SELF-RULE TO ADAPT: GENERALIZED MULTI-SOURCE FEATURE LEARNING USING UNSUPERVISED DOMAIN ADAPTATION FOR COLORECTAL CANCER TISSUE DETECTION

A PREPRINT

Christian Abbet*
Signal Processing Lab 5 (LTS5)
EPFL
Switzerland
christian.abbet@epfl.ch

Linda Studer*
Documents, Image and Video Analysis (DIVA)
University of Fribourg
Switzerland
linda.studer@unifr.ch

Andreas Fischer
Documents, Image and Video Analysis (DIVA)
University of Fribourg
Switzerland

Heather Dawson
Institute of Pathology
University of Bern
Switzerland

Inti Zlobec
Institute of Pathology
University of Bern
Switzerland

Behzad Bozorgtabar
Signal Processing Lab 5 (LTS5)
EPFL
Switzerland

Jean-Philippe Thian
Signal Processing Lab 5 (LTS5)
EPFL
Switzerland

August 23, 2021

ABSTRACT

Supervised learning is constrained by the availability of labeled data, which are especially expensive to acquire in the field of digital pathology. Making use of open-source data for pre-training or using domain adaptation can be a way to overcome this issue. However, pre-trained networks often fail to generalize to new test domains that are not distributed identically due to variations in tissue stainings, types, and textures. Additionally, current domain adaptation methods mainly rely on fully-labeled source datasets. In this work, we propose Self-Rule to Adapt (SRA), which takes advantage of self-supervised learning to perform domain adaptation and removes the necessity of a fully-labeled source dataset. SRA can effectively transfer the discriminative knowledge obtained from a few labeled source domain's data to a new target domain without requiring additional tissue annotations. Our method harnesses both domains' structures by capturing visual similarity with intra-domain and cross-domain self-supervision. Moreover, we present a generalized formulation of our approach that allows the architecture to learn from multi-source domains. We show that our proposed method outperforms baselines for domain adaptation of colorectal tissue type classification and further validate our approach on our in-house clinical cohort. The code and models are available open-source: https://github.com/christianabbet/SRA

Keywords Computational pathology · self-supervised learning · unsupervised domain adaptation · colorectal cancer

*Co-first author. Christian Abbet and Linda Studer contributed equally.

1 Introduction

Colorectal cancer (CRC) is one of the most common cancers worldwide and its understanding through computational pathology techniques can significantly improve the chances of effective treatment [Geessink et al., 2019, Smit and Mesker, 2020] by refining disease prognosis and assisting pathologists in their daily routine. The data used in computational pathology most often consists of Hematoxylin and Eosin (H&E) stained whole slide images (WSIs) [Hegde et al., 2019, Lu et al., 2020] and tissue microarrays (TMAs) [Arvaniti et al., 2018, Nguyen et al., 2021]

Although fully supervised deep learning models have been widely used for a variety of tasks, including tissue classification [Kather et al., 2019] and semantic segmentation [Qaiser et al., 2019, Chan et al., 2019], in practice it is time-consuming and expensive to obtain fully-labeled data as it involves expert pathologists. This hinders the applicability of supervised machine learning models to real-world scenarios. Weakly supervised learning is a less demanding approach that does not depend on large labeled cohorts. Examples of this approach applied in digital pathology include WSIs classification [Tellez et al., 2018, Silva-Rodríguez et al., 2021] and Multiple-Instance Learning (MIL) based on diagnostic reports [Campanella et al., 2019]. However, these methods still need an adequate training set to initialize the learning process, which limits the gain that can be achieved from using unlabeled samples.

Self-supervised learning was proposed to address these limitations. It involves a training scheme where "*the data creates its own supervision*"[Pieter et al., 2020] to learn rich features from structured unlabeled data. Applications of this approach in computational pathology include multiple tasks such as survival analysis [Abbet et al., 2020], WSIs classification [Li et al., 2020], or even a image retrieval [Gildenblat and Klaiman, 2019].

Over the years, various large data banks have been made available online containing samples from a variety of organs [Weinstein et al., 2013, Litjens et al., 2018, Veta et al., 2019], such as the colon and rectum [Kather et al., 2016, Shanah et al., 2016a,b, Kather et al., 2019], which opens up possibilities for transfer learning and domain adaptation. Yet, using these data banks to develop computational pathology-based models for real-world scenarios remains challenging because of the domain gap, as these images were created under different imaging scenarios. A tissue sample's visual appearance can be largely affected by the staining procedure [Otálora et al., 2019], the type of scanner used [Cheng et al., 2019], or other artifacts such as folded tissues [Komura and Ishikawa, 2018].

To tackle this issue, color normalization techniques [Macenko et al., 2009, Zanjani et al., 2018, Anand et al., 2019] have been widely adopted. Nevertheless, these techniques solely rely on image color information, while the morphological structure of the tissue is not taken into account [Tam et al., 2016, Zarella et al., 2017]. This could lead to unpredictable results in the presence of substantial staining variations and dark staining due to densely clustered tumor cells.

Another field of research that aims to improve the classification of heterogeneous WSIs is unsupervised domain adaptation (UDA). These methods address the issue by learning from a rich source domain together with the label-free target domain to have a well-performing model on the target domain at inference time. UDA allows models to include a large variety of constraints to match relevant morphological features across the source and target domains.

DANN [Ganin and Lempitsky, 2015] for example uses gradient reversal layers to learn domain-invariant features. Self-Path [Koohbanani et al., 2020] combines the DANN approach and self-supervised auxiliary tasks. The selected tasks reflect the inner properties of the tissue and are assumed to improve the stability of the framework when working with histopathological images. The auxiliary tasks include hematoxylin channel prediction, Jigsaw puzzle-solving, and magnification prediction. Another example is CycleGAN [Zhu et al., 2017], which takes advantage of adversarial learning to cyclically map images between the source and target domain. However, adversarial approaches can fall short because they do not consider task-specific decision boundaries and only try to distinguish the features as either coming from the source or target domain [Saito et al., 2018a].

A further issue is that most methods consider fully-labeled source datasets [Dou et al., 2019] for domain adaptation. However, digital pathology mostly relies on unlabeled or partly-labeled data as the acquisition of fully labeled cohorts is often unfeasible. In addition, recent approaches tend to treat domain adaptation as a closed-set scenario [Carlucci et al., 2019], which assumes that all target samples belong to classes present in the source domain, even though this is often not the case in a real-world scenario. To overcome this concern, OSDA [Saito et al., 2018b] proposes an adversarial open-set domain adaptation approach, where the feature generator has the option to reject mistrusted or unknown target samples as an additional class. In another recent work, SSDA [Xu et al., 2019] uses self-supervised domain adaptation methods that combine auxiliary tasks, adversarial loss, and batch normalization calibration across the source and target domains. Finally, some approaches take advantage of multiple source datasets to learn features that are discriminant under varying modalities. In Matsuura and Harada [2020], domain-agnostic features are generated by combining a domain discriminator as well as a hard clustering approach.

In this work, we propose a label-efficient framework called Self-Rule to Adapt (SRA) for tissue type recognition in histological images and attempt to overcome the above-mentioned issues by combining self-supervised learning

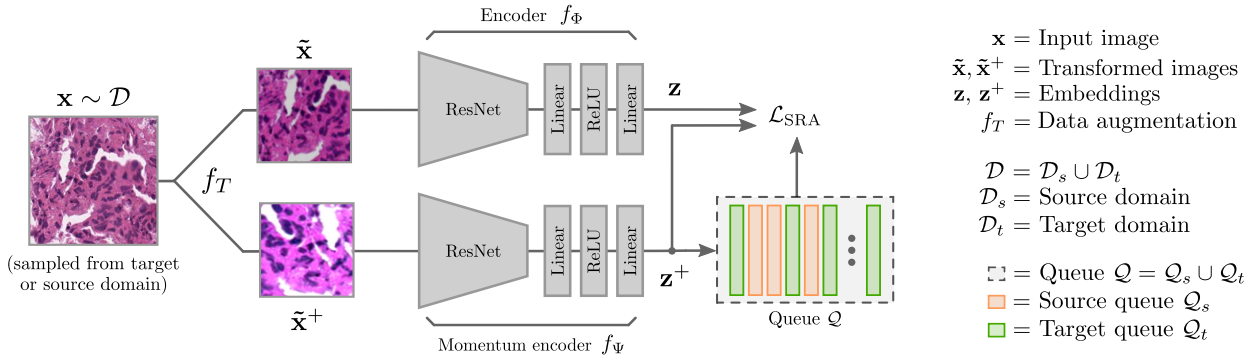


Figure 1: Schematic overview of the proposed Self-Rule to Adapt (SRA) architecture for a given input image $\mathbf{x}$. The loss $\mathcal{L}_{\text{SRA}}$ is the composition of the in-domain $\mathcal{L}_{\text{IND}}$ and cross-domain $\mathcal{L}_{\text{CRD}}$ losses. The queue $\mathcal{Q}$ keeps track of previous samples' embeddings and their set of origin (source or target).

approaches with UDA. We present an entropy-based approach that progressively learns domain invariant features, thus making our model more robust to class definition inconsistencies as well as the presence of unseen tissue classes when performing domain adaptation. SRA is able to accurately identify tissue types in H&E stained images, which is an important step for many downstream tasks. Our proposed method achieves this by using few labeled open-source datasets and unlabeled data that are abundant in digital pathology, thus reducing the annotation workload for pathologists. We show that our method outperforms previous domain adaptation approaches in a few-label setting and highlight the potential use for clinical application in the diagnostics of CRC.

This study is an extension of the work we presented at the Medical Imaging with Deep Learning (MIDL) 2021 conference [Abbet et al., 2021]. In this work, we provide a more in-depth explanation and analysis of our proposed entropy-based easy-to-hard (E2H) learning strategy. Furthermore, we reformulate the entropy-based cross-domain matching used by the E2H learning strategy to improve prediction robustness when dealing with complex tissue structures. Moreover, we also provide the generalization of SRA to multi-source domain adaptation by including an additional public dataset and perform additional experiments to assess the model's performance.

2 Methods

In our unsupervised domain adaptation setting, we have access to a small set of labeled data sampled from a source domain distribution and a set of unlabeled data from a target distribution. The goal is to learn a hypothesis function (for example a classifier) on the source domain that provides a good generalization in the target domain. To this end, we propose a novel self-supervised cross-domain adaptation setting called SRA, which is described in more detail below. Figure 1 gives an overview of the proposed network architecture.

To train our architecture, we rely on a set of images $\mathcal{D} = \mathcal{D}_s \cup \mathcal{D}_t$ that is the aggregation of a set of source images $\mathcal{D}_s$ and a set of target images $\mathcal{D}_t$. The model takes as input an RGB image $\mathbf{x} \in \mathbb{R}^{H \times W \times 3}$ sampled from $\mathcal{D}$ where H and W denote the height and width of the image, respectively. When sampling from $\mathcal{D}$, there is an equal probability to draw a sample from either the source or the target domain. After sampling, two sets of random transformations are applied to the image $\mathbf{x}$ using an image transformer $f_T : \mathbb{R}^{H \times W \times 3} \to \mathbb{R}^{H \times W \times 3}$. This generates a pair of augmented views $\tilde{\mathbf{x}}$, $\tilde{\mathbf{x}}^+ \in \mathbb{R}^{H \times W \times 3}$ that are assumed to share similar content as they are both different augmentations of the same source image. Each image of the pair $\tilde{\mathbf{x}}$, $\tilde{\mathbf{x}}^+$ is then fed to its respective encoder $f_\Phi : \mathbb{R}^{H \times W \times 3} \to \mathbb{R}^d$ and $f_\Psi : \mathbb{R}^{H \times W \times 3} \to \mathbb{R}^d$ to compute the query $\mathbf{z} \in \mathbb{R}^d$ and key $\mathbf{z}^+ \in \mathbb{R}^d$ embeddings of the input image. Here d represents the dimension of the embedding space. For simplicity, we will denote the embedding of an image drawn from the source and target domain as $\mathbf{z}_s$ and $\mathbf{z}_t$, respectively. In addition, when sampling from a dataset, we directly assume the image embedding as $\mathbf{z}, \mathbf{z}^+ \in \mathcal{D}$.

