

CellTranspose: Few-shot Domain Adaptation for Cellular Instance Segmentation

Matthew R. Keaton Ram J. Zaveri Gianfranco Doretto
West Virginia University
Morgantown, WV 26506
{mrkeaton, rz0012, gidoretto}@mix.wvu.edu

Abstract

Automated cellular instance segmentation is a process utilized for accelerating biological research for the past two decades, and recent advancements have produced higher quality results with less effort from the biologist. Most current endeavors focus on completely cutting the researcher out of the picture by generating highly generalized models. However, these models invariably fail when faced with novel data, distributed differently than the ones used for training. Rather than approaching the problem with methods that presume the availability of large amounts of target data and computing power for retraining, in this work we address the even greater challenge of designing an approach that requires minimal amounts of new annotated data as well as training time. We do so by designing specialized contrastive losses that leverage the few annotated samples very efficiently. A large set of results show that 3 to 5 annotations lead to models with accuracy that: 1) significantly mitigate the covariate shift effects; 2) matches or surpasses other adaptation methods; 3) even approaches methods that have been fully retrained on the target distribution. The adaptation training is only a few minutes, paving a path towards a balance between model performance, computing requirements and expert-level annotation needs.

1. Introduction

Automating the analysis of scientific imaging data via computer vision techniques is becoming increasingly convincing as our field matures. In order to accelerate scientific discovery, neural network-based methods have recently been developed to automatically segment and count individual instances of cells in laboratory-produced imaging data [44, 51, 14]. Such type of data acquisitions exhibit a remarkable variabil-

ity, which is due to the large variety of imaging modalities being used, the different types of tissues and how they are processed.

Current approaches for the specific task of cell instance segmentation are mainly based on supervised learning. They are trained on large datasets in an attempt to compensate for the diversity of the new data they are meant to be used on. However, new data to process will very likely not be distributed in the same way as the data used for training the models, so, they will perform the task, often with disappointing accuracy. To address this covariate shift problem [42] the obvious solution is retraining the models, which is costly and time consuming because it requires manual annotation of large amounts of the new *target* data we are seeking to automatically process. An alternative is to use domain adaptation methods, which attempt to adapt the model to the target data distribution. Current domain adaptation methods for segmentation are by and large tailored to imaging modalities, or specific tasks or applications that are very different than cell instance segmentation [15, 53]. Some promising work has addressed the problem in an unsupervised manner [26, 27]. But these approaches assume that a large fraction and amount of target data is available to undergo a relatively intense training to adapt the model.

In this work, we advocate for a more practical and scalable solution to address the need to generalize well out of distribution. We assume that a model for segmenting instances, such as cell bodies, membranes, or nuclei, has already been trained on a *source* dataset. Then, by annotating only a handful of samples of the *target* dataset we *adapt* the model, with a low training budget, to generalize well on the new distribution. We introduce *CellTranspose*, a new approach that implements the paradigm just described, for the few-shot supervised adaption of cell instance segmentation. The approach builds on a state-of-the-art model, and we introduce new losses and a training procedure for the

quick adaptation of the model to new data.

Our framework allows for a broad range of data to be properly segmented beyond the capabilities of current generalist approaches. We show that only a small number of annotations on the target dataset are required for the model to learn to produce high-fidelity segmentations and demonstrate this both on 2-D and 3-D data. In particular, few annotated samples are sufficient also to reach adaptation levels comparable to the unsupervised adaptation models. Additionally, CellTranspose affords a much faster training scheme as compared to training a model with similar accuracy from scratch.

2. Related Work

2.1. Cellular Instance Segmentation

The current top approach for cell instance segmentation is Cellpose [44]. Like many recent segmentation algorithms, Cellpose’s underlying model is a variant of U-Net [38], outputting a pixel-wise mask prediction. By itself, this only produces a semantic segmentation: the output mask merely determines the class of each pixel (foreground/cell or background), making it impossible to delineate between individual cells when they are clustered together. Deriving inspiration from the traditional watershed algorithm [3] and a gradient-based deep pose recognition algorithm called OpenPose [5], Cellpose contains two additional outputs corresponding to predicted gradients towards each pixel’s associated cell source location, one in the x -direction and one in the y -direction. Through an iterative process, the gradient of each pixel is followed to neighboring pixels until a “source” pixel representing a cell center is encountered. Each pixel directly linked to the source is then considered a part of the cell instance, in this way constructing segmentations of individual cells.

Other cellular segmentation methods have been proposed recently. Similar to Cellpose, Mesmer [14] represents another variant of “deep watershed” algorithms, also producing a pixel-wise mask prediction, but instead of the two flow outputs it generates an “inner distance transform” to predict the distance of each pixel to a cell’s center. StarDist [40] and StarDist-3D [51] predict object centers and then approximate the distance from the given location to cell boundaries at fixed angles from the center. The points defined by these predictions are then connected to each neighboring point to produce the outline of the cell’s mask. Other approaches include NuSeT [57], which also builds on the watershed algorithm, and DenoiSeg [4], which utilizes a joint learning strategy in concert with a denoising task to produce better accuracy on noisy samples.

2.2. Domain Adaptation for Medical Imaging

A common issue in deep learning practice is the scarcity of labeled data for many tasks and the often unrealistic requirement to produce vast amounts of costly annotations in order to appropriately train a model. Although many efforts have been made to produce “generalist” models which are invariant across different datasets for a specific task, doing so is intractable in real-world applications where data has intrinsically high variability. Domain adaptation allows for taking a high-performing model trained on a large dataset and adapting it to work on new data representing some target domain. This enables the model to take advantage of learned low-level features acquired from the larger source dataset while tuning to the specific features of the target dataset, requiring less target data.

Due to this lower dependence on manual annotation from experts, a vast amount of research on domain adaptation-based approaches has been unleashed on the medical imaging field [15, 53]. In addition to a few domain adaptation approaches on semantic segmentation for various imaging modalities including brain tumors [10, 47, 12], whole tissue [65], and organelles [2], several unsupervised mechanisms for semantic segmentation [55, 22, 19, 49, 36, 9, 1, 11, 21, 20, 63, 62, 7, 54, 56, 37, 41, 35, 23] and the more challenging instance segmentation task [26, 27, 58, 24] have been proposed. Generally, these unsupervised methods aim to learn representations of a given target domain without the use of any annotations. This greatly reduces the need for effort from medical experts while producing better segmentations than non-adapted models. However, this is accompanied by a clear trade-off with model accuracy, and the training scheme, in addition to more complex model architectures, causing training time to increase heavily. There is also an intrinsic expectation for the shape and relative size of segmentations from the target data to match closely with those of the source dataset, constraining such approaches to nuclei segmentation rather than more heterogeneous full-cell segmentations. These methods are also notably not “generalist,” and tend to focus on and perform best on a particular cell or imaging type.

