Image-level Harmonization of Multi-Site Data using Image-and-Spatial Transformer Networks

R. Robinson[1], Q. Dou[1], D. C. Castro[1], K. Kamnitsas[1], M. de Groot[2], R.M. Summers[3], D. Rueckert[1], B. Glocker[1]

[1] BioMedIA, Department of Computing, Imperial College London, UK
[2] Research & Development, GlaxoSmithKline, UK
[3] Department of Radiology and Imaging Sciences, Clinical Center, National Institutes of Health, USA

Abstract. We investigate the use of image-and-spatial transformer networks (ISTNs) to tackle domain shift in multi-site medical imaging data. Commonly, domain adaptation (DA) is performed with little regard for explainability of the inter-domain transformation and is often conducted at the feature-level in the latent space. We employ ISTNs for DA at the image-level which constrains transformations to explainable appearance and shape changes. As proof-of-concept we demonstrate that ISTNs can be trained adversarially on a classification problem with simulated 2D data. For real-data validation, we construct two 3D brain MRI datasets from the Cam-CAN and UK Biobank studies to investigate domain shift due to acquisition and population differences. We show that age regression and sex classification models trained on ISTN output improve generalization when training on data from one and testing on the other site.

1 Introduction

Domain shift (DS) concerns the problem of mismatch between the statistics of the training data used for model development and the statistics of the test data seen after model deployment. DS can cause significant drops in predictive performance, which has been observed in almost all recent imaging challenges when final test data was coming from different clinical sites [1]. DS is a major hurdle for successfully translating predictive models into clinical routine.

Acquisition and population shift are two common forms of DS that appear in medical image analysis [2]. Acquisition shift is observed due to differences in imaging protocols, modalities or scanners. Such a shift will be observed even if the same subjects are scanned. Population shift occurs when cohorts of subjects under investigation exhibit different statistics, e.g., varying demographics or disease prevalence. It is not uncommon for both types of DS to occur simultaneously, in particular in multi-center studies. It is essential to tackle DS in machine learning to perform reliable analysis of large populations across sites and to avoid introducing biases into results. Recent work has shown that even after careful pre-processing, site-specific differences remain in the images [3,4].

While methods like ComBat [5] aim to harmonize image-derived measurements, we focus on the images themselves.

One solution is domain adaptation (DA), a transductive [6] transfer learning technique that aims to modify the source domain's marginal distribution of the feature space such that it resembles the target domain. In medical imaging, labelled data is scarce and typically unavailable for the target domain. It is also unlikely to have the same subjects in both domains. Thus, we focus on 'unsupervised' and 'unpaired' DA, wherein labelled data is available only in the source domain and no matching samples exist between source and target.

Many DA approaches focus on learning domain-invariant feature representations, by either forcing latent representations of the inputs to follow similar distributions, or 'disentangling' domain-specific features from generic features [7]. This can be achieved with some divergence measure based on data statistics or by training adversarial networks to model the divergence between the feature representations [8]. These methods have been applied to brain lesions [9] and tumours [10] in MRI, and in contrast to non-contrast CT segmentation [11]

While these approaches seem appealing and have shown some success, they lack a notion of explainability as it is difficult to know what transformations are applied to the feature space. Additionally, although the learned task model may perform equally well on both domains, it is not guaranteed to perform as well as separate models trained on the individual domains.

We explore model-agnostic DA by working at the image level. Our approach is based on domain mapping (DM), which aims to learn the pixel-level transformations between two image domains, and includes techniques such as style transfer. Pix2Pix [12] (supervised) and CycleGAN [13] (unsupervised) take images from one domain through some encoder-decoder architecture to produce images in the new domain. The method in [8] uses CycleGAN to improve segmentation across scanners and applies DA at both image and feature levels, thus losing interpretability. It does not decompose the image and spatial transformations.

Methods for DM primarily use UNet-like architectures to learn image-to-image transformations that are easier to interpret, as one can visually inspect the output. For medical images of the same anatomy, but from different scanners, we assume that domain shift manifests primarily in appearance changes (contrast, signal-to-noise, resolution) and anatomical variation (shape changes), plus further subtle variations caused by image reconstruction or interpolation.

