# Unsupervised Domain Transfer with Conditional Invertible Neural Networks

Kris K. Dreher[1,2(✉)], Leonardo Ayala[1,3], Melanie Schellenberg[1,4], Marco Hübner[1,4], Jan-Hinrich Nölke[1,4], Tim J. Adler[1], Silvia Seidlitz[1,4,5,6], Jan Sellner[1,4,5], Alexander Studier-Fischer[7], Janek Gröhl[8,9], Felix Nickel[7], Ullrich Köthe[4], Alexander Seitel[1,6], and Lena Maier-Hein[1,3,4,5,6]

[1] Intelligent Medical Systems, German Cancer Research Center (DKFZ), Heidelberg, Germany
{k.dreher,l.maier-hein}@dkfz-heidelberg.de
[2] Faculty of Physics and Astronomy, Heidelberg University, Heidelberg, Germany
[3] Medical Faculty, Heidelberg University, Heidelberg, Germany
[4] Faculty of Mathematics and Computer Science, Heidelberg University, Heidelberg, Germany
[5] Helmholtz Information and Data Science School for Health, Karlsruhe, Heidelberg, Germany
[6] National Center for Tumor Diseases (NCT) Heidelberg a Partnership Between DKFZ and Heidelberg University Hospital, Heidelberg, Germany
[7] Department of General, Visceral, and Transplantation Surgery, Heidelberg University Hospital, Heidelberg, Germany
[8] Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK
[9] Department of Physics, University of Cambridge, Cambridge, UK

**Abstract.** Synthetic medical image generation has evolved as a key technique for neural network training and validation. A core challenge, however, remains in the domain gap between simulations and real data. While deep learning-based domain transfer using Cycle Generative Adversarial Networks and similar architectures has led to substantial progress in the field, there are use cases in which state-of-the-art approaches still fail to generate training images that produce convincing results on relevant downstream tasks. Here, we address this issue with a domain transfer approach based on conditional invertible neural networks (cINNs). As a particular advantage, our method inherently guarantees cycle consistency through its invertible architecture, and network training can efficiently be conducted with maximum likelihood training. To showcase our method's generic applicability, we apply it to two spectral imaging modalities at different scales, namely hyperspectral imaging (pixel-level) and photoacoustic tomography (image-level). According to

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43907-0_73.


H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14220, pp. 770–780, 2023.
https://doi.org/10.1007/978-3-031-43907-0_73

comprehensive experiments, our method enables the generation of realistic spectral data and outperforms the state of the art on two downstream classification tasks (binary and multi-class). cINN-based domain transfer could thus evolve as an important method for realistic synthetic data generation in the field of spectral imaging and beyond. The code is available at https://github.com/IMSY-DKFZ/UDT-cINN.

**Keywords:** Domain transfer · invertible neural networks · medical imaging · photoacoustic tomography · hyperspectral imaging · deep learning

## 1 Introduction

The success of supervised learning methods in the medical domain led to countless breakthroughs that might be translated into clinical routine and have the potential to revolutionize healthcare [6,13]. For many applications, however, labeled reference data (ground truth) may not be available for training and validating a neural network in a supervised manner. One such application is spectral imaging which comprises various non-interventional, non-ionizing imaging techniques that can resolve functional tissue properties such as blood oxygenation in real time [1,3,23]. While simulations have the potential to overcome the lack of ground truth, synthetic data is not yet sufficiently realistic [9]. Cycle Generative Adversarial Networks (GAN)-based architectures are widely used for domain transfer [12,24] but may suffer from issues such as unstable training, hallucinations, or mode collapse [15]. Furthermore, they have predominantly been used for conventional RGB imaging and one-channel cross-modality domain adaptation, and may not be suitable for other imaging modalities with more channels. We address these challenges with the following contributions:

**Domain Transfer Method:** We present an entirely new sim-to-real transfer approach based on conditional invertible neural networks (cINNs) (cf. Fig. 1) specifically designed for data with many spectral channels. This approach inherently addresses weaknesses of the state of the art with respect to the preservation of spectral consistency and, importantly, does not require paired images.

**Instantiation to Spectral Imaging:** We show that our method can generically be applied to two complementary modalities: photoacoustic tomography (PAT; image-level) and hyperspectral imaging (HSI; pixel-level).