2.3. Few-shot Domain Adaptation

In order to accommodate many real-world settings, one set of techniques known as “few-shot” domain adaptation aims to balance the demand for annotations with overall model performance. Since at least one large dataset is often available for a given task, many of the weaknesses posed by few-shot learning on its own can be avoided. Some popular techniques for addressing domain adaptation with a limited number of

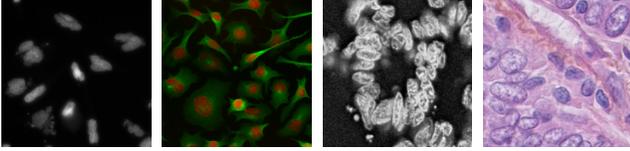


Figure 1. **Variability of microscopy data.** Image samples highlighting the variability of cell images. From left to right: Human U2OS cells with Hoechst and phalloidin stains from BBBC006 [29]; Neuroblastoma cells labelled with phalloidin and DAPI stains from the Cell Image Library [60]; GI tissue cells imaged by co-detection by indexing (CODEX) from Tissuenet [14]; Breast cancer cells, using hematoxylin and eosin (H&E) stain from TNBC [34].

target samples includes utilizing adversarial techniques to increase the confusion between the source and target domain [32], using meta-learning approaches to use tasks from one domain for adversarially learning tasks in a smaller, novel domain [39], and generating prototypes based on learned embeddings from the few target samples [64], a technique common to few-shot learning.

One which has proven effective across multiple paradigms is the use of contrastive losses [16]. A contrastive loss function encourages each sample from the target domain to be represented similarly to one or more samples of the same class (positive samples) in the source domain while simultaneously pushing its representation away from samples from other classes (negative samples) within the source domain. This approach has proven successful in a number of scenarios including domain adaptation [25, 50, 28], unsupervised learning [48, 66, 30, 45, 18, 6], self-supervised learning [8, 59, 31], meta-learning [39], and even few-shot domain adaptation [33].

3. Problem Definition

We are interested in the segmentation of instances such as cell bodies, membranes, or nuclei which are present in cell images. Such image structures have a remarkable variability because they are acquired with a variety of microscopy techniques, come from a diverse range of tissues, and present very different shapes. See Figure 1. Therefore, even state-of-the-art supervised instance-based segmentation approaches [44, 14], which claim to offer a *generalist* solution, unfortunately experience rapid performance degradation when facing the challenge of data distribution shift [42]. In essence, when the so-called generalist model is tasked to operate on data distributed differently than the data it was trained on, expectations are unmet.

Given how frequent the need to generalize out-of-distribution is for the task at hand, and given how costly and time consuming, if not completely impractical, the process of collecting and annotating sufficient

target data for retraining the model can be, we reframe the problem as one of *few-shot* learning. The intent is to significantly widen the range of image variability handled by current generalist solutions. Specifically, given a *source* dataset \mathcal{D}^s made of labeled images with which a generalist model could be trained, the learning task is to *adapt* such a model to generalize well on a *target* dataset \mathcal{D}^t , which is distributed differently than \mathcal{D}^s . To perform such adaptation, the user is required to label only a *minimal* amount of target data, let us say K data instances for a K -shot learning. In Section 4.2.3 we clarify the meaning of a single shot.

Note that simple fine-tuning of a generalist model on very few target samples is not a feasible solution due to the obvious overfitting challenges [13], which is why we introduce a method that aims at a very low budget in terms of the amount of target data to annotate, as well as the time for adaptation training. Given the specific setting of our problem, to the best of our knowledge, this is the first work to propose a method for few-shot supervised domain adaptation for the task of cell instance segmentation.

4. Approach

Our approach, shown in Figure 2, assumes that some large labeled source dataset \mathcal{D}^s can be used to initially train the model. However, regardless of the variety of cell images within \mathcal{D}^s , there will always conceivably be new data samples outside of the source distribution. This is due to limitations in both model and dataset size. We also assume that given some out-of-distribution set of target samples, annotation is costly, and therefore only a few samples will be labeled. In the following subsections, we first describe the pretrained method we rely on, and then introduce the proposed few-shot supervised adaptation model.

4.1. Pretrained Model

To provide a robust starting point, we use the state-of-the-art cell instance segmentation approach Cellpose [44], which we summarize here to introduce notation and make the paper self-contained. Given an image I , the model uses a network f to produce a dense, pixel-wise feature $\mathbf{Z} = f(I)$, where $\mathbf{Z} = [Z_1, Z_2, Z_3] \in \mathbb{R}^{h \times w \times 3}$. Given the feature $\mathbf{z} = [z_1, z_2, z_3] \in \mathbf{Z}$ for some pixel i , then $\mathbf{z} \doteq (z_1, z_2)$ has the meaning of gradient pointing towards the center of the cell structure to which pixel i belongs. Collectively, (Z_1, Z_2) form a *gradient flow*. Instead, $z \doteq z_3$ represents the unnormalized score indicating the probability of pixel i to belong to a cell structure. Note that with this notation, the feature \mathbf{z} can be written as $\mathbf{z} = [z, z]$. The network f is trained in a supervised manner with pixel-wise instance

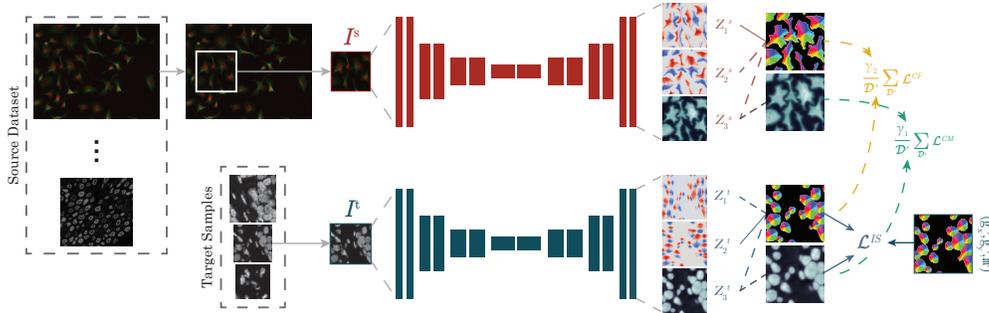


Figure 2. **Architecture approach.** Comprehensive illustration of our contrastive learning-based few-shot cellular instance segmentation approach.

segmentation loss

$$\mathcal{L}_i^{IS} = (z_1 - \mathbf{g}_x)^2 + (z_2 - \mathbf{g}_y)^2 + \nu H(\mathbf{m}, \sigma(z)), \quad (1)$$

where, for pixel i , $(\mathbf{g}_x, \mathbf{g}_y)$ represents the ground-truth gradient label with unit ℓ_2 -norm, $\mathbf{m} \in \{0, 1\}$ is the binary mask label indicating absence/presence of a cell structure, $\sigma(z) \doteq 1/(1 + \exp(-z))$, H represents the binary cross-entropy, and ν is a hyperparameter set to 0.04. The pixel-wise loss contributions are then aggregated into a final loss $\mathcal{L}^{IS} = \sum \mathcal{L}_i^{IS}$ for the image I .