Contributions: We propose the use of image-and-spatial transformer networks (ISTNs) [14] to tackle domain shift at image-feature level in multi-site imaging data. ISTNs separate and compose the transformations for adapting appearance and shape differences between domains. We believe our approach is the first to use such an approach with retraining of the downstream task model on images transferred from source to target. We show that ISTNs can be trained adversarially in a task model-agnostic way. The transferred images can be visually inspected, and thus, our approach adds explainability to domain adaptation— which is important for validating the plausibility of the learned transformations.
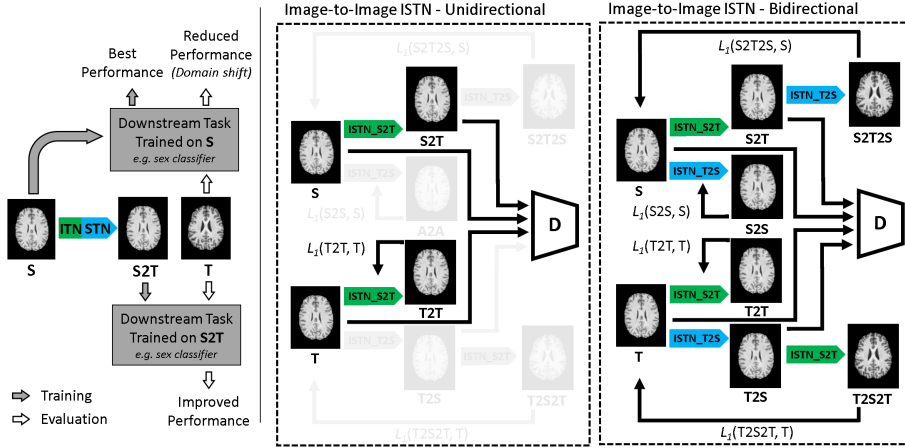
Fig. 1: (left) The domain shift problem can be mitigated by retraining or finetuning a task model on images $S2T$. (Middle) The ISTN is trained adversarially such that the discriminator D becomes better at identifying real (S and T) and transformed ($S2T$) images. The ISTN simultaneously produces better transformations $S2T$ of S that look more like the images T. The training process can also be done bidirectionally (right).

Our results demonstrate the successful recovery of performance on classification and regression tasks when using ISTNs to tackle domain shift. We explore both unidirectional and bidirectional training schemes and compare retraining the task model from scratch versus finetuning. We present proof-of-concept results on synthetic images generated with Morpho-MNIST [15] for a 3-class classification task. Our method is then validated on real multi-site data with 3D T1-weighted brain MRI. Our results indicate that ISTNs improve generalization and predictive performance can be recovered close to single-site accuracy.

2   Method

We propose adversarial training of ISTNs to perform model-agnostic DA via explicit appearance and shape transformations between the domains. We explore unidirectional and bidirectional training schemes as illustrated in Figure 1.

Models. ISTNs have two components: an image transformer network (ITN) and a spatial transformer network (STN) [16,14]. Here, we additionally require a discriminator model for adversarial training of the ISTN.

   ITN: The ITN performs appearance transformations such as contrast and brightness changes, and other localised adaptations at the image-level. A common image-to-image (I2I) translation network based on UNet with residual skip connections can be employed. We use upsample-convolutions to reduce chequerboard artifacts compared with transposed convolution. We use batch normalization, dropout layers and ReLU activations with a final tanh activation for the output. All input images are pre-normalized to the $[-1, 1]$ intensity range.

STN: We experiment with both the affine and B-spline STNs described in the original ISTN paper. Affine STNs learn to regress the parameters of linear spatial transforms with translation, rotation, scaling, and shearing. B-spline STNs regress control point displacements. Linear interpolation is used throughout. Note that in this work, Affine and B-Spline STNs are considered independently and are not composed.

Discriminator: In both Morpho-MNIST and brain MRI experiments, we use a standard fully-convolutional classification network with instance normalization, dropout layers and a sigmoid output.

Task models: The employed classifiers and regressors follow the same fully-convolutional structure as the discriminator, reducing the dimensions of the input images to a multi-class or continuous value prediction, depending on the task. We use cross-entropy or mean-squared error loss functions, respectively.

Appendices C and D provide details about the architectures of different networks. All implementations are in PyTorch [17] with code available online.[1]

Training. The output from the ITN is directly fed into the STN. They are then composed into a single ISTN unit, and are trained jointly end-to-end. *Discriminator*: The images S (from the source domain) are passed through the ISTN to generate images $S2T$, where T indicates images from the target domain. Next, the $S2T$ are passed through the discriminator D_T to yield a score in the range $(0, 1)$ denoting whether the image is a real sample from domain T or a transformed one. The discriminator is trained by minimizing the binary cross-entropy loss $\mathcal{L}_{bce}$ between the predicted and true domain labels. Eq. (1) shows the total discriminator loss. Soft labels for the true domain are used to stabilize early training of the discriminator. We replace the hard '0' and '1' domain labels by random uniform values in the ranges $[0.00, 0.03]$ and $[0.97, 1.00]$, respectively.

ISTN: The ISTN is trained as a generator. The ISTN output $S2T$ is passed through the discriminator and forced to be closer to domain T by computing the adversarial loss $\mathcal{L}_{adv} = \mathcal{L}_{bce}(D_T(S2T), 1)$. Soft labels are also used here. We expect that when images T are passed through the ISTN, the output $T2T$ should be unchanged as it is already in domain T. This is enforced by the identity loss $\mathcal{L}_{idt} = \ell_1(T, T2T)$ acting on image intensities of T and $T2T$. A weighting factor λ is applied to L_{idt} giving the total loss function for the ISTN in Eq. (3)c.

