Robust Cardiac MRI Segmentation with Data-Centric Models to Improve Performance via Intensive Pre-training and Augmentation

Shizhan Gong[1], Weitao Lu[2], Jize Xie[2], Xiaofan Zhang[2,3], Shaoting Zhang[2,3], Qi Dou[1]

[1]Dept. of Computer Science and Engineering, The Chinese University of Hong Kong
[2]School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University
[3]Shanghai Artificial Intelligence Laboratory

Abstract. Segmentation of anatomical structures from Cardiac Magnetic Resonance (CMR) is central to the non-invasive quantitative assessment of cardiac function and structure, and deep-learning-based automatic segmentation models prove to have satisfying performance. However, patients' respiratory motion during the scanning process can greatly degenerate the quality of CMR images, resulting in a serious performance drop for deep learning algorithms. Building a robust cardiac MRI segmentation model is one of the keys to facilitating the use of deep learning in practical clinic scenarios. To this end, we experiment with several network architectures and compare their segmentation accuracy and robustness to respiratory motion. We further pre-train our network on large publicly available CMR datasets and augment our training set with adversarial augmentation, both methods bring significant improvement. We evaluate our methods on the cine MRI dataset of the CMRxMotion challenge and obtain promising performance for the segmentation of the left ventricle, left ventricular myocardium, and right ventricle.

Keywords: Cardiac segmentation · Network Pre-train · Robust learning

1 Introduction

Cardiac Magnetic Resonance (CMR) imaging is the golden standard sequence for non-invasive evaluation of cardiac anatomical structures and functionalities [1]. Anatomical segmentation allows for analysis of what is of interest to the clinicians. By discarding irrelevant information, the images can be smaller in size, which can reduce the post-processing time and computing power for the downstream analysis. It is also a crucial pre-requisite for the calculation of several image-based bio-markers [2–4] with diagnostic value. Recently, with the development of artificial intelligence, fully-automatic segmentation algorithms based on deep learning start to surpass manual segmentation with faster speed, less subjective bias, and comparable or even higher accuracy. However, in clinical

practice, the model performance highly relies on image qualities. For CMR acquisition, respiration motion is one of the major causes of degenerated image qualities, as it may be difficult for certain patients with acute symptoms to follow the instructions and hold their breath for a long time during the scan. Images contaminated by respiratory motion have seen a significant drop in model performance and result in obvious failure cases.

Currently, most automatic cardiac segmentation models are based on deep learning methods, which learn a function to map the input images to the segmentation masks. This method highly relies on the quantity and quality of the training data, and is based on the assumption that the hold-out data has alike distribution to the training data. In practical clinic scenarios, the number of available training data is limited, and the quality of most images is high to guarantee diagnostic value. With such training data as supervision, the model is trained to only perform well on clean images and can easily fail when encountering images with low quality.

Pre-training and data augmentation are two important data-driven methods which have proven effect in improving model robustness. These two methods focus on enlarging the training data and better utilizing the existing data respectively. Pre-training exposes the model to a large dataset other than the training set, so as to broaden the model's horizon. It is recognized to yield better results than training from scratch [21, 22]. Although some researchers argue that pre-training offers little benefit for certain tasks or light weight architectures [23, 24], it is still undeniable that pre-training can enhance the model robustness and improve its performance on hold-out individuals [25, 26]. Data augmentation is another standard trial to build robust segmentation models. It exposes the network to higher variability through perturbations of the training data. This includes native approaches such as cropping, rotation, and flipping. As respiration motion will cause spatial transformation of the anatomic structure, deformation-based augmentation [6] is also rewarding for training. A recently proposed adversarial data augmentation method [5] can generate plausible and realistic signal corruptions that are difficult for the models to analyze and therefore increase the model's adversarial robustness.

In this work, we explore these two data-driven approaches in the context of the CMRxMotion challenge. The main goal of this challenge is to build a model for segmentation of the left ventricle (LV), left ventricular myocardium (MYO), and right ventricle (RV), based on limited training data. This model should be robust under different levels of respiration motion as the test data composes of images with diverse quality levels. We pre-train our model on large publicly available datasets with the same tasks, and increase the data variability through both random augmentation and adversarial augmentation. We find that, through intensive pre-training and strong data augmentation, even without novel DCNN architectures, the model can still reveal high robustness towards the qualities of the images, regardless of reasons for low-quality images, such as respiration motion.

