Unsupervised Domain Adaptation for Cardiac Segmentation: Towards Structure Mutual Information Maximization

Changjie Lu, Shen Zheng, Gaurav Gupta

College of Science and Technology, Wenzhou-Kean University

Outline

- Introduction
- Related Works
- Methodology
- Experiment
- Conclusion

Introduction

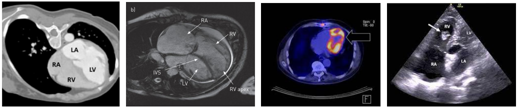


Figure 1: Four types of Cardiac Imaging. From left to right: Computerized Tomography (CT), Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), and Ultrasound (US).[MRI,]

Motivation

- Different statistical distribution problems (MRI->CT)[Kalogeiton et al., 2016, Tommasi et al., 2016]
- MRI, CT play complementary roles in cardiac disease diagnosis.
- Manual annotations consumes 2-4 hours[Zhuang, 2013]

How to deal with Domain Shift?

Possible Solution

- Deep learning-based methods on detection[Liu et al., 2017, Yan et al., 2019] segmentation[Ronneberger et al., 2015, Dou et al., 2017]
  Effectively training and testing images from the **same** modality
- Unsupervised Domain adaptation (UDA)[Dou et al., 2018, Dou et al., 2019]
  Transfers knowledge from the source domain to the target domain (e.g., MRI to CT) without paired images
- Source medical image with the ground truth segmentation
- Target medical image without the ground truth

Related Work

- UDA with a GAN-based[Goodfellow et al., 2014].
  [Zhu et al., 2017, Isola et al., 2017]
  [Zhang et al., 2018, Chen et al., 2020, Liu and Du, 2020]
  Good performance but Unstable, Large Model
- UDA with a VAE-based[Kingma and Welling, 2013]
  [Purushotham et al., 2016, Wu and Zhuang, 2021]
  [Ouyang et al., 2019, Gu et al., 2022]
  Ingenious design with fast training

Method Analysis

- UDA-GAN
  - Two stages: Translation and Adaptation
- UDA-VAE
  - Two tasks: Reconstruction and Segmentation
- Our focus: UDA-VAE method
  problems of UDA-VAE:
  Information from the reconstructed output cannot be directly delivered to the segmentation.
  Utilize parallel reparameterization for latent space with different resolutions.

Our Solution

- deduce a compact loss function lower bound in which each term is orthogonal, discovering a new mutual information term.
- design a novel, plug-and-play style, Structure Mutual Information Estimation (SMIE) block.
- convert parallel reparameterization to sequential reparameterization,

Symbols	Discription
S	Source domain
T	Target domain
z	Latent variable
x	Input image data point
$p_\theta()$	PDF of variables with parameter θ
$q_\phi()$	Neural network with parameter ϕ
$D(\phi_S, \phi_T)$	Domain distance between source and target
$\hat{y}$	Predicted segmentation
y	Ground truth Segmentation
R_S	Reconstructed image in the source domain
R_T	Reconstructed image in the target domain
D_{KL}	KL Divergence

Lower Bound Deduction

The UDA-VAE model maximizes the joint log-likelihood of the complete data:

$$\text{JLL} = \log p_{\theta_S}\left(\left(x_S^1, y_S^1\right), \cdots, \left(x_S^{N_S}, y_S^{N_S}\right)\right) \tag{1}$$