**Comprehensive Validation:** In comprehensive validation studies based on more than 2,000 PAT images (real: ∼1,000) and more than 6 million spectra for HSI (real: ∼6 million) we investigate and subsequently confirm our two main hypotheses: (H1) Our cINN-based models can close the domain gap between simulated and real spectral data better than current state-of-the-art methods

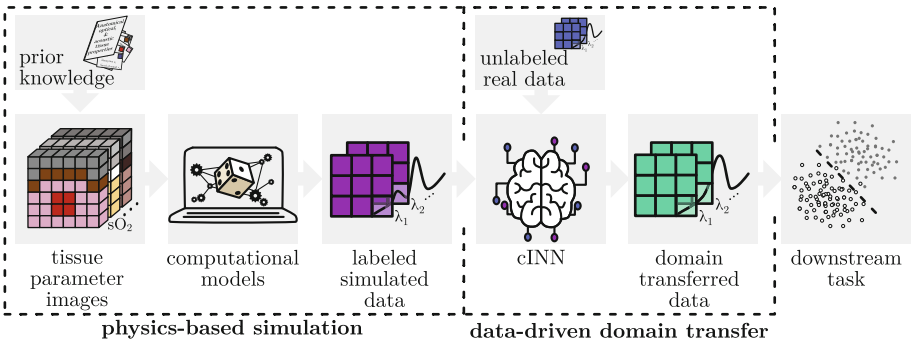


**Fig. 1. Pipeline for data-driven spectral image analysis in the absence of labeled reference data.** A physics-based simulation framework generates simulated spectral images with corresponding reference labels (e.g. tissue type or oxygenation ($sO_2$)). Our domain transfer method based on cINNs leverages unlabeled real data to increase their realism. The domain-transferred data can then be used for supervised training of a downstream task (e.g. classification).

regarding spectral plausibility. (H2) Training models on data transferred by our cINN-based approach can improve their performance on the corresponding (clinical) downstream task without them having seen labeled real data.

## 2 Materials and Methods

### 2.1 Domain Transfer with Conditional Invertible Neural Networks

**Concept Overview.** Our domain transfer approach (cf. Fig. 2) is based on the assumption that data samples from both domains carry domain-invariant information (e.g. on optical tissue properties) and domain-variant information (e.g. modality-specific artifacts). The invertible architecture, which inherently guarantees cycle consistency, transfers both simulated and real data into a shared latent space. While the domain-invariant features are captured in the latent space, the domain-variant features can either be filtered (during encoding) or added (during decoding) by utilizing a domain label $D$. To achieve spectral consistency, we leverage the fact that different tissue types feature characteristic spectral signatures and condition the model on the tissue label $Y$ if available. For unlabeled (real) data, we use randomly generated proxy labels instead. To achieve high visual quality beyond spectral consistency, we include two discriminators $Dis_{sim}$ and $Dis_{real}$ for their respective domains. Finally, as a key theoretical advantage, we avoid mode collapse with maximum likelihood optimization. Implementation details are provided in the following.

**cINN Model Design.** The core of our architecture is a cINN [2] (cf. Fig. 2), comprising multiple ($i$) scales of $N_i$-chained affine conditional coupling (CC)

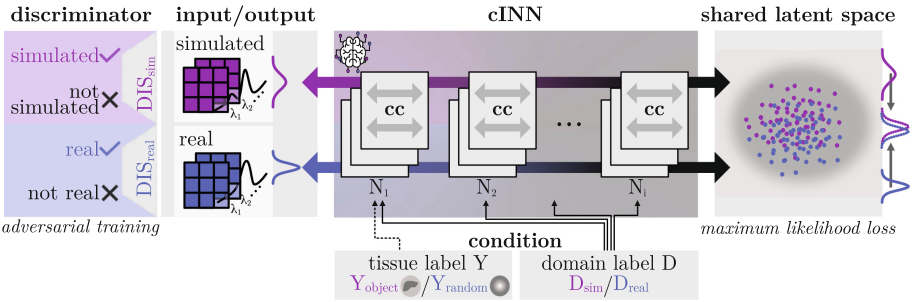


**Fig. 2. Proposed architecture based on cINNs.** The invertible architecture transfers both simulated and real data into a shared latent space (right). By conditioning on the domain $D$ (bottom), a latent vector can be transferred to either the simulated or the real domain (left) for which the discriminator $\text{Dis}_{\text{sim}}$ and $\text{Dis}_{\text{real}}$ calculate the losses for adversarial training.

