ADAPTATION OF A MULTI-SITE NETWORK TO A NEW CLINICAL SITE VIA BATCH-NORMALIZATION SIMILARITY

Shira Kasten Serlin[⋆] *Jacob Goldberger*[†] *Hayit Greenspan*[⋆]

[⋆] Department of Biomedical Engineering, Tel Aviv University, Tel Aviv, Israel
[†] Faculty of Engineering, Bar-Ilan University, Ramat-Gan, Israel

ABSTRACT

This paper tackles the challenging problem of medical site adaptation; i.e., learning a model from multi-site source data such that it can be modified and adapted to a new site using only unlabeled data from the new site. The method is based on Domain Specific Batch Normalization architecture and uses the Batch Normalization statistics of the new site to find the most similar internal site. The similarity measure is computed in an embedded space of the BN parameters. We evaluated our method on the task of MRI prostate segmentation. Public datasets from six different institutions were used, containing distribution shifts. The experimental results show that the proposed approach outperforms other generalization and adaptation methods.

Index Terms— multi-site, batch-normalization, domain adaptation, prostate segmentation

1. INTRODUCTION

Deep learning methods have produced remarkable achievements in many medical imaging analysis tasks. The ability of deep learning solutions to generalize across multiple sites is critical for clinical adoption of these methods. Due to the scarcity of medical images, it is important to effectively aggregate data from multiple sites to ensure robust model training. However, medical imaging data, especially MRI, usually present heterogeneities and can have highly variable intensity characteristics across different sites due to the differences in scanners and imaging protocols. Fig. 1 shows the diversity we can find among prostate MRI images from six different medical sites [1].

The clinical deployment of existing models to new medical sites suffers from performance degradation due to distribution shifts across different clinical sites. It is not practical to train algorithms using data from all imaging equipment sources at all possible sites. Learning a robust and accurate model that can be deployed in new sites by taking advantage of multiple sources of data is clearly a non-trivial challenging and practically important task.

This research was supported by the Ministry of Science & Technology, Israel.

Tackling non-iid data with feature shifts has been explored in the general machine learning context. There are two main solution types for the domain shift problem: Domain Generalization focuses on learning a site invariant model that is robust to unseen sites (see e.g. [1, 2, 3, 4]); Domain Adaptation tries to adapt the multi-site model to a new site given examples (labeled or unlabeled). In this case better performance on the given site can be achieved than when using a single site invariant universal model (see e.g. [5, 6, 7]). In most cases, assuming the availability of labeled data on each installation of the system in a new medical site is not realistic.

Recent work has proposed Batch Normalization (BN) [8] as a tool to mitigate domain shifts via Domain Specific Batch Normalization (DSBN) network architecture that shares most of the parameters across the sites and assigns individual BN layers to each domain site [4, 9, 10, 11]. Finding the best way to implement domain generalization and domain adaptation within the setup of DSBN networks is an active research topic.

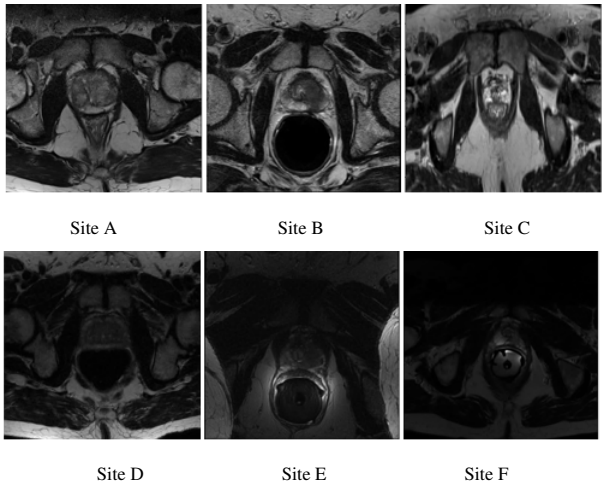


Fig. 1: Prostate MRI example images from six different sites; site details appear in Table 1.

In this study we tackle the problem of site adaptation given unlabeled data from the new site. The adaptation

method is based on the DSBN architecture and uses Batch Normalization statistics of images from the new site to find the most similar internal site. The similarity measure is computed in an embedded space of the BN parameters. We evaluated our method on a segmentation task of prostate MRI data from six different institutions with distribution shifts acquired from public datasets. The experimental results show that our approach outperforms other generalization and adaptation methods across almost all six settings of unseen domains.

