Enhancing Cross-Modal Medical Image Segmentation through Compositionality

Aniek Eijpe[1,2,3], Valentina Corbetta[1,3,4], Kalina Chupetlovska[3], Regina Beets-Tan[3,4], and Wilson Silva[1,3]

[1] AI Technology for Life, Department of Information and Computing Sciences, Department of Biology, Utrecht University, Utrecht, The Netherlands
{a.eijpe}@uu.nl
[2] University of Amsterdam, Amsterdam, The Netherlands
[3] Department of Radiology, The Netherlands Cancer Institute, Amsterdam, The Netherlands
[4] GROW School for Oncology and Developmental Biology, Maastricht University Medical Center, Maastricht, The Netherlands

Introduction Disentangled Representation Learning (DRL) techniques have found significant application in the context of cross-modal medical image segmentation [6], often focusing on Content-Style Disentanglement (CSD) [1] which separates a *content* (e.g., spatial information) from a *style* (e.g., appearance information) representation. However, current methods have high computational costs and lack interpretability as they require complex architectures [14,16,10,13] and it is often unclear whether the learned representations are effectively disentangled or not [9,6]. We address these challenges by introducing compositionality [12] as an inductive bias into a cross-modal segmentation framework, to reduce complexity and enhance interpretability. Our proposed end-to-end network[1] enforces compositionality on the learned representations using learnable von Mises-Fisher (vMF) kernels [7], facilitating CSD and producing interpretable representations that effectively separate different anatomical structures.

Methodology Figure 1 shows our proposed framework. We aim to segment images from a target domain $y \in Y$, using images from a source domain $x \in X$ with corresponding labels $m_x \in M$. In the context of CSD, we adopt a two-stage disentanglement approach, by first roughly aligning the deep features Z and then filtering out all remaining target-specific domain information with the vMF kernels [7] K_{vMF}. The model is trained with a cross-cycle consistency objective [17] to learn the bi-directional mapping between the two domains and a generative adversarial learning objective [8], to generate plausible, translated images. Moreover, we employ a cluster loss to fit the vMF kernels to be the style representations of several compositional components of human anatomy [7]. Lastly, we use a standard dice loss on the predicted masks of the translated target images with the corresponding source labels. Code and checkpoints are publicly available at: https://github.com/Trustworthy-AI-UU-NKI/Cross-Modal-Segmentation.

[1] Accepted at the Deep Generative Models workshop at MICCAI 2024. Proceedings available at https://doi.org/10.1007/978-3-031-72744-3_5.

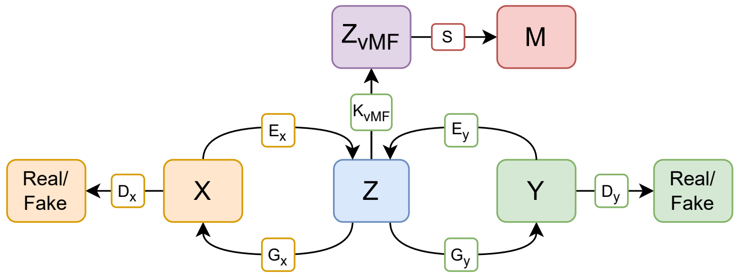

Fig. 1: Overview of the proposed framework. X and Y denote the source and target domain from which the encoders E_x and E_y extract the deep features into Z. From Z, the deep features can be translated to either domain with the generators G_x and G_y, or compositional representations Z_{vMF} can be obtained via the vMF kernels (K_{vMF}). From Z_{vMF}, the segmentation model S predicts the final segmentation masks. D_x and D_y denote the domain discriminators.

Table 1: Quantitative results on the cardiac dataset with CT as target domain.

MRI → CT	MYO		LV		RV	
	DSC(%)	ASSD(mm)	DSC(%)	ASSD(mm)	DSC(%)	ASSD(mm)
UNet-FS [11]	$87.1_{2.4}$	$1.3_{0.3}$	$92.1_{1.1}$	$1.4_{0.2}$	$90.2_{2.4}$	$1.8_{0.4}$
UNet-NA [11]	$5.3_{4.7}$	$26.7_{6.2}$	$37.2_{23.4}$	$19.7_{16.6}$	$25.4_{22.5}$	$22.9_{9.8}$
vMFNet [7]	$2.3_{1.4}$	$26.7_{4.0}$	$52.4_{15.3}$	$10.8_{3.7}$	$40.4_{9.7}$	$12.5_{1.3}$
DRIT [5]+UNet [11]	$47.5_{8.5}$	$5.3_{2.0}$	$69.5_{3.3}$	$6.0_{1.4}$	$67.9_{5.8}$	$\mathbf{5.5_{0.8}}$
DRIT [5]+RUNet [4]	$58.4_{3.8}$	$3.9_{0.2}$	$75.1_{3.1}$	$5.1_{0.5}$	$71.5_{2.5}$	$6.7_{1.5}$
Proposed	$\mathbf{65.1_{4.8}}$	$\mathbf{3.0_{0.6}}$	$\mathbf{80.2_{4.7}}$	$\mathbf{4.7_{1.5}}$	$\mathbf{77.3_{3.6}}$	$5.6_{2.0}$

Results We tested our model on an unpaired public cardiac CT & MRI dataset (MMWHS [18,15]), and an abdominal multi-modal MRI dataset (CHAOS [3,2]) and compared it with several segmentation baselines. Table 1 presents the quantitative results for segmenting the Myocardium (MYO), Left Ventricle cavity (LV), and Right Ventricle cavity (RV) with CT as the target domain. Overall, our proposed method outperforms all cross-modal baselines. Furthermore, the different channels of the learned compositional representations demonstrated distinct activation patterns corresponding to several anatomical structures. Lastly, introducing compositionality reduced training times significantly.

