Secondary Publication

Doerrich, Sebastian; Di Salvo, Francesco; Ledig, Christian

unORANIC : Unsupervised Orthogonalization of Anatomy and Image-Characteristic Features

Date of secondary publication: 11.03.2025

Submitted Version (Preprint), Conferenceobject

Persistent identifier: urn:nbn:de:bvb:473-irb-1069345

Primary publication

Doerrich, S.; Di Salvo, F.; Ledig, C. (2023): unORANIC : Unsupervised Orthogonalization of Anatomy and Image-Characteristic Features, in: X. Cao, X. Xu, I. Rekik, u. a. (Hrsg.), Machine Learning in Medical Imaging : 14th International Workshop, MLMI 2023, Held in Conjunction with MICCAI 2023, Vancouver, BC, Canada, October 8, 2023 ; Proceedings, Part I, Cham: Springer Nature Switzerland, S. 62–71, doi: 10.1007/978-3-031-45673-2_7.

Legal Notice

This work is protected by copyright and/or the indication of a licence. You are free to use this work in any way permitted by the copyright and/or the licence that applies to your usage. For other uses, you must obtain permission from the rights-holders.

This document is made available under a Creative Commons license.

The license information is available online:
https://creativecommons.org/licenses/by/4.0/legalcode

arXiv:2308.15507v1 [eess.IV] 29 Aug 2023

unORANIC: Unsupervised Orthogonalization of Anatomy and Image-Characteristic Features

Sebastian Doerrich, Francesco Di Salvo, and Christian Ledig

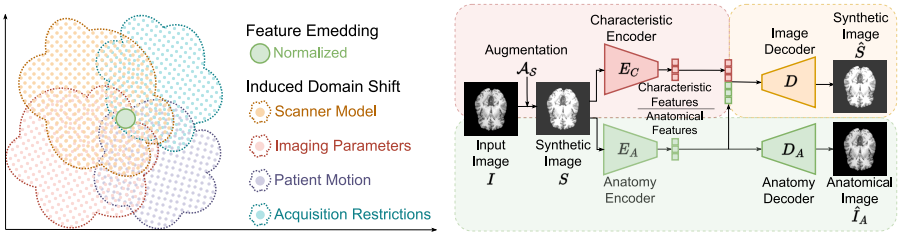
xAILab, University of Bamberg, Germany
sebastian.doerrich@uni-bamberg.de

Abstract. We introduce unORANIC, an unsupervised approach that uses an adapted loss function to drive the orthogonalization of anatomy and image-characteristic features. The method is versatile for diverse modalities and tasks, as it does not require domain knowledge, paired data samples, or labels. During test time unORANIC is applied to potentially corrupted images, orthogonalizing their anatomy and characteristic components, to subsequently reconstruct corruption-free images, showing their domain-invariant anatomy only. This feature orthogonalization further improves generalization and robustness against corruptions. We confirm this qualitatively and quantitatively on 5 distinct datasets by assessing unORANIC's classification accuracy, corruption detection and revision capabilities. Our approach shows promise for enhancing the generalizability and robustness of practical applications in medical image analysis. The source code is available at github.com/sdoerrich97/unORANIC.

Keywords: Feature Orthogonalization · Robustness · Corruption Revision · Unsupervised learning · Generalization

1 Introduction

In recent years, deep learning algorithms have shown promise in medical image analysis, including segmentation [24,23], classification [8,17], and anomaly detection [19,10]. However, their generalizability across diverse imaging domains remains a challenge [7] especially for their adoption in clinical practice [25] due to domain shifts caused by variations in scanner models [12], imaging parameters [13], corruption artifacts [21], or patient motion [20]. A schematic representation of this issue is presented in Fig. 1a. To address this, two research areas have emerged: domain adaptation (DA) and domain generalization (DG). DA aligns feature distributions between source and target domains [14], while DG trains models on diverse source domains to learn domain-invariant features [15]. Within this framework, some methods aim to disentangle anatomy features from modality factors to improve generalization. Chartsias et al. introduce SDNet, which uses segmentation labels to factorize 2D medical images into spatial anatomical and non-spatial modality factors for robust cardiac image segmentation [4]. Robert et al. present HybridNet that utilizes a two-branch encoder-decoder architecture to learn invariant class-related representations via a supervised train-