2 Methods

We first give a brief introduction to the dataset of the CMRxMotion challenge and then explain our proposed approaches in detail.

2.1 Dataset

The CMRxMotion dataset consists of short-axis (SA) cine MRI acquisitions of 45 healthy volunteers. Each volunteer is trained to act in 4 manners during the scanning process, namely a) stick to the breath-hold instructions, b) halve the breath-hold period, c) breathe freely, and d) conduct intensive breathing. The pixel sizes vary from ~ 0.66 to ~ 0.76 mm, the image resolution ranges from 400 to 512 pixels, the number of slices is between 9 and 13, and the slice thickness ranges from ~ 9.6 mm to 10 mm. For images with diagnostic qualities, LV, MYO, and RV at the end of systole (ES) and end of diastole (ED) are manually segmented by radiologists. Exams of 20 volunteers with both images and ground-truth segmentation as well as 5 volunteers with only images are released for training and validation respectively. The remaining 20 volunteers are withheld for testing.

2.2 Network Architecture

We try three variants of U-Net: nnU-Net [14], Swin-UNETR [13], Swin-UNet [8].

nnU-Net [14] is a pure CNN-based method, which is good at catching the local pattern of the image. One modification we make is to replace convolution operation at the bottleneck layer with deformable convolution [10], which has shown increases in performance since it allows for a flexible receptive field [20]. The sampling offset learned by deformable convolutions is expected to counteract some of the shifts caused by respiratory motion.

On the contrary, Swin-UNet [8] is a pure transformer-based network, which is better at capturing global features through self-attention and shifted windows [11]. However, the segmentation result of raw Swin-UNet usually contains zigzag margins since Swin-UNet uses patch rather than pixel as the smallest unit to operate on. To overcome this limitation, we add two convolutional layers with layer normalization and leaky-relu at the end of the network, which helps produce smooth segmentation results.

Swin-UNETR [13] is a combination of transformer-based encoder and CNN-based decoder. which is expected to utilize the global information of the images and generate a refined segmentation map.

The input volumes only contain around 10 slices, which is not enough to be divided into multiple patches and windows. Therefore, for nnU-Net, we try both 2D and 3D variants, while for Swin-UNETR and Swin-UNet, only the 2D version is used.

2.3 Pre-training

Our training data is only composed of 20 healthy volunteers, which is too few for the model to even exhibit strong robustness towards the inherent variations of anatomical structures among different people, let alone unstable image quality or perturbations. Pre-training has proven to be effective in improving model performance as well as enhancing its robustness. To increase the model robustness towards unseen data with different appearances and qualities, we collect five public datasets of cine MRI with the same segmentation tasks listed in Table 1 for pre-training. We fuse these datasets and use labeled images from both the training phase and testing phase for pre-training. Moreover, pure transformer-based methods (i.e. Swin-Unet) is additionally pre-trained with ImageNet [19]. Although no deliberate respiration motion is conducted during the acquisition of these images, they exhibit significant variations in many other aspects including but not limited to scanner type, acquisition center, protocols, and health condition of the subjects. We believe these variations allow the model to ignore pixel-wise noisy signals and turn to catch the essential and high-level features useful for cardiac segmentation.

Table 1. Public CMR dataset used for pre-training

Dataset	Number of cases	Number of labeled SA images per case	Note
ACDC [15]	100	2	
M&Ms [16]	344	2	
M&Ms-2 [16]	360	2	use only SA images
MyoPS [17, 18]	25	1	use only bSSFP
MS-CMRSeg [17, 18]	45	1	use only bSSFP

2.4 Data Augmentation

We use both random data augmentation and adversarial data augmentation.

For random data augmentation, we follow the same schema as the default mode of nnU-Net [14], which contains rotation, scaling, Gaussian noise, Gaussian blur, brightness, contrast, simulation of low resolution, gamma correction, and mirroring.