Each network's branch consists of a residual encoder followed by two linear layers based on the state-of-the-art (SOTA) architecture proposed in Chen et al. [2020a] (MoCoV2). We use the key embeddings $\mathbf{z}^+$ to maintain a queue of negative samples $\mathcal{Q} = \{\mathbf{q}_l \in \mathbb{R}^d\}_{l=1}^{|\mathcal{Q}|}$ in a first-in, first-out fashion. When updating the queue with a new negative sample, not only the sampled image's embedding is stored but also its domain of origin (source or target).

The queue provides a large number of examples which alleviates the need for a large batch size [Chen et al., 2020b] or the use of a memory bank [Kim et al., 2020]. Moreover, f_Ψ is updated using a momentum approach, combining its weights with those of f_Φ. This approach ensures that f_Ψ generates a slowly shifting and therefore coherent embedding.

Motivated by Ge et al. [2020], Kim et al. [2020], we extend the domain adaptation learning procedure to our model definition and task. Hence, we split the loss terms into two distinct tasks, namely in-domain $\mathcal{L}_{\text{IND}}$ and cross-domain $\mathcal{L}_{\text{CRD}}$ representation learning. The objective loss $\mathcal{L}_{\text{SRA}}$ is the summation of both terms which are described in more detail below.

$$\mathcal{L}_{\text{SRA}} = \mathcal{L}_{\text{IND}} + \mathcal{L}_{\text{CRD}}, \tag{1}$$

2.1 In-domain Loss

The first objective $\mathcal{L}_{\text{IND}}$ aims at learning the individual distribution of each the source and the target domain features. We want to keep the two domains independent as we will optimize their alignment later. For each embedding vector $\mathbf{z}$, there is a paired embedding $\mathbf{z}^+$ that is generated from the same tissue image and therefore is, by definition, similar. As a result, their similarity can be jointly optimized using a contrastive learning approach [Oord et al., 2018]. Here, we strongly benefit from data augmentation to create discriminant features that match both $\mathbf{z}$ and $\mathbf{z}^+$, making them more robust to outliers. By selecting the proper data augmentations, we can guide the model toward meaningful histopathological feature representations. This approach differs from Kim et al. [2020] where a memory bank is used instead of a queue, and no data augmentation is used to create negative and positive examples. The contrastive loss, as expressed in Equations 2 and 3, is therefore used to constrain the representation of the embedding space for each domain separately.

$$p_{\text{IND}}(\mathbf{z}, \mathbf{z}^+, \mathcal{Q}) = \frac{\exp(\mathbf{z}^\top \mathbf{z}^+/\tau)}{\exp(\mathbf{z}^\top \mathbf{z}^+/\tau) + \sum_{\mathbf{q}_l \in \mathcal{Q}} \exp(\mathbf{z}^\top \mathbf{q}_l/\tau)}. \tag{2}$$

$$\mathcal{L}_{\text{IND}} = \frac{-1}{|\mathcal{D}_s| + |\mathcal{D}_t|} \left(\sum_{\mathbf{z}_s, \mathbf{z}_s^+ \in \mathcal{D}_s} \log\left[p_{\text{IND}}(\mathbf{z}_s, \mathbf{z}_s^+, \mathcal{Q}_s)\right] + \sum_{\mathbf{z}_t, \mathbf{z}_t^+ \in \mathcal{D}_t} \log\left[p_{\text{IND}}(\mathbf{z}_t, \mathbf{z}_t^+, \mathcal{Q}_t)\right] \right). \tag{3}$$

We denote $\mathcal{Q}_s, \mathcal{Q}_t \subset \mathcal{Q}$ as the sets of indexed samples of the queue that were previously drawn from the corresponding domain $\mathcal{D}_s, \mathcal{D}_t \subset \mathcal{D}$, and $\tau \in \mathbb{R}$ as the temperature. The temperature is typically small ($\tau \ll 1$) thus sharpening the signal and helping the model to make confident predictions. For all images of each dataset $\mathcal{D}_s, \mathcal{D}_t$, we want to minimize the distance between $\mathbf{z}$ and $\mathbf{z}^+$ while maximizing the distance to the previously generated negative samples from the corresponding sets $\mathcal{Q}_s, \mathcal{Q}_t$. The queue samples are considered reliable negative candidates as they are generated by f_Ψ whose weights slowly vary due to its momentum update procedure.

2.2 Cross-domain Loss

We can see the cross-domain matching task as the generation of features that are discriminative across both sets. In other words, two samples that are visually similar but are drawn from the source $\mathcal{D}_s$ and target $\mathcal{D}_t$ domain, respectively, should have a similar embedding. On the other hand, when comparing these samples to the remaining candidates of the opposite domain, their resulting embeddings should be far apart. Practically, performing cross-domain matching using the number of available candidates within a batch might deteriorate the quality of the domain matching process due to the limited amount of negative samples. Therefore, we use the queue for negative samples mining the domain matching. Hence, we compute the similarity and entropy of each query pair $\mathbf{z}$, $\mathbf{z}^+$ drawn from one set (for example $\mathcal{D}_s$) to the stored queue samples from the other set (for example $\mathcal{Q}_t$):

$$p_{\text{CRD}}(\mathbf{z}, \mathbf{q}, \mathcal{Q}) = \frac{\exp(\mathbf{z}^\top \mathbf{q}/\tau)}{\sum_{\mathbf{q}_l \in \mathcal{Q}} \exp(\mathbf{z}^\top \mathbf{q}_l/\tau)}, \tag{4}$$

$$H(\mathbf{z}, \mathbf{z}^+, \mathcal{Q}) = -\sum_{\mathbf{q} \in \mathcal{Q}} p_{\text{CRD}}(\mathbf{z}, \mathbf{q}, \mathcal{Q}) \log\left[p_{\text{CRD}}(\mathbf{z}^+, \mathbf{q}, \mathcal{Q})\right], \tag{5}$$

$$\bar{H}(\mathbf{z}, \mathbf{z}^+, \mathcal{Q}) = \frac{1}{2}\left[H(\mathbf{z}\,, \mathbf{z}^+, \mathcal{Q}) + H(\mathbf{z}^+, \mathbf{z}\,, \mathcal{Q})\right]. \tag{6}$$

Algorithm 1: Generalized SRA pseudo code

Initialize queue $\mathcal{Q}$ with normal distribution;
Normalize queue entries $\{q_i\} \in Q$;
for $\mathrm{e} = 0$ **to** $N_{\text{epochs}} - 1$ **do**
  Create $\mathcal{D}$ by uniformly sampling from $\mathcal{D}_s^k$ and $\mathcal{D}_t$;
  Update easy-to-hard coefficient r using Equation 8;
  for batch $\{\mathbf{x}_i\}_{i=1}^{B}$ in $\mathcal{D}$ **do**
    Get augmented samples $\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_i^+$ using f_T;
    Perform forward pass $\mathbf{z}_i = f_\phi(\tilde{\mathbf{x}}_i),\ \mathbf{z}_i^+ = f_\psi(\tilde{\mathbf{x}}_i^+)$;
    Normalize vectors $\mathbf{z}_i$, $\mathbf{z}_i^+$;
    Compute in-domain loss $\mathcal{L}_{\text{IND}}$ using Equation 10;
    Calculate cross-domain entropy $\bar{H}$ using Equation 6 ;
    Compute easy-to-hard $\mathcal{R}_s^k$, $\mathcal{R}_t$ sets using Equation 12 ;
    Determine cross-domain loss $\mathcal{L}_{\text{CRD}}$ by replacing $\mathcal{D}_s^k$, $\mathcal{D}_t$ with $\mathcal{R}_s^k$, $\mathcal{R}_t$ in Equation 11 respectively;
    Compute $\mathcal{L}_{\text{SRA}} = \mathcal{L}_{\text{IND}} + \mathcal{L}_{\text{CRD}}$;
    Update f_Φ weights with backpropagation ;
    Update f_Ψ weights with momentum ;
    Update queue $\mathcal{Q}$ by appending $\mathbf{z}_i^+$;
  end
end

A low entropy $\bar{H}$ means that the selected query from one domain matches with a limited number of samples from another domain. Moreover, we penalize the model when the predictions from $\mathbf{z}$, $\mathbf{z}^+$ of the same image are different to improve the consistency of the domain matching [Assran et al., 2021]. This differs from our initial definition [Abbet et al., 2021] where solely $\mathbf{z}$ is used. As a result, the loss aims to minimize the average entropy of the similarity distributions, assisting the model in making confident predictions:

$$\mathcal{L}_{\text{CRD}} = \frac{1}{|\mathcal{D}_s| + |\mathcal{D}_t|} \left[\sum_{\mathbf{z}_s, \mathbf{z}_s^+ \in \mathcal{D}_s} \bar{H}(\mathbf{z}_s, \mathbf{z}_s^+, \mathcal{Q}_t) + \sum_{\mathbf{z}_t, \mathbf{z}_t^+ \in \mathcal{D}_t} \bar{H}(\mathbf{z}_t, \mathbf{z}_t^+, \mathcal{Q}_s) \right]. \tag{7}$$

2.3 Easy-to-hard Learning

There are two main pitfalls that can hamper the performance of the cross-domain entropy minimization.

Firstly, at the start of the learning process, the similarity measure between samples and the queue is unclear as the model weights are initialized randomly, which does not guarantee proper feature descriptions. As a result, the optimization of their relative entropy and the loss term $\mathcal{L}_{\text{CRD}}$ is ambiguous in the first epochs.

Secondly, being able to find matching samples for all input queries across datasets is a strong assumption. In clinical applications, we often rely on open-source datasets with a limited number of classes to annotate complex tissue databases. More specifically, challenging tissue types such as complex stroma are often not present in public datasets while being frequent in daily diagnostics. This example is illustrated in Figure 2. The top row shows the case where for a given target query $\mathbf{z_t}$ there are samples with a similar pattern in the source queue, i.e. the distribution of similarities p_{CRD} has low entropy. The second row highlights the opposite scenario where no queue elements match the query, generating a quasi-uniform distribution of similarities and therefore a high entropy. In other words, optimizing Equation 6 for all samples will result in a performance drop as the loss will try to find cross-domain candidates even if there are none to be found.

To tackle both of these issues, we introduce an easy-to-hard (E2H) learning scheme. The model starts with easy (low entropy) samples and progressively includes harder (high entropy) samples as the training progresses. We assume that the model becomes more robust after each iteration and therefore is more likely to properly process harder examples in later stages. Formally, we substitute the domains $\mathcal{D}_s, \mathcal{D}_t$ in Equation 7 with the corresponding set of candidates $\mathcal{R}_s, \mathcal{R}_t$ defined as:

$$r = \left\lfloor \frac{e}{N_{\text{epochs}} \cdot s_w} \right\rfloor \cdot s_h, \tag{8}$$

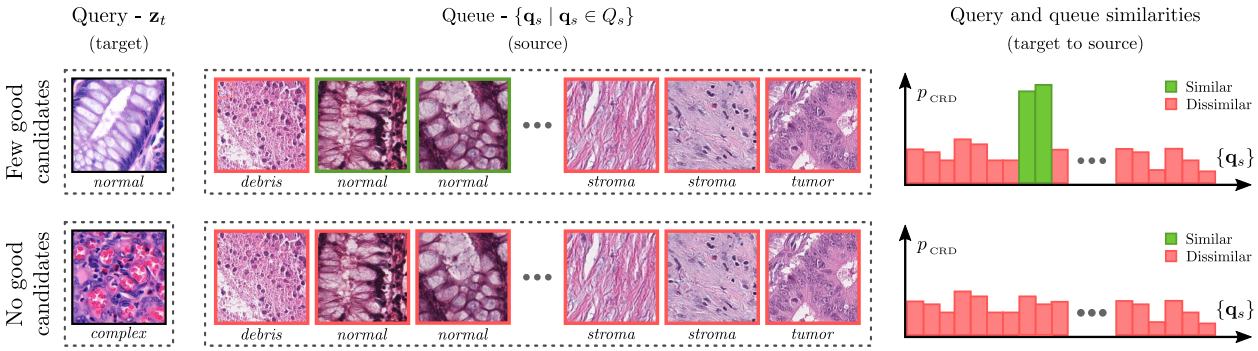


Figure 2: Toy example of cross-domain matching of different target queries to a fixed source queue. The first column shows two example target query images with computed embedding $\mathbf{z}_t$, the second column the source queue images maintained by the model and their corresponding embeddings $\{\mathbf{q}_s\}$, and the third column the computed similarities p_{CRD} between the queries and each queue sample. Similar and dissimilar patterns with respect to the query are displayed in green and red. The top row highlights the case where the model is able to find at least a subset of elements of the queue that match the query (low entropy), as opposed to the bottom row where none of the queue samples match the presented query (high entropy). The class labels in this figure have been added for ease of reading and are not available during training.

$$\mathcal{R}_s = \{\mathbf{z}_s, \mathbf{z}_s^+ \in \mathcal{D}_s \mid \bar{H}(\mathbf{z}_s, \mathbf{z}_s^+, \mathcal{Q}_t) \text{ is reverse top-}r\}, \tag{9}$$

$$\mathcal{R}_t = \{\mathbf{z}_t, \mathbf{z}_t^+ \in \mathcal{D}_t \mid \bar{H}(\mathbf{z}_t, \mathbf{z}_t^+, \mathcal{Q}_s) \text{ is reverse top-}r\},$$

where the ratio r is gradually increased during training using a step function. We denote s_w, s_h as the width and height of the step, respectively, N_{epochs} as the total number of epochs, and e the current epoch. This definition ensures that as long as $r = 0$ (e.i. $e < N_{\text{epochs}} \cdot s_w$) we only use the inner-domain loss $\mathcal{L}_{\text{IND}}$ for backpropagation and the cross-domain loss term $\mathcal{L}_{\text{CRD}}$ is not considered. This allows us to first learn the representation of the feature based on the inner-domain feature distribution only. Moreover, with the tuning of the parameter s_h, we can limit the maximal value of r and therefore avoid cross-domain matching when no candidates are available.