(a) (b)

Fig. 1: **a)** Schematic visualization of a domain disparity caused by variations in the imaging process. **b)** Visualization of unORANIC's fundamental idea.

ing concept [22]. Dewey et al. propose a deep learning-based harmonization technique for MR images under limited supervision to standardize across scanners and sites [6]. In contrast, Zuo et al. propose unsupervised MR image harmonization to address contrast variations in multi-site MR imaging [29] by using T1- and T2-weighted image pairs of the same patient. However, all of these approaches are constrained by either requiring a certain type of supervision, precise knowledge of the target domain, or specific inter-/intra-site paired data samples.

To address these limitations, we present unORANIC, an approach to orthogonalize anatomy and image-characteristic features in an unsupervised manner, without requiring domain knowledge, paired data, or labels of any kind. The method enables bias-free anatomical reconstruction and works for a diverse set of modalities. For that scope, we jointly train two encoder-decoder branches. One branch is used to extract true anatomy features and the other to model the characteristic image information discarded by the first branch to reconstruct the input image in an autoencoder objective. A high-level overview of the proposed approach is provided in Fig. 1b. Experimental results on a diverse dataset demonstrate its feasibility and potential for improving generalization and robustness in medical image analysis.

2 Methodology

For the subsequent sections, we consider the input images as bias-free and uncorrupted (I). We further define $\mathcal{A}_\mathcal{S}$ as a random augmentation that distorts an input image I for the purpose to generate a synthetic, corrupted version S of that image. As presented in Fig. 1b, such a synthetic image S is obtained via the augmentation $\mathcal{A}_\mathcal{S}$ applied to I and subsequently fed to both the anatomy encoder E_A and the characteristic encoder E_C simultaneously. The resulting embeddings are concatenated ($\oplus$) and forwarded to a convolutional decoder D to create the reconstruction $\hat{S}$ with its specific characteristics such as contrast level or brightness. By removing these characteristic features in the encoded embeddings of E_A, we can reconstruct a distortion-free version ($\hat{I}_A$) of the original input image I. To allow this behavior, the anatomy encoder, E_A, is actively

enforced to learn acquisition- and corruption-robust representations while the characteristic encoder E_C retains image-specific details.

2.1 Training

During training the anatomy encoder is shared across the corrupted image S as well as additional corrupted variants v_i of the same input image I to create feature embeddings for all of them. By applying the consistency loss $\mathcal{L}_{\mathrm{C}}$ given as:

$$\mathcal{L}_{\mathrm{C}} = \frac{1}{Z}\left(\sum_{\forall\,\{v_i,v_j\}\in V,\ v_i\neq v_j} \left|\left|E^A(v_i) - E^A(v_j)\right|\right|_2\right) \quad \text{with } Z = \binom{V}{2} \tag{1}$$

on the resulting feature maps, the anatomy encoder is forced to learn distortion-invariant features. Here, V is the set of all variants including S, v_i and v_j are two distinct variants as well as $E^A(v_i)$ and $E^A(v_j)$ are the corresponding anatomy feature embeddings. To further guide the anatomy encoder to learn representative anatomy features, the consistency loss is assisted by a combination of the reconstruction loss of the synthetic image $\mathcal{L}_{\mathrm{R}_S}$ and the reconstruction loss of the original, distortion-free image $\mathcal{L}_{\mathrm{R}_I}$ given as:

$$\mathcal{L}_{\mathrm{R}_S} = \frac{1}{NM}\left|\left|S - \hat{S}\right|\right|_2 = \frac{1}{NM}\left|\left|S - D\left(E_A(S) \oplus E_C(S)\right)\right|\right|_2 \tag{2}$$

$$\mathcal{L}_{\mathrm{R}_I} = \frac{1}{NM}\left|\left|I - \hat{I}_A\right|\right|_2 = \frac{1}{NM}\left|\left|I - D_A(E_A(I))\right|\right|_2 \tag{3}$$

with N and M being the image height and width respectively. Using $\mathcal{L}_{\mathrm{R}_I}$ to update the encoder E_A, as well as the decoder D_A, allows the joint optimization of both. This yields more robust results than updating only the decoder with that loss term. The complete loss for the anatomy branch is therefore given as:

$$\mathcal{L}_{\mathrm{total}} = \lambda_{\mathrm{reconstruction}}\left(\mathcal{L}_{\mathrm{R}_I} + \mathcal{L}_{\mathrm{R}_S}\right) + \lambda_{\mathrm{consistency}}\,\mathcal{L}_{\mathrm{C}} \tag{4}$$

where $\lambda_{\mathrm{reconstruction}}$ and $\lambda_{\mathrm{consistency}}$ are two non-negative numbers that control the trade-off between the reconstruction and consistency loss. In contrast, the characteristic encoder E_C and decoder D are only optimized via the reconstruction loss of the synthetic image $\mathcal{L}_{\mathrm{R}_S}$.

2.2 Implementation

The entire training pipeline is depicted in Fig. 2. Both encoders consist of four identical blocks in total, where each block comprises a residual block followed by a downsampling block. Each residual block itself consists of twice a convolution followed by batch normalization and the Leaky ReLU activation function. The downsampling blocks use a set of a strided convolution, a batch normalization, and a Leaky ReLU activation to half the image dimension while doubling the

channel dimension. In contrast, the decoders mirror the encoder architecture by replacing the downsampling blocks with corresponding upsampling blocks for which the convolutions are swapped with transposed convolutional layers. During each iteration, the input image I is distorted using augmentation $\mathcal{A}_S$ to generate S and subsequently passed through the shared anatomy encoder E_A in combination with two different distorted variants V_1 and V_2. These variants comprise the same anatomical information but different distortions than S. The consistency loss $\mathcal{L}_{\mathrm{C}}$ is computed using the encoded feature embeddings, and the reconstruction loss of the original, distortion-free image $\mathcal{L}_{\mathrm{R}_I}$ is calculated by passing the feature embedding of S through the anatomy decoder D_A to generate the anatomical reconstruction $\hat{I}_A$. Furthermore, the anatomy and characteristic feature embeddings are concatenated and used by the image decoder D to reconstruct $\hat{S}$ for the calculation of $\mathcal{L}_{\mathrm{R}_S}$. $\mathcal{A}_S$, $\mathcal{A}_{v_1}$ and $\mathcal{A}_{v_2}$ are independent of each other and can be chosen randomly from a set of augmentations such as the Albumentations library [3]. The network is trained until convergence using a batch size of 64 as well as the Adam optimizer with a cyclic learning rate. Experiments were performed on 28×28 pixel images with a latent dimension of 256, but the approach is adaptable to images of any size. The number of variants used for the anatomy encoder training is flexible, and in our experiments, three variants were utilized.

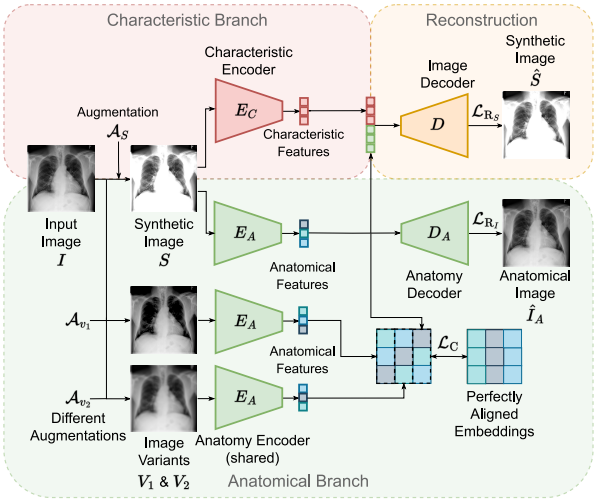


Fig. 2: Training pipeline of unORANIC for CT images.

3 Experiments and Results

We evaluate the model's performance in terms of its reconstruction ability as well as its capability to revise existing corruptions within an image. Further,

we assess its applicability for the execution of downstream tasks by using the anatomy feature embedding, as well as its capacity to detect corruptions by using the characteristic feature embedding. Last, we rate the robustness of our model against different severity and types of corruption.