As the inherent property of our data is that it is contaminated by respiratory motion, which can result in diffeomorphic deformation or spatial transformation of the raw image. Adversarial data augmentation proves to be more effective than random data augmentation in terms of improving model robustness towards a certain type of perturbations [7]. In adversarial data augmentation, given a model, the optimal perturbation of certain types is learned which will impair the model performance to the utmost extent. The model is then trained

to resist this perturbation, which in our case, can be deformation caused by respiration motion. In this work, we apply AdvChain [5] to improve our model's robustness towards diffeomorphic deformation and spatial transformation. To be more specific, we first train the network with random data augmentation, then fine-tune it with AdvChain. In each iteration of the fine-tuning phase, we turn off Gaussian noise, rotation, and scaling from the random data augmentation while keeping the rest. A perturbation consists of a series of Gaussian noise, spatial transformation, and diffeomorphic deformation is randomly generated, whose parameters are trainable. We freeze the network parameters and optimize the perturbation parameters to increase the consistency loss (MSE and contour loss [9] in our case) between two predicted segmentation maps before and after the perturbation. Finally, we fix this perturbation and update the network parameters by propagating supervised loss together with consistency loss. The above steps are conducted for multiple iterations until convergence.

2.5 Training Protocol

Pre-processing methods follow a similar framework to the default mode of nnU-Net [14], which consists of adjusting the pixel size of all images to ~ 0.66 with third-order spline interpolation, cropping or padding the resulting images to the resolution of 512×512 pixels, normalization, and intensity clipping (0.5 and 99.5 percentiles). For transformer-based methods, we crop the images to the same resolution of 224×224 pixels. We posit the heart in the middle of the images using ground-truth segmentation labels (training set) or pseudo labels predicted by nnU-Net (validation and test set) as reference. We also apply min-max normalization to each image. For inference, we apply test time augmentation by mirroring along all axes. In post-processing steps, we select only the largest connected component of each structure in the predicted segmentation mask. We conduct this operation twice, in both a slice-wise manner and a volume-wise manner. And then we remove the RV from slices predicted to have no LV nor MYO. We train our model on an NVIDIA A100 GPU with 80GB memory. All networks are implemented using the PyTorch framework. The optimization of nnU-Net follows its default setting. The Swin-UNETR is trained using the AdamW optimizer with initial learning rate of 0.0004 and weight decay of 0.00005. The Swin-UNet is trained using the SGD optimizer with initial learning rate of 0.05, weight decay of 0.0001 and momentum of 0.9. A weighted sum of cross-entropy loss and dice loss is used for back-propagation to optimize our model.

3 Experiments

We conduct a series of comparison and ablation studies so as to select the best network architecture (Sec. 3.1) and evaluate the effectiveness of different data-driven approaches (Sec. 3.2). We randomly split the 20 volunteers in the training set into five non-overlapping folders. For quick comparisons of proposed methods,

we use four folders to train networks and use the rest folder to determine the convergence of the training process. The Dice coefficients and 95% Hausdorff distance for three anatomical structures of the online validation set are reported.

3.1 Architectural Variants

We first compare the performance of 2D nnU-Net, 3D nnU-Net, Swin-UNETR, and Swin-UNet. For a fair comparison, all models are pre-trained with public datasets. The results on the hold-out dataset are depicted in Table 2.

From the results, we find that 2D nnU-Net performs better than 3D nnU-Net, which agrees with existing literature and may be due to the large between-slice distance of the SA sequence [12]. We also find that transformer-based methods have comparable (slightly lower) results to CNN-based methods on LV segmentation and RV segmentation, but are inferior to CNN on MYO segmentation. Through assessing the performance on individual cases and slices, we further observe that the segmentation map generated by nnU-Net is very smooth and regular, while the results of the transformer have twisted margins, especially when the edge in the raw image is vague. We infer that the patch division and window partition process of the swin-transformer is the reason for rough segmentation maps. As the raw input is discretized into multiple windows, it can perform well in segmenting large and convex-shaped structures such as LV and RV but is hard to segment exquisite structures such as MYO.

Moreover, when comparing results of different respiratory motion intensities, we find that nnU-Net is more robust. The respiratory motion can result in an unclear margin between RV and its adjacent tissue, to which transformer-based methods are more vulnerable. The transformer is prone to include nearby objects with similar pixel intensities in the predicted RV segmentation map. In addition, when the margin of MYO is ambiguous, nnU-Net seems to follow a population-based prior to give a smooth and rounded segmentation map, while the results of transformation and be more distorted and angular, which are more faithful to the raw pixel intensities.