2.4 Generalization to Multiple Source Scenario

Our proposed SRA model can be generalized to consider multiple datasets in the source domain. This is especially useful if the available source datasets overlap in terms of class definitions, but increase the diversity of the visual appearance. More formally, we rely on K source datasets denoted $\mathcal{D}_s^k$ where $\bigcup_{k=0}^{K-1} \mathcal{D}_s^k = \mathcal{D}_s$, and $\mathcal{D} = \mathcal{D}_s \cup \mathcal{D}_t$. The same is valid for the source queues $\mathcal{Q}_s^k$ where $\bigcup_{k=0}^{K-1} \mathcal{Q}_s^k = \mathcal{Q}_s$, and $\mathcal{Q} = \mathcal{Q}_s \cup \mathcal{Q}_t$. The in-domain and cross-domain losses are now defined as:

$$\mathcal{L}_{\text{IND}} = \frac{-1}{\sum_{k=0}^{K-1} |\mathcal{D}_s^k| + |\mathcal{D}_t|} \left(\sum_{k=0}^{K-1} \sum_{\mathbf{z}_s, \mathbf{z}_s^+ \in \mathcal{D}_s^k} \log \left[p_{\text{IND}}(\mathbf{z}_s, \mathbf{z}_s^+, \mathcal{Q}_s^k) \right] + \sum_{\mathbf{z}_t, \mathbf{z}_t^+ \in \mathcal{D}_t} \log \left[p_{\text{IND}}(\mathbf{z}_t, \mathbf{z}_t^+, \mathcal{Q}_t) \right] \right), \tag{10}$$

$$\mathcal{L}_{\text{CRD}} = \frac{1}{\sum_{k=0}^{K-1} |\mathcal{D}_s^k| + K|\mathcal{D}_t|} \sum_{k=0}^{K-1} \left[\sum_{\mathbf{z}_s \in \mathcal{D}_s^k} \bar{H}(\mathbf{z}_s, \mathbf{z}_s^+, \mathcal{Q}_t) + \sum_{\mathbf{z}_t \in \mathcal{D}_t} \bar{H}(\mathbf{z}_t, \mathbf{z}_t^+, \mathcal{Q}_s^k) \right]. \tag{11}$$

The definition of the E2H learning procedure has to be updated to the multi-source domain definition as well. So far, the model seeks to match the target domain to the source domain without taking into consideration that there are multiple available source domains. In this scenario, the model might match the samples exclusively to one of the source sets while ignoring more complex examples in the other ones. To overcome this, we impose cross-domain matching for each source set individually. Our assumption is that for a given target sample, the architecture should be able to retrieve

good candidates in each of the source domains and vise-versa. We replace the domains $\mathcal{D}_s^k, \mathcal{D}_t$ in Equation 11 with the corresponding set of candidates $\mathcal{R}_s^k, \mathcal{R}_t$ defined as:

$$\mathcal{R}_s^k = \{\mathbf{z}_s, \mathbf{z}_s^+ \in \mathcal{D}_s^k \mid \bar{H}(\mathbf{z}_s, \mathbf{z}_s^+, \mathcal{Q}_t) \text{ is reverse top-}r\}, \tag{12}$$

$$\mathcal{R}_t = \{\mathbf{z}_t, \mathbf{z}_t^+ \in \mathcal{D}_t \mid \bar{H}(\mathbf{z}_t, \mathbf{z}_t^+, \mathcal{Q}_s^k) \text{ is reverse top-}r\},$$

Algorithm 1 presents the pseudo-code of our generalized SRA method.

3 Datasets

In this study, we use three publicly available datasets, Kather-16 (K16), Kather-19 (K19) and Colorectal Cancer Tissue Phenotyping (CRC-TP), that contain patches extracted from H&E-stained WSIs of different tissue types found in the human gastrointestinal tract. We also use an in-house CRC cohort which does not have patch-level labels and evaluate our method on three regions of interest (ROIs). More details on the datasets can be found below.

Figure 3 shows the occurrence and relationship of different tissue types across all four datasets. The displayed crops of the in-house WSI datasets are cherry-picked for comparison purposes.

3.1 Kather-16 Dataset

The K16 dataset [Kather et al., 2016] contains $5,000$ patches (150×150 pixels, $74\mu m \times 74\mu m$) from multiple H&E WSIs. All images are digitized using a scanner magnification of 20x ($0.495\mu m$ per pixel). There are eight classes of tissue phenotypes, namely tumor epithelium, simple stroma (homogeneous composition, and smooth muscle), complex stroma (stroma containing single tumor cells and/or few immune cells), immune cells, debris (including necrosis, erythrocytes, and mucus), normal mucosal glands, adipose tissue, and background (no tissue). The dataset is balanced with 625 patches per class.

3.2 Kather-19 Dataset

The K19 dataset [Kather et al., 2019] consists of patches depicting nine different tissue types: cancerous tissue, stroma, normal colon mucosa, adipose tissue, lymphocytes, mucus, smooth muscle, debris, and background. Each class is roughly equally represented in the dataset. In total, there are $100,000$ patches (224×224 pixels, $112\mu m \times 112\mu m$) in the training set. All images are digitized using a scanner at a magnification of 20x ($0.5\mu m$ per pixel).

3.3 Colorectal Cancer Tissue Phenotyping Dataset

The CRC-TP [Javed et al., 2020] dataset contains a total of $196,000$ patches depicting seven different tissue phenotypes (tumor, inflammatory, stroma, complex stroma, necrotic, benign, and smooth muscle). The different phenotypes are roughly equally represented in the dataset. For tumor, complex stroma, stroma, and smooth muscle there are 35,000 patches per class, for benign and inflammatory there are 21,000, and for debris, there are 14,000. The patches (150×150 pixels) are extracted at 20x resolution from 20 H&E WSIs, each one coming from a different patient. For each class, only a subset of the WSIs is used to extract the patches. The annotations are made by two expert pathologists. Out of the two dataset splits available, we use the training set of the patient-level separation.

3.4 In-house Dataset

Our cohort is composed of 665 H&E-stained WSIs from our local CRC patient cohort at the Institute of Pathology, University of Bern, Switzerland. The slides originate from 378 unique patients diagnosed with adenocarcinoma and are scanned at a resolution of $0.248\mu m$ per pixel (40x). None of the selected slides originated from patients that underwent preoperative treatment.

From each WSI we uniformly sample 300 (448×448 pixels, $111\mu m \times 111\mu m$) regions from the foreground masks to reduce the computational complexity of the proposed approach. This creates a dataset with a total of $199,500$ unique and unlabeled patches. We assume that these randomly selected samples are a good estimation of the tissue complexity and heterogeneity of our cohort.

We also select three ROIs of size $5 \times 5mm$ ($\simeq 20,000 \times 20,000$ pixel), which are annotated by an expert pathologist according to the definitions used in the K19 dataset, and used for evaluation.

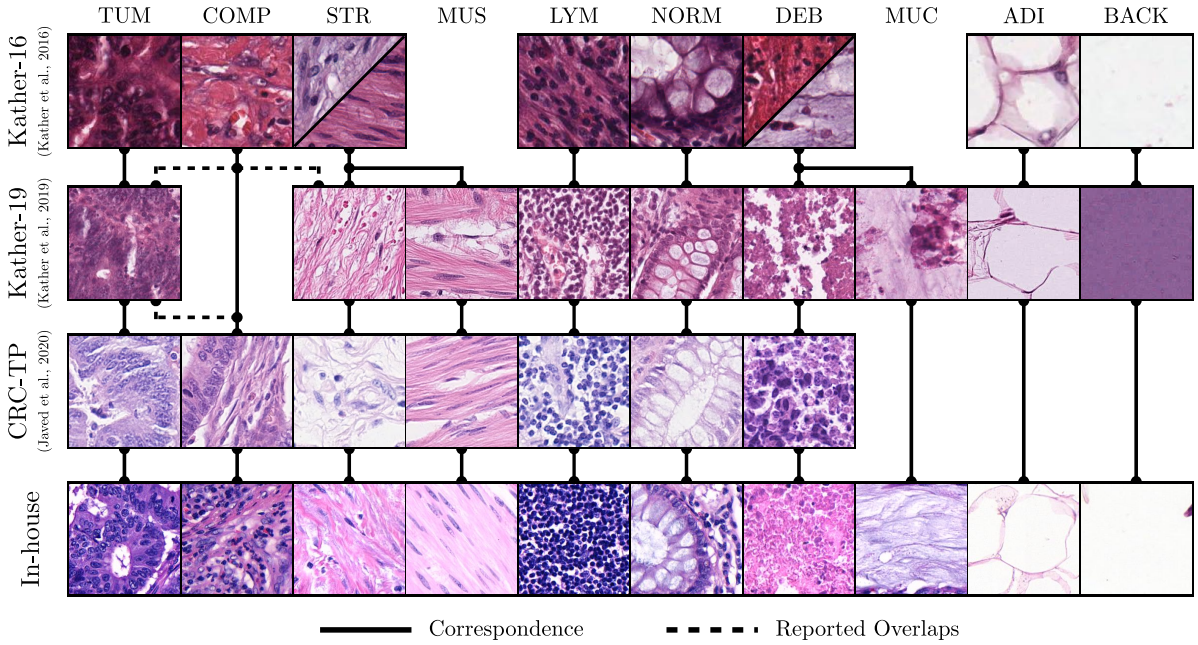


Figure 3: Example images of the different tissue types present in the used datasets and their association. The labeled datasets Kather-16 (K16), Kather-19 (K19), and Colorectal Cancer Tissue Phenotyping (CRC-TP) are publicly available. Examples from the in-house dataset are manually picked for comparison but are not labeled. We use the following abbreviations: TUM: tumor epithelium, STR: simple stroma, COMP: complex stroma, LYM: lymphocytes, NORM: normal mucosal glands, DEB: debris/necrosis, MUS: muscle, MUC: mucus, ADI: adipose tissue, BACK: background. The solid and dashed lines indicate classes correspondences and reported overlaps, respectively.

The regions are selected such that, overall, they represent all tissue types, as well as challenging cases such as late cancer stage (ROI 1), mucinous carcinoma (ROI 2), and torn tissue (ROI 3).

3.5 Discrepancies in Class Definitions Between Datasets

The class definitions are not homogeneous across the datasets and they also do not contain the same number of tissue classes. Following a discussion with expert pathologists, we group stroma/muscle and debris/mucus as stroma and debris, respectively, to create a corresponding adaptation between K19 and K16. Complex stroma which is only present in K16 is kept for training but excluded from the evaluation process when performing adaptation from K19 to K16. With this problem definition, we fall into an open-set scenario where the class distribution of the two domains does not rigorously match, as opposed to a closed set adaptation scheme.