3.1 Dataset

To demonstrate the versatility of our method, all experiments were conducted on a diverse selection of datasets of the publicly available MedMNIST v2 benchmark[1][27,28]. The MedMNIST v2 benchmark is a comprehensive collection of standardized, 28 x 28 dimensional biomedical datasets, with corresponding classification labels and baseline methods. The selected datasets for our experiments are described in detail in Table 1.

Table 1: Details of the selected set of datasets

MedMNIST Dataset	Modality	Task (# Classes)	# Train / Val / Test
Blood [1]	Blood Cell Microscope	Multi-Class (8)	11,959 / 1,712 / 3,421
Breast [2]	Breast Ultrasound	Binary-Class (2)	546 / 78 / 156
Derma [26,5]	Dermatoscope	Multi-Class (7)	7,007 / 1,003 / 2,005
Pneumonia [11]	Chest X-Ray	Binary-Class (2)	4,708 / 524 / 624
Retina [16]	Fundus Camera	Ordinal Regression (5)	1,080 / 120 / 400

3.2 Reconstruction

We compared unORANIC's reconstruction results with a vanilla autoencoder (AE) architecture to assess its encoding and reconstruction abilities. The vanilla AE shares the same architecture and latent dimension as the anatomy branch of our model. The average peak signal-to-noise ratio (PSNR) values for the reconstructions of both methods on the selected datasets are presented in Table 2. Both models demonstrate precise reconstruction of the input images, with our model achieving a slight improvement across all datasets. This improvement is attributable to the concatenation of anatomy and characteristic features, resulting in twice the number of features being fed to the decoder compared to the AE model. The reconstruction quality of both models is additionally depicted visually in Fig. 3 using selected examples.

[1] Yang, J., et al. *MedMNIST v2-A large-scale lightweight benchmark for 2D and 3D biomedical image classification.* Scientific Data. 2023. License: CC BY 4.0. Zenodo. https://zenodo.org/record/6496656

Table 2: PSNR values of the reconstructions

Methods	Blood	Breast	Derma	Pneumonia	Retina
AE	30.837	28.725	38.125	34.787	35.872
unORANIC	**31.700**	**29.390**	**38.569**	**36.040**	**36.309**

(a) (b) (c)

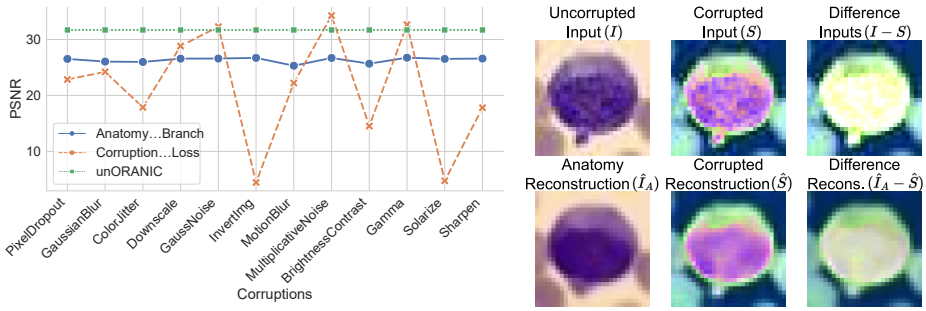
Fig. 3: **a)** Reconstruction unORANIC **b)** Original input **c)** Reconstructions AE.

3.3 Corruption Revision

We now evaluate the capability of unORANIC's anatomical branch to revise existing corruptions in an input image. For this, all input images in the test set $\{I_1, I_2, ..., I_n\}$ are intentionally corrupted using a set of corruptions from the Albumentations library [3] to generate synthetic distorted versions $\{S_1, S_2, ..., S_n\}$ first, before passing those through the unORANIC network afterward. Each corruption is hereby uniformly applied to all test images. Despite those distortions, the anatomy branch of the model should still be able to reconstruct the original, uncorrupted input images $\hat{I}_i$. To assess this, we compute the average PSNR value between the original images I_i and their corrupted versions S_i. Afterward, we compare this value to the average PSNR between the original uncorrupted input images I_i and unORANIC's anatomical reconstructions $\hat{I}_{A_i}$. Both of these values are plotted against each other in Fig. 4a, for the BloodMNIST dataset. For reference purposes, the figure contains the PSNR value for uncorrupted input images as well. It can be seen that the anatomical reconstruction quality is close to the uncorrupted one and overall consistent across all applied corruptions which proves unORANIC's corruption revision capability. As it can be seen in Fig. 4b, this holds true even for severe corruptions such as solarization.