Table 2. Quantitative comparison of proposed architectures on the online validation set of 5 volunteers, in terms of Dice and Hausdorff distance

Methods	Dice score			95% Hausdorff distance		
	LV	MYO	RV	LV	MYO	RV
2D nnU-Net [14]	**0.9184**	0.8284	**0.8994**	**8.4581**	4.3903	**5.0750**
3D nnU-Net [14]	0.9164	**0.8295**	0.8879	9.1120	**3.9550**	6.9939
Swin-UNETR [13]	0.9145	0.8133	0.8925	9.1525	4.9728	5.2112
Swin-UNet [8]	0.9167	0.8206	0.8962	8.5567	4.3373	5.4207

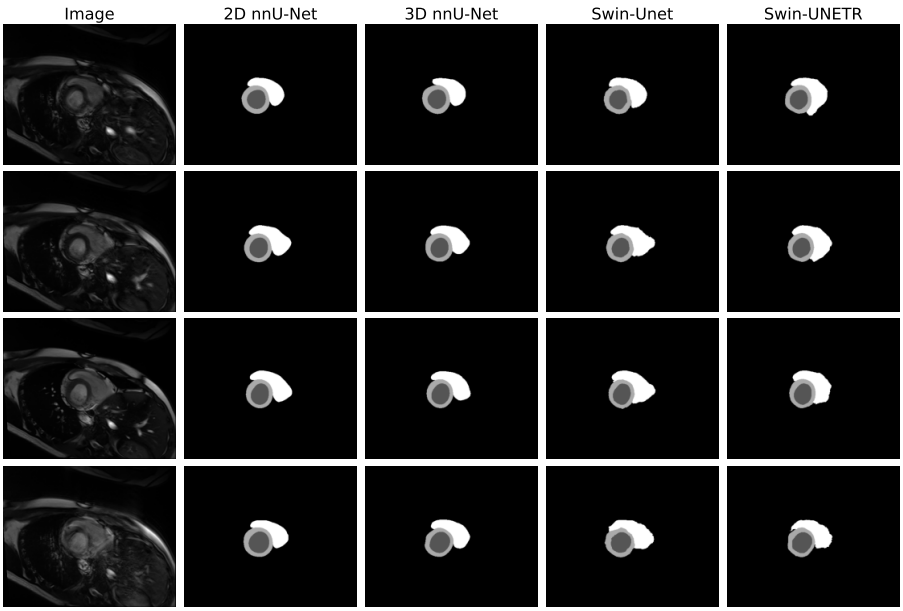

Fig. 1. Qualitative segmentation results of different network architectures. From top to bottom are images from a single volunteer with breath-holding, half breath-holding, regular breathe and intensive breathe. Cases are selected from online validation set.

3.2 Data-Driven Methods with Pre-training and Augmentation

To evaluate the effectiveness of pre-training and adversarial data augmentations, we choose the vanilla 2D nnU-Net as our baseline. We then compare it to both a nnU-Net pre-trained on the public dataset (Sec. 2.3) and a nnU-Net with AdvChain (Sec. 2.4). The results are reported in Table 3.

From the results, we can conclude that both pre-training and adversarial augmentations improve performance compared to the baseline. Pre-training can significantly improve the segmentation accuracy of LV and MYO, while AdvChain promises considerable gain in RV segmentation. Composition of both methods can further enhance the segmentation accuracy.

3.3 Ensemble

Based on the above comparison and analysis, we combine the network architecture and data-driven methods proven to bring out improvements over their baseline in an ensemble manner as our final submission to the CMRxMotion challenge. To this end, we train four networks mentioned in Sec. 3.1 with pre-training and adversarial augmentations in a 5-fold cross-validation setting on the training dataset, and average the outputs of each network to obtain the ensemble

Table 3. Segmentation performance of baseline, pre-training, and AdvChain, in terms of Dice and Hausdorff distance.

Methods	Dice score			95% Hausdorff distance		
	LV	MYO	RV	LV	MYO	RV
Baseline	0.9141	0.8264	0.8915	8.4561	4.2720	5.2156
+ Pre-training	0.9184	0.8284	0.8994	8.4781	4.3903	5.0750
+ AdvChain [5]	0.9160	0.8296	0.9019	8.6097	4.0094	4.8959
+ Both	**0.9201**	**0.8306**	**0.9062**	**8.0947**	**4.0088**	**4.6297**

prediction. Finally, in the online validation set, our model achieves dice scores of 0.9220, 0.8352, and 0.9069 for segmentation of LV, MYO, and RV respectively and 95% Hausdorff distance of 8.0736, 3.6968, and 4.6864.