We compare with the CycleGAN [18] training approach, which trains both directions simultaneously using two ISTNs (ISTN$_{S2T}$ and ISTN$_{T2S}$) and two discriminators (D_S and D_T). The CycleGAN introduces the cycle-consistency term to $\mathcal{L}_{istn}$ such that when ISTN$_{T2S}$ is used to transform $S2T$, the result $S2T2S$ is forced to be close to S. Figure 1 shows the two ISTNs, their outputs and associated losses. The loss functions for ISTN$_{S2T}$ are shown in Eq. (3). Optimization is done using the Adam optimizer.

Downstream Tasks: The goal of our work is to demonstrate that such explicit appearance and spatial transformations via ISTNs can successfully tackle DS in

---

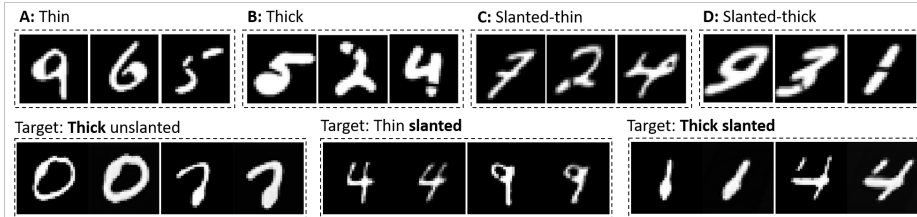[1] https://github.com/mlnotebook/domain_adapation_istn

Fig. 2: (Top) Examples from Morpho-MNIST datasets from domains (left-to-right) A thin un-slanted digits; B thickened digits; C slanted digits; D thickened *and* slanted digits. Each contains 'healthy', 'fractured' and 'swollen' classes. (Bottom) Examples of source domain images before (left) and after (right) ISTN-transformation showing ISTN recovery of appearance and shape changes.

certain applications. Ideally, we would like to observe that the performance of a predictor trained on $S2T$ and tested on T can recover to single-site performance. To demonstrate this, prior to training the ISTN, we train a task model (*e.g.* classifier or regressor) $\mathcal{T}_S$ on domain S. The performance of $\mathcal{T}_S(S)$ is likely to be our 'best performance' whilst $\mathcal{T}_S(T)$ will degrade due to DS. During ISTN training, we simultaneously re-train $\mathcal{T}_S$ on the ISTN output of $S2T$. This model $\mathcal{T}_{S2T}$ is trained to achieve maximum performance on the transformed images $\mathcal{T}_{S2T}(S2T)$ using labels from S. We assess the performance 'recovery' of $\mathcal{T}_{S2T}$ by comparing $\mathcal{T}_S(T)$ with $\mathcal{T}_{S2T}(T)$. In practice, data from T would be unlabelled. Our approach ensures that test data from the new domain T is not modified in any way. Additionally, in scenarios where the original model $\mathcal{T}_S$ is deployed, it is likely to have been trained on a large, well-curated, high-quality dataset; we cannot assume similar would be available for each new test domain. Our model-agnostic unsupervised DA is validated on two problems: i) proof-of-concept showing recovery of a classifier's performance on digit recognition, ii) classification and regression tasks with real-world, multi-site T1-weighted brain MRI.

$$\mathcal{L}_{disT} = \tfrac{1}{2}\left[\mathcal{L}_{bce}(D_T(S2T),0) + \mathcal{L}_{bce}(D_T(T),1)\right]. \tag{1}$$

$$\mathcal{L}_{istn}^{S2T} = \mathcal{L}_{bce}(D_T(S2T),1) + \tfrac{1}{2}\lambda\left\|T2T - T\right\|_1. \tag{2}$$

$$\mathcal{L}_{istn}^{S2T} = \mathcal{L}_{bce}(D_S(T2S),0) + \tfrac{1}{2}\lambda\left\|S2S - S\right\|_1 + \lambda\left\|S2T2S - S\right\|_1. \tag{3}$$

3   Materials

3.1   Proof-of-concept: Morpho-MNIST Experiments

Data. Morpho-MNIST is a framework that enables applying medically-inspired perturbations, such as local swellings and fractures, to the well-known MNIST dataset [15]. The framework also allows us to control transformations to obtain thickening and shearing of the original digits. We first create a dataset with three classes: 'healthy' digits with no transformations; 'fractured' digits with a single thin disruption and 'swollen' digits which exhibit a localized, tumor-like

abnormal growth. A digit is only either fractured or swollen, not both. We specify a set of 'thin' digits (2.5 pixels across) to be source domain A. To simulate domain shift, we create three more datasets—domain B: thickened, 5.0 pixels digits; domain C: slanted digits created by shearing the image by 20–25° and domain D: thickened-slanted digits at 5.0 pixels and 20–25° shearing. Datasets B–D contain the same three classes as A, while each set has its own data characteristics simulating different types of domain shift. All images are single-channel and 28×28 pixels. Figure 2 shows some visual examples.

