter R Project* 2002;2(3):18–22.
- [28] Smola A, Schölkopf B. A tutorial on support vector regression. *Stat Comput* 2004; 14:199–222 [Kluwer Academic Publishers].
- [29] Geiß C, Brzoska E, Aravena Pelizari P, Lautenbach S, Taubenböck H. Multi-target regressor chains with repetitive permutation scheme for characterization of built environments with remote sensing. *Int J Appl Earth Obs Geoinf* 2022;106:102657.