3.4 Classification

We proceed to assess the representativeness of the encoded feature embeddings, comprising both anatomical and characteristic information, and their suitability for downstream tasks or applications. To this end, we compare our model with the ResNet-18 baseline provided by MedMNIST v2 on two distinct tasks. The first task involves the classification of each dataset, determining the type of disease (e.g., cancer or non-cancer). For the second task, the models are assigned to detect whether a corruption (the same ones used in the previous revision experiment) has or has not been applied to the input image in the form of a



(a) (b)

Fig. 4: **a)** Corruption revision results of unORANIC for a set of corruptions. **b)** Visualization of unORANIC's corruption revision for a solarization corruption. The method reconstructs I via $\hat{I}_A$ while seeing only the corrupted input S.

binary classification. To do so, we freeze our trained network architecture and train a single linear layer on top of each embedding for its respective task. The same procedure is applied to the AE architecture, while the MedMNIST baseline architecture is retrained specifically for the corruption classification task. Results are summarized in Table 3 and Table 4, with AUC denoting the Area under the ROC Curve and ACC representing the Accuracy of the predictions. Overall, our model's classification ability is not as strong as that of the baseline model. However, it is important to note that our model was trained entirely in an unsupervised manner, except for the additional linear layer, compared to the supervised baseline approach. Regarding the detection of corruptions within the input images, our model outperforms the baseline method. Furthermore, it should be emphasized that our model was trained simultaneously for both tasks, while the reference models were trained separately for each individual task.

Table 3: Classification results on each selected dataset

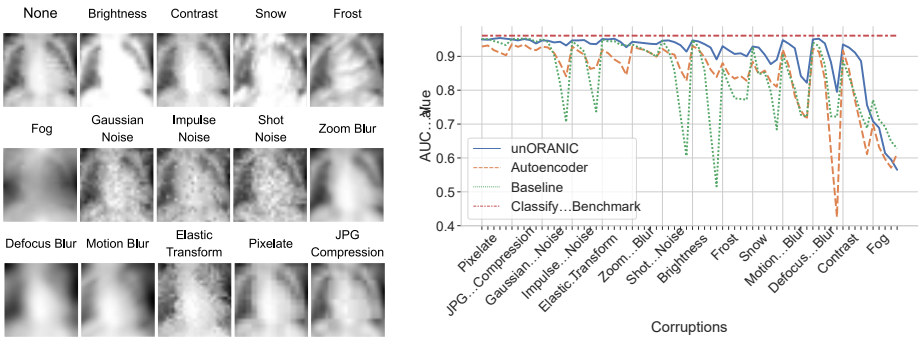
Methods	Blood		Breast		Derma		Pneumonia		Retina	
	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
Baseline	**0.998**	**0.958**	**0.901**	**0.863**	**0.917**	**0.735**	0.944	0.854	**0.717**	0.524
AE	0.966	0.806	0.819	0.801	0.787	0.694	0.904	0.840	0.678	0.470
unORANIC	0.977	0.848	0.812	0.808	0.776	0.699	**0.961**	**0.862**	0.691	**0.530**

3.5 Corruption Robustness

Lastly, we assess unORANIC's robustness against unseen corruptions during training. To accomplish this, we adopt the image corruptions introduced by

Table 4: Corruption detection on each selected dataset

Methods	Blood		Breast		Derma		Pneumonia		Retina	
	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
Baseline	0.548	0.184	0.565	0.571	0.517	0.046	0.614	0.524	0.500	0.283
AE	0.746	0.736	0.576	0.545	**0.655**	**0.640**	0.648	0.657	0.827	0.785
unORANIC	**0.755**	**0.746**	**0.612**	**0.590**	0.643	0.620	**0.667**	**0.660**	**0.847**	**0.823**



(a) (b)

Fig. 5: **a)** Visualization of the applied corruptions. **b)** Classification results of unORANIC compared to the vanilla AE and the baseline when exposed to the applied corruptions with increasing severity.