All of the data are considered as $i.i.d.$ random variables. Similar as VAE, we approximate $p_{\theta_S}(z \mid x)$ by a parameterized model $q_{\phi_S}(z \mid x)$. Moreover, we follow the assumption of distribution independence:
$q_{\phi_S}(y, z \mid x) = q_{\phi_S}(y \mid x) \cdot q_{\phi_S}(z \mid x)$. To estimate the JLL, we deduce lower bound of UDA-VAE:

$$\log p_{\theta_S}(x, y) \geq LB_{VAE}\left(\theta_S, \phi_S\right) \tag{2}$$

where $LB_{VAE}\left(\theta_S, \phi_S\right)$ is formulated by

$$
\begin{aligned}
LB_{VAE}(\theta_S, \phi_S) = & \left(\mathcal{R} + I_{q_{\phi_S}}(x, y, z) - H_{q_{\phi_S}}(z) + \log \frac{p_{\theta_S}(x, y)}{q_{\phi_S}(x, y)}\right) - \\
& D_{KL}\left(q_{\phi_S}(z \mid x) \| p_{\theta_S}(z)\right) + E_{q_{\phi_S}(z|x)} \log p_{\theta_S}(x \mid y, z) \\
& + E_{q_{\phi_S}(z|x)} \log p_{\theta_S}(y \mid z)
\end{aligned} \tag{3}
$$

Proof:

$$
\begin{aligned}
& \log p_{\theta_S}(x, y) \\
= & \int q_{\phi_S}(z \mid x, y) \log \left[\frac{q_{\phi_S}(z \mid x, y)}{p_{\theta_S}(z \mid x, y)} \cdot \frac{p_{\theta_S}(z)}{q_{\phi_S}(z \mid x, y)} \cdot p_{\theta_S}(x, y \mid z)\right] dz \\
= & D_{KL}\left(q_{\phi_S}(z \mid x, y) \| p_{\theta_S}(z \mid x, y)\right) - D_{KL}\left(q_{\phi_S}(z \mid x, y) \| p_{\theta_S}(z)\right) \\
& + E_{q_{\phi_S}(z|x,y)} \log \left[p_{\theta_S}(x, y \mid z)\right]
\end{aligned} \tag{4}
$$

Note that UDA-VAE[Wu and Zhuang, 2021] neglects the term $D_{KL}\left(q_{\phi_S}(z \mid x, y) \| p_{\theta_S}(z \mid x, y)\right)$ as it is greater than 0.
In comparison, we deduce a compact lower bound with the following term.

$$
\begin{aligned}
& D_{KL}\left(q_{\phi_S}(z \mid x, y) \| p_{\theta_S}(z \mid x, y)\right) \\
& = \int q_{\phi_S}(z \mid x, y) \log \frac{q_{\phi_S}(z \mid x, y)}{p_{\theta_S}(z \mid x, y)} dz \\
& = \int \frac{q_{\phi_S}(x, y, z)}{q_{\phi_S}(x, y)} \log \frac{q_{\phi_S}(x, y, z)}{p_{\theta_S}(x, y, z)} \frac{p_{\theta_S}(x, y)}{q_{\phi_S}(x, y)} dz \\
& = \frac{1}{q_{\phi_S}(x, y)} [\int q_{\phi_S}(x, y, z) \log \frac{q_{\phi_S}(x, y, z)}{p_{\theta_S}(x, y, z)} \\
& + q_{\phi_S}(x, y, z) \log \frac{p_{\theta_S}(x, y)}{q_{\phi_S}(x, y)} dz] \\
& = \frac{1}{q_{\phi_S}(x, y)} \int q_{\phi_S}(x, y, z) \log \frac{q_{\phi_S}(x, y, z)}{p_{\theta_S}(x, y, z)} dz \\
& + \log \frac{p_{\theta_S}(x, y)}{q_{\phi_S}(x, y)}
\end{aligned}
\tag{5}
$$

Consider the reconstruction error[Belghazi et al., 2018]:

$$\begin{aligned}\mathcal{R} =& \mathbb{E}_{(x,y,z)\sim} q_{\phi_S}(x,y,z) \log \frac{q_{\phi_S}(x,y,z)}{p_{\theta_S}(x,y,z)} - \\ & \mathbb{E}_{(x,y,z)\sim q_{\phi_S}(x,y,z)} \log q_{\phi_S}(x,y,z) + \mathbb{E}_{z\sim q_{\phi_S}(z)} \log p_{\theta_S}(z)\end{aligned} \tag{6}$$