Task. The downstream task in this experiment is a 3-class classification problem: 'healthy' vs. 'fractured' vs. 'swollen'. We train a small, fully-convolutional classifier to perform the classification on domain A. We use ISTNs to retrain the classifier on transformed images $A2B$, $A2C$, and $A2D$, and evaluate each on their corresponding test domains B, C, and D.

We run training for 100 epochs and perform grid search to find suitable hyperparameters including learning rate, trade-off λ and the control-point spacing of the B-spline STN. We conduct experiments using ITN only, STN only and combinations of affine and B-spline ISTNs to determine the best model for the task. We also consider both transfer directions, switching the roles of source and target domains.

3.2   Application to Brain MRI Experiments

We apply the same methodology to a real-world domain shift problem where we observe a significant drop in prediction accuracy when naively training on one site and testing on another without any DA. We utilise 3D brain MRI from two sites that employ similar but not identical imaging protocols.

Data. We construct two datasets of T1-weighted brain MRI from subjects with no reported pathology, where $n = 565$ are taken from the Cambridge Centre for Ageing and Neuroscience study (Cam-CAN) [19,20] and $n = 689$ from the UK Biobank imaging study (UKBB) [21,22,23]. From each site, 450 subjects are used for training and the remainder for testing. The UKBB dataset contains equal numbers of male and female subjects between the ages of 48 and 71 ($\mu = 59.5$). In the classification task, to simulate the effect of population shift our Cam-CAN dataset has a wider age range (30–87, $\mu = 57.9$) but maintains the male-to-female ratio. We match the age range of both datasets in the regression task, limiting DS only to the more subtle scanner effects. UKBB images were acquired at the UKBB imaging centre, and Cam-CAN images were acquired at the Medical Research Council Cognition and Brain Sciences Unit in Cambridge, UK. Both sites acquire 1 mm isotropic images using the 3D MPRAGE pulse sequence on Siemens 3 T scanners with a 32-channel receiver head coil and in-plane acceleration factor 2. Appendix A presents the acquisition parameters that differ between the two sites. We note that generally the acquisition parameters of both sites are similar, and the images cannot be easily distinguished visually. For pre-processing, all images are affinely aligned to MNI space, skull-stripped,

Table 1: 3-class classification results on MorphoMNIST. Images transferred from classifier domain A: 'thin unslanted' to three target domains. Accuracies shown for classifiers retrained on the ISTN output from scratch (Acc_s) and finetuned (Acc_f). Δ is model improvement from baseline. Control-point spacings indicated for B-Spline STNs. First row is the original classifier without DA.

Target		Thick Unslanted				Thin Slanted				Thick Slanted			
ITN	STN	Acc_s	Δ	Acc_f	Δ	Acc_s	Δ	Acc_f	Δ	Acc_s	Δ	Acc_f	Δ
no	no	41.2				45.7				32.8			
yes	no	**79.0**	**37.8**	**83.3**	**42.1**	83.4	37.7	83.3	37.6	**82.4**	**49.6**	**84.6**	**51.8**
no	**Affine**	52.4	11.2	68.9	27.7	92.4	46.7	93.0	47.3	54.8	22.0	64.8	32.0
no	**B-spline (4)**	39.0	-2.2	54.4	13.2	92.1	46.4	93.1	47.4	36.0	3.2	57.2	24.4
no	**B-spline (8)**	49.2	8.0	61.5	20.3	92.5	46.8	92.3	46.6	37.0	4.2	61.8	29.0
yes	**Affine**	78.8	37.6	77.1	35.9	86.7	41.0	88.4	42.7	81.9	49.1	83.1	50.3
yes	**B-spline (4)**	66.3	25.1	75.8	34.6	**92.7**	**47.0**	91.0	45.3	79.3	46.5	82.7	49.9
yes	**B-spline (8)**	69.5	28.3	77.2	36.0	91.8	46.1	**93.4**	**47.7**	79.0	46.2	80.8	48.0

Table 2: Sex classification results on 3D Brain MRI

Source		UKBB								Cam-CAN							
Method		Uni-ISTN				CycleGAN Bi-ISTN				Uni-ISTN				CycleGAN Bi-ISTN			
ITN	STN	Acc_s	Δ	Acc_f	Δ	Acc_s	Δ	Acc_f	Δ	Acc_s	Δ	Acc_f	Δ	Acc_s	Δ	Acc_f	Δ
no	no	54.8				54.8				64.3				64.3			
yes	no	79.1	24.3	72.2	17.4	**80.0**	**25.2**	80.8	26.0	**86.2**	21.9	78.2	13.9	80.8	16.5	79.9	15.6
yes	**Affine**	**80.9**	**26.1**	75.7	20.9	70.4	15.6	**82.4**	**27.6**	79.9	15.6	79.1	14.8	**82.4**	**18.1**	72.0	7.7
yes	**B-spline (8)**	78.3	23.5	76.5	21.7	79.1	24.3	78.7	23.9	80.3	16.0	**84.5**	**20.2**	78.7	14.4	**80.8**	**16.5**
yes	**B-spline (16)**	80.0	25.2	**78.3**	**23.5**	73.0	18.2	67.8	13.0	85.4	21.1	84.1	19.8	67.8	3.5	68.6	4.3

bias-field-corrected, and intensity-normalised to zero mean unit variance within a brain mask. Voxels outside the mask are set to 0. Images are passed through a tanh function before being consumed by the networks.