According to expert pathologists, the complex stroma class in the CRC-TP dataset contains tiles from the tumor border region, and is also consistent with the tumor class in the K16 and K19 dataset. In K16, complex stroma is not limited to the tumor border surroundings and is defined as the desmoplastic reaction area which is usually composed of a mixture of debris, lymphocytes and single tumor cells and tumor cell clusters.

4 Results and Discussion

In this section, we present the results of the experiments. The general experimental setup is described in Section 4.1. We validate our proposed self-supervised domain adaptation approach using publicly available datasets and compare it to current SOTA methods for UDA in Section 4.2. Additionally, we assess the performance in a clinically relevant use case by validating our model on WSI sections from our in-house cohort in Section 4.3. These experiments are further extended to a multi-source application in Section 4.4. Finally, we perform an ablation study in Section 4.5 as well as additional experiments on the importance of the E2H learning procedure in Section 4.6. To help future research, the implementation is available open-source[2].

[2]Code available on `https://github.com/christianabbet/SRA`.

4.1 General Experimental Setup

In this section, we present the general setup that is common to all experiments. First, the architecture is trained in an unsupervised fashion. In a second step, a linear classifier is trained on top as described by Chen et al. [2020b].

For the unsupervised learning step, the architecture of the feature extractors, f_Φ and f_Ψ, are composed of a ResNet18 [He et al., 2016] followed by two fully connected layers (projection head) using rectified linear activation units (ReLUs). The output dimension of the multi-layer projection head is $d = 128$. We update the weights of f_Φ as θ_Φ using standard backpropagation and f_Ψ as θ_Ψ with momentum $m = 0.999$, as described in He et al. [2020].

The model is trained from scratch for $N_{\text{epochs}} = 200$ epochs until convergence using the stochastic gradient descent (SGD) optimizer (momentum $= 0.9$, weight decay $= 10^{-4}$), a learning rate of $\lambda = 0.03$ and a batch size of $B = 128$. The size of the queue is fixed to $|\mathcal{Q}| = 2^{16} = 65,536$ samples. For the similarity learning we set $\tau = 0.2$. We apply random cropping, gray transform, horizontal/vertical flipping, and color jittering as data augmentations f_T. At each epoch, we sample $50,000$ examples with replacement from both the source and target dataset to create $\mathcal{D}$ with a total of $N = 100,000$ samples.

During the second phase, the momentum encoder branch is discarded as it is not used for inference. The classification performance is evaluated using a linear layer, which is placed on top of the frozen feature extractor. The linear classifier directly matches the output of the embedding d to the number of classes. It is trained for $N_{\text{epochs}} = 100$ epochs until convergence using the SGD optimizer (momentum $= 0.9$, weight decay $= 0$), a batch size of $B = 128$, and a learning rate of $\lambda = 1$. We use only 1% of randomly selected source labels to train this classification layer in order to simulate the clinical application, where we usually rely on a large quantity of unlabeled data and only have access to few labeled samples. The training of the linear classifier is multi-run 10 times to obtain statistically relevant results. The selected fraction of labels (1%) differs between each run. We typically use $s_w = 0.25$ and $s_h = 0.15$ for E2H learning.

For a fair comparison, we also use a ResNet18 backbone for the presented baselines.

4.2 Cross-Domain Patch Classification

In this task, we use the larger dataset K19 as the source dataset and adapt it to K16. We motivate the selection of K19 as the source set by the fact that it is closer to the clinical scenario where we mainly rely on a large quantity of unlabeled data and only a few labeled ones, by using only 1% of the labels in K19. We evaluate the performance of the model with the patch classification task on the K16 dataset. To allow comparison with K16 definition, the mucin and muscle in K19 are grouped with debris and stroma, respectively. We use 70% of K16 to train the unsupervised domain adaptation. The remaining 30% are used to test the performance of the linear classifier trained on top of the self-supervised model.

The results of our proposed SRA method are presented in Table 1, in comparison with SOTA algorithms for domain adaption. As the lower bound, we consider direct transfer learning, where the model is trained in a supervised fashion on the source data only. We use the same logic for the upper bound by training on the target domain data (fully supervised approach). The performances on complex stroma are not reported as the class is undefined in K19. Figure 4 shows the t-SNE projection and alignment of the domain adaptation for the source only, top-performing baselines (OSDA, SSDA with jigsaw solving), and our method (SRA). Complementary results can be found in A and B.

Stain normalization slightly decreases the performance as it introduces color artifacts that are very challenging for the network classifier. This mainly comes from the distribution of target samples, namely K16, that are composed of dark stained patches which are difficult to normalize properly.

CycleGAN suffers from performance degradation for the lymphocytes predictions. Like color normalization, it tends to create saturated images. In addition, the model alters the shape of the lymphocytes nuclei, thus fooling the classifier toward either debris or tumor classification.

In our setup, we observe that the use of the gradient reversal layer leads to an unstable loss optimization for both Self-Path and DANN which explains the large performance drops when training. Heavier data augmentations partially solve this issue.

OSDA benefits from the open-set definition of the approach and achieves very good performance for lymphocytes detections.

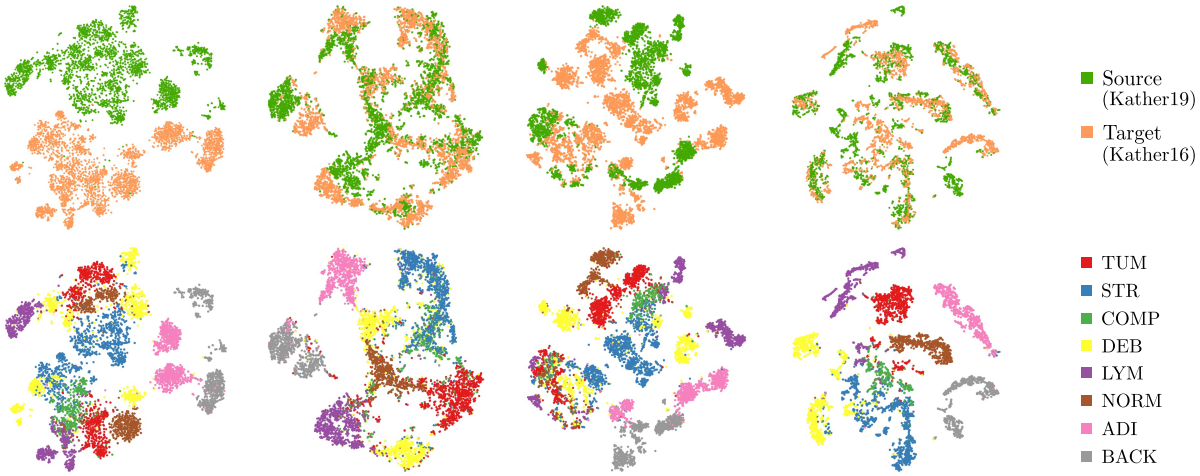
SSDA achieves similar results when using either rotation or jigsaw puzzle-solving as an auxiliary task. Due to the rotational invariance structure of the tissue and selected large magnification for tilling, rotation and jigsaw puzzle-solving are not optimal auxiliary tasks for digital pathology.

Our proposed SRA method shows an excellent alignment between the same class clusters of the source and target distributions and outperforms SOTA approaches in terms of weighted F1 score. Notably, our approach is even able to

Table 1: Results of the domain adaptation from K19 (source) to K16 (target). 1% of the source domain labels are known and the target domain labels are unknown. Complex stroma is excluded as the class is not present in K19 definition. Moreover, the mucin and muscle in K19 are respectively grouped with debris and stroma to respectively. The top results for the domain adaptation methods are highlighted in bold. We report the F1 score for each class as well as the overall weighted F1 score, averaged over 10 runs.

Methods	TUM	COMP	STR	LYM	DEB	NORM	ADI	BACK	ALL
Source only‡	74.0**	-	77.4**	75.3**	50.5**	66.9**	87.0**	93.1**	75.1**
Stain norm. [Macenko et al., 2009]	77.8**	-	75.9**	68.2**	42.1**	75.1**	77.4**	87.6**	72.2**
CylceGAN [Zhu et al., 2017]	70.7**	-	71.6**	62.3**	47.6**	75.5**	89.0**	88.2**	72.4**
DANN [Ganin and Lempitsky, 2015]	65.8**	-	60.8**	42.3**	47.8**	61.9**	64.1**	62.3**	57.8**
SelfPath [Koohbanani et al., 2020]	71.5**	-	68.8**	68.1**	57.6**	77.6**	82.3**	85.5**	73.1**
OSDA [Saito et al., 2018b]	82.0**	-	78.2*	**83.6**	63.8**	80.3**	90.8**	93.2*	81.7**
SSDA - Rot [Xu et al., 2019]	85.1**	-	78.5**	81.3*	**68.2**	88.7**	93.9**	**96.5**$^{+}$	84.7**
SSDA - Jigsaw Xu et al. [2019]	90.0**	-	**81.2**	79.5**	64.4**	88.3**	94.2**	95.7*	84.9**
SRA (ours)	**93.4**	-	72.9**	**82.7**$^{+}$	**67.9**$^{+}$	**96.5**	**97.0**	**97.2**	**86.9**
Target only†	94.6^{+}	-	83.6**	92.6**	88.7**	95.4^{+}	97.8*	98.5*	93.0**

‡ Direct transfer learning: trained on the source domain only, no adaptation (lower bound).
† Fully supervised: trained knowing all labels of the target domain (upper bound).
$^{+}$ $p \geq 0.05$; * $p < 0.05$; ** $p < 0.001$; unpaired t-test with respect to the top result.

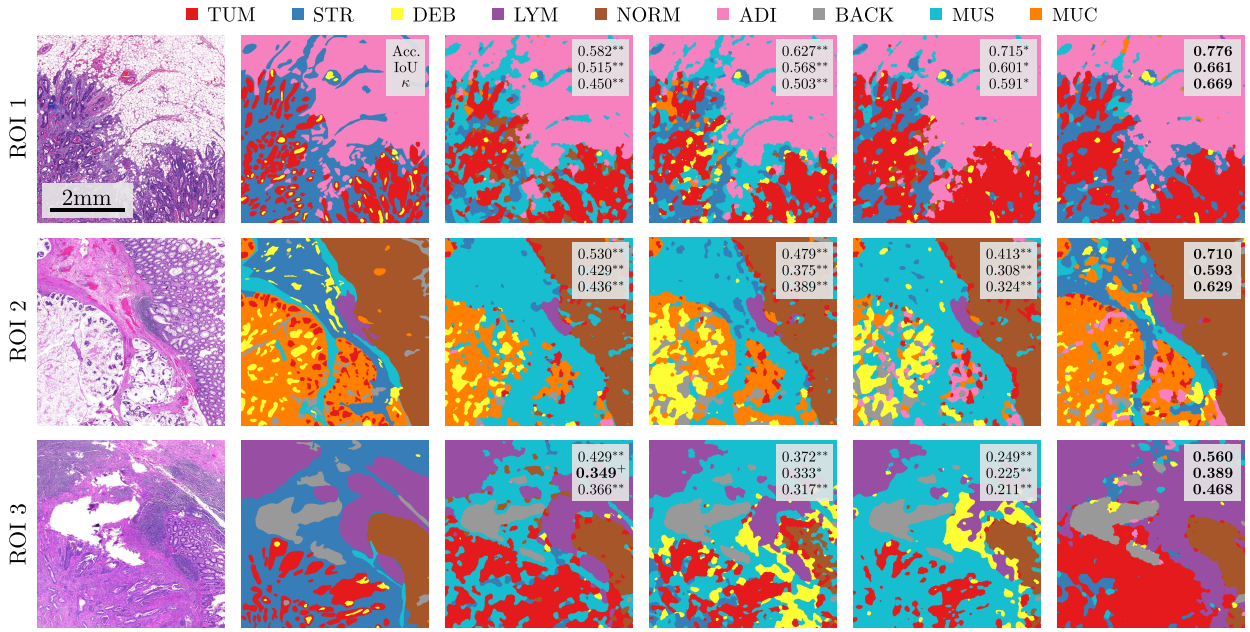


(a) Source Only (b) OSDA (Saito et al., 2018b) (c) SSDA - Jigsaw (Xu et al., 2019) (d) SRA (ours)

Figure 4: The t-SNE projection of the source (K19) and target (K16) domain embeddings. The top row shows the alignment between the source and target domain, while the bottom row highlights the representations of the different classes. We compare our approach (d) to other UDA methods (a-c).

match the upper bound model for normal and tumor tissue identification. The embedding of complex stroma, which only exists in the target domain, is represented as a single cluster with no matching candidates, which highlights the model's ability to reject unmatchable samples from domain alignment.