Given the feature \mathbf{Z} , the cell instance segmentation head g produces the mask $Y = g(\mathbf{Z})$, where for pixel i , the predicted label y is a number in the set $\{0, 1, \dots, N\}$, with N being the total number of cell instances segmented. Pixels belonging to the same cell instance have same label, and $y = 0$ indicates absence of a cell instance. For details on how g is implemented the reader can consult [44]. See also Figure 2.

4.2. Adaptation Model

Consider now a target image $I \in \mathcal{D}^t$. If I was drawn from the same distribution from which source images composing \mathcal{D}^s were drawn, we would expect the following to be true: For a target pixel i with label $(\mathbf{g}_x^t, \mathbf{g}_y^t, \mathbf{m}^t)$, its feature \mathbf{z}^t should be very close to the features of pixels in the source dataset that have the same label. However, in presence of a *domain shift* this is generally untrue, which leads to performance deterioration of the cell instance segmentation process, when we use the feature network f on I , followed by the instance segmentation head g . Therefore, provided that the labels of some target image pixels are available, we design a method for reversing the effects of domain shift, by adapting f to generalize well on the target dataset \mathcal{D}^t . Given the distinct predictive nature (continuous vs. discrete) of the gradient flow features $\mathbf{z} = (z_1, z_2)$, from the mask feature z , we proceed by designing adaptation losses for each case.

4.2.1 Contrastive Flow Loss

In order to align \mathbf{z}^t with the gradient flow features of source pixels with same label, we set up a contrastive prediction task [8]. We identify a *positive* source pixel with binary label $\mathbf{m}_+^s = \mathbf{m}^t = 1$, and with gradient flow features \mathbf{z}_+^s that best matches the label $(\mathbf{g}_x^t, \mathbf{g}_y^t)$, according to a similarity measure, like cosine similarity $s(\mathbf{u}, \mathbf{v}) \doteq \mathbf{u}^\top \mathbf{v} / \|\mathbf{u}\| \|\mathbf{v}\|$, where $\|\cdot\|$ denotes ℓ_2 -norm. Then, we compose a set of *negative* source gradient flow features $\mathcal{N}_i = \{\mathbf{z}_-^s \mid s(\mathbf{z}_+^s, \mathbf{z}_-^s) < \delta, \mathbf{m}_-^s = 1\}$, where δ is a suitable constant hyperparameter threshold that we choose. Now we can use a contrastive loss function for pixel i that attempts to pull the positive pair $(\mathbf{z}^t, \mathbf{z}_+^s)$ closer, while pushing apart every negative pair $(\mathbf{z}^t, \mathbf{z}_-^s)$, for $\mathbf{z}_-^s \in \mathcal{N}_i$. The loss function we used, which we name *contrastive flow loss*, is

$$\mathcal{L}_i^{CF} = -\log \frac{\exp(s(\mathbf{z}^t, \mathbf{z}_+^s)/\tau)}{\exp(s(\mathbf{z}^t, \mathbf{z}_+^s)/\tau) + \sum_{\mathbf{z}_-^s \in \mathcal{N}_i} \exp(s(\mathbf{z}^t, \mathbf{z}_-^s)/\tau)} \quad (2)$$

where τ denotes a temperature parameter. Note that (2) addresses only the directional alignment of the gradient flow features. This is also what matters most, because the segmentation head g uses only that information for assigning pixels to instances, not the gradient magnitude. However, in Section 4.2.3 we will see that (1) is also applied on target data, which does encourage gradients to have unit magnitude.

An important component of the loss is how we mine the negative source features. Specifically, \mathcal{N}_i is composed by selecting the features closest to the positive source feature \mathbf{z}_+^s . This hard-mining strategy removes the need to consider large amounts of negative source features, because it makes them less informative. This leads to faster training, and usually more rapid convergence and better model performance, as it was shown also in [48]. Moreover, for a given target pixel i we extract one positive feature, and all the $|\mathcal{N}_i|$ negative features from one source image. We set $|\mathcal{N}_i|$ to be the same for every target pixel, and generally, there

is a sufficient number of negative features such that the similarity between z_+^s and each negative feature is approximately equal to δ . Therefore, the \mathcal{N}_i features are roughly equally divided between those at $\cos^{-1} \delta$ radians clockwise and those at $\cos^{-1} \delta$ radians counter-clockwise from z_+^s . Thus, when z^t is off, the negative samples nearer to z^t will produce a greater gradient than those further away.

Given a target image patch and a source image patch in the minibatch, for this pair we compute a loss contribution by aggregating all the components coming from the target pixels that belong to cell structures, so they have label $m^t = 1$. Let us indicate with \mathcal{M} this set of pixels, then the loss for the given pair will be

$$\mathcal{L}^{CF} = \frac{1}{|\mathcal{M}|} \sum_{\mathcal{M}} \mathcal{L}_i^{CF}. \quad (3)$$

4.2.2 Contrastive Mask Loss

The contrastive flow loss (2) is a specialized version of a loss used in [43, 46, 8, 48], and when $\tau \rightarrow +\infty$ it converges to

$$\lim_{\tau \rightarrow +\infty} \mathcal{L}_i^{CF} = -s(z^t, z_+^s) + \lambda \sum_{z_-^s \in \mathcal{N}_i} s(z^t, z_-^s), \quad (4)$$

where λ is a hyperparameter. The loss (4) was originally proposed in [17], and a specialized form of it was introduced in [33] for supervised domain adaptation and generalization for multi-class visual classification.

Since we are interested in aligning the unnormalized binary classification score z^t with the scores of the source pixels with same label, we derive a contrastive loss for segmentation, inspired by (4) and [33]. This is also motivated by the fact that losses like (2) have been proven effective with cosine similarity, which we cannot use for this task and we need to replace.

When (4) is optimized, the first term aims at maximizing similarity. In our case this would be the similarity between z^t and the unnormalized score of the pixel of a source image z_+^s with label $m^s = m^t$. We can replace that term with an alignment term based on a distance, leading to a squared loss

$$d(z^t, z_+^s) = \frac{1}{2}(z^t - z_+^s)^2. \quad (5)$$

The second term of (4), instead, aims at minimizing similarity. In our case this would be the similarity between z^t and the unnormalized score of pixels in source images with label different than m^t . This would lead to separation between scores with opposite labels. Given one such pixel with score z_-^s , we can measure the similarity with z^t with

$$k(z^t, z_-^s) = \frac{1}{2} \max(0, m - |z^t - z_-^s|)^2, \quad (6)$$

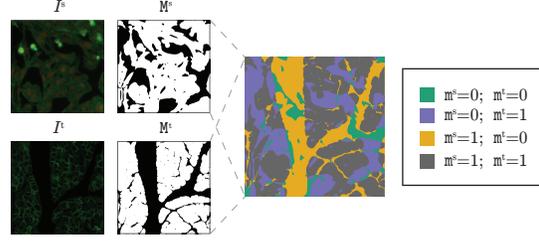


Figure 3. **Contrastive Mask Loss.** Representation of the Contrastive Mask Loss for a given source-target sample patch pair. The mask of a source and target sample are overlapped to compare representations on a per-pixel basis. Pixels belonging to the same class are encouraged to be represented similarly while pixels from different classes have their representations separated.

where m is a margin, and the loss provides gradient contributions when z^t is within the margin m of z_-^s .

