

Semi-supervised Meta-learning with Disentanglement for Domain-generalised Medical Image Segmentation

By Zhen Huang

1. Introduction

《基于半监督解耦元学习的领域泛化的医学图像分割》，出自MICCAI 2021 oral。论文lib: <https://arxiv.org/abs/2106.13292>。code: <https://github.com/vios-s/DGNet>。

本篇论文作为一篇oral论文，重点聚焦于Domain generalisation (DG, 领域泛化) 和 Disentanglement (解耦)，以及Meta-learning (元学习)，并应用在医学图像分割场景中。文中提到的概念都很新颖，也巧妙的结合了前人的一些工作，最终在当前的公开数据集M&M以及SCGM中都取得了比baseline更优的结果。

首先，文章开篇提到了当前迁移学习中存在的困难--将模型推广到新domain上的新数据仍然存在困难，泛化能力往往达不到，从而最终在未知的数据集 (unseen) 上也达不到非常理想的效果。这很大程度上是因为源数据和未知域的数据上存在domain shifts (数据统计上的变化)。

最新的研究表明，基于梯度的元学习策略 (gradient-based meta-learning) 显示出了较好的泛化性能。大概的做法是将训练数据分为meta-train以及meta-test两个集合来模拟并处理训练中可能出现的domain shifts问题。但是同时存在问题：**1.**目前的元学习策略都是基于全监督的，并不适合拓展到**医学图像分割领域** (Medical image segmentation, 后面称**MIS**)。因为在MIS中，创建像素级别的注释是非常time-effort的。**2.** MIS中数据量往往是一个问题，在数据不足的情况下，模拟的域迁移可能不能很好地接近真实域迁移。

接着文章提出了一个新颖的**半监督元学习解耦模型**来解决问题：首先对domain shifts的相关表示进行建模，通过解耦这些表示并重组来重建输入图像，使得可以使用未标记的数据来更好地贴合元学习的实际域迁移从而获得更好的泛化性能，尤其是标记样本有限的情况下。

2. 知识整理

2.1 Domain generalisation

域泛化，也称DG。是迁移学习中的一种方法，它研究的问题是从若干个具有不同数据分布的数据集 (领域) 中学习一个泛化能力强的模型，以便在「未知(Unseen)」的测试集上取得较好的效果，

domain adaptation: 领域自适应。是迁移学习中的一种方法。旨在利用源域中标注好的数据，学习一个精确的模型，运用到无标注或只有少量标注的目标域中。本质上是一种数据增强的迁移方法 [1]。

主要区别:

Domain Adaptation (DA) : 需要源域 (train set) 和目标域 (test set) 都有数据。

Domain Generalization (DG) : 只需要源域 (train set)

DG有两个主要的难点:

1. 目标域不可见 (unseen)

2.对任何目标域都有作用。

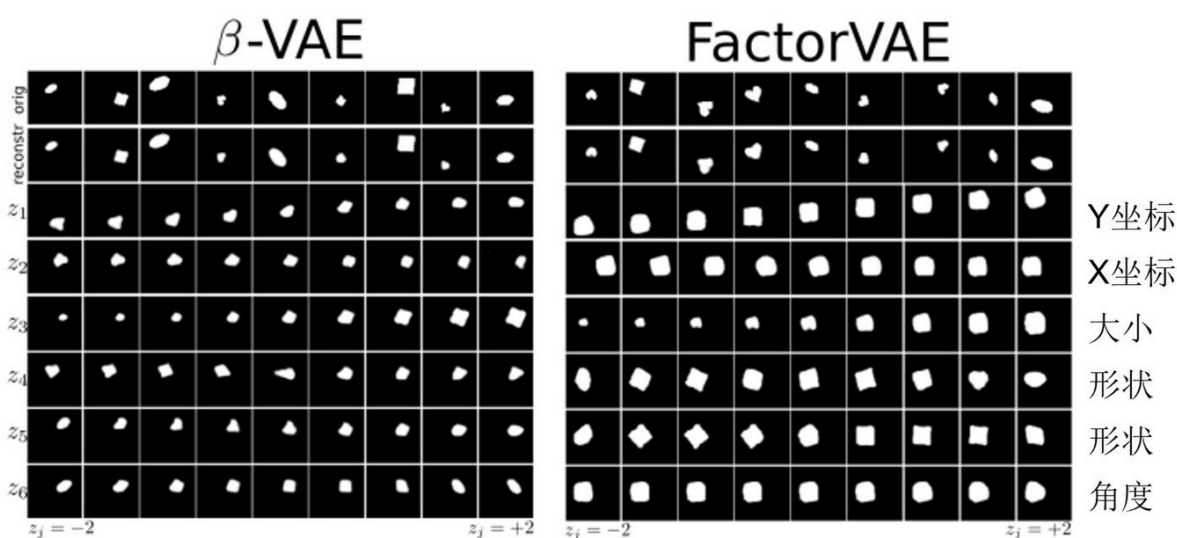
按照我个人的理解，“任何”二字应该还是有局限的，比如说源域为cv中的图像分类任务，目标域（target domain）也应该是类似的任务而不能是NLP任务，所以在这里的“任何”并不是广义上的任何。