Furthermore, the cluster representation is more compact compared to other presented methods, where for example, normal mucosa tends to be aligned with complex stroma and tumor. Our approach suffers a drop in performance for stroma detection, which can be explained by the presence of lymphocytes in numerous stroma tissue examples, causing a higher rate of misclassification. Moreover, the presence of loose tissue that has a similar structure as stroma in the debris class is challenging. The overlap is even observed in the embedding projection.



(a) Original (b) Ground truth (c) Source only (d) OSDA (Saito et al., 2018b) (e) SSDA - Jigsaw (Xu et al., 2019) (f) SRA (ours)

$^{+}$ $p \geq 0.05$; $^{*}p < 0.05$; $^{**}p < 0.001$; unpaired t-test with respect to the top result.

Figure 5: Examples of domain adaptation from K19 to our in-house dataset. (a-b) show the original regions of interest (ROIs) from the WSIs and their ground truth, respectively. We compare the performance of our Self-Rule to Adapt (SRA) algorithm (f) to the lower bound and the top-performing SOTA methods (c-e). We report the pixel-wise accuracy, the weighted intersection over union, and the pixel-wise Cohen's kappa (κ) score averaged over 10 runs.

4.3 Use Case: Cross-Domain Segmentation of WSIs

To further validate our approach in a real case scenario, we perform domain adaptation using our proposed model from K19 to our in-house dataset and validate it on WSIs regions of interest (ROIs).

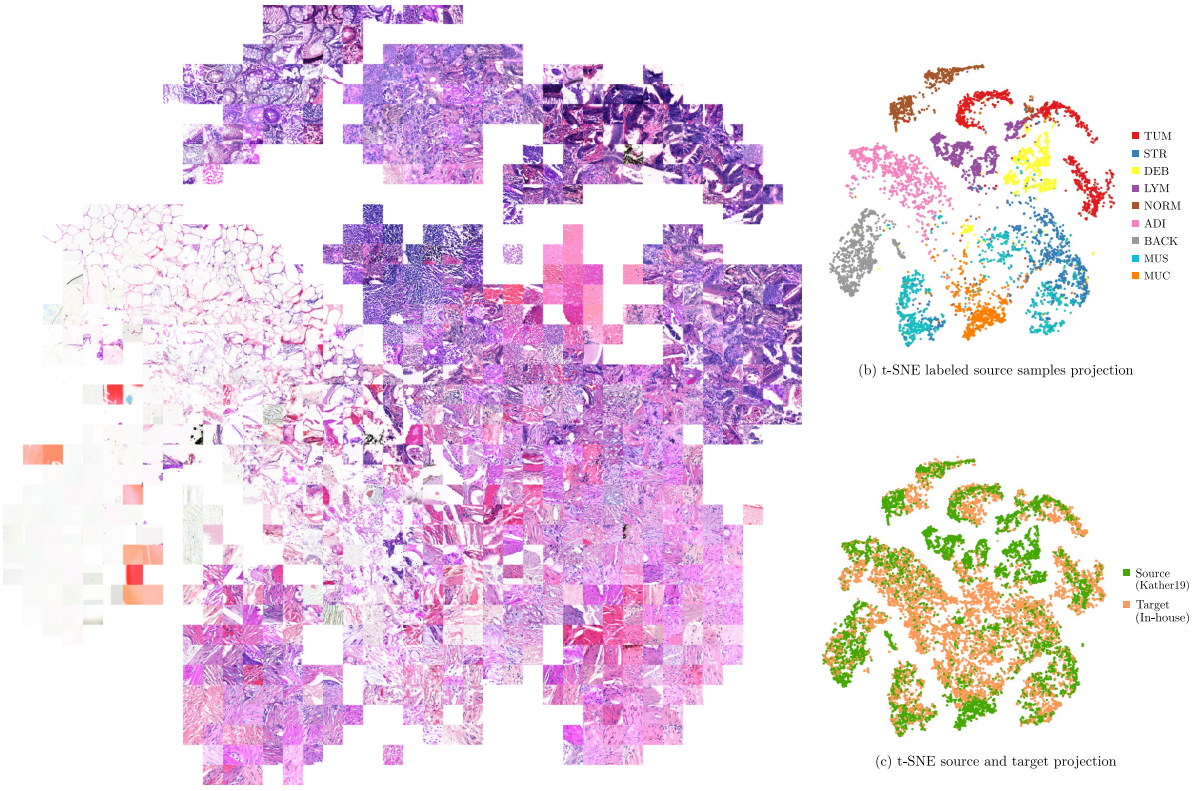
The results are presented in Figure 5, alongside the original H&E ROIs, their corresponding ground truth annotations, as well as comparative results of the top-scoring SOTA approaches. We use a tile-based approach to predict classes on each ROIs and use conditional random fields as in Chan et al. [2019] to smooth the prediction map. The number of available labeled tissue regions is limited to the presented ROIs.

For all models, stroma and muscle are poorly differentiated as both have similar visual features without contextual information. This phenomenon is even more apparent in the source only setting, where muscle tissue is almost systematically interpreted as stroma. Moreover, due to the lack of domain adaptation, the boundary between tumor and normal tissues is not well defined, leading to incorrect predictions of these classes.

OSDA, on the other hand, fails to adapt and generalize to new tumor examples while trying to reject mistrusted samples. This phenomenon is most visible in ROI 3, where the model interprets the surroundings of the cancerous region as a mixture of debris, stroma, and muscle.

SSDA tends to predict lymphocyte aggregates as debris. This can be explained by the model's sensitivity to staining variations as well as both classes' similarly dotted structure. Moreover, the model struggles to properly embed the representations of mucin. The scarcity of mucinous examples in the target domain makes the representation of the class difficult.

Our approach outperforms the other SOTA domain adaptation methods in terms of pixel-wise accuracy, weighed intersection over union (IoU) and pixel-wise Cohen's kappa score κ. Regions with mixtures of tissue types (e.g., lymphocytes + stroma or stroma + isolated tumor cells) are challenging cases because the samples available in public cohorts mainly contain homogeneous tissue textures and few examples of class mixtures. Subsequently, domain adaptation methods naturally struggle to align features resulting in a biased classification. We observe that thinner or torn stroma regions, where the background behind is well visible, are often misclassified as adipose tissue by SRA, which is most likely due to their similar appearance.



(a) t-SNE visualization of target patches distribution

(b) t-SNE labeled source samples projection

(c) t-SNE source and target projection

Figure 6: t-SNE visualization of the SRA model trained on K19 and our in-house data. All sub-figures depict the same embedding. (a) Patch-based visualization of the embedding. (b) Distribution of the labeled source samples. (c) The relative alignment of the source and target domain samples.

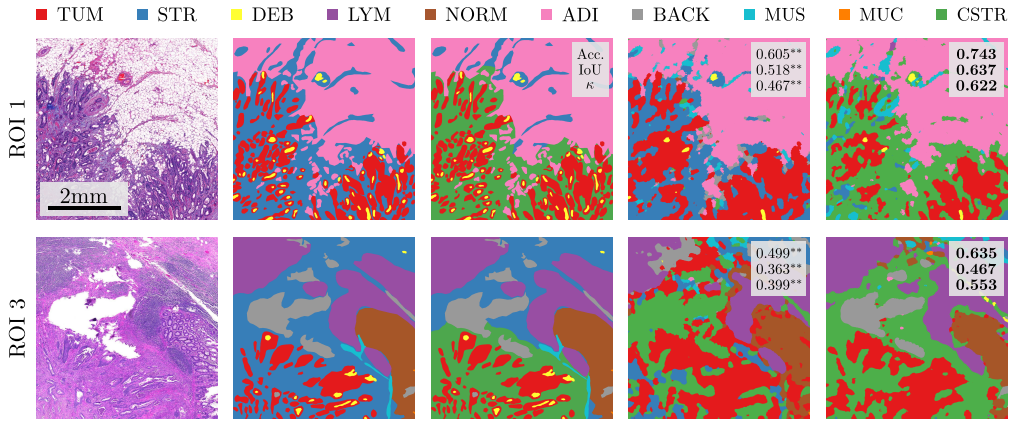
However, our SRA model is able to correctly distinguish between normal mucosa and tumor, which are tissue regions with relevant information for survival analysis.

Figure 6 presents a qualitative visualization of the model's embedding space. The figure shows the actual visual distribution of the target patches, the source domain label arrangement, and the overlap of the source and target domain. The patch visualization also shows a smooth transition between class representations where for example, neighboring samples of the debris cluster include a mixture of tissue and debris. The embedding reveals a large area in the center of the visualization that does not match with the source domain. The area mostly includes loose connective tissue and stroma, which are both under-represented in the training examples. Also, mucin is improperly matched to the loose stroma, which explains the misclassification of stromal tissue in the ROI 2. The scarcity of mucinous examples in our in-house cohort makes it difficult for the model to find good candidates.

4.4 Use Case: Multi-source Segmentation of WSI

In this section, we extend the results to multi-source domain adaptation. More specifically, we are interested in the detection of desmoplastic reactions (complex stroma), which is a prognostic factor in CRC [Ueno et al., 2021]. We combine K19 and CRC-TP to add complex stroma examples to our source data representation. Our in-house dataset is used as the target domain. The probability to draw a sample $\mathbf{x}$ from the source or the target domain is the same and is given by $p(\mathbf{x} \in \mathcal{D}_s) = Kp(\mathbf{x} \in \mathcal{D}_s^k) = p(\mathbf{x} \in \mathcal{D}_t)$, where K is the number of source domains.

To assess the quality of the prediction, we use the same ROIs as in the single-source setting. However, the previously provided annotations do not include complex stroma. To overcome this issue, we use the already existing annotation and define a margin around the tumor tissue that is considered as the interaction area, where the stroma will be interpreted as complex stroma. The margin is fixed to $500\mu m$ such that it includes the close tumor neighborhood and matches previous complex stroma definitions in the field [Berben et al., 2020, Nearchou et al., 2021].



(a) Original (b) Original ground truth (c) Extended ground truth (d) DeepAll (e) SRA (ours)

$^{+}$ $p \geq 0.05$; $^{*}p < 0.05$; $^{**}p < 0.001$; unpaired t-test with respect to the top result.

Figure 7: Results of the multi-source domain adaptation from K19 and CRC-TP to our in-house dataset. (a-c) show the original regions of interest (ROIs) from the WSIs, their original ground truth (without CSTR), and the extended ground truth (with CSTR), respectively. We compare the performance of our Self-Rule to Adapt (SRA) algorithm (e) to the DeepAll baseline (d). We report the pixel-wise accuracy, the weighted intersection over union, and the pixel-wise Cohen's kappa (κ) score averaged over 10 runs.

Table 2: Ablation study for the proposed Self-Rule to Adapt (SRA) approach. We denote $\mathcal{L}_{\text{IND}}$ as the in-domain loss, $\mathcal{L}_{\text{CRD}}$ as the cross-domain loss, and E2H as easy-to-hard. We train the domain adaptation from Kather-19 to our in-house dataset. Only 1% of the source domain Kather-19 labels are used, and no labels for the target domain. We report the pixel-wise accuracy, the weighted intersection over union, and the pixel-wise Cohen's kappa (κ) score averaged over 10 runs.

				ROI 1			ROI 2			ROI 3		
Model	$\mathcal{L}_{\text{IND}}$	$\mathcal{L}_{\text{CRD}}$	E2H	Acc.	IoU	κ	Acc.	IoU	κ	Acc.	IoU	κ
SRA†	-	-	-	0.619^{**}	0.525^{**}	0.476^{**}	0.291^{**}	0.189^{**}	0.212^{**}	0.304^{**}	0.242^{**}	0.226^{**}
SRA	-	✓	-	0.586^{**}	0.462^{**}	0.428^{**}	0.070^{**}	0.014^{**}	0.012^{**}	0.209^{**}	0.087^{**}	0.068^{**}
SRA	✓	-	-	0.749^{*}	$\mathbf{0.645}^{+}$	0.637^{*}	0.682^{*}	0.559^{**}	0.598^{*}	0.502^{**}	0.358^{*}	0.417^{**}
SRA	✓	✓	-	0.722^{**}	0.626^{*}	0.604^{**}	$\mathbf{0.702}^{+}$	$\mathbf{0.590}^{+}$	$\mathbf{0.625}^{+}$	0.519^{**}	0.319^{**}	0.415^{**}
SRA	✓	✓	✓	**0.776**	**0.661**	**0.669**	**0.710**	**0.593**	**0.629**	**0.560**	**0.389**	**0.468**

† Model jointly trained. Both source and target dataset are merged assuming a similar distribution.
$^{+}$ $p \geq 0.05$; * $p < 0.05$; ** $p < 0.001$; unpaired t-test with respect to top result.