The second term is the joint entropy $H_q(x,y,z)$.
The third term can be written as:

$$\mathbb{E}_{z\sim q_{\phi_S}(z)} \log p_{\theta_S}(z) = -D_{KL}(q_{\phi_S(z)} \| p_{\theta_S}) - H_{q_{\phi_S}}(z) \tag{7}$$

With

$$H_{q_{\phi_S(z)}}(x,y,z) - H_{q_{\phi_S}}(z) = H_{q_{\phi_S}}(z) - I_{q_{\phi_S}}(x,y,z) \tag{8}$$

where I is mutual information.
The reconstruction error can be written as:

$$\mathcal{R} \leq D_{KL}(q_{\phi_S(x,y,z)} \| p_{\theta_S(x,y,z)}) - I_{q_{\phi_S}}(x,y,z) + H_{q_{\phi_S}}(z) \tag{9}$$

which is compact when $q_{\phi_S(z)}$ matches the prior distribution $p_{\theta_S}(z)$.

$$D_{KL}(q_{\phi_S(x,y,z)} \| p_{\theta_S(x,y,z)}) \geq \mathcal{R} + I_{q_{\phi_S}}(x,y,z) - H_{q_{\phi_S}}(z) \tag{10}$$

Thus, we obtain the bound,

$$\begin{aligned}
&D_{KL}\left(q_{\phi_S}(z \mid x,y) \| p_{\theta_S}(z \mid x,y)\right) \\
&\geq D_{KL}(q_{\phi_S}(x,y,z) \| p_{\theta_S}(x,y,z)) + \log \frac{p_{\theta_S}(x,y)}{q_{\phi_S}(x,y)} \\
&\geq \mathcal{R} + I_{q_{\phi_S}}(x,y,z) - H_{q_{\phi_S}}(z) + \log \frac{p_{\theta_S}(x,y)}{q_{\phi_S}(x,y)}
\end{aligned} \tag{11}$$

From, Eq.4 and Eq.11,

$$
\begin{aligned}
&\log p_{\theta_S}(x,y) \\
\geq &(\mathcal{R} + I_{q_{\phi_S}}(x,y,z) - H_{q_{\phi_S}}(z) + \log \frac{p_{\theta_S}(x,y)}{q_{\phi_S}(x,y)}) - \\
&D_{KL}\left(q_{\phi_S}(z \mid x) \| p_{\theta_S}(z)\right) + E_{q_{\phi_S}(z|x)} \log p_{\theta_S}(x,y \mid z) \\
= &(\mathcal{R} + I_{q_{\phi_S}}(x,y,z) - H_{q_{\phi_S}}(z) + \log \frac{p_{\theta_S}(x,y)}{q_{\phi_S}(x,y)}) - \\
&D_{KL}\left(q_{\phi_S}(z \mid x) \| p_{\theta_S}(z)\right) + E_{q_{\phi_S}(z|x)} \log p_{\theta_S}(x \mid y,z) \\
&+ E_{q_{\phi_S}(z|x)} \log p_{\theta_S}(y \mid z)
\end{aligned}
\tag{12}
$$

where R, $\log \frac{p_{\theta_S}(x,y)}{q_{\phi_S}(x,y)}$ and $H_{q_{\phi_S}}(z)$ are constant. The equation holds, as $p_{\theta_S}(x,y \mid z) = p_{\theta_S}(y \mid z) \cdot p_{\theta_S}(x \mid y,z)$. Meanwhile, y_S and z_s are conditionally independent on x_S for distribution q_{ϕ_S}, so that $q_{\phi_S}(z \mid x,y) = q_{\phi_S}(z \mid x)$.
Finally, We get the compact lower bound (plus red terms) than UDA-VAE .
The UDA-VAE++ maximizes the mutual information of $I_{q_{\phi_S}}(x,y,z)$.
Proved.

