Image-to-image translation trained on unrelated histopathology data helps for Domain Generalization

Marin Scalbert[1,2] MARIN.SCALBERT@CENTRALESUPELEC.FR
[1] *MICS, CentraleSupélec - Université Paris-Saclay, Gif-sur-Yvette, France*
[2] *VitaDX International, Paris, France*

Maria Vakalopoulou[*1] MARIA.VAKALOPOULOU@CENTRALESUPELEC.FR
Florent Couzinié-Devy[2] F.COUZINIE-DEVY@VITADX

Abstract

Histopathology Whole Slide Images (WSIs) present large illumination or color variations due to protocol variability (scanner, staining). This can strongly harm the generalization performances of deep learning algorithms. To address this problem, we propose to train a multi-domain image-to-image translation (I2IT) model on WSIs from The Cancer Genome Atlas Program (TCGA) and use it for data augmentation. Using TCGA WSIs from different cancer types has several advantages: our data augmentation method can be used for tasks where data is small, the I2IT model does not need to be relearned for each task and the variability of TCGA protocols is high leading to better robustness. The method efficiency is assessed on the Camelyon17 WILDS dataset where we outperform sophisticated data augmentations and domain generalization methods. Results also confirms that training the I2IT model on unrelated histopathology data is much more efficient for generalization than training it on the training data of the domain generalization (DG) task.

Keywords: Multi-domain Image-to-image translation, Domain Generalization, Data augmentation, Histopathology

1. Introduction

Domain shift remains one of the most challenging problems in computational pathology. Indeed, there can be large discrepancies in appearances between WSIs from different hospitals, scanners, stains or patients making generalization of deep learning models on unseen data difficult. One of the most studied setting within this problem is domain generalization (DG). In DG, the goal is to learn a model on labeled data from source domains and to generalize well on data from unseen target domains. To tackle this problem in histopathology, a straightforward approach consists in aligning features by domain to obtain domain invariant features which can be achieved through classic DG methods such as (Shi et al., 2021) or H&E specific data augmentation (Tellez et al., 2018). In this work, we propose to train a StarGanV2 (Choi et al., 2020) on WSIs from TCGA to be able to stain any image into one of the stains of TCGA. The trained StarGanV2 is then used to perform data augmentation to enhance generalization performances on unseen data. Our method differs from other because we do not consider source training data to train the I2IT model but rather external WSIs from TCGA that may represent tissues from different cancer types than those in the DG task. It is evaluated on the Camelyon17 WILDS (Koh et al., 2021) DG task and outperforms conventional DG methods and H&E tailored data augmentations.

2. StarGanV2 based data augmentation: StyleAugMix

StarGanV2 is a multi-domain I2IT model made of a mapping network F and a generator G. F takes as input a random latent vector z and a target domain label d and outputs a style vector $s = F(z, d)$. Given an input image x and the style s, G translates x into $\tilde{x} = G(x, s)$ exhibiting same appearance as real images from domain d. StarGanV2 is also able to perform style vectors interpolation. Our data augmentation called StyleAugMix is performed by: (1) transforming geometrically (flips, rotations, translations, crop and resize) an input image x into two independent versions x_1 and x_2. (2) x_1 and x_2 are then translated independently into $\tilde{x_1}$ and $\tilde{x_2}$ using G and a random interpolated style vector made of 2 style vectors from random domains.

3. Experiments and Results

For StyleAugMix, StarGanV2 is trained either on Camelyon17 WILDS `train` split or on 1000 WSIs sampled randomly from TCGA. The two variants are respectively referred to as StyleAugMix Camelyon17/TCGA. For both, StarGanV2 is trained for 150000 steps, with a batch size of 16, Adam optimizer and learning rate of 10^{-4}. We also include Domain Invariant Perceptual loss term into the overall loss of StarGanV2. Some translated image examples from Camelyon WILDS are shown on Figure 1. To train the classifier on the