2. SITE ADAPTATION BASED ON BN SIMILARITY

In this section we present a site adaptation method for a multi-site network where batch normalization is involved in both the multi-site architecture and in the adaptation to a new site. Batch normalization [8] has been widely used in CNNs for reducing internal covariate shift, and therefore helps to improve feature discrimination capability and to speed up the learning process. The central idea is to normalize the internal representations along the channel dimension, and then apply affine transformation on the whitened feature maps with trainable parameters. Let x be a certain channel in a certain layer of the multi-channel feature maps. The corresponding normalized representation y is computed as:

$$y = \gamma \cdot \hat{x} + \beta \quad \text{where} \quad \hat{x} = \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}}.$$

μ and σ^2 are the empirical mean and variance of x that are computed on mini-batches during training and ϵ is an infinitesimal number. The parameters γ and β are learned in the training phase. A moving average of the BN statistics is also collected in the training phase to capture the global statistics, and is used in the BN layer for inference.

In case of multi-site data, rather than learning an independent CNN for each source site, we can share most of the parameters among the source sites. Domain Specific Batch Normalization (DSBN) [4, 9, 10, 11] is a network architecture that shares most of the parameters across the domains (sites) and assigns individual BN layers to each domain (site) to effectively deal with the inter-site discrepancy. The DSBN network scheme is shown in Fig. 2. Compared to using the same BN layers across different sites, the DSBN utilizes domain-specific variables to handle domain-specific nuances and task-irrelevant inter-site variations by performing individual feature normalization. Specifically, the DSBN layer assigns to each site s (in each layer and filter) domain-specific trainable parameters γ_s and β_s and trainable data statistics μ_s and σ_s. Let x_s be a certain channel of the feature maps of a sample from site s. The corresponding normalized representation is expressed as:

$$y_s = \gamma_s \cdot \hat{x}_s + \beta_s \quad \text{where} \quad \hat{x}_s = \frac{x_s - \mu_s}{\sqrt{\sigma_s^2 + \epsilon}}.$$

We use the data from all the given sites to train the network, and the site-specific BN parameters are trained using data from the corresponding site. In the testing phase, a DSBN layer applies the learned domain-specific BN parameters for the data from the corresponding site.

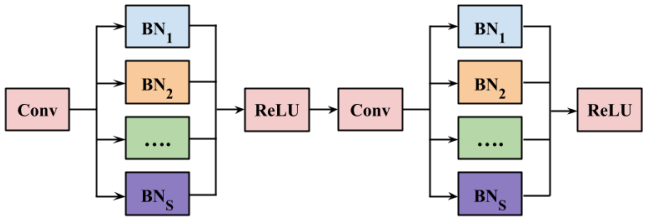


Fig. 2: A scheme of a DSBN network architecture.

Next, we address the problem of generalization of the DSBN model to a new site that was not available during the model learning. The design of DSBN limits the generalization ability since the corresponding statistic information and the trainable variables of the new site are unknown. Liu et al. [4] suggested averaging the site-based predictions on a new site image. Alternatively, we can average the trained BN parameters to form a site-invariant network that can be applied to a new site [11].

A very challenging and real-world scenario is one in which images from a new site are introduced to the system, with no pre-annotated training set provided. In the current work we propose to shift from a site-invariant approach to a site-adaptation methodology. To accomplish this, we search for the internal site that is the most similar to the new site such that applying the BN parameters of the internal site to images of the new site yields the best results. If manually annotated images from the new site are available, finding the best internal site is easy. Here we aim to find the most similar internal site without any annotation.

A simple way to assess the similarity of the new site to each one of the internal sites is to use a site classification network that classifies each image to one of the internal sites and is trained on images from the internal sites [12]. However, this similarity measure is not necessarily correlated with the task at hand, as seen in the next section where it indeed yields inferior adaptation results.