Task. We consider two prediction tasks, namely sex classification and age regression using the UKBB and Cam-CAN sets, each once as source and once as target domain. The task networks are retrained on the transformed images produced by the ISTN and evaluated on the corresponding target domain.
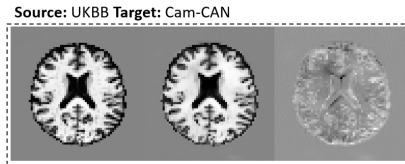
4 Experimental Results

Morpho-MNIST. Quantitative results for the synthetic experiments are summarized in Table 1. ITNs are able to harmonize local appearance such as thickness between source and target domains, while STNs perform well in recovering shape variations such as slant. Where both thickness and slant are varied between source and target domains, we note an ITN-only performs as well (or slightly better) than a joint ISTN, suggesting that thickness is more important for the classification task. In Fig. 2 we show visual results on how the ISTNs are able to recover both appearance and shape differences between domains.

Brain MRI. Quantitative results are summarized in Tables 2 and 3. The sex classifier trained and tested on UKBB achieves 84.3% accuracy. This drops to 54.8% when tested on Cam-CAN. Similarly, training and testing on Cam-CAN yields 91.6%, dropping to 64.3% when testing on UKBB. Using ISTNs for domain

Table 3: Age regression results on 3D Brain MRI. MAE_s is the task model retrained from scratch.

Source		UKBB		Cam-CAN	
	Method	Uni-ISTN		Uni-ISTN	
ITN	STN	MAE_s	Δ	MAE_s	Δ
no	no	5.13		4.61	
yes	no	4.71	0.42	**4.57**	**0.04**
yes	**Affine**	**4.58**	**0.55**	5.00	-0.39
yes	**B-spline (16)**	5.06	0.07	4.90	-0.29

Fig. 3: Examples of (left-to-right) source domain, transformed ISTN output and difference image.

Source: UKBB **Target:** Cam-CAN

adaptation, and retraining the classifiers increases the accuracy substantially on Cam-CAN from 54.8% to 80.9%, and on UKBB from 64.3% to 86.2%, which is close to the single-site performance. Training the classifier from scratch performs similarly well to fine-tuning. Bidirectional training with CycleGAN seems not to provide substantial improvements over the simpler unidirectional scheme. The ISTNs are able to overcome some of the acquisition and population shifts between the two domains. The age regressor trained and tested on UKBB achieves mean absolute error (MAE) of 4.25 years increasing to 5.13 when evaluated on Cam-CAN. The regressor trained and tested on Cam-CAN yields 4.10 years MAE increasing to 4.61 when tested on UKBB. Despite the initially smaller drop in performance for age regression, ISTNs still improve performance. The UKBB-trained regressor recovers to 4.58 years MAE and the Cam-CAN-trained one to 4.56 years. Note, we had limited the population shift here by constraining the age range, thus the recovery is likely due to a reduction in acquisition shift.

5   Conclusion

We explored adversarially-trained ISTNs for model-agnostic domain adaptation. The learned image-level transformations help explainability, as the resulting images can be visually inspected and checked for plausibility (cf. Fig. 3). Further interrogation of deformations fields also adds to explainability, *e.g.* Appendix B. Image-level DA seems suitable in cases of subtle domain shift caused by acquisition and population differences in multi-center studies. Predictive performance approached single-site accuracies. The choice of STN and control-point spacings may need to be carefully considered for specific use cases. An extension of our work to many-sites may be possible by simultaneously adapting to multiple sites. A quantitative comparison to feature-level DA would be a natural next step for future work. Another interesting direction could be to integrate the ISTN component in a fully end-to-end task-driven optimisation, where the ISTN and the task network are trained jointly.