目前成熟的domain generalization方法基本可以分为三类，分别是：

- 1.feature-based method
- 2.classifier-based method
- 3.instance-reweighting method

2.2 Disentanglement

解耦的特征中，每一个维度都表示具体的、不相干的意义。其中最重要的是要让学到的表征具备人类可理解的意义。实际上在这个领域中，并没有一个标准的定义。概念本身也是不断发展的，甚至每个学者发表论文的时候还要专门提出一个指标来定义解耦。但是最核心且不变的诉求是要让表征可读。



遍历隐变量可视化（latent traversal）。 z_i 表示特征的第*i*个维度，横坐标表示该变量的取值。在绘制重构图像时，固定 z_j 以外的值为0，遍历 z_j （从-2到2）得到一行图像。遍历 z_1 可以观察到图像沿着Y轴移动，类似的每一个维度都有具体的意义。（图来自FactorVAE）

2.3 Meta-Learning

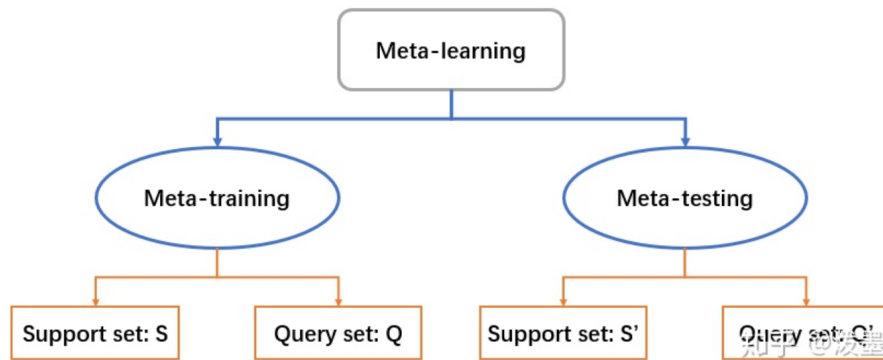
这也是本文所重点提到的概念，但我个人之前没有接触过，查阅了相关文献进行了一定的了解。

元学习（Meta-Learning），又称“学会学习”（Learning to learn），即利用以往的知识经验来指导新任务的学习，使网络具备学会学习的能力，是解决小样本问题（Few-shot Learning）常用的方法之一

这里面有几个点是需要注意的。

1.元学习中的元，即meta的含义：meta-learning的本质是增加学习器在multi-task中的泛化能力，其对于数据和任务都需要采样。因此最终学习到的 $F(x)$ 可以在unseen的task中迅速建立起对应的mapping，达到“学会学习”的目的。具体的，“meta”体现在网络对于每个task的学习，不断适应具体的task，使得我们的网络最终具备一种抽象的学习能力。

2.meta-learning中的train和test：训练过程定义为“meta-training”，测试过程则定义为“meta-testing”。具体如下图：



其训练和测试过程都需要两类数据集 (Support set & Query set) 。

构建S和Q主要还是采用randomly choose的模式进行随机选出N个类，再按照类别随机选出sample。

训练则通常采用一种称为Episodic Training的方法。

3.和迁移学习的区别：如paper中所述，迁移学习和元学习都是本文设计到的重点。所以有必要了解一下二者的difference&connection。目标上看，二者都是为了增加学习器在multi-task中的泛化能力。但meta-learning更偏重于task和data的双重采样。for instance，对于一个N分类的task，meta-learning只会建立一个N/2的classifier，每轮train的episode都能被视为一个子任务，学习到的 $F(x)$ 则可以帮助其在unseen task中建立mapping。相比之下，迁移学习更多强调从A任务到B任务的能力迁移，较为不强调任务空间的概念。