[9,18], following a similar methodology as outlined in Section 3.3. The used corruptions are presented in Fig. 5a for an example of the PneumoniaMNIST dataset. Sequentially, we apply each corruption to all test images first and subsequently pass these corrupted images through the trained unORANIC, vanilla AE and baseline model provided by the MedMNIST v2 benchmark for the datasets associated classification task. This process is repeated for each individual corruption. Additionally, we vary the severity of each corruption on a scale of 1 to 5, with 1 representing minor corruption and 5 indicating severe corruption. By collecting the AUC values for each combination of corruption and severity on the PneumoniaMNIST dataset and plotting them for each method individually, we obtain the results presented in Fig. 5b. Notably, our model demonstrates greater overall robustness to corruption, particularly for noise, compared to both the baseline and the simple autoencoder architecture.

4 Conclusion

The objective of this study was to explore the feasibility of orthogonalizing anatomy and image-characteristic features in an unsupervised manner, without requiring specific dataset configurations or prior knowledge about the data. The

conducted experiments affirm that unORANIC's simplistic network architecture can explicitly separate these two feature categories. This capability enables the training of corruption-robust architectures and the reconstruction of unbiased anatomical images. The findings from our experiments motivate further investigations into extending this approach to more advanced tasks, architectures as well as additional datasets, and explore its potential for practical applications.

References

1. Acevedo, A., Merino, A., Alférez, S., Ángel Molina, Boldú, L., Rodellar, J.: A dataset of microscopic peripheral blood cell images for development of automatic recognition systems. Data in Brief **30**, 105474 (2020)
2. Al-Dhabyani, W., Gomaa, M., Khaled, H., Fahmy, A.: Dataset of breast ultrasound images. Data in Brief **28**, 104863 (2020)
3. Buslaev, A., Iglovikov, V.I., Khvedchenya, E., Parinov, A., Druzhinin, M., Kalinin, A.A.: Albumentations: Fast and flexible image augmentations. Information 2020, Vol. 11, Page 125 **11**, 125 (2020)
4. Chartsias, A., Joyce, T., Papanastasiou, G., Semple, S., Williams, M., Newby, D.E., Dharmakumar, R., Tsaftaris, S.A.: Disentangled representation learning in cardiac image analysis. Medical Image Analysis **58** (2019)
5. Codella, N., Rotemberg, V., Tschandl, P., Celebi, M.E., Dusza, S., Gutman, D., Helba, B., Kalloo, A., Liopyris, K., Marchetti, M., Kittler, H., Halpern, A.: Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic) (2019)
6. Dewey, B.E., Zuo, L., Carass, A., He, Y., Liu, Y., Mowry, E.M., Newsome, S., Oh, J., Calabresi, P.A., Prince, J.L.: A disentangled latent space for cross-site mri harmonization (2020)
7. Eche, T., Schwartz, L.H., Mokrane, F.Z., Dercle, L.: Toward generalizability in the deployment of artificial intelligence in radiology: Role of computation stress testing to overcome underspecification. Radiology: Artificial Intelligence **3** (2021)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition **2016-December**, 770–778 (2016)
9. Hendrycks, D., Dietterich, T.: Benchmarking neural network robustness to common corruptions and perturbations. Proceedings of the International Conference on Learning Representations (2019)
10. Jeong, J., Zou, Y., Kim, T., Zhang, D., Ravichandran, A., Dabeer, O.: Winclip: Zero-/few-shot anomaly classification and segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 19606–19616 (2023)
11. Kermany, D.S., et al.: Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell **172**, 1122–1131.e9 (2018)
12. Khan, A., Janowczyk, A., Müller, F., Blank, A., Nguyen, H.G., Abbet, C., Studer, L., Lugli, A., Dawson, H., Thiran, J.P., Zlobec, I.: Impact of scanner variability on lymph node segmentation in computational pathology. Journal of Pathology Informatics **13**, 100127 (2022)
13. Lafarge, M.W., Pluim, J.P., Eppenhof, K.A., Moeskops, P., Veta, M.: Domain-adversarial neural networks to address the appearance variability of histopathology images. Lecture Notes in Computer Science **10553 LNCS**, 83–91 (2017)