blocks [7]. These scales are necessary in order to increase the receptive field of the network and are achieved by Haar wavelet downsampling [11]. A CC block consists of subnetworks that can be freely chosen depending on the data dimensionality (e.g. fully connected or convolutional networks) as they are only evaluated in the forward direction. The CC blocks receive a condition consisting of two parts: domain label and tissue label, which are then concatenated to the input along the channel dimension. In the case of PAT, the tissue label is a full semantic and random segmentation map for the simulated and real data, respectively. In the case of HSI, the tissue label is a one-hot encoded vector for organ labels.

**Model Training.** In the following, the proposed cINN with its parameters $\theta$ will be referred to as $f(x, DY, \theta)$ and its inverse as $f^{-1}$ for any input $x \sim p_D$ from domain $D \in \{D_{sim}, D_{real}\}$ with prior density $p_D$ and its corresponding latent space variable $z$. The condition $DY$ is the combination of domain label $D$ as well as the tissue label $Y \in \{Y_{sim}, Y_{real}\}$. Then the maximum likelihood loss $\mathcal{ML}$ for a training sample $x_i$ is described by

$$\underset{D}{\mathcal{ML}} = \mathbb{E}_i \left[ \frac{||f(x_i, DY, \theta)||_2^2}{2} - log|J_i| \right] \text{ with } J_i = det \left( \frac{\partial f}{\partial x} \Big|_{x_i} \right). \tag{1}$$

For the adversarial training, we employ the least squares training scheme [18] for generator $Gen_D = f_D^{-1} \circ f_{D'}$ and discriminator $Dis_D$ for each domain with $x_{D'}$ as input from the source domain and $x_D$ as input from the target domain:

$$\underset{Gen_D}{\mathcal{L}} = \underset{x_{D'} \sim p_{D'}}{\mathbb{E}} \left[ (Dis_D(Gen_D(x_{D'}) - 1))^2 \right] \tag{2}$$

$$\underset{Dis_D}{\mathcal{L}} = \underset{x_D \sim p_D}{\mathbb{E}} \left[ (Dis_D(x_D) - 1)^2 \right] + \underset{x_{D'} \sim p_{D'}}{\mathbb{E}} \left[ (Dis_D(Gen_D(x_{D'})))^2 \right]. \tag{3}$$

Finally, the full loss for the proposed model comprises the following:

$$\mathcal{L}_{Total_{Gen}} = \mathcal{ML}_{real} + \mathcal{ML}_{sim} + \mathcal{L}_{Gen_{real}} + \mathcal{L}_{Gen_{sim}} \quad \text{and} \quad \mathcal{L}_{Total_{Dis}} = \mathcal{L}_{Dis_{real}} + \mathcal{L}_{Dis_{sim}} . \tag{4}$$

**Model Inference.** The domain transfer is done in two steps: 1) A simulated image is encoded in the latent space with conditions $D_{sim}$ and $Y_{sim}$ to its latent representation $z$, 2) $z$ is decoded to the real domain via $D_{real}$ with the simulated tissue label $Y_{sim}$: $x_{sim \to real} = f^{-1}(\cdot, D_{real} Y_{sim}, \theta) \circ f(\cdot, D_{sim} Y_{sim}, \theta)(x_{sim})$.

### 2.2 Spectral Imaging Data

**Photoacoustic Tomography Data.** PAT is a non-ionizing imaging modality that enables the imaging of functional tissue properties such as tissue oxygenation [22]. The **real PAT data** (cf. Fig. 3) used in this work are images of human forearms that were recorded from 30 healthy volunteers using the MSOT Acuity Echo (iThera Medical GmbH, Munich, Germany) (all regulations followed under study ID: S-451/2020, and the study is registered with the German Clinical Trials Register under reference number DRKS00023205). In this study, 16 wavelengths from 700 nm to 850 nm in steps of 10 nm were recorded for each image. The resulting 180 images were semantically segmented into the structures shown in Fig. 3 according to the annotation protocol provided in [20]. Additionally, a full sweep of each forearm was performed to generate more unlabeled images, thus