In the site adaptation process, we want to select the most suitable set of BN parameters for the new site. Hence, it makes sense to carry out this decision based on the BN statistics of the new site images. Next, we define a site similarity measure based on the BN statistics of the images. We use the trained DSBN model and average the trained BN parameters to form a site-invariant model with a single set of BN parameters for all the multi-site training data. For each image in the training set we can compute the BN statistics of the mean and variance of each filter in each layer using the filter responses of all the patches. We can also apply the same procedure to

each image of the new site. We thus obtain an image representation by concatenating its BN statistics. In this study, we focus on a segmentation task and use a Residual-Unet architecture [13] that is comprised of an encoder and decoder. We use the mean values of the BN layer input at the encoder's last layer of the site-invariant model as we found that it contains the traits of the data domain. In the experiments section we validate that this is indeed the case. We next extract a low dimension embedding by applying the t-SNE algorithm [14] on all the images' BN representations. We first apply PCA on the original representation to decrease the dimensionality from 512 to 100 and then apply the t-SNE algorithm to obtain a 2D representation. In the 2D space we compute the site centers by calculating the means for observations assigned to each site and finally calculate the Euclidean distance from the new site center to each one of the internal site centers. The internal site with the nearest center is considered as the most similar to the new site and its trained BN parameters are used to adapt the model to the new site. The algorithm is summarized in Algorithm Box 1.

Algorithm 1 BN based model adaptation of DSBN

input: Images from internal sites, DSBN model trained on the internal sites and unlabeled images from the new site.
- Average the trained BN parameters of the DSBN model to form a single site-invariant model.
- Extract the BN parameters for each image.
- Embed the BN feature vector in a low-dimensional space.
- Compute the average embedded BN parameters for each internal site and for the new site.
- Find the internal site s that is the closest to the average parameters of the new site.

output: A DSBN model where the BN parameters are taken from site s.

3. EXPERIMENTAL RESULTS

In this section we demonstrate the performance of our approach and compare to other methods for multi-site model adaptation.

Multi-site dataset. We evaluated the model's performance on a publicly available multi-site dataset for prostate MRI segmentation which contains prostate T2-weighted MRI data (with segmentation masks) collected from six different data sources with a distribution shift from public datasets (Table 1 summarizes their sample numbers and scanning protocols). Samples of sites A and B are from the NCI-ISBI13 dataset [15], samples of site C are from the I2CVB dataset [16], and samples of sites D, E and F are from PROMISE12 dataset [17]. Note that the NCI-ISBI13 and PROMISE12 actually include multiple data sources, and hence were decomposed by [1, 4]. For pre-processing,

Table 1: Details of data and imaging protocols from six different sites [1].

Site	# Cases	Field strength (T)	Resolution (mm) (in/through plane)	Endorectal Coil	Company
A - RUNMC	30	3	0.6-0.625/3.6-4	Surface	Siemens
B - BMC	30	1.5	0.4/3	Endorectal	Philips
C - HCRUDB	19	3	0.67-0.79/1.25	No	Siemens
D - UCL	13	1.5,3	0.325-0.625/3-3.6	No	Siemens
E - BIDMC	12	3	0.25/2.2-3	Endorectal	GE
F - HK	12	1.5	0.625/3.6	Endorectal	Siemens

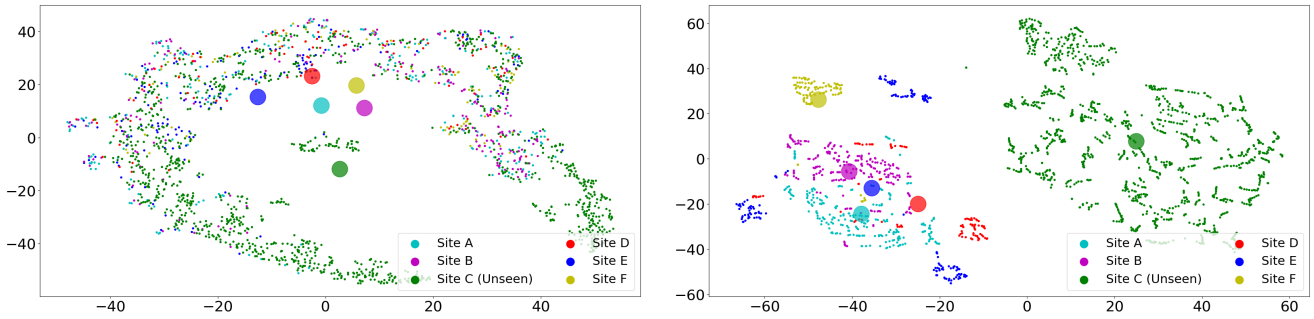
we normalized each sample to have zero mean and unit variance in intensity value before inputting to the network.