4 Discussion and Conclusion

In this work, we propose a data-centric model for the tasks of cardiac segmentation which can achieve high performance even when the image quality is degenerated by intensive repository motion. We compare multiple networks or methods and find that pre-training and adversarial augmentation are two effective data-driven approaches that can significantly improve the model performance. Deep learning is essentially a data-driven methodology, whose performance highly relies on data quantity and quality. Collecting sufficient data and making utmost use of the data is second to nothing in terms of improving model performance and robustness. We believe this philosophy can guide the real application of deep learning. Instead of designing novel DCNN architectures or fancy training schema, rethinking how to utilize the data can be far more important.

Intensive respiratory motion impairs the images by making its margin indistinct so that it becomes hard for the model to decide which structure each pixel belongs to. During the experiments, we find even if the boundary between LV and MYO or between MYO and RV is ambiguous, the model can always predict it well. However, it is hard for the model to delineate the boundary between MYO, RV, and their surrounding tissues, especially for RV, which shows great irregularity and variability in terms of shape and pixel intensity. Currently, we are mainly using some common data augmentations such as rotation or flip. In the future, we may design some exclusive perturbations that can better reflect this fuzzy-boundary property.

Another future direction entails the use of inter-slice information. Although the performance of 3D nnU-Net is inferior to that of 2D nnU-Net due to discontinuity between slices, we find that the ensemble of both models results in a better performance. Therefore, we believe the inter-slice information is beneficial for the cardiac segmentation task. Furthermore, the influence of respiratory motion on slices may differ from each other. A well-designed architecture may utilize the information of clean slices to help do segmentation on dirty slices.

References

1. Schulz-Menger, J., Bluemke, DA., Bremerich, J., Flamm, SD., Fogel, MA., Friedrich, MG., Kim, RJ., von Knobelsdorff-Brenkenhoff, F., Kramer, CM., Pennell, DJ., Plein, S., Nagel, E.: Standardized image interpretation and post-processing in cardiovascular magnetic resonance - 2020 update : Society for Cardiovascular Magnetic Resonance (SCMR): Board of Trustees Task Force on Standardized Post-Processing. J Cardiovasc Magn Reson **22**(1), 19 (2022). https://doi.org/10.1186/s12968-020-00610-6
2. Alfakih, K., Plein, S., Thiele, H., Jones, T., Ridgway, JP., Sivananthan, MU.: Normal human left and right ventricular dimensions for MRI as assessed by turbo gradient echo and steady-state free precession imaging sequences. J Magn Reson Imaging. **17**(3), 323-9 (2003). https://doi.org/10.1002/jmri.10262
3. Bai, W, Shi, W, de Marvao, A, Dawes, TJ, O'Regan, DP, Cook, SA, Rueckert, D.: A bi-ventricular cardiac atlas built from 1000+ high resolution MR images of healthy subjects and an analysis of shape and motion. Med Image Anal. **26**(1), 133-45 (2015). https://doi.org/10.1016/j.media.2015.08.009
4. Bai, W., et al.: Biventricular surface reconstruction from cine MRI contours using point completion networks. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), pp. 105-109 (2021). https://doi.org/10.1109/ISBI48211.2021.9434040
5. Chen, C., Qin, C., Ouyang, C., Li, Z., Wang, S., Qiu, H., Chen, L., Tarroni, G., Bai, W., Rueckert, D.: Enhancing MR image segmentation with realistic adversarial data augmentation. arXiv preprint (2022) https://arxiv.org/abs/2108.03429
6. Corral Acero, J., et al.: SMOD - data augmentation based on statistical models of deformation to enhance segmentation in 2d cine cardiac MRI. In: Coudiere, Y., Ozenne, V., Vigmond, E., Zemzemi, N. (eds.) FIMH 2019. LNCS, vol. 11504, pp. 361-369. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21949-9_39
7. Madry, A., Makelov, A., Schmidt, L., Tsipras, D., Vladu, A.: Towards deep learning models resistant to adversarial attacks. In: International Conference on Learning Representations, pp. 1–23 (2017). http://arxiv.org/abs/1706.06083
8. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-Unet: Unet-like pure transformer for medical image segmentation. (2021) https://arxiv.org/abs/2105.05537
9. Chen, C., Qin, C., Qiu, H., Ouyang, C., Wang, S., Chen, L., Tarroni, G., Bai, W., Rueckert, D.: Realistic adversarial data augmentation for MR image segmentation. In: Martel, A.L., Abolmaesumi, P., Stoyanov, D., Mateus, D., Zuluaga, M.A., Zhou, S.K., Racoceanu, D., Joskowicz, L. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2020 - 23rd International Conference, Lima, Peru, October 4-8, 2020, Proceedings, Part I, Springer, pp. 667–677 (2020). https://doi.org/10.1007/978-3-030-59710-8_65
10. Zhu, X., Hu, H., Lin, S., Dai, J.: Deformable ConvNets V2: More deformable, better results, In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9300-9308 (2019). http://doi.org/10.1109/CVPR.2019.00953
11. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin Transformer: hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 3464-3473 (2021). http://doi.org/10.48550/arXiv.2103.14030
12. Isensee, F., Jaeger, P.F., FUll, P.M., Wolf, I., Engelhardt, S., Maier-Hein, K.H.: Automatic cardiac disease assessment on cine-MRI via time-series segmentation and