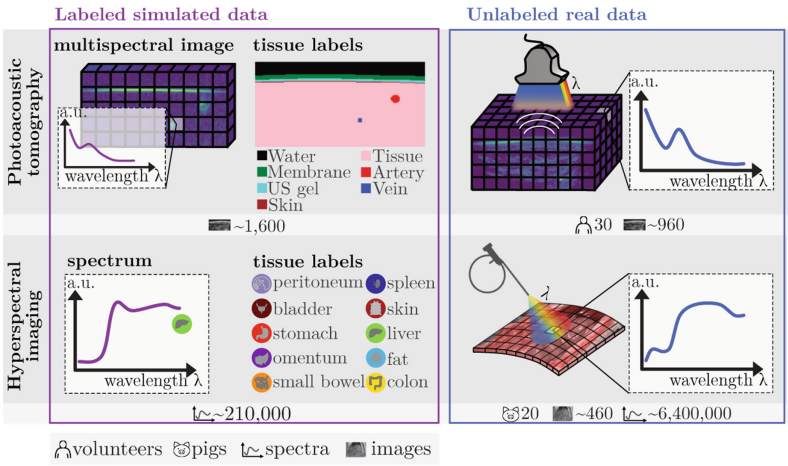


**Fig. 3. Training data used for the validation experiments.** For PAT, 960 real images from 30 volunteers were acquired. For HSI, more than six million spectra corresponding to 460 images and 20 individuals were used. The tissue labels PAT correspond to 2D semantic segmentations, whereas the tissue labels for HSI represent 10 different organs. For PAT, ∼1600 images were simulated, whereas around 210,000 spectra were simulated for HSI.

amounting to a total of 955 real images. The **simulated PAT data** (cf. Fig. 3) used in this work comprises 1,572 simulated images of human forearms. They were generated with the toolkit for Simulation and Image Processing for Photonics and Acoustics (SIMPA) [8] based on a forearm literature model [21] and with a digital device twin of the MSOT Acuity Echo.

**Hyperspectral Imaging Data.** HSI is an emerging modality with high potential for surgery [4]. In this work, we performed pixel-wise analysis of HSI images. The **real HSI data** was acquired with the Tivita® Tissue (Diaspective Vision GmbH, Am Salzhaff, Germany) camera, featuring a spectral resolution of approximately 5 nm in the spectral range between 500 nm and 1000 nm nm. In total, 458 images, corresponding to 20 different pigs, were acquired (all regulations followed under study IDs: 35-9185.81/G-161/18 and 35-9185.81/G-262/19) and annotated with ten structures: bladder, colon, fat, liver, omentum, peritoneum, skin, small bowel, spleen, and stomach (cf. Fig. 3). This amounts to 6,410,983 real spectra in total. The **simulated HSI data** was generated with a Monte Carlo method (cf. algorithm provided in the supplementary material). This procedure resulted in 213,541 simulated spectra with annotated organ labels.

## 3 Experiments and Results

The purpose of the experiments was to investigate hypotheses H1 and H2 (cf. Sect. 1). As comparison methods, a CycleGAN [24] and an unsupervised image-to-image translation (UNIT) network [16] were implemented fully convolutionally for PAT and in an adapted version for the one-dimensional HSI data. To make the comparison fair, the tissue label conditions were concatenated with the input, and we put significant effort into optimizing the UNIT on our data.

*Realism of Synthetic Data (H1)* : According to qualitative analyses (Fig. 4) our domain transfer approach improves simulated PAT images with respect to key properties, including the realism of skin, background, and sharpness of vessels.

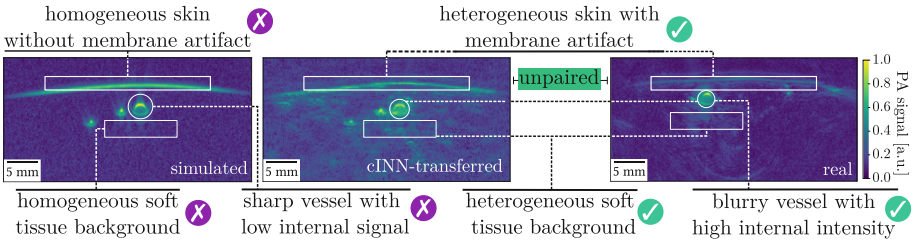


**Fig. 4. Qualitative results.** In comparison to simulated PAT images (left), images generated by the cINN (middle) resemble real PAT images (right) more closely. All images show a human forearm at 800 nm.