To compare our results on multi-source domain adaptation, we use DeepAll as the baseline. DeepAll is defined as the aggregation of all the source tissue data into a single training set [Dou et al., 2019]. The model is trained in an unsupervised fashion using a standard contrastive loss to optimize the data representation of the features [Chen et al., 2020a]. In this case, no domain adaption is performed across the sets.

The results of our proposed SRA approach are presented in Figure 7, including the reference images, the original ground truth labels, the extended ground truth labels with complex stroma, and the DeepAll baseline. We emphasize the results on ROI 1 and 3, where desmoplastic reactions can be identified.

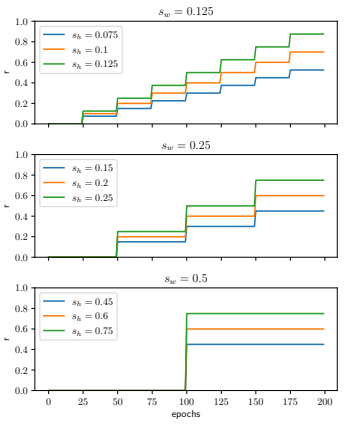
We observe that our model outperforms the baseline in terms of pixel-wise accuracy, Jaccard index (IoU), and Cohen's kappa score κ. Notably, the detection of the tumor appears more detailed compared to the single-source approach. This is visible on all presented ROIs. Parts of the tissue that were previously considered as tumor, due to the absence of a complex stroma class, can now be properly matched.

Another interesting result is the presence of the complex stroma in ROI 3 that replaces all the stromal area in both the baseline and the SRA predictions. It highlights how challenging complex stroma classification is without high-level context, which is even difficult for pathologists. They do not only rely on the tissue morphology for this assessment but

Images	Acc.	IoU	κ	Acc.	IoU	κ	Acc.	IoU	κ
	Metrics								
	$s_w = 0.125$, $s_h = 0.075$			$s_w = 0.125$, $s_h = 0.1$			$s_w = 0.125$, $s_h = 0.125$		
ROI 1	**0.771**$^{+}$	**0.649**$^{+}$	**0.661**$^{+}$	0.755*	0.636*	0.642*	0.750*	**0.644**$^{+}$	0.639*
ROI 2	0.677**	0.560**	0.592**	0.655**	0.533**	0.564**	**0.704**$^{+}$	0.578*	**0.623**$^{+}$
ROI 3	0.543*	**0.374**$^{+}$	0.452*	**0.548**$^{+}$	**0.387**$^{+}$	**0.461**$^{+}$	**0.547**$^{+}$	0.370*	**0.455**$^{+}$
	$s_w = 0.25$, $s_h = 0.15$			$s_w = 0.25$, $s_h = 0.2$			$s_w = 0.25$, $s_h = 0.25$		
ROI 1	**0.776**$^{+}$	**0.661**$^{+}$	**0.669**$^{+}$	**0.757**$^{+}$	0.633*	**0.643**$^{+}$	0.751*	0.623**	0.634*
ROI 2	**0.710**$^{+}$	**0.593**	**0.629**$^{+}$	**0.712**$^{+}$	**0.588**$^{+}$	**0.630**$^{+}$	**0.710**$^{+}$	**0.586**$^{+}$	**0.630**$^{+}$
ROI 3	**0.560**	**0.389**	**0.468**	**0.554**$^{+}$	**0.380**$^{+}$	**0.465**$^{+}$	0.544*	**0.386**$^{+}$	**0.461**$^{+}$
	$s_w = 0.5$, $s_h = 0.45$			$s_w = 0.5$, $s_h = 0.6$			$s_w = 0.5$, $s_h = 0.75$		
ROI 1	**0.777**	**0.667**	**0.672**	0.754*	0.633*	0.641*	0.743*	0.618**	0.625*
ROI 2	**0.714**	**0.592**$^{+}$	**0.633**	0.700*	0.570**	0.615*	0.702*	0.580**	**0.621**$^{+}$
ROI 3	0.536*	**0.379**$^{+}$	0.448*	0.526**	0.368*	0.439**	0.525**	0.369*	0.443*

$^{+}$ $p \geq 0.05$; * $p < 0.05$; ** $p < 0.001$; unpaired t-test with respect to top result.

Classification performance for different s_w, s_h values on the ROIs.



Profile of the easy-to-hard ratio r.

Figure 8: Study of the importance of the s_w and s_h parameter tuning. (left) Performance of the model (averaged over 10 runs) on the three regions of interest for each parameter pair. (right) Corresponding profiles of the step function r (Equation 8) as a function of the current epoch. The variable r represents the fraction of the trusted examples included for cross-domain matching.

also the spatial relations. Here, according to our extended ground truth, the complex stroma only surrounds the tumor region. However, the torn tissue at the center of the crop indicates that the whole surrounding regions were connected, which suggests that the complex stroma area spans even further. It, therefore, correlates with the prediction of both models that identify all the regions as complex stroma.

Complementary results on the quality of visualization of the multi-source domain embedding are available in C.

4.5 Ablation Study of the Proposed Loss Function

We present the ablation study of our approach in Table 2. We denote $\mathcal{L}_{\text{IND}}$ as the in-domain loss, $\mathcal{L}_{\text{CRD}}$ as the cross-domain loss, and E2H as the easy-to-hard learning scheme. We report the performance of our model with the task of matching K19 (source) to our in-house (target) domain (see Section 4.3). For the baseline, we do not take into consideration the information relative to the target and the source domain. Both domains are merged to a single domain $\mathcal{D}$ that is trained following the approach of Chen et al. [2020a].

The baseline fails to learn discriminant features that match both sets leading to poor cross-domain performances. If not constrained, the model is not able to generalize the knowledge and end up naturally learning two distinct features space as one for source and one for the target.

Using $\mathcal{L}_{\text{CRD}}$ alone does not help and creates an unstable model. As we do not impose domain representation, the model converges toward incorrect solutions where random sets of samples are matched between the source and target datasets. Moreover, it can create degenerated solutions where examples from the source and target domain are perfectly matched even though they do not present any visual similarity.

$\mathcal{L}_{\text{IND}}$ achieves relatively good performances but fails to generalize knowledge to classes where textures and staining strongly differ as in background or tumor.

The combination of the in-domain and cross-domain loss is not sufficient to improve the capability of the model. When performing a class-wise analysis, we observe that the score on tumor and normal drastically dropped. Both classes were forced to match samples from other classes, thus worsening the representation of the embedding.

The introduction of the E2H procedure improves the classification performances across all metrics. A detailed explanation of the importance of the E2H is discussed in the next section.

4.6 Evaluation of the E2H Learning Scheme

In this section, we discuss the usefulness and robustness of the E2H learning. The learning procedure is based on r (see Equations 8), and the two contributing variables, namely s_w and s_h.

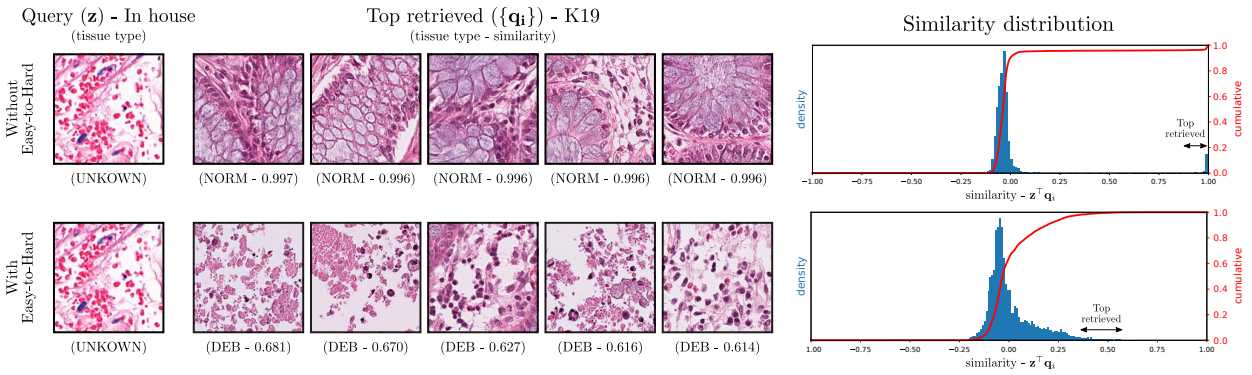


Figure 9: Cross-domain image retrieval based on an input query from our in-house cohort. The first column shows a target query $\mathbf{z}$ image example, the second column presents the retrieved samples from K19 that have the highest similarity in the source domain $\{\mathbf{q}_i\}$, and the third column shows the density distribution (blue) of similarities across the source domain as well as its cumulative profile (red). We list the retrieved examples with their assigned classes. The query class is unknown. The top and bottom rows highlight the result of training without and with E2H learning, respectively. Without E2H, the model tries to optimize $\mathcal{L}_{\text{CRD}}$ at any cost, which creates out-of-distribution samples (seen at the very right). With E2H the model predicts samples with lower confidence, but that are still visually similar.

In Figure 8, we show the impact of different combinations of these parameters on the single cross-domain segmentation task (see Section 4.3). We report the pixel-wise accuracy, the weighted intersection over union, and the pixel-wise Cohen's kappa (κ) score over the presented ROIs. For each parameters pair, we also display the profile of the variable r as a function of the number of epochs.

We can observe two primary outcomes of the experiments. Firstly, the model is more robust when s_h is low. The variable is an indicator of the ratio of samples used for cross-domain matching. In other words, the architecture benefits from a small s_h that allows it to focus on examples with high similarity/confidence while avoiding complex samples without properly matching candidates. Secondly, the selection of s_w is also crucial to the stability of the prediction. This quantity measures the number of epochs to wait before considering more complex examples in the cross-domain matching optimization. For small s_w values, the model has no time to learn the feature representation properly. This is especially true for the first few epochs after initialization, where the architecture is not yet able to embed features optimally. Furthermore, using large s_w weakens the model capability from progressively learning from less complex samples.

Figure 9 shows a practical example that is observed during the training phase and highlights the usefulness of the E2H scheme. When dealing with a heterogeneous target data cohort, some tissue types might not have relevant candidates in the other set (open-set scenario). The presented example shows an example composed of a vein and blood cells. Such a tissue structure is absent from the source cohort, thus making the cross-domain sample matching impossible.

Without the E2H learning, the model is still forced to find matching candidates, here normal mucosa (NORM), for the query $\mathbf{z}$. We observe that the matched samples form an out-of-distribution cluster with high similarity to the query ($\mathbf{z}^\top \mathbf{q}_i \simeq 1$) which directly comes from the optimization of the entropy term $\bar{H}$ (see Equation 6). This phenomenon is even more visible with the cumulative function (red) that correlates to the step function.

After introducing the E2H scheme, we can observe a continuous transition in the distribution of samples similarities. Here, the top retrieved samples share the same granular structure as the query. Still, we have to be careful as they do not represent the same type of tissue. The retrieved samples are examples of necrosis, whereas the query shows red blood cells. The fact that the architecture is less confident (i.e., a lower similarity for the top retrieved) is a good indicator of its robustness and ability to process complex queries.

As a result, the introduction of the E2H process prevents the model from learning degenerated solutions. The same behavior is observed in other open-set tissue classes, such as the absence of complex stroma and loose connective tissue in the source domain.

5 Conclusion and Future Work

In this work, we explore the usefulness of self-supervised learning and UDA for the identification of histological tissue types. Motivated by the difficulty of obtaining expert annotations, we explore diverse UDA models in various label-scarce histopathology datasets.

As our main contribution, we present a new label transferring approach from a partially labeled source domain to an unlabeled target domain. This is more practical than most previous UDA approaches tailored to fully annotated source domain data or tied to additional network branches dedicated to auxiliary tasks. Instead, we perform progressive entropy minimization based on the similarity distribution among the unlabeled target and source domain samples yielding discriminative and domain-agnostic features for domain adaptation.