Given a target image patch and a source image patch in a minibatch, we aggregate the losses (5) and (6) as follows. Let \mathcal{P} be the set of positive pairs of scores (z^t, z_+^s) corresponding to pixels in the same relative positions in the target and source patches, and with same label, i.e., $m^s = m^t$; and let \mathcal{N} be the set of negative pairs of scores (z^t, z_-^s) corresponding to pixels in the same relative positions in the target and source patches but with different label, i.e., $m^s \neq m^t$. Then, the loss for a pair of target and source patches is

$$\mathcal{L}^{CM} = \frac{1}{|\mathcal{P}|} \sum_{\mathcal{P}} d(z^t, z_+^s) + \lambda \frac{1}{|\mathcal{N}|} \sum_{\mathcal{N}} k(z^t, z_-^s), \quad (7)$$

and we refer to this as the *contrastive mask loss*. Note that each term of the loss is normalized with respect to the area covered by the positive and negative pairs, respectively. Also, the formation of the sets \mathcal{P} and \mathcal{N} is based on the comparison of the mask labels of the target and source patches, as in Figure 3, and the pairs originate from pixels with same relative position. This is not a strict requirement, but it is convenient because it allows forming sufficiently large sets \mathcal{P} and \mathcal{N} , while implementing (7) is easier and faster since it takes advantage of the parallel architecture of GPUs.

4.2.3 Few-shot Adaptation

Assuming that \mathcal{D}_K^t is the target dataset portion of \mathcal{D}^t with K labeled samples, the *K-shot adaptation learning* aims at minimizing the loss

$$\mathcal{L}^{ISA} = \sum_{\mathcal{D}_K^t} \left(\mathcal{L}^{IS} + \frac{\gamma_1}{|\mathcal{D}^s|} \sum_{\mathcal{D}^s} \mathcal{L}^{CM} + \frac{\gamma_2}{|\mathcal{D}^s|} \sum_{\mathcal{D}^s} \mathcal{L}^{CF} \right). \quad (8)$$

The training assumes that the generalist model has already been pretrained. The Cellpose model we used

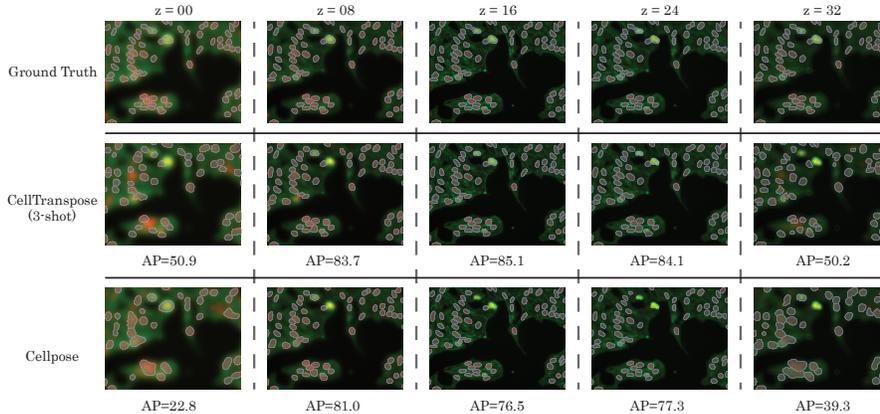


Figure 4. **BBBC006 dataset**. Qualitative results on an image from the BBBC006 dataset at different levels of focus. $z=16$ is considered in-focus, while values further from this are more out of focus in either direction. Segmentations are shown in gray with cell borders shown in blue. AP values are computed at 0.5 IoU.

was pretrained for 500 epochs [44]. The adaptation with (8) lasts 5 epochs where source images are continuously randomly paired with one of the K target samples without replacement. Pairing the target samples with $1/K$ -th of \mathcal{D}^s in every epoch ensures that even a 1-shot adaptation can operate a significant pull of the model towards the target distribution. The adapted model is then fine-tuned on \mathcal{D}_K^t with \mathcal{L}^{IS} only for 5 more epochs at a very low constant learning rate.

Since this is the first attempt at few-shot domain adaptation for cellular instance segmentation, we need to define what constitutes one “shot”. A target exemplar cell is selected manually such that its size and the relative density of cells nearby is representative of the average scenario across the target dataset. If $K > 1$ cells are selected, they should better capture the variability of the dataset, and will lead to the K -shot scenario. For a given cell sample the cell is measured, giving the value m_c in pixel units. The one shot patch is center-cropped around the cell and has size $\beta_{max} m_c w / m_n$, where m_n is the nominal cell size, w is the size of the patch passed into the network, and β_{max} is the largest scaling factor admissible. Data augmentation including translation and random cropping with factor 0.75 to $\beta_{max} = 1.25$ is part of the training to promote invariance to cell size.

5. Experiments

We evaluate the proposed approach, which we name *CellTranspose*, with several datasets and multiple settings. In absence of additional directions, the following setup is used for each experiment. The model in Section 4.1 is pretrained with the “generalized” dataset of [44] as source data \mathcal{D}^s . Adaptation is done as described in Section 4.2.3. In general, we follow the data splitting guidelines laid out by previous works to en-

sure fair comparisons. In order to fit our approach to these guidelines, we draw our K sample patches, to produce \mathcal{D}_K^t , from the training split of the dataset of the target distribution. Instead, the testing split of the target distribution dataset is used as our remaining, unlabeled portion of the target dataset \mathcal{D}^t . We use SGD with initial learning rate 10^{-2} , momentum 0.9, weight decay 10^{-5} , and batch size of 2 provides optimal results. For the first five epochs, the learning rate decreases by a factor of 10 each epoch, and is kept constant for the remaining five. We take square patches with side length $h = w = 112$, use a minimum overlap of 84 during evaluation, and enforce a nominal cell size $m_n = 30$. Additional hyperparameters are set as: $|\mathcal{N}_i| = 20$, $\tau = 0.1$, $m = 10$, $\lambda = 1$, $\gamma_1 = 0.05$, $\gamma_2 = 2$, and $\delta = 0.05$. Because the source dataset used is always significantly larger than the target data, adaptation takes roughly the same amount of time regardless of the size of K or the target data to which the model is adapted. Using a singular NVIDIA TITAN Xp GPU, adaptation takes approximately 5 minutes to complete for each experiment.