References

1. Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., van Walsum, T., eds.: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer International Publishing (2019)
2. Castro, D.C., Walker, I., Glocker, B.: Causality matters in medical imaging. arXiv:1912.08142 (2019)
3. Wachinger, C., Becker, B.G., Rieckmann, A., Pölsterl, S.: Quantifying confounding bias in neuroimaging datasets with causal inference. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2019, Springer (2019) 484–492
4. Glocker, B., Robinson, R., Castro, D.C., Dou, Q., Konukoglu, E.: Machine learning with multi-site imaging data: An empirical study on the impact of scanner effects. In: Medical Imaging meets NeurIPS. (2019)
5. Yu, M., Linn, K.A., Cook, P.A., Phillips, M.L., McInnis, M., Fava, M., Trivedi, M.H., Weissman, M.M., Shinohara, R.T., Sheline, Y.I.: Statistical harmonization corrects site effects in functional connectivity measurements from multi-site fmri data. Human Brain Mapping **39**(11) (2018) 4213–4227
6. Pan, S., Yang, Q.: A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering **22**(10) 13451359,
7. Yang, J., Dvornek, N.C., Zhang, F., Chapiro, J., Lin, M., Duncan, J.S.: Unsupervised domain adaptation via disentangled representations: Application to cross-modality liver segmentation. In Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.T., Khan, A., eds.: Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, Cham, Springer International Publishing (2019) 255–263
8. Yan, W., Wang, Y., Gu, S., Huang, L., Yan, F., Xia, L., Tao, Q.: The domain shift problem of medical image segmentation and vendor-adaptation by unet-gan. In Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.T., Khan, A., eds.: Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, Cham, Springer International Publishing (2019) 623–631
9. Kamnitsas, K., Baumgartner, C., Ledig, C., Newcombe, V., Simpson, J., Kane, A., Menon, D., Nori, A., Criminisi, A., Rueckert, D., et al.: Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. In: International Conference on Information Processing in Medical Imaging, Springer (2017) 597–609
10. Dai, L., Li, T., Shu, H., Zhong, L., Shen, H., Zhu, H.: Automatic brain tumor segmentation with domain adaptation. In Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., van Walsum, T., eds.: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, Cham, Springer International Publishing (2019) 380–392
11. Sandfort, V., Yan, K., Pickhardt, P.J., Summers, R.M.: Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. Scientific Reports **9**(1) (November 2019)
12. Isola, P., Zhu, J., Zhou, T., Efros, A.: Image-to-image translation with conditional adversarial networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR, Honolulu, HI 59675976

13. Zhu, J., Park, T., Isola, P., Efros, A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: 2017 IEEE International Conference on Computer Vision (ICCV, Venice 22422251

14. Lee, M.C.H., Oktay, O., Schuh, A., Schaap, M., Glocker, B.: Image-and-spatial transformer networks for structure-guided image registration. In Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.T., Khan, A., eds.: Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, Cham, Springer International Publishing (2019) 337–345

15. Castro, D.C., Tan, J., Kainz, B., Konukoglu, E., Glocker, B.: Morpho-MNIST: Quantitative assessment and diagnostics for representation learning. Journal of Machine Learning Research **20**(178) (2019)

16. Jaderberg, M., Simonyan, K., Zisserman, A., et al.: Spatial transformer networks. In: Advances in Neural Information Processing Systems 28. (2015) 2017–2025

17. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: Pytorch: An imperative style, high-performance deep learning library. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché Buc, F., Fox, E., Garnett, R., eds.: Advances in Neural Information Processing Systems 32. Curran Associates, Inc. (2019) 8024–8035

18. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Computer Vision (ICCV), 2017 IEEE International Conference on. (2017)

19. Shafto, M.A., Tyler, L.K., Dixon, M., Taylor, J.R., Rowe, J.B., Cusack, R., Calder, A.J., Marslen-Wilson, W.D., Duncan, J., Dalgleish, T., et al.: The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study protocol: a cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing. BMC Neurology **14**(1) (2014) 204

20. Taylor, J.R., Williams, N., Cusack, R., Auer, T., Shafto, M.A., Dixon, M., Tyler, L.K., Henson, R.N., et al.: The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data repository: structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample. NeuroImage **144** (2017) 262–269

21. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., Collins, R.: UK Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Medicine **12**(3) (2015) e1001779

22. Miller, K.L., Alfaro-Almagro, F., Bangerter, N.K., Thomas, D.L., Yacoub, E., Xu, J., Bartsch, A.J., Jbabdi, S., Sotiropoulos, S.N., Andersson, J.L.R., Griffanti, L., Douaud, G., Okell, T.W., Weale, P., Dragonu, I., Garratt, S., Hudson, S., Collins, R., Jenkinson, M., Matthews, P.M., Smith, S.M.: Multimodal population brain imaging in the UK Biobank prospective epidemiological study. Nature Neuroscience **19**(11) (2016) 1523–1536

23. Alfaro-Almagro, F., Jenkinson, M., Bangerter, N.K., Andersson, J.L., Griffanti, L., Douaud, G., Sotiropoulos, S.N., Jbabdi, S., Hernandez-Fernandez, M., Vallee, E., Vidaurre, D., Webster, M., McCarthy, P., Rorden, C., Daducci, A., Alexander, D.C., Zhang, H., Dragonu, I., Matthews, P.M., Miller, K.L., Smith, S.M.: Image processing and quality control for the first 10,000 brain imaging datasets from UK Biobank. NeuroImage **166** (feb 2018) 400–424

Supplementary Material

A   Acquisition Parameters.

Table 4: Acquisition parameters for the multi-site brain MRI datasets.