Through adaptation experiments, we show that our proposed Self-Rule to Adapt method can discover the relevant semantic information even in the presence of few labeled source samples and yields a better generalization on different target domain datasets. Moreover, we show that our model definition can be generalized to a multi-source setting. As a result, the proposed model is able to learn rich data representation using multi-source domains.

The distribution of WSI patches are usually imbalanced in their class labels (categories) and exhibit some rare categories, posing significant challenges for trained models to generalize. We can cite the example of mucin that is frequent when dealing with mucinous carcinoma but is scarcely found in adenocarcinomas. The same logic applies to the complex stroma class that can be further divided into three subcategories as immature, intermediate, or mature and whose distributions are linked to patients' prognostic factor [Okuyama et al., 2020]. A possible extension of this work would be to design a framework that is able to take into account the distribution of classes across WSI to improve the quality and variety of the provided positive and negative examples.

In addition, publicly available datasets are solely composed of homogeneous patches. Such patches, however, do not capture the heterogeneity of complex images present in the diagnosis routine, which can lead to erroneous detections (e.g., background and stroma interaction interpreted as adipose). Thus, another future extension of this work is the definition of a self-supervised learning approach that can properly embed such mixed patches.

Lastly, the patch-based segmentation achieved by our method can be used for clinically relevant applications, such as tumor-stroma ratio calculation, disease-free survival prediction, or adjuvant treatment decision-making.

Acknowledgments

This work was supported by the Personalized Health and Related Technologies grant number 2018-327, and the Rising Tide foundation with the grant number CCR-18-130. The authors would like to thank Dr. Felix Müller, M. Med. Philipp Zens, and Guillaume Vray for the annotation of the WSI crops, the feedback on complex stroma detection and the computation of a few baselines that greatly helped the evaluation of our method.

References

Oscar GF Geessink, Alexi Baidoshvili, Joost M Klaase, Babak Ehteshami Bejnordi, Geert JS Litjens, Gabi W van Pelt, Wilma E Mesker, Iris D Nagtegaal, Francesco Ciompi, and Jeroen AWM van der Laak. Computer aided quantification of intratumoral stroma yields an independent prognosticator in rectal cancer. *Cellular Oncology*, 42(3):331–341, 2019.

Marloes A Smit and Wilma E Mesker. The role of artificial intelligence to quantify the tumour-stroma ratio for survival in colorectal cancer. *EBioMedicine*, 61, 2020.

Narayan Hegde, Jason D Hipp, Yun Liu, Michael Emmert-Buck, Emily Reif, Daniel Smilkov, Michael Terry, Carrie J Cai, Mahul B Amin, Craig H Mermel, et al. Similar image search for histopathology: Smily. *NPJ digital medicine*, 2 (1):1–9, 2019.

Ming Y Lu, Drew FK Williamson, Tiffany Y Chen, Richard J Chen, Matteo Barbieri, and Faisal Mahmood. Data efficient and weakly supervised computational pathology on whole slide images. *arXiv preprint arXiv:2004.09666*, 2020.

Eirini Arvaniti, Kim S Fricker, Michael Moret, Niels Rupp, Thomas Hermanns, Christian Fankhauser, Norbert Wey, Peter J Wild, Jan H Rueschoff, and Manfred Claassen. Automated gleason grading of prostate cancer tissue microarrays via deep learning. *Scientific reports*, 8(1):1–11, 2018.

Huu-Giao Nguyen, Annika Blank, Heather E Dawson, Alessandro Lugli, and Inti Zlobec. Classification of colorectal tissue images from high throughput tissue microarrays by ensemble deep learning methods. *Scientific Reports*, 11(1): 1–11, 2021.

Jakob Nikolas Kather, Johannes Krisam, Pornpimol Charoentong, Tom Luedde, Esther Herpel, Cleo-Aron Weis, Timo Gaiser, Alexander Marx, Nektarios A Valous, Dyke Ferber, et al. Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. *PLoS medicine*, 16(1), 2019.

Talha Qaiser, Yee-Wah Tsang, Daiki Taniyama, Naoya Sakamoto, Kazuaki Nakane, David Epstein, and Nasir Rajpoot. Fast and accurate tumor segmentation of histology images using persistent homology and deep convolutional features. *Medical image analysis*, 55:1–14, 2019.

Lyndon Chan, Mahdi S Hosseini, Corwyn Rowsell, Konstantinos N Plataniotis, and Savvas Damaskinos. Histosegnet: Semantic segmentation of histological tissue type in whole slide images. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 10662–10671, 2019.

David Tellez, Jeroen van der Laak, and Francesco Ciompi. Gigapixel whole-slide image classification using unsupervised image compression and contrastive training. *Medical Imaging with Deep Learning*, 2018.

Julio Silva-Rodríguez, Adrián Colomer, and Valery Naranjo. Weglenet: A weakly-supervised convolutional neural network for the semantic segmentation of gleason grades in prostate histology images. *Computerized Medical Imaging and Graphics*, 88:101846, 2021.

Gabriele Campanella, Matthew G Hanna, Luke Geneslaw, Allen Miraflor, Vitor Werneck Krauss Silva, Klaus J Busam, Edi Brogi, Victor E Reuter, David S Klimstra, and Thomas J Fuchs. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature medicine*, 25(8):1301–1309, 2019.

Chen Pieter, Abbeel anf Xi, Ho Jonathan, Srinivas Aravind, Li Alex, and Yan Wilson. Cs 294-158. deep unsupervised learning, February 2020.

Christian Abbet, Inti Zlobec, Behzad Bozorgtabar, and Jean-Philippe Thiran. Divide-and-rule: Self-supervised learning for survival analysis in colorectal cancer. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 480–489. Springer, 2020.

Bin Li, Yin Li, and Kevin W Eliceiri. Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. *arXiv preprint arXiv:2011.08939*, 2020.

Jacob Gildenblat and Eldad Klaiman. Self-supervised similarity learning for digital pathology. *2nd COMPAY Workshop at MICCAI 2019*, 2019.

John N Weinstein, Eric A Collisson, Gordon B Mills, Kenna R Mills Shaw, Brad A Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander, Joshua M Stuart, Cancer Genome Atlas Research Network, et al. The cancer genome atlas pan-cancer analysis project. *Nature genetics*, 45(10):1113, 2013.

Geert Litjens, Peter Bandi, Babak Ehteshami Bejnordi, Oscar Geessink, Maschenka Balkenhol, Peter Bult, Altuna Halilovic, Meyke Hermsen, Rob van de Loo, Rob Vogels, et al. 1399 h&e-stained sentinel lymph node sections of breast cancer patients: the camelyon dataset. *GigaScience*, 7(6):giy065, 2018.

Mitko Veta, Yujing J Heng, Nikolas Stathonikos, Babak Ehteshami Bejnordi, Francisco Beca, Thomas Wollmann, Karl Rohr, Manan A Shah, Dayong Wang, Mikael Rousson, et al. Predicting breast tumor proliferation from whole-slide images: the tupac16 challenge. *Medical image analysis*, 54:111–121, 2019.

Jakob Nikolas Kather, Cleo-Aron Weis, Francesco Bianconi, Susanne M Melchers, Lothar R Schad, Timo Gaiser, Alexander Marx, and Frank Gerrit Zöllner. Multi-class texture analysis in colorectal cancer histology. *Scientific reports*, 6:27988, 2016.

Kirk Shanah, Sadow Cheryl A., Smith J. Keith, Levine Seth, Roche Charles, Bonaccio Ermalinda, and Filippini Joe. Radiology data from the cancer genome atlas colon adenocarcinoma [tcga-coad] collection, 2016a.

Kirk Shanah, Sadow Cheryl A., and Levine Seth. Radiology data from the cancer genome atlas rectum adenocarcinoma [tcga-read] collection, 2016b.

Sebastian Otálora, Manfredo Atzori, Vincent Andrearczyk, Amjad Khan, and Henning Müller. Staining invariant features for improving generalization of deep convolutional neural networks in computational pathology. *Frontiers in bioengineering and biotechnology*, 7:198, 2019.

Wei-Chung Cheng, Firdous Saleheen, and Aldo Badano. Assessing color performance of whole-slide imaging scanners for digital pathology. *Color Research & Application*, 44(3):322–334, 2019.

Daisuke Komura and Shumpei Ishikawa. Machine learning methods for histopathological image analysis. *Computational and structural biotechnology journal*, 16:34–42, 2018.

Marc Macenko, Marc Niethammer, James S Marron, David Borland, John T Woosley, Xiaojun Guan, Charles Schmitt, and Nancy E Thomas. A method for normalizing histology slides for quantitative analysis. In *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1107–1110. IEEE, 2009.

Farhad Ghazvinian Zanjani, Svitlana Zinger, et al. Deep convolutional gaussian mixture model for stain-color normalization of histopathological images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 274–282. Springer, 2018.

Deepak Anand, Goutham Ramakrishnan, and Amit Sethi. Fast gpu-enabled color normalization for digital pathology. In *2019 International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 219–224. IEEE, 2019.

Allison Tam, Jocelyn Barker, and Daniel Rubin. A method for normalizing pathology images to improve feature extraction for quantitative pathology. *Medical physics*, 43(1):528–537, 2016.

Mark D Zarella, Chan Yeoh, David E Breen, and Fernando U Garcia. An alternative reference space for h&e color normalization. *PloS one*, 12(3):e0174489, 2017.

Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *International conference on machine learning*, pages 1180–1189, 2015.

Navid Alemi Koohbanani, Balagopal Unnikrishnan, Syed Ali Khurram, Pavitra Krishnaswamy, and Nasir Rajpoot. Self-path: Self-supervision for classification of pathology images with limited annotations. *arXiv preprint arXiv:2008.05571*, 2020.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.

Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, and Tatsuya Harada. Maximum classifier discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3723–3732, 2018a.

Qi Dou, Daniel Coelho de Castro, Konstantinos Kamnitsas, and Ben Glocker. Domain generalization via model-agnostic learning of semantic features. *Advances in Neural Information Processing Systems*, 32:6450–6461, 2019.

Fabio M Carlucci, Antonio D'Innocente, Silvia Bucci, Barbara Caputo, and Tatiana Tommasi. Domain generalization by solving jigsaw puzzles. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2229–2238, 2019.

Kuniaki Saito, Shohei Yamamoto, Yoshitaka Ushiku, and Tatsuya Harada. Open set domain adaptation by backpropagation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 153–168, 2018b.

Jiaolong Xu, Liang Xiao, and Antonio M López. Self-supervised domain adaptation for computer vision tasks. *IEEE Access*, 7:156694–156706, 2019.

Toshihiko Matsuura and Tatsuya Harada. Domain generalization using a mixture of multiple latent domains. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 11749–11756, 2020.

Christian Abbet, Linda Studer, Andreas Fischer, Heather Dawson, Inti Zlobec, Behzad Bozorgtabar, and Jean-Philippe Thiran. Self-rule to adapt: Learning generalized features from sparsely-labeled data using unsupervised domain adaptation for colorectal cancer tissue phenotyping. In *Medical Imaging with Deep Learning*, 2021. URL `https://openreview.net/forum?id=VO7asaS5GUk`.

Xinlei Chen, Haoqi Fan, Ross Girshick, and Kaiming He. Improved baselines with momentum contrastive learning. *arXiv preprint arXiv:2003.04297*, 2020a.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. *arXiv preprint arXiv:2002.05709*, 2020b.

Donghyun Kim, Kuniaki Saito, Tae-Hyun Oh, Bryan A Plummer, Stan Sclaroff, and Kate Saenko. Cross-domain self-supervised learning for domain adaptation with few source labels. *arXiv preprint arXiv:2003.08264*, 2020.

Yixiao Ge, Dapeng Chen, Feng Zhu, Rui Zhao, and Hongsheng Li. Self-paced contrastive learning with hybrid memory for domain adaptive object re-id. *arXiv preprint arXiv:2006.02713*, 2020.

Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

Mahmoud Assran, Mathilde Caron, Ishan Misra, Piotr Bojanowski, Armand Joulin, Nicolas Ballas, and Michael Rabbat. Semi-supervised learning of visual features by non-parametrically predicting view assignments with support samples. *arXiv preprint arXiv:2104.13963*, 2021.