Conclusion We introduced compositionality into a cross-modal segmentation network to address the lack of interpretability and high computational costs in the current models. By enforcing the learned representations to be compositional, we effectively disentangle the style and content features of different anatomical structures. The qualitative and quantitative experiments demonstrated enhanced performance and interpretability while reducing computational costs.

References

1. Gatys, L.A., Ecker, A.S., Bethge, M.: Image style transfer using convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2414–2423 (2016)
2. Kavur, A.E., Gezer, N.S., Barış, M., Aslan, S., Conze, P.H., Groza, V., Pham, D.D., Chatterjee, S., Ernst, P., Özkan, S., Baydar, B., Lachinov, D., Han, S., Pauli, J., Isensee, F., Perkonigg, M., Sathish, R., Rajan, R., Sheet, D., Dovletov, G., Speck, O., Nürnberger, A., Maier-Hein, K.H., Bozdağı Akar, G., Ünal, G., Dicle, O., Selver, M.A.: CHAOS Challenge - combined (CT-MR) healthy abdominal organ segmentation. Medical Image Analysis **69**, 101950 (Apr 2021). https://doi.org/https://doi.org/10.1016/j.media.2020.101950, http://www.sciencedirect.com/science/article/pii/S1361841520303145
3. Kavur, A.E., Selver, M.A., Dicle, O., Barış, M., Gezer, N.S.: CHAOS - Combined (CT-MR) Healthy Abdominal Organ Segmentation Challenge Data (Apr 2019). https://doi.org/10.5281/zenodo.3362844, https://doi.org/10.5281/zenodo.3362844
4. Kerfoot, E., Clough, J., Oksuz, I., Lee, J., King, A.P., Schnabel, J.A.: Left-ventricle quantification using residual u-net. In: Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges: 9th International Workshop, STACOM 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers 9. pp. 371–380. Springer (2019)
5. Lee, H.Y., Tseng, H.Y., Mao, Q., Huang, J.B., Lu, Y.D., Singh, M., Yang, M.H.: Drit++: Diverse image-to-image translation via disentangled representations. International Journal of Computer Vision **128**, 2402–2417 (2020)
6. Liu, X., Sanchez, P., Thermos, S., O'Neil, A.Q., Tsaftaris, S.A.: Learning disentangled representations in the imaging domain. Medical Image Analysis **80**, 102516 (2022)
7. Liu, X., Thermos, S., Sanchez, P., O'Neil, A.Q., Tsaftaris, S.A.: vmfnet: Compositionality meets domain-generalised segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 704–714. Springer (2022)
8. 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)
9. Ouyang, J., Adeli, E., Pohl, K.M., Zhao, Q., Zaharchuk, G.: Representation disentanglement for multi-modal brain mri analysis. In: Information Processing in Medical Imaging: 27th International Conference, IPMI 2021, Virtual Event, June 28–June 30, 2021, Proceedings 27. pp. 321–333. Springer (2021)
10. Pei, C., Wu, F., Huang, L., Zhuang, X.: Disentangle domain features for cross-modality cardiac image segmentation. Medical Image Analysis **71**, 102078 (2021)
11. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. pp. 234–241. Springer (2015)
12. Stone, A., Wang, H., Stark, M., Liu, Y., Scott Phoenix, D., George, D.: Teaching compositionality to cnns. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 5058–5067 (2017)

13. Wang, R., Zheng, G.: Unsupervised cross-modality cardiac image segmentation via disentangled representation learning and consistency regularization. In: Machine Learning in Medical Imaging: 12th International Workshop, MLMI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings 12. pp. 517–526. Springer (2021)
14. Wang, R., Zheng, G.: Cycmis: Cycle-consistent cross-domain medical image segmentation via diverse image augmentation. Medical Image Analysis **76**, 102328 (2022)
15. Wu, F., Zhuang, X.: Cf distance: a new domain discrepancy metric and application to explicit domain adaptation for cross-modality cardiac image segmentation. IEEE Transactions on Medical Imaging **39**(12), 4274–4285 (2020)
16. Xie, Q., Li, Y., He, N., Ning, M., Ma, K., Wang, G., Lian, Y., Zheng, Y.: Unsupervised domain adaptation for medical image segmentation by disentanglement learning and self-training. IEEE Transactions on Medical Imaging (2022)
17. 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)
18. Zhuang, X., Li, L., Payer, C., Štern, D., Urschler, M., Heinrich, M.P., Oster, J., Wang, C., Smedby, Ö., Bian, C., et al.: Evaluation of algorithms for multi-modality whole heart segmentation: an open-access grand challenge. Medical image analysis **58**, 101537 (2019)