Methodology

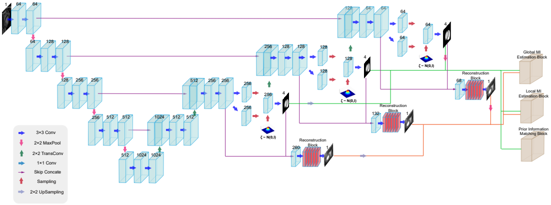


Figure 2: The Model Architecture of UDA-VAE++. The backbone: U-Net (blue boxes) with three scales of variational blocks. The reconstruction blocks (red boxes) contain seven convolution layers. The grey box refers to the MI estimation block detailed in Fig.

Mutual Information Neural Estimation

To estimate the mutual information between the segmentation outcome $\hat{y}$ and the reconstruction output R:

$$\widehat{\mathcal{I}}\left(\hat{y}; R\right) = D_{KL}\left(\mathbb{P}_{\hat{y}R} || \mathbb{P}_{\hat{y}} \otimes \mathbb{P}_{R}\right) \tag{13}$$

which can be written as its dual representation[Donsker and Varadhan, 1975] as below:

$$D_{KL}(\mathbb{P}_{\hat{y}R} || \mathbb{P}_{\hat{y}} \otimes \mathbb{P}_{R}) = \sup_{T:\Omega \to \mathbb{R}} (\mathbb{E}_{\mathbb{P}_{\hat{y}R}}[T] - \log\left(\mathbb{E}_{\mathbb{P}_{\hat{y}} \otimes \mathbb{P}_{R}}\left[e^{T}\right]\right)) \tag{14}$$

where T is the set of all possible neural network. [Belghazi et al., 2018]

Deep InfoMax

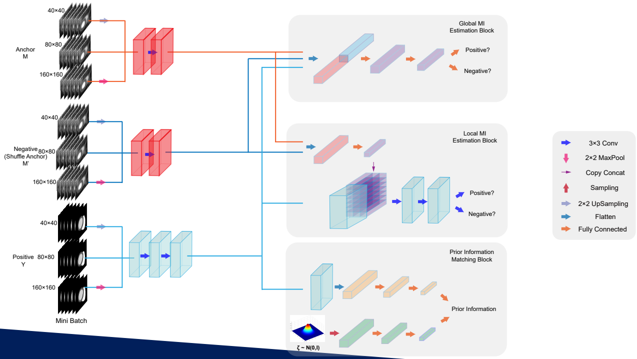
We are interested in maximizing the mutual information rather than obtaining the exact value. So the mutual information maximization process can be formulated as:

$$\widehat{\mathcal{I}}\left(\hat{y};R\right)=\sup_{T:\Omega\to\mathbb{N}}\mathbb{E}_{\mathbb{P}_{\hat{y}R}}\left[-\mathrm{sp}\left(-T\left(\hat{y},R\right)\right)\right]-\mathbb{E}_{\mathbb{P}_{\hat{y}}\otimes\mathbb{P}_{R}}\left[\mathrm{sp}\left(T\left(\hat{y},R'\right)\right)\right] \tag{15}$$

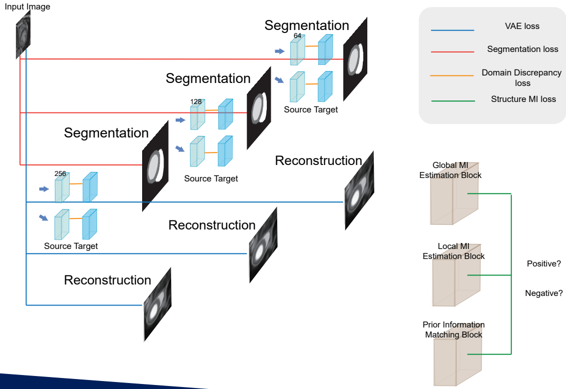
where R' is an input sampled from R, N contains all possible function, and $\mathrm{sp}(z)=\log\left(1+e^{z}\right)$ is the softplus function. [Hjelm et al., 2018]