domain specific features. In: Pop, M., et al. (eds.) STACOM 2017. LNCS, vol.10663, pp. 120-129. Springer, Cham (2018). http://doi.org/10.1007/978-3-319-75541-0_13
13. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images. In: Crimi, A., Bakas, S. (eds) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2021. Lecture Notes in Computer Science, vol 12962. Springer, Cham, (2022). https://doi.org/10.1007/978-3-031-08999-2_22
14. Isensee, F., Jaeger, P.F., Kohl, S.A.A. et al.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods **18**, 203–211 (2021). https://doi.org/10.1038/s41592-020-01008-z
15. Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., et al.: Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the Problem Solved? In: IEEE Transactions on Medical Imaging, **37**(11), 2514-2525 (2018). http://doi.org/10.1109/TMI.2018.2837502
16. Campello, V. M. et al.: Multi-centre, multi-vendor and multi-disease cardiac segmentation: The M&Ms Challenge. In: IEEE Transactions on Medical Imaging **40**(12), 3543 - 3554 (2021). http://doi.org/10.1109/TMI.2021.3090082
17. Zhuang, X: Multivariate mixture model for myocardial segmentation combining multi-source images. In: IEEE Transactions on Pattern Analysis and Machine Intelligence **41**(12), 2933-2946 (2019). http://doi.org/10.1109/TPAMI.2018.2869576
18. Zhuang, X.: Multivariate mixture model for cardiac segmentation from multi-sequence MRI. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp.581-588 (2016). https://doi.org/10.1007/978-3-319-46723-8_67
19. Deng J., Dong, W., Socher, R., Li, L.J., Li, K., Li, F.: ImageNet: A large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 248-255 (2009). https://doi.org/10.1109/CVPR.2009.5206848
20. Fulton, M.J., Heckman, C.R., Rentschler, M.E.: Deformable Bayesian convolutional networks for disease-robust cardiac MRI segmentation. In: Anton, E.P., et al.: Statistical Atlases and Computational Models of the Heart. Multi-Disease, Multi-View, and Multi-Center Right Ventricular Segmentation in Cardiac MRI Challenge. STACOM 2021. Lecture Notes in Computer Science, vol 13131. Springer, Cham, (2022). https://doi.org/10.1007/978-3-030-93722-5_32
21. Yosinski, J., Clune, J., Bengio, Y., Lipson, H.: How transferable are features in deep neural networks? In: Advances in neural information processing systems, pp.3320–3328 (2014). https://doi.org/10.48550/arXiv.1411.1792
22. Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., He. Q.: A comprehensive survey on transfer learning. arXiv preprint (2019). https://arxiv.org/abs/1911.02685
23. He, K., Girshick, R., Dollar, P.: Rethinking imagenet pre-training. arXiv preprint (2018). https://arxiv.org/abs/1811.08883
24. Raghu, M., Zhang, C., Kleinberg, J., Bengio, S.: Transfusion: Understanding transfer learning for medical imaging. In: Advances in Neural Information Processing Systems, pp.3347–3357 (2019). https://doi.org/10.48550/arXiv.1902.07208
25. Hendrycks, D., Lee, K., Mazeika, M.: Using pre-training can improve model robustness and uncertainty. In: Proceedings of the International Conference on Machine Learning, (2019). https://doi.org/10.48550/arXiv.1901.09960
26. Mathis, A., Biasi, T., Schneider, S., Yuksekgonul, M., Rogers, B., Bethge, M., Mathis, M. W.: Pretraining boosts out-of-domain robustness for pose estimation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 1859-1868 (2019). https://doi.org/10.48550/arXiv.1909.11229