Sajid Javed, Arif Mahmood, Muhammad Moazam Fraz, Navid Alemi Koohbanani, Ksenija Benes, Yee-Wah Tsang, Katherine Hewitt, David Epstein, David Snead, and Nasir Rajpoot. Cellular community detection for tissue phenotyping in colorectal cancer histology images. *Medical Image Analysis*, page 101696, 2020.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9729–9738, 2020.

Hideki Ueno, Yoshiki Kajiwara, Yoich Ajioka, Tamotsu Sugai, Shigeki Sekine, Megumi Ishiguro, Atsuo Takashima, and Yukihide Kanemitsu. Histopathological atlas of desmoplastic reaction characterization in colorectal cancer. *Japanese Journal of Clinical Oncology*, 51(6):1004–1012, 2021.

Lieze Berben, Hans Wildiers, Lukas Marcelis, Asier Antoranz, Francesca Bosisio, Sigrid Hatse, and Giuseppe Floris. Computerised scoring protocol for identification and quantification of different immune cell populations in breast tumour regions by the use of qupath software. *Histopathology*, 77(1):79–91, 2020.

Ines P Nearchou, Hideki Ueno, Yoshiki Kajiwara, Kate Lillard, Satsuki Mochizuki, Kengo Takeuchi, David J Harrison, and Peter D Caie. Automated detection and classification of desmoplastic reaction at the colorectal tumour front using deep learning. *Cancers*, 13(7):1615, 2021.

Takashi Okuyama, Shinichi Sameshima, Emiko Takeshita, Takashi Mitsui, Takuji Noro, Yuko Ono, Tamaki Noie, Shinichi Ban, and Masatoshi Oya. Myxoid stroma is associated with postoperative relapse in patients with stage ii colon cancer. *BMC cancer*, 20(1):1–11, 2020.

Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *arXiv preprint arXiv:2004.11362*, 2020.

A Selection of Self-supervised Model

Table 3: Classification results of the different self-supervised approaches, as well as the supervised baseline on the Kather-19 and Kather-16 patch classification tasks. We present the results for different percentages of available data. The top results are highlighted in bold. We report the weighted F1 score.

	Kather-16 Labels fraction			Kather-19 Labels fraction		
Methods	10%	20%	50%	1%	2%	5%
Supervised[‡]	85.8**	86.5**	87.9**	**89.2**$^{+}$	**89.9**$^{+}$	**90.5**$^{+}$
SimCLR Chen et al. [2020b]	79.6**	78.9**	78.6**	76.9**	79.4**	80.7**
SupContrast Khosla et al. [2020]	60.8**	73.2**	80.8**	78.7**	81.6**	85.0**
MoCoV2 Chen et al. [2020a]	**88.5**	**90.2**	**91.1**	**89.9**	**90.3**	**90.6**

[‡] Model initialized with ImageNet pre-trained weights.
[+] $p \geq 0.05$; * $p < 0.05$; ** $p < 0.001$; unpaired t-test with respect to the top result.

To assess which self-supervised model we should use as the backbone for the UDA, we compare the performances of the SOTA self-supervised methods, including SimCLR Chen et al. [2020b], SupContrast Khosla et al. [2020], and MoCoV2 Chen et al. [2020a], as well as the standard supervised learning approach when facing different levels of data availability. The results are presented in Table 3. We report the performance of the single domain classification on K16 and K19. The supervised approach uses ImageNet pre-trained weights. The self-supervised baselines are trained from scratch. After self-supervised training, we freeze the weights, add a linear classifier on top, and train it until convergence. For SupContrast Khosla et al. [2020] we jointly train the representation and the classification as described in the original paper.

We find that MoCoV2 Chen et al. [2020a] outperforms the two other SOTA approaches. On K16, the model gains up to 10% in terms of the F1-score with respect to the other self-supervised baselines. In addition, MoCoV2 gives competitive results with the supervised baseline that is initialized with ImageNet weights. It shows that MoCoV2 is able to learn from unlabeled data to create generalizable feature spaces efficiently. This mainly comes from the combination of the momentum encoder and the access to a large number of negative samples. Hence, we choose to adapt MoCoV2 for our proposed UDA method.

B Patch Classification - t-SNE Projection

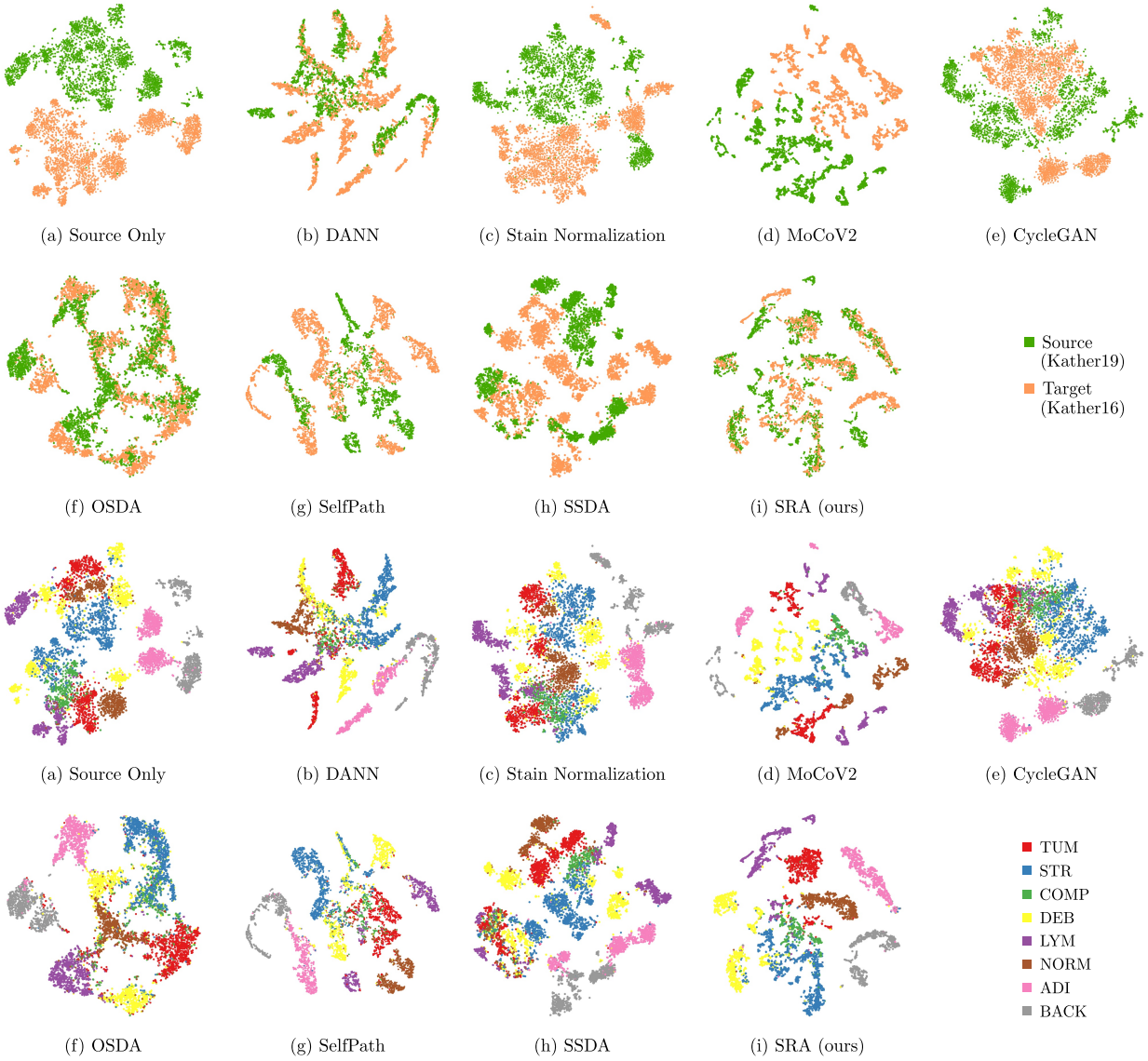
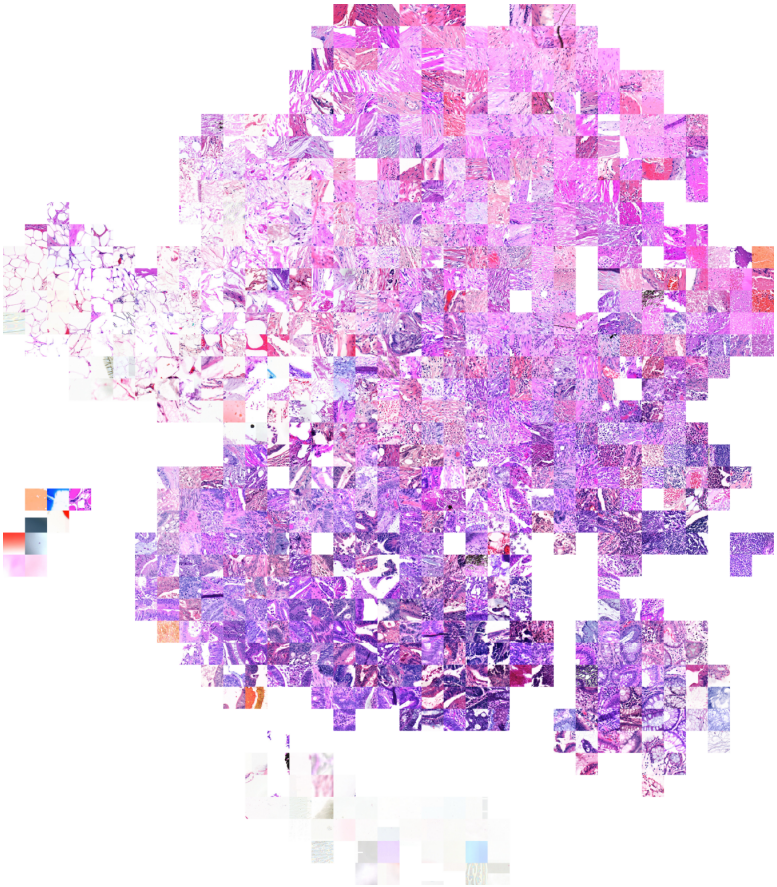


Figure 10: t-SNE projection of the source (Kather-19) and target (Kather-16) domain embeddings. We show the alignment of the embedding space for all presented models between the source and target domain, and the classes. The classes of Kather-19 are merged and relabeled according to the Kather-16 definition. The standard supervised approach is depicted in (a). We compared our approach (i) to other domain adaptation methods (b-h). Our approach (i) qualitatively shows the best alignment between the source and target domain.

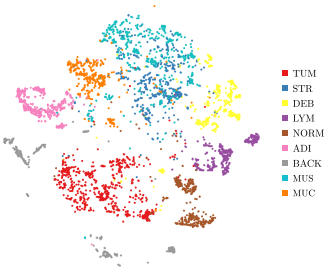
In this section, we display the complementary results to the ones presented in Section 4.2. The embedding for all baselines and the presented approach are displayed in Figure 10. We show the alignment between the source (K19) and target (K16) embedding domain-wise, as well as classes-wise.

C Multi-source - t-SNE Projection

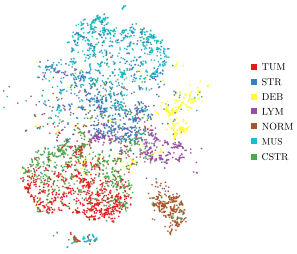
Figure 11 shows the visualization of the embedding for the proposed multi-source domain adaptation in Section 4.4. It highlights the alignment of the feature space between the two source sets (K19, CRC-TP) and our in-house dataset.



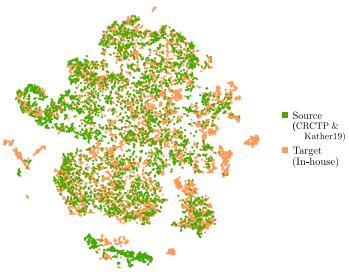
(a) t-SNE visualization of target patches distribution



(b) t-SNE Kather19 labeled source sample projection



(c) t-SNE CRCTP labeled source sample projection



(d) t-SNE source and target projection

Figure 11: t-SNE visualization of the SRA model trained on CRC-TP, K19 and in-house data. All sub-figures depict the same embedding. (a) Patch-based visualization of the embedding. (b) Distribution of the labeled source samples. (c) Relative alignment of the source and target domain samples.