Mutual Information Neural Estimation

- Inspired by MINE[Belghazi et al., 2018], Deep InfoMax[Hjelm et al., 2018]



Loss Function



Loss Function

- Reconstruction Loss

$$D_{KL}\left(q_\phi(z|x)\|p_\theta(z|x)\right) \quad = D_{KL}\left(q_\phi(z|x)\|p_\theta(z)\right) - E_{z\sim q_\phi}(\log p_\theta(x|z)). \tag{16}$$

First term:

$$D_{KL}\left(q_\phi(z|x)\|p_\theta(z)\right) = \frac{1}{2}\left(\sigma^2 + u^2 - \log\sigma^2 - 1\right) \tag{17}$$

Second term:

$$\mathcal{L}_{ce} = -(x\log(R) + (1-x)\log(1-R)) \tag{18}$$

Final:

$$\mathcal{L}_{recon} = D_{KL} + \mathcal{L}_{ce} \tag{19}$$

Loss function

- Segmentation Loss

$$\mathcal{L}_{seg} = -(y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})) \tag{20}$$

- Domain Discrepancy Loss

$$\begin{aligned}
\mathcal{L}_D &= D(q_{\phi_S}(z), q_{\phi_T}(z)) \\
&= \int [q_{\phi_S}(z) - q_{\phi_T}(z)]^2 dz \\
&= \frac{1}{M^2} \sum_{i=1}^{M} \sum_{j=1}^{M} \left[k\left(x_{S_i}, x_{S_j}\right) + k\left(x_{T_i}, x_{T_j}\right) - 2k\left(x_{S_i}, x_{T_j}\right) \right]
\end{aligned} \tag{21}$$

Loss function

kernel function:

$$k\left(x_{S_i}, x_{T_j}\right) = (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}\left[\frac{(u_{S_i}-u_{T_j})^2}{\sigma_{S_i}^2+\sigma_{T_j}^2}+\log(\sigma_{S_i}^2+\sigma_{T_j}^2)\right]} \tag{22}$$

- Structure Mutual Information Loss

$$\mathcal{L}_{MI} = -(\alpha\widehat{\mathcal{I}}(\hat{y}; R)_{Global} + \beta\widehat{\mathcal{I}}(\hat{y}; R)_{Local} + \gamma\widehat{\mathcal{I}}_{Prior}) \tag{23}$$

where α, β, γ are set as 0.5, 1.0, 0.1. $\widehat{\mathcal{I}}_{Prior} = log(\mathcal{N}) + log(1-\hat{y})$, where $\mathcal{N}$ is the standard normal distribution.

Loss function

Total loss:

$$\begin{aligned}\mathcal{L}_{total} &= (c1\mathcal{L}_{recon} + c2\mathcal{L}_{seg} + c3\mathcal{L}_{MI})_{source} \\ &+ (c1\mathcal{L}_{recon} + c2\mathcal{L}_{seg} + c3\mathcal{L}_{MI})_{target} \\ &+ c4\mathcal{L}_{D}\end{aligned} \tag{24}$$

where c1, c2, c3, c4 are empirically set as 1e-2, 1, 1e-1, 1e-5, respectively.