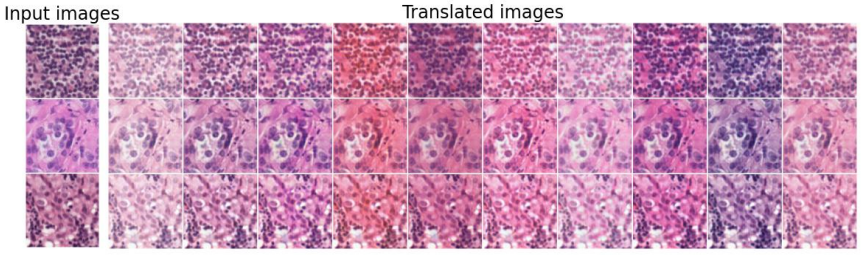

Figure 1: Input images from Camelyon17 WILDS translated by the trained StarGanV2 into one of the domains of TCGA.

Camelyon17 WILDS `train` split, we use the same setting as the official challenge along with our StyleAugMix data augmentation. The classifier is trained by minimizing cross entropies and a consistency regularization term (Jensen-Shannon divergence) between predictions on translated images $\tilde{x_1}$ and $\tilde{x_2}$. The classifier is evaluated on the remaining splits: `id_val` (source domains) and `val`/`test` (unseen domains). The performances are reported on Table 1, along with a baseline Deep-All, the best performing official DG method (FISH) and data augmentations (RandAugment, H&E color jitter). FISH reports better results on `val` and `test` than the baseline Deep-All. Data augmentation methods report much higher results than FISH. This is not surprising as both RandAugment and H&E color jitter use a prior information on the domain shift as part of their transformations (color jitter, stain variations). StyleAugMix TCGA, without using any prior information on the domain shift, performs similarly well on `val` and better on `test` than tailored data augmentations. The

Table 1: Comparison of performances for different methods on the Camelyon17 WILDS DG task. Performances are averaged over 10 independent runs.

Method	`id_val`	`val`	`test`
Deep-All	98.50(0.10)	80.30(4.40)	62.70(7.40)
FISH (Shi et al., 2021)		83.90(1.20)	74.70(7.10)
†RandAugment (Cubuk et al., 2020)		**90.60(1.20)**	82.00(7.40)
H&E color jitter (Tellez et al., 2018)		88.00(4.20)	**91.60(1.90)**
StyleAugMix Camelyon17	98.40(0.10)	89.60(0.70)	76.40(4.50)
StyleAugMix TCGA	97.77(0.26)	**89.68(1.12)**	**92.74(1.56)**

† Hyperparameters finetuned via random search on `val`.

most unexpected results of this study is that exploiting I2IT to transfer shifts present in unrelated histopathology data (StyleAugMix TCGA) usually results in much better generalization than I2IT model that transfer shifts present in the DG training data (StyleAugMix Camelyon17).

4. Conclusion

In this work, we have proposed to use an I2IT model trained on unrelated histopathology WSIs to perform data augmentation and obtain models robust to protocols variation. It has been evaluated on the histopathology Camelyon17 WILDS dataset and has shown superior results over previous state-of-the-art methods. Finally, this study has revealed that an I2IT model trained on other cancer types histopathology WSIs and use for data augmentation can be exploited to achieve strong generalization on out-of-distribution data.

References

Y. Choi, Y.g Uh, J. Yoo, and J.-W. Ha. Stargan v2: Diverse image synthesis for multiple domains. In *Proceedings of CVPR*, 2020.

E.D. Cubuk, B. Zoph, J. Shlens, and Q. V Le. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of CVPR Workshops*, 2020.

P. W. Koh, S. Sagawa, H. Marklund, S. M. Xie, M. Zhang, A. Balsubramani, W. Hu, et al. Wilds: A benchmark of in-the-wild distribution shifts. In *International Conference on Machine Learning*. PMLR, 2021.

Y. Shi, J. Seely, P. HS Torr, N Siddharth, A. Hannun, N. Usunier, and G. Synnaeve. Gradient matching for domain generalization. *arXiv preprint arXiv:2104.09937*, 2021.

D. Tellez, M. Balkenhol, I. Otte-Höller, R. van de Loo, et al. Whole-slide mitosis detection in h&e breast histology using phh3 as a reference to train distilled stain-invariant convolutional networks. *IEEE transactions on medical imaging*, 37(9), 2018.