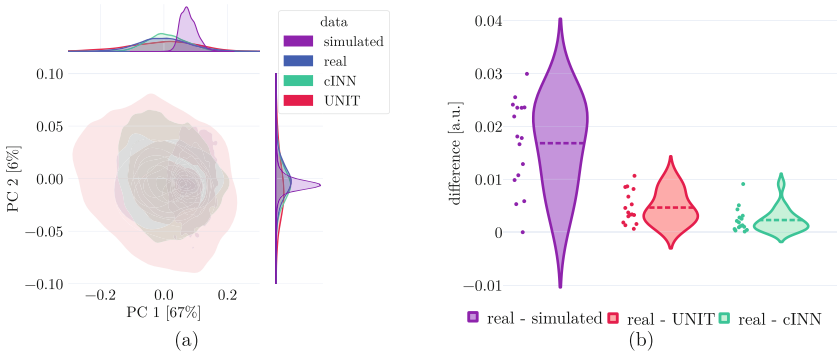


**Fig. 5. Our domain transfer approach yields realistic spectra (here: of veins).** The PCA plots in a) represent a kernel density estimation of the first and second components of a PCA embedding of the real data, which represent about 67% and 6% of the variance in the real data, respectively. The distributions on top and on the right of the PCA plot correspond to the marginal distributions of each dataset's first two components. b) Violin plots show that the cINN yields spectra that feature a smaller difference to the real data compared to the simulations and the UNIT-generated data. The dashed lines represent the mean difference value, and each dot represents the difference for one wavelength.

A principal component analysis (PCA) performed on all artery and vein spectra of the real and synthetic datasets demonstrates that the distribution of the synthetic data is much closer to the real data after applying our domain transfer approach (cf. Fig. 5a)). The same holds for the absolute difference, as shown in Fig. 5b). Slightly better performance was achieved with the cINN compared to the UNIT. Similarly, our approach improves the realism of HSI spectra, as illustrated in Fig. 6, for spectra of five exemplary organs (colon, stomach, omentum, spleen, and fat). The cINN-transferred spectra generally match the real data very closely. Failure cases where the real data has a high variance (translucent band) are also shown.

*Benefit of Domain-Transferred Data for Downstream Tasks (H2):* We examined two classification tasks for which reference data generation was feasible: classification of veins/arteries in PAT and organ classification in HSI. For both modalities, we used the completely untouched real test sets, comprising 162 images in the case of PAT and $\sim$ 920,000 spectra in the case of HSI. For both tasks, a calibrated random forest classifier (sklearn [19] with default parameters) was trained on the simulated, the domain-transferred (by UNIT and cINN), and real spectra. As metrics, the balanced accuracy (BA), area under receiver operating characteristic (AUROC) curve, and F1-score were selected based on [17].

As shown in Table 1, our domain transfer approach dramatically increases the classification performance for both downstream tasks. Compared to physics-based simulation, the cINN obtained a relative improvement of 37% (BA), 25% (AUROC), and 22% (F1 Score) for PAT whereas the UNIT only achieved a

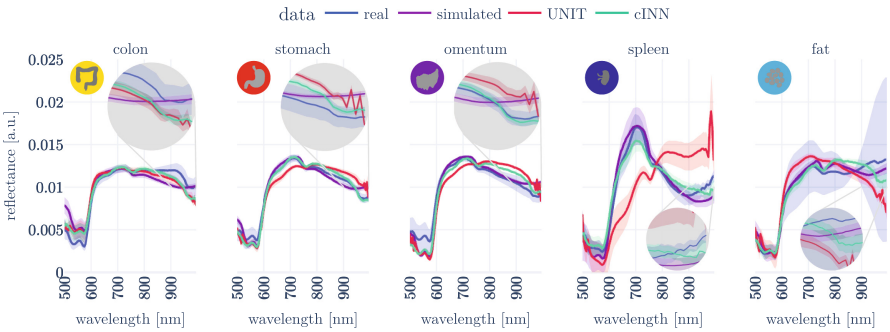


**Fig. 6. The cINN-transferred spectra are in closer agreement with the real spectra than the simulations and the UNIT-transferred spectra**. Spectra for five exemplary organs are shown from 500 nm to 1000 nm. For each subplot, a zoom-in for the near-infrared region (>900 nm) is shown. The translucent bands represent the standard deviation across spectra for each organ.