Experiment Design

- Adam optimizer [Kingma and Ba, 2014] and Pytorch framework [Paszke et al., 2019] 30 epochs
- learning rate is initialized at 1e-4, reduced by 10 % after every epoch
- batch size is 12
- takes about 1 hour to converge on a single NVIDIA Tesla V100 GPU
- Xavier initialization [Glorot and Bengio, 2010]

Dataset

- Multi-Modality Whole Heart Segmentation (MM-WHS) Challenge dataset [Zhuang and Shen, 2016]: contains 20 labeled CT images and 20 labeled LGE-MRI images
- Multi-Sequence Cardiac MR Segmentation (MS-CMRSeg) Challenge dataset: contains 35 labeled CT images and 45 labeled LGE-MRI images [Zhuang, 2018]

Result

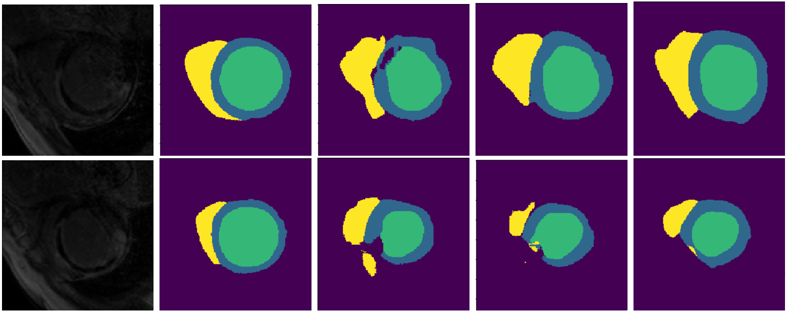


Figure 5: Segmentation output from MS-CMRSeg Dataset (CT to MRI). From left to right: MRI, Ground truth, CFDNet[Wu and Zhuang, 2020], UDA-VAE[Wu and Zhuang, 2021],

Result

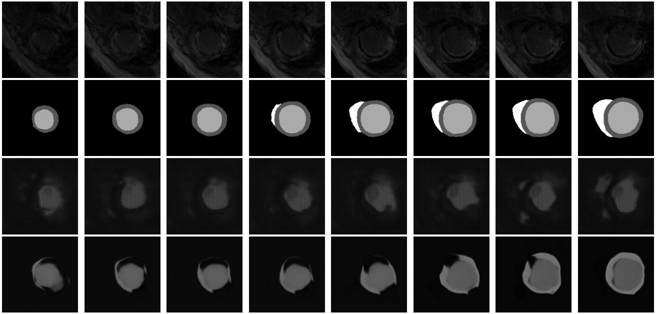


Figure 6: Reconstruction Images from MS-CMRSeg Dataset (CT to MRI). From top to bottom row: MRI images, corresponding segmentation ground truth, UDA-VAE,

Result

	Dice (%)			ASSD (mm)		
	MYO	LV	RV	MYO	LV	RV
NoAdapt	12.32	30.24	37.25	24.9	10.4	16.7
CFDNet [Wu and Zhuang, 2020]	57.41	78.44	77.63	3.61	3.87	2.49
SIFA [Chen et al., 2020]	60.89	79.32	**82.39**	3.44	3.65	1.80
UDA-VAE [Wu and Zhuang, 2021]	58.58	79.43	80.43	3.53	3.27	2.04
UDA-VAE++	**68.74**	**85.08**	81.42	**2.34**	**2.61**	**1.71**

Table 1: Unsupervised Domain Adaptation for MS-CMRSeg Dataset from **MRI to CT**. The best score for Dice ↑ and ASSD↓ are in **bold**.

Result

	Dice (%)			ASSD (mm)		
	MYO	LV	RV	MYO	LV	RV
NoAdapt	14.50	34.51	31.10	21.6	11.3	14.5
CFDNet [Wu and Zhuang, 2020]	64.21	81.39	72.30	2.81	3.41	4.91
SIFA [Chen et al., 2020]	67.69	83.31	**79.04**	2.56	3.44	**2.13**
UDA-VAE [Wu and Zhuang, 2021]	68.42	84.41	72.59	2.39	2.59	3.97
UDA-VAE++	**70.75**	**88.64**	75.82	**2.02**	**2.27**	3.62

Table 2: Unsupervised Domain Adaptation for MS-CMRSeg Dataset from **CT to MRI**. The best score for Dice$\uparrow$ and ASSD$\downarrow$ are in **bold**.