Broad Bioimage Benchmark Collection – 006.

We evaluated CellTranspose on the target dataset BBBC006 [29], hosted by the Broad Institute. This dataset is composed of human U2OS cells which are fairly homogeneous and easy to segment in ideal settings. However, the same tissue samples have been imaged with different focus settings, generating different images, which allow us to observe the effect of the associated covariate shift on generalist and our approaches. We consider images taken with five focal planes, specifically at $z=00, 08, 16, 24, \text{ and } 32$. The optimal focal plane is at $z=16$. Figure 4 shows qualitative segmentation results. As we move away from the optimal focal plane the generalist Cellpose model exposes greater performance deterioration than CellTranspose.

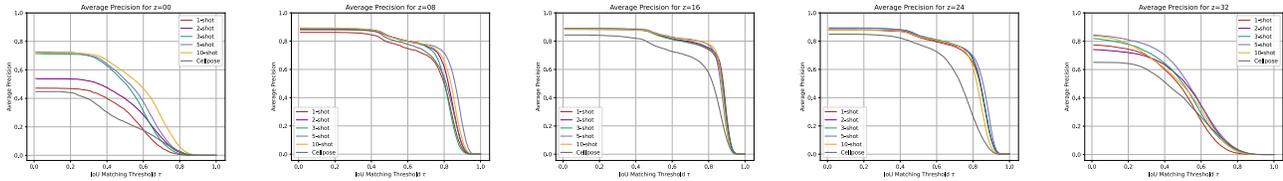


Figure 5. **BBBC006 dataset.** Average precision of K -shot adaptation with $K = 1, 2, 3, 5, 10$ target samples. Experiments were completed using five BBBC006 datasets.

Table 1. **TissueNet dataset.** F-1 score from state-of-the-art “generalist” approaches and CellTranspose.

TissueNet Results	Platform-specific				Tissue-specific			
	CODEX	CyCIF	MIBI	Vectra	Breast	GI	Imm.	Panc.
Cellpose	0.785	0.548	0.479	0.609	0.670	0.523	0.350	0.797
Mesmer	0.88	0.80	0.76	0.72	0.74	0.82	0.82	0.92
CellTranspose	0.940	0.940	0.932	0.918	0.911	0.906	0.934	0.955
Cellpose-UB	0.962	0.967	0.945	0.960	0.950	0.940	0.953	0.958

We also test CellTranspose with different numbers $K = 1, 2, 3, 5, 10$, of few-shot adaptation. Figure 5 plots the average precision (AP) against the intersection over union (IoU) for the different focal settings. CellTranspose consistently achieves high performance levels with as few as three annotated shots. Beyond this, results begin to exhibit diminished returns. So, unless otherwise specified, other results have been obtained with a 3-shot adaptation, given the balance it gives between performance and annotation needs.

TissueNet. We evaluated CellTranspose on TissueNet [14], a dataset developed alongside a generalist method called Mesmer. TissueNet is comprised of samples from various imaging platforms and tissue types, providing a wide-spanning array of cellular images. In [14] one set of experiments split TissueNet into subsets of the four most common imaging types, each of which was further divided into different tissue types. Similarly, four other subsets were composed of the four most common tissue types, each being further split into the imaging types that make up the samples for that tissue type. Table 1 shows the F-1 scores on these eight data splits, computed for the generalist Cellpose, Mesmer trained on each of the splits, and CellTranspose 3-shot adapted to each of the splits. Again, CellTranspose consistently shows improved performance. Since CellTranspose essentially adapts Cellpose to the target domain, we can interpret it as a *lower bound* of performance. On the other hand, we also re-trained Cellpose with each target training dataset, obtaining what could be interpreted as an *upper bound* for CellTranspose, and we indicated that as *Cellpose-UB*.

Triple Negative Breast Cancer. Among the most challenging types of cellular data to segment is that of hematoxylin and eosin-stained (H&E) images. This is in part due to the fact that multiple cell types often appear within an individual sample, in addition to the high variability of the background. The Triple Nega-

Table 2. **TNBC dataset.** Comparison between the top unsupervised approach and CellTranspose. Best results are in bold, and second best results are underlined.

BBBC039 → TNBC	AJI	Pixel-F1	Object-F1
Cellpose	0.3815 ± 0.0794	0.5829 ± 0.0689	0.5408 ± 0.1124
CyC-PDAM [26]	0.5672 ± 0.0646	0.7593 ± 0.0566	0.7478 ± 0.0417
CellTranspose 3-shot	0.4916 ± 0.0771	0.6702 ± 0.0710	0.7092 ± 0.0818
CellTranspose 5-shot	<u>0.5693 ± 0.0576</u>	0.7377 ± 0.0431	<u>0.7825 ± 0.0625</u>
CellTranspose 10-shot	0.5906 ± 0.0617	<u>0.7568 ± 0.0493</u>	0.7879 ± 0.0687
Cellpose-UB	0.5498 ± 0.0860	0.7216 ± 0.0704	0.7760 ± 0.0640

tive Breast Cancer (TNBC) dataset [34], gathered by the Curie Institute, is comprised of 50 images obtained from 11 distinct tissue types, furthering the inherent difficulty of the dataset.

We compare CellTranspose with the top-performing unsupervised domain adaptive cellular instance segmentation approach, CyC-PDAM [26]. This is the closest approach to ours we found in terms of problem settings and data, since the only other supervised domain adaptive instance segmentation approaches [61, 52] were tuned to very specific and different applications. Following the lead of CyC-PDAM, we divert from the protocol used above and pretrain CellTranspose on the BBBC039 dataset. We then adapt the model with 3, 5, and 10 shot samples selected from the target dataset of 40 images and 8 tissue types from TNBC. Each model is then tested on the 10 images from the remaining 3 tissue types. Results, using the same metrics as CyC-PDAM, are shown in Table 2. It can be noted that a 5-shot adaptation leads to performance metrics comparable with those of CyC-PDAM, that has used all the training target data available to adapt the model in an unsupervised manner. Note that the settings of this experiment is still different and disadvantageous from the protocol for which CellTranspose was designed, because the target testing data distribution is different from the target training data distribution. We also tested Cellpose-UB in this settings, which appears to be affected more than CellTranspose by the domain shift still present in the evaluation protocol.