Site	Scanner	TR (ms)	TE (ms)	TI (ms)	TA (s)	FOV (mm)
Cam-CAN	Siemens TIM Trio	2250	2.99	900	272	256x240x192
UKBB	Siemens Skyra	2000	2.01	880	294	208x256x256

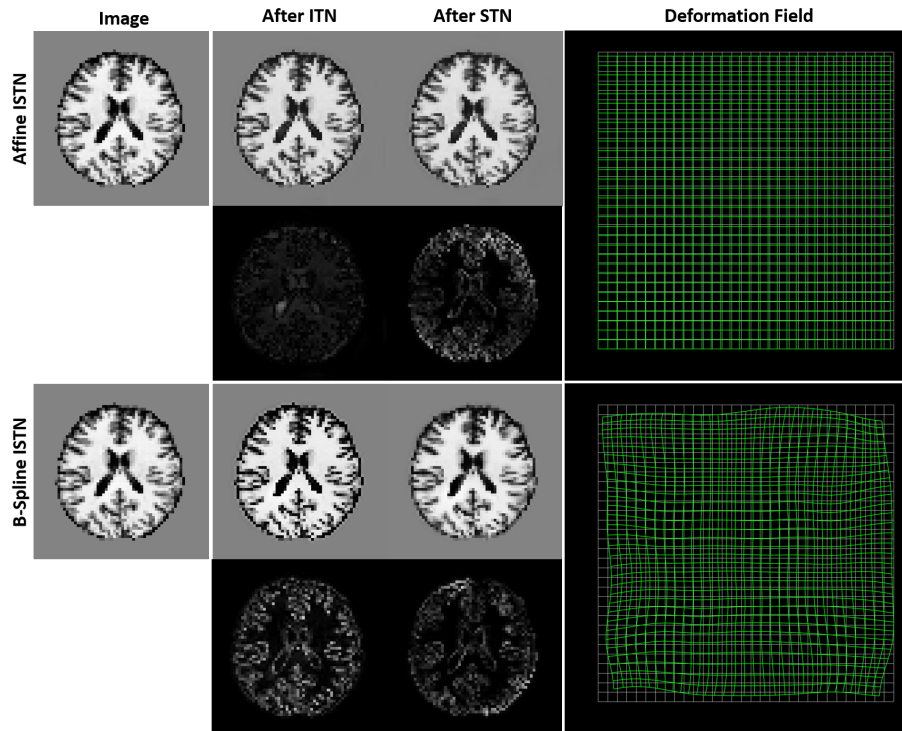# B   ISTN Transformation Visualization.

Fig. 4: The original image (left) passes through the ISTN. The transformations applied by the ITN and subsequently by the STN are visualized by showing difference images. The transformation applied by the STN can also be visualized as a spatial deformation field (right). This is shown for the Affine (top) and B-Spline (bottom) STNs.

C   Morpho-MNIST Architectures.

Table 5: ITN (left) and discriminator (right) architectures for Morpho-MNIST experiments. nf: number of channels, k: square kernel size, s: stride, in and out: layer input and output dimensions, N: normalization (BN: batch normalization, IN: instance normalization), D: Dropout keep-rate, A: activation function. 'up' is composed of bilinear upsampling followed by zero-padding of 1 and convolution shown in the table.

ITN Architecture - Morpho-MNIST

layer	nf	k	s	p	in	out	N	D	A
in	-	-	-	-	[1,28,28]	-	-	-	-
conv	16	3	1	1	[1,28,28]	[16,28,28]	BN	-	ReLU
conv	32	3	2	1	[16,28,28]	[32,14,14]	BN	-	ReLU
conv	64	3	2	1	[32,14,14]	[64,7,7]	BN	-	ReLU
conv	128	3	1	1	[64,7,7]	[128,7,7]	BN	-	ReLU
conv	64	3	1	1	[128,7,7]	[64,7,7]	BN	-	ReLU
up	32	3	1	1	[64,7,7]	[32,14,14]	BN	-	ReLU
up	16	3	1	1	[32,14,14]	[16,28,28]	BN	-	ReLU
up	1	3	1	1	[16,28,28]	[1,28,28]	-	-	tanh
out	-	-	-	-	-	[1,28,28]	-	-	-

Discriminator Architecture - Morpho-MNIST

layer	nf	k	s	p	in	out	N	D	A
in	-	-	-	-	[1,28,28]	-	-	-	-
conv	32	3	1	1	[1,28,28]	[32,28,28]	-	-	ReLU
conv	64	3	2	1	[32,28,28]	[64,14,14]	IN	-	ReLU
conv	128	3	2	1	[64,14,14]	[128,7,7]	IN	-	ReLU
conv	256	3	2	1	[128,7,7]	[256,4,4]	IN	0.5	ReLU
conv	1	3	2	1	[256,4,4]	[1,1,1]	-	-	sigmoid
out	-	-	-	-	-	[1,1,1]	-	-	-

3-Class Classifier Architecture - Morpho-MNIST

layer	nf	k	s	p	in	out	N	D	A
in	-	-	-	-	[1,64,64,64]	-	-	-	-
conv	16	3	1	1	[1,24,24]	[16,24,24]	-	-	ReLU
conv	32	3	2	1	[16,14,14]	[32,7,7]	BN	-	ReLU
conv	64	3	2	1	[32,7,7]	[64,4,4]	BN	-	ReLU
conv	128	3	2	1	[64,4,4]	[128,1,1]	BN	0.5	ReLU
conv	3	3	2	0	[128,1,1]	[3,1,1]	-	-	sigmoid
out	-	-	-	-	-	[3,1,1]	-	-	-

D   Brain MRI Architectures.