Result

Methods	Dice(%)					ASSD(mm)				
	MYO	LA	LV	RA	RV	MYO	LA	LV	RA	RV
NoAdapt	0.08	3.08	0.00	0.74	23.9	–	–	–	–	–
PnP-AdaNet	32.7	49.7	48.4	62.4	44.2	6.89	22.6	9.56	20.7	20.0
SIFA	37.1	65.7	61.2	51.9	18.5	11.8	5.47	16.0	14.7	21.6
UDA-VAE	47.0	63.1	73.8	71.1	73.4	4.73	5.33	4.30	6.97	4.56
UDA-VAE++	**51.4**	**65.9**	**76.5**	**73.0**	**75.5**	**3.88**	**5.23**	**3.78**	**6.25**	**4.06**

Table 3: Unsupervised Domain Adaptation for MM-WHS Dataset from **CT to MRI**. The best score for Dice↑ and ASSD↓ are in **bold**.

Ablation Study

Model Components						Dice (%)		
Base	SR	Att	Global	Local	Prior	MYO	LV	RV
✓						68.42	84.41	72.59
✓	✓					68.56	84.07	74.06
✓	✓	✓				68.30	84.91	74.72
✓	✓	✓	✓			69.25	84.70	75.63
✓	✓	✓	✓	✓		68.49	87.50	**77.37**
✓	✓		✓	✓	✓	**70.75**	**88.64**	75.82
✓	✓	✓	✓	✓	✓	69.81	87.54	77.13

Conclusion

- This paper introduces UDA-VAE++, an unsupervised domain adaptation framework for cardiac segmentation.
- We deduce a compact loss function lower bound in which each term is orthogonal, and design a MINE block to evaluate it.
- We introduce the sequential reparameterization design, allowing information flow between multi-scale latent space features.
- Our model achieved **state-of-the-art** performances on benchmark datasets.
- Future work: self-supervised domain adaptation methods, extend to other medical image segmentation tasks (e.g., brain image segmentation)

References I

https://err.ersjournals.com/content/22/130/526, https://www.researchgate.net/publication, https://www.cureus.com/articles/, https://www.sciencephoto.com/.

Belghazi, M. I., Baratin, A., Rajeshwar, S., Ozair, S., Bengio, Y., Courville, A., and Hjelm, D. (2018).
Mutual information neural estimation.
In *International conference on machine learning*, pages 531–540. PMLR.

Chen, C., Dou, Q., Chen, H., Qin, J., and Heng, P. A. (2020).
Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation.
IEEE transactions on medical imaging, 39(7):2494–2505.

Donsker, M. D. and Varadhan, S. S. (1975).
Asymptotic evaluation of certain markov process expectations for large time, i.
Communications on Pure and Applied Mathematics, 28(1):1–47.

References II

Dou, Q., Ouyang, C., Chen, C., Chen, H., Glocker, B., Zhuang, X., and Heng, P.-A. (2019).
Pnp-adanet: Plug-and-play adversarial domain adaptation network at unpaired cross-modality cardiac segmentation.
IEEE Access, 7:99065–99076.

Dou, Q., Ouyang, C., Chen, C., Chen, H., and Heng, P.-A. (2018).
Unsupervised cross-modality domain adaptation of convnets for biomedical image segmentations with adversarial loss.
arXiv preprint arXiv:1804.10916.

Dou, Q., Yu, L., Chen, H., Jin, Y., Yang, X., Qin, J., and Heng, P.-A. (2017).
3d deeply supervised network for automated segmentation of volumetric medical images.
Medical image analysis, 41:40–54.

References III

Glorot, X. and Bengio, Y. (2010).
Understanding the difficulty of training deep feedforward neural networks.
In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256. JMLR Workshop and Conference Proceedings.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014).
Generative adversarial nets.
Advances in neural information processing systems, 27.