**Table 1. Classification scores for different training data.** The training data refers to real data, physics-based simulated data, data generated by a CycleGAN, by a UNIT without and with tissue labels ($\text{UNIT}_\text{Y}$), and by a cINN without ($\text{cINN}_\text{D}$) and with (proposed $\text{cINN}_\text{DY}$) tissue labels as condition. Additionally, $\text{cINN}_{\text{DY without GAN}}$ refers to a $\text{cINN}_\text{DY}$ without the adversarial training. The best-performing methods, except if trained on real data, are printed in **bold**.

| Classifier training data | PAT | | | HSI | | |
|---|---|---|---|---|---|---|
| | BA | AUROC | F1-Score | BA | AUROC | F1-Score |
| **Real** | 0.75 | 0.84 | 0.82 | 0.40 | 0.81 | 0.44 |
| **Simulated** | 0.52 | 0.64 | 0.64 | 0.24 | 0.75 | 0.18 |
| **CycleGAN** | 0.39 | 0.20 | 0.16 | 0.11 | 0.57 | 0.06 |
| **UNIT** | 0.50 | 0.44 | 0.65 | 0.20 | 0.72 | 0.20 |
| **$\text{UNIT}_\text{Y}$** | 0.64 | **0.81** | 0.77 | 0.24 | 0.74 | 0.25 |
| **$\text{cINN}_\text{D}$** | 0.66 | 0.73 | 0.72 | 0.25 | 0.72 | 0.20 |
| **$\text{cINN}_{\text{DY without GAN}}$** | 0.65 | 0.78 | 0.76 | 0.28 | 0.75 | **0.26** |
| **$\text{cINN}_\text{DY}$ (proposed)** | **0.71** | 0.80 | **0.78** | **0.29** | **0.76** | 0.24 |

relative improvement in the range of 20%-27% (depending on the metric). For HSI, the cINN achieved a relative improvement of 21% (BA), 1% (AUROC), and 33% (F1 Score) and it scored better in all metrics except for the F1 Score than the UNIT. For all metrics, training on real data still yields better results.

## 4 Discussion

With this paper, we presented the first domain transfer approach that combines the benefits of cINNs (exact maximum likelihood estimation) with those

of GANs (high image quality). A comprehensive validation involving qualitative and quantitative measures for the remaining domain gap and downstream tasks suggests that the approach is well-suited for sim-to-real transfer in spectral imaging. For both PAT and HSI, the domain gap between simulations and real data could be substantially reduced, and a dramatic increase in downstream task performance was obtained - also when compared to the popular UNIT approach.

The only similar work on domain transfer in PAT has used a cycle GAN-based architecture on a single wavelength with only photon propagation as PAT image simulator instead of full acoustic wave simulation and image reconstruction [14]. This potentially leads to spectral inconsistency in the sense that the spectral information either is lost during translation or remains unchanged from the source domain instead of adapting to the target domain. Outside the spectral/medical imaging community, Liu et al. [16] and Grover et al. [10] tasked variational autoencoders and invertible neural networks for each domain, respectively, to create the shared encoding. They both combined this approach with adversarial training to achieve high-quality image generation. Das et al. [5] built upon this approach by using labels from the source domain to condition the domain transfer task. In contrast to previous work, which used en-/decoders for each domain, we train a single network as shown in Fig. 2. with a two-fold condition consisting of a domain label ($D$) and a tissue label ($Y$) from the source domain, which has the advantage of explicitly aiding the spectral domain transfer.

The main limitation of our approach is the high dimensionality of the parameter space of the cINN as dimensionality reduction of data is not possible due to the information and volume-preserving property of INNs. This implies that the method is not suitable for arbitrarily high dimensions. Future work will comprise the rigorous validation of our method with tissue-mimicking phantoms for which reference data are available.

In conclusion, our proposed approach of cINN-based domain transfer enables the generation of realistic spectral data. As it is not limited to spectral data, it could develop into a powerful method for domain transfer in the absence of labeled real data for a wide range of image modalities in the medical domain and beyond.