3D Segmentation. As biological applications often produce and require the analysis of 3-D data, following the approach in [44], we extend CellTranspose to operate in 3-D, named *CellTranspose3D*, by making adaptation and combining the predictions along the xy , yz , and zx volume sections. We test our approach on two

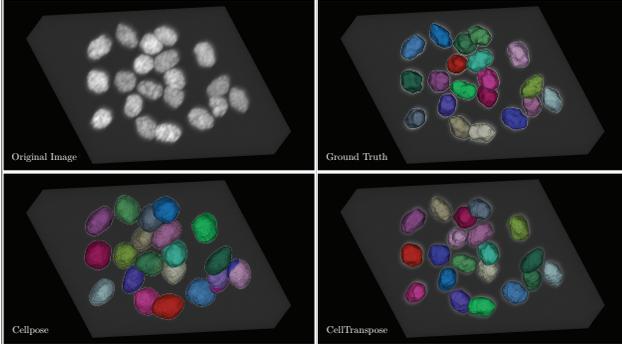


Figure 6. **3D Segmentation.** Qualitative segmentation results on samples from the BBBC024 dataset. Note that Cellpose3D tends to oversample; this is likely due to the fact that the particular characteristics of the data Cellpose3D was trained on necessitated pixels of even lower intensities to be segmented. More importantly, on the bottom right of the results obtained with Cellpose3D, it is possible to observe some over-segmentation effects, which are not present in the results of CellTranspose3D.

Table 3. **3D datasets.** Average Precision at IoU threshold 0.5.

3-D Results	Worm	BBBC024
Cellpose3D	0.575	0.822
CellTranspose3D	0.648	0.994
Cellpose3D-UB	0.675	1.0
StarDist-3D	0.765	—

3-D datasets. The BBBC024 is composed of synthetic annotated cells from the Broad Institute. The Worm dataset is a series of nuclei images from larval stage *C. elegans*. Similar to the 2-D experiments, only three 2-D sample patches, taken from the XY-plane from volumes in each training set, are used as target data for the model.

StarDist-3D [51] and Cellpose3D-UB have been trained and tested on Worm, and Cellpose3D-UB on BBBC024 also, and serve here as upper bounds. Table 3 shows the average precision results, highlighting the improvement of CellTranspose3D over the generalist Cellpose3D. Figure 6 provides a qualitative comparison between 3D segmentations of the BBBC024 dataset.

Ablation Study. Table 4 shows ablation results computed on the BBBC006 dataset with $z=00$ and 3-shot adaptation. The addition of both contrastive losses improves the overall AP by more than 15%. Interestingly, however, the removal of only one adaptation loss tends to decrease the performance to below that of removing both, which follows the same training scheme outlined in Section 4.2.3, but without either adaptation loss. This indicates that the flow and mask losses are intrinsically tied to one another, which is consistent with the fact that weights are shared between both outputs

Table 4. **Ablation results on BBBC006.** AP is calculated for an IoU threshold of 0.5.

Ablation Results	AP_{50}
CellTranspose	0.509
No Contrastive Flow Loss	0.414
No Contrastive Mask Loss	0.434
No Adaptation Losses	0.441
No Adaptation Losses & No cell size	0.390
Cellpose	0.233

until the final layer.

Additionally, the cell size calculation method seems to play an important role in accurate segmentation. Cellpose computes the cell size based upon the number of pixels corresponding to a particular cell, but non-spherical cells have the potential to provide a similar diameter while spanning a much larger area. Thus, a more robust cell size calculation could be based on computing the total rectangular area enclosing a cell, which is how it is done in CellTranspose. The second row from bottom in Table 4 shows that when this second strategy is replaced with the one used by Cellpose, performance further deteriorates.

6. Conclusions

In this work, we have exemplified the need for adaptation of cell instance segmentation methods when working in new domains. Our approach is effective at achieving this with very few samples, thanks to the proposed new losses and training procedures we implemented. CellTranspose shows a level of performance that effectively tackles the effects of the covariate shift, and that can be comparable to models fully retrained on the target distribution. We have also shown that few samples are sufficient to reach performance similar to unsupervised approaches that rely on larger datasets and training simulations. We found that 3 to 5 annotated samples usually allowed to match or surpass the state of the art. The adaptation procedure takes a few minutes on a single GPU, striking an attractive balance in terms of time and resources used for training and annotation.

Acknowledgements

Research reported in this publication was supported by the National Institute Of Mental Health of the National Institutes of Health under Award Number R44MH125238. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This material is also based upon work supported by the National Science Foundation under Grant No. 1920920.

References

- [1] Mathilde Bateson, Hoel Kervadec, Jose Dolz, Hervé Lombaert, and Ismail Ben Ayed. Constrained domain adaptation for segmentation. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, pages 326–334. Springer International Publishing, 2019.
- [2] Róger Bermúdez-Chacón, Pablo Márquez-Neila, Mathieu Salzmann, and Pascal Fua. A domain-adaptive two-stream U-Net for electron microscopy image segmentation. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 400–404, Apr. 2018.
- [3] S. Beucher. Use of watersheds in contour detection. *Proceedings of the International Workshop on Image Processing*, 1979.
- [4] Tim-Oliver Buchholz, Mangal Prakash, Deborah Schmidt, Alexander Krull, and Florian Jug. DenoiSeg: Joint denoising and segmentation. In *Computer Vision – ECCV 2020 Workshops*, pages 324–337. Springer International Publishing, 2020.
- [5] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.
- [6] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Un-supervised learning of visual features by contrasting cluster assignments. *Adv. Neural Inf. Process. Syst.*, 33:9912–9924, 2020.
- [7] Cheng Chen, Qi Dou, Hao Chen, Jing Qin, and Pheng-Ann Heng. Synergistic image and feature adaptation: Towards Cross-Modality domain adaptation for medical image segmentation. *AAAI*, 33(01):865–872, July 2019.
- [8] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In Hal Daumé Iii and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 1597–1607. PMLR, 2020.
- [9] Qi Dou, Cheng Ouyang, Cheng Chen, Hao Chen, Ben Glocker, Xiahai Zhuang, and Pheng-Ann Heng. PnP-AdaNet: Plug-and-Play adversarial domain adaptation network at unpaired Cross-Modality cardiac segmentation. *IEEE Access*, 7:99065–99076, 2019.
- [10] Mohsen Ghafoorian, Alireza Mehrtash, Tina Kapur, Nico Karssemeijer, Elena Marchiori, Mehran Pesteie, Charles R G Guttman, Frank-Erik de Leeuw, Clare M Tempany, Bram van Ginneken, Andriy Fedorov, Purang Abolmaesumi, Bram Platel, and William M Wells. Transfer learning for domain adaptation in MRI: Application in brain lesion segmentation. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2017*, pages 516–524. Springer International Publishing, 2017.
- [11] Amir Gholami, Shashank Subramanian, Varun Shenoy, Naveen Himthani, Xiangyu Yue, Sicheng Zhao, Peter Jin, George Biros, and Kurt Keutzer. A novel domain adaptation framework for medical image segmentation. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 289–298. Springer International Publishing, 2019.
- [12] Michael Goetz, Christian Weber, Franciszek Binczyk, Joanna Polanska, Rafal Tarnawski, Barbara Bobek-Billewicz, Ullrich Koethe, Jens Kleesiek, Bram Stieltjes, and Klaus H Maier-Hein. DALSA: Domain adaptation for supervised learning from sparsely annotated MR images. *IEEE Trans. Med. Imaging*, 35(1):184–196, Jan. 2016.
- [13] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.
- [14] Noah F Greenwald, Geneva Miller, Erick Moen, Alex Kong, Adam Kagel, Thomas Dougherty, Christine Camacho Fullaway, Brianna J McIntosh, Ke Xuan Leow, Morgan Sarah Schwartz, Cole Pavelchek, Sunny Cui, Isabella Camplisson, Omer Bar-Tal, Jaiveer Singh, Mara Fong, Gautam Chaudhry, Zion Abraham, Jackson Moseley, Shiri Warshawsky, Erin Soon, Shirley Greenbaum, Tyler Risom, Travis Hollmann, Sean C Bendall, Leeat Keren, William Graf, Michael Angelo, and David Van Valen. Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nat. Biotechnol.*, Nov. 2021.
- [15] Hao Guan and Mingxia Liu. Domain adaptation for medical image analysis: A survey. *IEEE Trans. Biomed. Eng.*, 69(3):1173–1185, Mar. 2022.
- [16] R Hadsell, S Chopra, and Y LeCun. Dimensionality reduction by learning an invariant mapping. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’06)*, volume 2, pages 1735–1742, June 2006.
- [17] R Hadsell, S Chopra, and Y LeCun. Dimensionality reduction by learning an invariant mapping. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’06)*, volume 2, pages 1735–1742, June 2006.
- [18] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2020.
- [19] Mehran Javanmardi and Tolga Tasdizen. Domain adaptation for biomedical image segmentation using adversarial training. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 554–558, Apr. 2018.
- [20] Jue Jiang, Yu-Chi Hu, Neelam Tyagi, Pengpeng Zhang, Andreas Rimmer, Gig S Mageras, Joseph O