Gu, M., Vesal, S., Kosti, R., and Maier, A. (2022).
Few-shot unsupervised domain adaptation for multi-modal cardiac image segmentation.
arXiv preprint arXiv:2201.12386.

References IV

Hjelm, R. D., Fedorov, A., Lavoie-Marchildon, S., Grewal, K., Bachman, P., Trischler, A., and Bengio, Y. (2018).
Learning deep representations by mutual information estimation and maximization.
arXiv preprint arXiv:1808.06670.

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017).
Image-to-image translation with conditional adversarial networks.
In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134.

Kalogeiton, V., Ferrari, V., and Schmid, C. (2016).
Analysing domain shift factors between videos and images for object detection.
IEEE transactions on pattern analysis and machine intelligence, 38(11):2327–2334.

Kingma, D. P. and Ba, J. (2014).
Adam: A method for stochastic optimization.
arXiv preprint arXiv:1412.6980.

References V

Kingma, D. P. and Welling, M. (2013).
Auto-encoding variational bayes.
arXiv preprint arXiv:1312.6114.

Liu, Y. and Du, X. (2020).
Duda: Deep unsupervised domain adaptation learning for multi-sequence cardiac mr image segmentation.
In *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, pages 503–515. Springer.

Liu, Y., Gadepalli, K., Norouzi, M., Dahl, G. E., Kohlberger, T., Boyko, A., Venugopalan, S., Timofeev, A., Nelson, P. Q., Corrado, G. S., et al. (2017).
Detecting cancer metastases on gigapixel pathology images.
arXiv preprint arXiv:1703.02442.

Ouyang, C., Kamnitsas, K., Biffi, C., Duan, J., and Rueckert, D. (2019).
Data efficient unsupervised domain adaptation for cross-modality image segmentation.
In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 669–677. Springer.

References VI

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. (2019).
Pytorch: An imperative style, high-performance deep learning library.
Advances in neural information processing systems, 32.

Purushotham, S., Carvalho, W., Nilanon, T., and Liu, Y. (2016).
Variational recurrent adversarial deep domain adaptation.

Ronneberger, O., Fischer, P., and Brox, T. (2015).
U-net: Convolutional networks for biomedical image segmentation.
In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.

Tommasi, T., Lanzi, M., Russo, P., and Caputo, B. (2016).
Learning the roots of visual domain shift.
In *European Conference on Computer Vision*, pages 475–482. Springer.

References VII

Wu, F. and Zhuang, X. (2020).
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.

Wu, F. and Zhuang, X. (2021).
Unsupervised domain adaptation with variational approximation for cardiac segmentation.
IEEE Transactions on Medical Imaging, 40(12):3555–3567.

Yan, K., Tang, Y., Peng, Y., Sandfort, V., Bagheri, M., Lu, Z., and Summers, R. M. (2019).
Mulan: multitask universal lesion analysis network for joint lesion detection, tagging, and segmentation.
In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 194–202. Springer.

References VIII

Zhang, Z., Yang, L., and Zheng, Y. (2018).
Translating and segmenting multimodal medical volumes with cycle-and shape-consistency generative adversarial network.
In *Proceedings of the IEEE conference on computer vision and pattern Recognition*, pages 9242–9251.

Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017).
Unpaired image-to-image translation using cycle-consistent adversarial networks.
In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232.

Zhuang, X. (2013).
Challenges and methodologies of fully automatic whole heart segmentation: a review.
Journal of healthcare engineering, 4(3):371–407.

Zhuang, X. (2018).
Multivariate mixture model for myocardial segmentation combining multi-source images.
IEEE transactions on pattern analysis and machine intelligence, 41(12):2933–2946.

References IX

Zhuang, X. and Shen, J. (2016).
Multi-scale patch and multi-modality atlases for whole heart segmentation of mri.
Medical image analysis, 31:77–87.