Implementation Details. We implemented an adapted Residual-UNet [13] as the segmentation network backbone, as was used in [4] and achieved remarkable performance in the prostate segmentation problem. Due to the large variance of slice thicknesses among different sites, we employed the 2D architecture which contains 4 down-sampling and 4 up-sampling blocks. Our framework was implemented in Python with PyTorch. We first trained the network by mixing multi-site data with regular BN layers, and afterwards all the BN layers in the encoder were converted to DSBN layers to handle the inter-site discrepancy. The pretrained BN parameters of the existing network were utilized as initialization of the DSBN parameters. For each training iteration, we fed the network with 5 batches of images, where each batch came from one internal site dataset. In post-processing, we conducted morphological operations in 3D to select the largest connective volume as the final segmentation mask.

Compared Methods. We compared the following methods for applying a multi-site DSBN model to a new site:

- **Mean Predictions** - averaging the single-site predictions of the multi-site network [4].
- **Mean BN** - a network obtained by averaging the multi-site BN parameters [11].
- **Mean BN weighted** - a weighted average of the BN parameters based on a site classification network that was trained on the internal sites [12]. The weights are the classifier results averaged on all the images from the new site.
- **Mean BN locally weighted** - similar to the Mean BN weighted approach, with weights computed separately per patient rather than for all the new site images together.
- **BN embedding** - our method which is based on site similarity in an embedded space of BN parameters of the encoder's last layer.

The Dice scores of the segmentation results are shown in Table 2. In each row, the site in the first column is the new site and the DSBN model was trained on the other five sites. It shows that predictions averaging and model averaging yielded the worst results, which means that adaptation to a new site



(a) Embedding of the BN first layer

(b) Embedding of the BN last layer

Fig. 3: BN parameters embedding in the first (a) and last (b) layers of the Residual-Unet encoder. The images of each site are shown in different colors and the large circles are the site means.

is better than a site invariant model. Of the model adaptation methods, the proposed embedding based approach yielded the best results across almost all the six experiments.

Table 2: Comparison of Dice score results on a new site.

New site	Mean predictions	Mean BN	Mean BN weighted	Mean BN locally weighted	BN embedding
Site A	85.39	86.22	87.94	87.95	**88.00**
Site B	78.72	78.27	79.03	79.00	**79.80**
Site C	72.98	73.26	73.99	74.02	**74.14**
Site D	79.65	81.25	82.55	82.58	**82.87**
Site E	72.10	71.82	72.72	73.70	**77.17**
Site F	84.58	**85.13**	84.93	85.03	85.04
Average	78.90	79.33	80.19	80.38	**81.17**

We next examined whether our method indeed selected the optimal internal site for each new site. Table 3 shows the segmentation results for a new site using the BN parameters of each of the other five internal sites. It is clear that selecting which internal site is used for adaptation significantly influence the segmentation performance on the new site. This underscores the importance of choosing the proper internal site for the adaptation process. The internal site whose BN parameters yielded the best Dice score on the new site is shown in bold. The internal site chosen by our algorithm is underlined. In most cases we chose the optimal internal site. In the case of new sites D and F we chose the second best internal site which was slightly worse than the best one.

We next illustrate the selection of the last layer of the encoder as the relevant representation of each site. Fig. 3 shows that the deep BN layer of the encoder is influenced by the domain shift and contains the traits of the data domain. In this case dots from different sites are more clearly separated into clusters unlike the case of the first layer. The green dots are the new site in this case (site C) and the closest center is the internal site D (red), whose BN parameters yielded the best Dice score on the new site.

Table 3: Dice scores of a DSBN source model applied to a new site. The best source site is bolded and the source site that was found by our algorithm is underlined.

New \ Internal	Site A	Site B	Site C	Site D	Site E	Site F
New site A	-	84.31	84.08	**88.00**	80.73	86.27
New site B	**79.80**	-	74.71	79.62	74.66	78.87
New site C	73.42	68.04	-	**74.14**	71.90	73.03
New site D	82.87	78.49	76.99	-	72.59	**83.01**
New site E	73.61	62.22	63.49	**77.17**	-	73.21
New site F	**85.46**	83.99	84.45	85.04	84.05	-

4. CONCLUSIONS

In this study we proposed a procedure for multi-site model adaptation to a new site which finds the internal site that is the most similar to the new site. The major challenge is how to define site similarity that is related to the image analysis task we want to conduct. We focused on a multi-site architecture based on allocating different BN layers to each internal site. We showed that in this case a Euclidean distance in a non-linear embedding of the BN statistics of each image constitutes a good similarity measure. The method is general and can be incorporated in any situation where multi-site or multi-domain modeling needs to be adapted to a new domain. In future work we plan to explore the methodology proposed to additional tasks. Another future research direction would be to address privacy issues and federated learning protocols within the proposed algorithm.

5. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using human subject data made available in open access. Ethical approval was not required as confirmed by the license attached with the open access data.

6. REFERENCES

[1] Quande Liu, Qi Dou, and Pheng-Ann Heng, "Shape-aware meta-learning for generalizing prostate MRI segmentation to unseen domains," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2020.

[2] Eli Gibson, Yipeng Hu, Nooshin Ghavami, Hashim U Ahmed, Caroline Moore, Mark Emberton, Henkjan J Huisman, and Dean C Barratt, "Inter-site variability in prostate segmentation accuracy using deep learning," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2018.

[3] Chris Yoon, Ghassan Hamarneh, and Rafeef Garbi, "Generalizable feature learning in the presence of data bias and domain class imbalance with application to skin lesion classification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2019, pp. 365–373.

[4] Quande Liu, Qi Dou, Lequan Yu, and Pheng Ann Heng, "MS-net: multi-site network for improving prostate segmentation with heterogeneous MRI data," *IEEE Trans. on Medical Imaging*, vol. 39, no. 9, pp. 2713–2724, 2020.

[5] Yanghao Li, Naiyan Wang, Jianping Shi, Xiaodi Hou, and Jiaying Liu, "Adaptive batch normalization for practical domain adaptation," *Pattern Recognition*, vol. 80, pp. 109–117, 2018.

[6] Cheng Chen, Qi Dou, Hao Chen, Jing Qin, and Pheng Ann Heng, "Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation," *IEEE Trans. on Medical Imaging*, 2020.

[7] Konstantinos Kamnitsas, Christian Baumgartner, Christian Ledig, et al., "Unsupervised domain adaptation in brain lesion segmentation with adversarial networks," in *Information Processing in Medical Imaging (IPMI)*, 2017.

[8] Sergey Ioffe and Christian Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning (ICML)*, 2015.

[9] Woong-Gi Chang, Tackgeun You, Seonguk Seo, Suha Kwak, and Bohyung Han, "Domain-specific batch normalization for unsupervised domain adaptation," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[10] Seonguk Seo, Yumin Suh, Dongwan Kim, Geeho Kim, Jongwoo Han, and Bohyung Han, "Learning to optimize domain specific normalization for domain generalization," in *Proc. of the European Conference on Computer Vision (ECCV)*, 2020.

[11] Xiaoxiao Li, Meirui Jiang, Xiaofei Zhang, Michael Kamp, and Qi Dou, "Fedbn: Federated learning on non-iid features via local batch normalization," in *International Conference on Learning Representations (ICLR)*, 2021.

[12] Massimiliano Mancini, Samuel Rota Bulo, Barbara Caputo, and Elisa Ricci, "Robust place categorization with deep domain generalization," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2093–2100, 2018.

[13] Lequan Yu, Xin Yang, Hao Chen, Jing Qin, and Pheng Ann Heng, "Volumetric convnets with mixed residual connections for automated prostate segmentation from 3D MR images," in *AAAI*, 2017.

[14] Laurens Van der Maaten and Geoffrey Hinton, "Visualizing data using t-sne.," *Journal of Machine Learning Research*, vol. 9, no. 11, pp. 2579–2605, 2008.

[15] Nicholas Bloch, Anant Madabhushi, Henkjan Huisman, John Freymann, Justin Kirby, Michael Grauer, Andinet Enquobahrie, Carl Jaffe, Larry Clarke, and Keyvan Farahani, "NCI-ISBI 2013 challenge: automated segmentation of prostate structures," *The Cancer Imaging Archive*, vol. 370, pp. 6, 2015.

[16] Guillaume Lemaître, Robert Martí, Jordi Freixenet, Joan C Vilanova, Paul M Walker, and Fabrice Meriaudeau, "Computer-aided detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: a review," *CBM*, vol. 60, pp. 8–31, 2015.

[17] Geert Litjens, Robert Toth, Wendy van de Ven, Caroline Hoeks, Sjoerd Kerkstra, Bram van Ginneken, Graham Vincent, Gwenael Guillard, Neil Birbeck, Jindang Zhang, et al., "Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge," *MIA*, vol. 18, no. 2, pp. 359–373, 2014.