- Deasy, and Harini Veeraraghavan. Tumor-Aware, adversarial domain adaptation from CT to MRI for lung cancer segmentation. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, pages 777–785. Springer International Publishing, 2018.
- [21] Jiang Jue, Hu Jason, Tyagi Neelam, Rimner Andreas, Berry L Sean, Deasy O Joseph, and Veeraraghavan Harini. Integrating cross-modality hallucinated MRI with CT to aid mediastinal lung tumor segmentation. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, pages 221–229. Springer International Publishing, 2019.
- [22] Konstantinos Kamnitsas, Christian Baumgartner, Christian Ledig, Virginia Newcombe, Joanna Simpson, Andrew Kane, David Menon, Aditya Nori, Antonio Criminisi, Daniel Rueckert, and Ben Glocker. Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. In *Information Processing in Medical Imaging*, pages 597–609. Springer International Publishing, 2017.
- [23] Neerav Karani, Krishna Chaitanya, Christian Baumgartner, and Ender Konukoglu. A lifelong learning approach to brain MR segmentation across scanners and protocols. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, pages 476–484. Springer International Publishing, 2018.
- [24] Chaoqun Li, Yitian Zhou, Tangqi Shi, Yenan Wu, Meng Yang, and Zhongyu Li. Unsupervised domain adaptation for the histopathological cell segmentation through Self-Ensembling. In Manfredo Atzori, Nikolay Burlutskiy, Francesco Ciompi, Zhang Li, Fayyaz Minhas, Henning Müller, Tingying Peng, Nasir Rajpoot, Ben Torben-Nielsen, Jeroen van der Laak, Mitko Veta, Yinyin Yuan, and Inti Zlobec, editors, *Proceedings of the MICCAI Workshop on Computational Pathology*, volume 156 of *Proceedings of Machine Learning Research*, pages 151–158. PMLR, Sept. 2021.
- [25] Shuang Li, Binhui Xie, Bin Zang, Chi Harold Liu, Xinjing Cheng, Ruigang Yang, and Guoren Wang. Semantic distribution-aware contrastive adaptation for semantic segmentation. *CoRR*, abs/2105.05013, May 2021.
- [26] Dongnan Liu, Donghao Zhang, Yang Song, Fan Zhang, Lauren O’Donnell, Heng Huang, Mei Chen, and Weidong Cai. Unsupervised instance segmentation in microscopy images via panoptic domain adaptation and task re-weighting. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2020.
- [27] Dongnan Liu, Donghao Zhang, Yang Song, Fan Zhang, Lauren O’Donnell, Heng Huang, Mei Chen, and Weidong Cai. PDAM: A Panoptic-Level feature alignment framework for unsupervised domain adaptive instance segmentation in microscopy images. *IEEE Trans. Med. Imaging*, 40(1):154–165, Jan. 2021.
- [28] Weizhe Liu, David Ferstl, Samuel Schuler, Lukas Zebedin, Pascal Fua, and Christian Leistner. Domain adaptation for semantic segmentation via Patch-Wise contrastive learning. *CoRR*, abs/2104.11056, Apr. 2021.
- [29] Vebjorn Ljosa, Katherine L Sokolnicki, and Anne E Carpenter. Annotated high-throughput microscopy image sets for validation. *Nat. Methods*, 9(7):637, June 2012.
- [30] Luke Melas-Kyriazi and Arjun K Manrai. PixMatch: Unsupervised domain adaptation via pixelwise consistency training. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2021.
- [31] Ishan Misra and Laurens van der Maaten. Self-supervised learning of pretext-invariant representations. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2020.
- [32] Saeid Motiian, Quinn Jones, Seyed Mehdi Iranmanesh, and Gianfranco Doretto. Few-shot adversarial domain adaptation. In *Adv. Neural Inf. Process. Syst. (NIPS)*. papers.nips.cc, 2017.
- [33] Saeid Motiian, Marco Piccirilli, Donald A Adjeroh, and Gianfranco Doretto. Unified deep supervised domain adaptation and generalization. In *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE, Oct. 2017.
- [34] Peter Naylor, Marick Lae, Fabien Reyat, and Thomas Walter. Segmentation of nuclei in histopathology images by deep regression of the distance map. *IEEE Trans. Med. Imaging*, 38(2):448–459, Feb. 2019.
- [35] Mauricio Orbes-Arteaga, Thomas Varsavsky, Carole H Sudre, Zach Eaton-Rosen, Lewis J Haddow, Lauge Sørensen, Mads Nielsen, Akshay Pai, Sébastien Ourselin, Marc Modat, Parashkev Nachev, and M Jorge Cardoso. Multi-domain adaptation in brain MRI through paired consistency and adversarial learning. In *Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data*, pages 54–62. Springer International Publishing, 2019.
- [36] Egor Panfilov, Aleksei Tiulpin, Stefan Klein, Miika T Nieminen, and Simo Saarakkala. Improving robustness of deep learning based knee MRI segmentation: Mixup and adversarial domain adaptation. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*. IEEE, Oct. 2019.
- [37] Christian S Perone, Pedro Ballester, Rodrigo C Barros, and Julien Cohen-Adad. Unsupervised domain adaptation for medical imaging segmentation with self-ensembling. *Neuroimage*, 194:1–11, July 2019.
- [38] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241. Springer International Publishing, 2015.