Table 6: Architectures for Brain MRI experiments. nf: number of channels, k: square kernel size, s: stride, in and out: layer input and output dimensions, N: normalization (BN: batch or IN: instance normalization), D: Dropout keep-rate, A: activation function. 'up' is composed of linear upsampling, zero-padding and convolution.

ITN Architecture - Brain MRI

layer	nf	k	s	p	in	out	N	D	A
in	-	-	-	-	[1,64,64,64]	-	-	-	-
conv	8	3	1	1	[1,64,64,64]	[8,64,64,64]	BN	-	ReLU
conv	16	3	2	1	[8,64,64,64]	[16,32,32,32]	BN	-	ReLU
conv	32	3	2	1	[16,32,32,32]	[32,16,16,16]	BN	-	ReLU
conv	64	3	2	1	[32,16,16,16]	[64,8,8,8]	BN	-	ReLU
conv	64	3	1	1	[64,8,8,8]	[64,8,8,8]	BN	-	ReLU
conv	64	3	1	1	[64,8,8,8]	[64,8,8,8]	BN	-	ReLU
up	32	3	1	1	[64,8,8,8]	[32,16,16,16]	BN	0.5	ReLU
up	16	3	1	1	[32,16,16,16]	[16,32,32,32]	BN	0.5	ReLU
up	8	3	1	1	[16,32,32,32]	[8,64,64,64]	BN	0.5	ReLU
up	1	3	1	1	[8,64,64,64]	[1,64,64,64]	-	-	tanh
out	-	-	-	-	-	[1,64,64,64]	-	-	-

Discriminator Architecture - Brain MRI

layer	nf	k	s	p	in	out	N	D	A
in	-	-	-	-	[1,64,64,64]	-	-	-	-
conv	32	3	1	1	[1,64,64,64]	[32,64,64,64]	-	-	ReLU
conv	64	3	2	1	[32,64,64,64]	[64,32,32,32]	IN	-	ReLU
conv	128	3	2	1	[64,32,32,32]	[128,16,16,16]	IN	-	ReLU
conv	256	3	2	1	[128,16,16]	[256,8,8]	IN	-	ReLU
conv	256	3	2	1	[256,8,8]	[256,4,4]	IN	0.5	ReLU
conv	1	3	2	1	[256,4,4]	[1,1,1]	-	-	sigmoid
out	-	-	-	-	-	[1,1,1]	-	-	-

Sex Classifier Architecture - Brain MRI

layer	nf	k	s	p	in	out	N	D	A
in	-	-	-	-	[1,64,64,64]	-	-	-	-
conv	8	5	2	2	[1,64,64,64]	[8,64,64,64]	-	-	ReLU
conv	16	5	2	2	[8,64,64,64]	[16,32,32,32]	BN	-	ReLU
conv	32	5	2	2	[16,32,32,32]	[32,16,16,16]	BN	-	ReLU
conv	64	5	2	2	[32,16,16]	[64,8,8]	BN	0.5	ReLU
conv	128	2	2	2	[64,8,8]	[128,4,4]	BN	0.5	ReLU
conv	128	2	2	2	[128,4,4]	[128,1,1]	BN	0.5	ReLU
conv	1	5	1	2	[128,1,1]	[1,1,1]	-	-	sigmoid
out	-	-	-	-	-	[1,1,1]	-	-	-

Age Regressor Architecture - Brain MRI

layer	nf	k	s	p	in	out	N	D	A
in	-	-	-	-	[1,64,64,64]	-	-	-	-
conv	16	3	1	1	[1,64,64,64]	[8,64,64,64]	-	-	ReLU
MaxPool	-	-	2	1	[8,64,64,64]	[8,32,32,32]	-	-	-
conv	32	3	2	1	[8,32,32,32]	[32,32,32,32]	-	-	ReLU
MaxPool	-	-	2	1	[32,32,32,32]	[32,16,16,16]	-	-	-
Linear	128	-	-	-	[32*32*32*32,1]	[128]	-	-	ReLU
Linear	64	-	-	-	[128]	[64]	-	-	ReLU
Linear	32	-	-	-	[64]	[32]	-	-	ReLU
Linear	1	-	-	-	[32]	[1]	-	-	
out	-	-	-	-	-	[1]	-	-	-