**Acknowledgements.** This project was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (NEURAL SPICING, 101002198) and the Surgical Oncology Program of the National Center for Tumor Diseases (NCT) Heidelberg.

## References

1. Adler, T.J., et al.: Uncertainty-aware performance assessment of optical imaging modalities with invertible neural networks. Int. J. Comput. Assist. Radiol. Surg. **14**(6), 997–1007 (2019). https://doi.org/10.1007/s11548-019-01939-9
2. Ardizzone, L., Lüth, C., Kruse, J., Rother, C., Köthe, U.: Conditional invertible neural networks for guided image generation (2020)

3. Ayala, L., et al.: Spectral imaging enables contrast agent-free real-time ischemia monitoring in laparoscopic surgery. Sci. Adv. (2023). https://doi.org/10.1126/sciadv.add6778
4. Clancy, N.T., Jones, G., Maier-Hein, L., Elson, D.S., Stoyanov, D.: Surgical spectral imaging. Med. Image Anal. **63**, 101699 (2020)
5. Das, H.P., Tran, R., Singh, J., Lin, Y.W., Spanos, C.J.: Cdcgen: cross-domain conditional generation via normalizing flows and adversarial training. arXiv preprint arXiv:2108.11368 (2021)
6. De Fauw, J., Ledsam, J.R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., et al.: Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat. Med. **24**(9), 1342–1350 (2018)
7. Dinh, L., Sohl-Dickstein, J., Bengio, S.: Density estimation using real nvp. arXiv preprint arXiv:1605.08803 (2016)
8. Gröhl, J., et al.: Simpa: an open-source toolkit for simulation and image processing for photonics and acoustics. J. Biomed. Opt. **27**(8), 083010 (2022)
9. Gröhl, J., Schellenberg, M., Dreher, K., Maier-Hein, L.: Deep learning for biomedical photoacoustic imaging: a review. Photoacoustics **22**, 100241 (2021)
10. Grover, A., Chute, C., Shu, R., Cao, Z., Ermon, S.: Alignflow: cycle consistent learning from multiple domains via normalizing flows. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 4028–4035 (2020)
11. Haar, A.: Zur theorie der orthogonalen funktionensysteme. Mathematische Annalen **71**(1), 38–53 (1911)
12. Hoffman, J., Tzet al.: Cycada: cycle-consistent adversarial domain adaptation. In: International Conference on Machine Learning, pp. 1989–1998 (2018)
13. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net a self-configuring method for deep learning-based biomedical image segmentation. Nat. Methods **18**(2), 203–211 (2021)
14. Li, J., et al.: Deep learning-based quantitative optoacoustic tomography of deep tissues in the absence of labeled experimental data. Optica **9**(1), 32–41 (2022)
15. Li, K., Zhang, Y., Li, K., Fu, Y.: Adversarial feature hallucination networks for few-shot learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13470–13479 (2020)
16. Liu, M.Y., Breuel, T., Kautz, J.: Unsupervised image-to-image translation networks. Adv. Neural Inf. Process. Syst. **30** (2017)
17. Maier-Hein, L., Reinke, A., Godau, P., Tizabi, M.D., Büttner, F., Christodoulou, E., et al.: Metrics reloaded: pitfalls and recommendations for image analysis validation (2022). https://doi.org/10.48550/ARXIV.2206.01653
18. Mao, X., Li, Q., Xie, H., Lau, R.Y., Wang, Z., Paul Smolley, S.: Least squares generative adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2794–2802 (2017)
19. Pedregosa, F., et al.: Scikit-learn: machine learning in python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
20. Schellenberg, M., et al.: Semantic segmentation of multispectral photoacoustic images using deep learning. Photoacoustics **26**, 100341 (2022). https://doi.org/10.1016/j.pacs.2022.100341
21. Schellenberg, M., et al.: Photoacoustic image synthesis with generative adversarial networks. Photoacoustics **28**, 100402 (2022)
22. Wang, X., Xie, X., Ku, G., Wang, L.V., Stoica, G.: Noninvasive imaging of hemoglobin concentration and oxygenation in the rat brain using high-resolution photoacoustic tomography. J. Biomed. Opt. **11**(2), 024015 (2006)

23. Wirkert, S.J., et al.: Physiological parameter estimation from multispectral images unleashed. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10435, pp. 134–141. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66179-7_16
24. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2223–2232 (2017)