- [39] Doyen Sahoo, Hung Le, Chenghao Liu, and Steven C H Hoi. Meta-Learning with domain adaptation for Few-Shot learning under domain shift. Sept. 2018.
- [40] Uwe Schmidt, Martin Weigert, Coleman Broaddus, and Gene Myers. Cell detection with star-convex polygons. In *Medical Image Computing and Computer Assisted Intervention - MICCAI 2018 - 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II*, pages 265–273, 2018.
- [41] Zahil Shanis, Samuel Gerber, Mingchen Gao, and Andinet Enquobahrie. Intramodality domain adaptation using self ensembling and adversarial training. In *Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data*, pages 28–36. Springer International Publishing, 2019.
- [42] Hidetoshi Shimodaira. Improving predictive inference under covariate shift by weighting the log-likelihood function. *J. Stat. Plan. Inference*, 90(2):227–244, Oct. 2000.
- [43] Kihyuk Sohn. Improved deep metric learning with multi-class n-pair loss objective. In *Proceedings of the 30th International Conference on Neural Information Processing Systems, NIPS’16*, pages 1857–1865, Red Hook, NY, USA, Dec. 2016. Curran Associates Inc.
- [44] Carsen Stringer, Tim Wang, Michalis Michaelos, and Marius Pachitariu. Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods*, 18(1):100–106, Jan. 2021.
- [45] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv e-prints*, page arXiv:1807.03748, July 2018.
- [46] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *CoRR*, abs/1807.03748, July 2018.
- [47] Annegreet van Opbroek, Meike W Vernooij, M Arfan Ikram, and Marleen de Bruijne. Weighting training images by maximizing distribution similarity for supervised segmentation across scanners. *Med. Image Anal.*, 24(1):245–254, Aug. 2015.
- [48] Feng Wang and Huaping Liu. Understanding the behaviour of contrastive loss. *CoRR*, abs/2012.09740, Dec. 2020.
- [49] Shujun Wang, Lequan Yu, Xin Yang, Chi-Wing Fu, and Pheng-Ann Heng. Patch-Based output space adversarial learning for joint optic disc and cup segmentation. *IEEE Trans. Med. Imaging*, 38(11):2485–2495, Nov. 2019.
- [50] Xinlong Wang, Rufeng Zhang, Chunhua Shen, Tao Kong, and Lei Li. Dense contrastive learning for self-supervised visual pre-training. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2021.
- [51] Martin Weigert, Uwe Schmidt, Robert Haase, Ko Sugawara, and Gene Myers. Star-convex polyhedra for 3d object detection and segmentation in microscopy. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, March 2020.
- [52] Wenqiang Xu, Yonglu Li, and Cewu Lu. SRDA: Generating instance segmentation annotation via scanning, reasoning and domain adaptation. In *ECCV*, pages 124–140. 2018.
- [53] Siyi Xun, Dengwang Li, Hui Zhu, Min Chen, Jianbo Wang, Jie Li, Meirong Chen, Bing Wu, Hua Zhang, Xi-angfei Chai, Zekun Jiang, Yan Zhang, and Pu Huang. Generative adversarial networks in medical image segmentation: A review. *Comput. Biol. Med.*, 140:105063, Nov. 2021.
- [54] Wenjun Yan, Yuanyuan Wang, Shengjia Gu, Lu Huang, Fuhua Yan, Liming Xia, and Qian Tao. The domain shift problem of medical image segmentation and Vendor-Adaptation by Unet-GAN. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, pages 623–631. Springer International Publishing, 2019.
- [55] Wenjun Yan, Yuanyuan Wang, Menghua Xia, and Qian Tao. Edge-Guided output adaptor: Highly efficient adaptation module for Cross-Vendor medical image segmentation. *IEEE Signal Process. Lett.*, 26(11):1593–1597, Nov. 2019.
- [56] Junlin Yang, Nicha C Dvornek, Fan Zhang, Julius Chapiro, Mingde Lin, and James S Duncan. Unsupervised domain adaptation via disentangled representations: Application to Cross-Modality liver segmentation. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, pages 255–263. Springer International Publishing, 2019.
- [57] Linfeng Yang, Rajarshi P Ghosh, J Matthew Franklin, Simon Chen, Chenyu You, Raja R Narayan, Marc L Melcher, and Jan T Liphardt. NuSeT: A deep learning tool for reliably separating and analyzing crowded cells. *PLoS Comput. Biol.*, 16(9):e1008193, Sept. 2020.
- [58] Siqi Yang, Jun Zhang, Junzhou Huang, Brian C Lovell, and Xiao Han. Minimizing labeling cost for nuclei instance segmentation and classification with cross-domain images and weak labels. *AAAI*, 35(1):697–705, May 2021.
- [59] Chun-Hsiao Yeh, Cheng-Yao Hong, Yen-Chi Hsu, Tyng-Luh Liu, Yubei Chen, and Yann LeCun. Decoupled contrastive learning. *CoRR*, abs/2110.06848, Oct. 2021.
- [60] Weimiao Yu, Hwee Kuan Lee, Srivats Hariharan, Wen Yu Bu, and Sohail Ahmed. Ccdb:6843, mus musculus, neuroblastoma. CIL. Dataset. <https://doi.org/doi:10.7295/W9CCDB6843>.
- [61] Hui Zhang, Yonglin Tian, Kunfeng Wang, Haibo He, and Fei-Yue Wang. Synthetic-to-Real domain adaptation for object instance segmentation. In *IJCNN*, pages 1–7, 2019.
- [62] Tianyang Zhang, Jun Cheng, Huazhu Fu, Zaiwang Gu, Yuting Xiao, Kang Zhou, Shenghua Gao, Rui Zheng, and Jiang Liu. Noise adaptation generative adversarial

- network for medical image analysis. *IEEE Trans. Med. Imaging*, 39(4):1149–1159, Apr. 2020.
- [63] Yue Zhang, Shun Miao, Tommaso Mansi, and Rui Liao. Task driven generative modeling for unsupervised domain adaptation: Application to x-ray image segmentation. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, pages 599–607. Springer International Publishing, 2018.
- [64] An Zhao, Mingyu Ding, Zhiwu Lu, Tao Xiang, Yulei Niu, Jiechao Guan, and Ji-Rong Wen. Domain-adaptive few-shot learning. In *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, Jan. 2021.
- [65] Qikui Zhu, Bo Du, and Pingkun Yan. Boundary-Weighted domain adaptive neural network for prostate MR image segmentation. *IEEE Trans. Med. Imaging*, 39(3):753–763, Mar. 2020.
- [66] Chengxu Zhuang, Alex Zhai, and Daniel Yamins. Local aggregation for unsupervised learning of visual embeddings. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, Oct. 2019.