14. Li, B., Wang, Y., Zhang, S., Li, D., Keutzer, K., Darrell, T., Zhao, H.: Learning invariant representations and risks for semi-supervised domain adaptation. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition pp. 1104–1113 (2021)
15. Li, D., Yang, Y., Song, Y.Z., Hospedales, T.M.: Learning to generalize: Meta-learning for domain generalization. Proceedings of the AAAI Conference on Artificial Intelligence **32**, 3490–3497 (2018)
16. Liu, R., Wang, X., Wu, Q., Dai, L., Fang, X., Yan, T., Son, J., Tang, S., Li, J., Gao, Z., Galdran, A., Poorneshwaran, J.M., Liu, H., Wang, J., Chen, Y., Porwal, P., Tan, G.S.W., Yang, X., Dai, C., Song, H., Chen, M., Li, H., Jia, W., Shen, D., Sheng, B., Zhang, P.: Deepdrid: Diabetic retinopathy—grading and image quality estimation challenge. Patterns **3**, 100512 (2022)
17. Manzari, O.N., Ahmadabadi, H., Kashiani, H., Shokouhi, S.B., Ayatollahi, A.: Medvit: A robust vision transformer for generalized medical image classification. Computers in Biology and Medicine **157**, 106791 (2023)
18. Michaelis, C., Mitzkus, B., Geirhos, R., Rusak, E., Bringmann, O., Ecker, A.S., Bethge, M., Brendel, W.: Benchmarking robustness in object detection: Autonomous driving when winter is coming. arXiv preprint arXiv:1907.07484 (2019)
19. Ngo, P.C., Winarto, A.A., Kou, C.K.L., Park, S., Akram, F., Lee, H.K.: Fence gan: Towards better anomaly detection. Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI **2019-November**, 141–148 (2019)
20. Oksuz, I., Clough, J.R., Ruijsink, B., Anton, E.P., Bustin, A., Cruz, G., Prieto, C., King, A.P., Schnabel, J.A.: Deep learning-based detection and correction of cardiac mr motion artefacts during reconstruction for high-quality segmentation. IEEE Transactions on Medical Imaging **39**, 4001–4010 (2020)
21. Priyanka, Kumar, D.: Feature extraction and selection of kidney ultrasound images using glcm and pca. Procedia Computer Science **167**, 1722–1731 (2020)
22. Robert, T., Thome, N., Cord, M.: Hybridnet: Classification and reconstruction cooperation for semi-supervised learning (2018)
23. Rondinella, A., Crispino, E., Guarnera, F., Giudice, O., Ortis, A., Russo, G., Lorenzo, C.D., Maimone, D., Pappalardo, F., Battiato, S.: Boosting multiple sclerosis lesion segmentation through attention mechanism. Computers in Biology and Medicine **161**, 107021 (2023)
24. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) **9351**, 234–241 (2015)
25. Stacke, K., Eilertsen, G., Unger, J., Lundstrom, C.: Measuring domain shift for deep learning in histopathology. IEEE Journal of Biomedical and Health Informatics **25**, 325–336 (2021)
26. Tschandl, P., Rosendahl, C., Kittler, H.: The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific Data 2018 5:1 **5**, 1–9 (2018)
27. Yang, J., Shi, R., Ni, B.: Medmnist classification decathlon: A lightweight automl benchmark for medical image analysis. Proceedings - International Symposium on Biomedical Imaging **2021-April**, 191–195 (2020)
28. Yang, J., Shi, R., Wei, D., Liu, Z., Zhao, L., Ke, B., Pfister, H., Ni, B.: Medmnist v2 - a large-scale lightweight benchmark for 2d and 3d biomedical image classification. Scientific Data 2023 10:1 **10**, 1–10 (2023)

29. Zuo, L., Dewey, B.E., Liu, Y., He, Y., Newsome, S.D., Mowry, E.M., Resnick, S.M., Prince, J.L., Carass, A.: Unsupervised mr harmonization by learning disentangled representations using information bottleneck theory. NeuroImage **243**, 118569 (2021)