3.实验做法

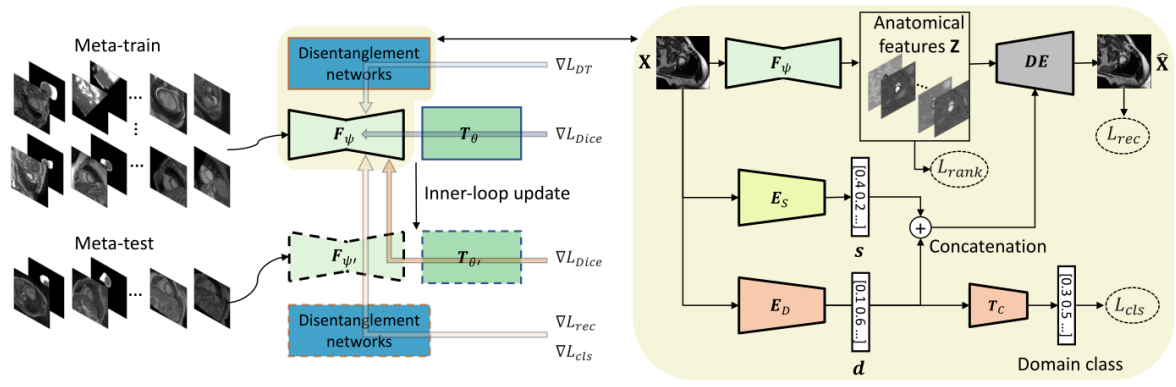
3.1 Summary

如第一节中提到的泛化、迁移、解耦存在的问题，以及在MIS中的扩展性不够的问题。文章提出采用meta-learning的方式，并解耦与domain shifts相关的表示。

通过重构 (reconstruction) 来学习这些完备的表示带来无监督学习的优势，因此我们可以通过使用任何其他源域的未标记数据来更好地模拟domain shifts。这里，我们主要考虑两种源的shifts：1.由于扫描仪和扫描采集设置方法不同。2.人种变化带来的差异。我们的task是MIS，所以我们希望对解剖学的变化更加敏感，对图像特征的变化不敏感。

在下图的Fig中，Z代表解剖学特征，为网格状的feature。s和d是两个vector，用来编码共同以及特定领域的图像特征。

同时，采用了一些特定设计以及学习偏差来解耦。比如，采用网格状的Z用于分割被证明可以提升效率。采用low-rank的正则化来促进Z和s与d的解耦。基于梯度的meta-learning也能提升其在unseen task上的性能，并支持其解耦。



3.2 Method

3.2.1 Denotation

$\mathcal{D} = \{\mathbf{X}_i^k, \mathbf{Y}_i^k\}_{i=1}^{N_k}, k \in \{1, 2, \dots, K\}$ 代表一个multi-domain的训练集，被定义在 $\mathcal{X} * \mathcal{Y}$ 的联合空间中。

\mathbf{X}_i^k ：从第k个源域中得到第i次训练的训练数据。

N_k ：第k个源域的训练样本数量。

我们试图去学习这样一个模型：包含一个特征网络 $F_\psi: \mathcal{X} \rightarrow \mathcal{Z}$ 来提取解剖学特征Z，以及一个任务网络 $T_\theta: \mathcal{Z} \rightarrow \mathcal{Y}$ 来预测分割mask，其中 ψ 和 θ 是网络参数。

3.2.2 面向域泛化的基于梯度的元学习

通过在一系列的episode上训练来模拟domain shift，每轮训练迭代中随机划分源域 \mathcal{D} 为meta-train Set和meta-test Set。在元-训练过程中，通过优化 $\mathcal{L}_{meta-train}$ 来确定参数：特征网络中的 ψ 以及任务网络中的 θ 。

$$(\psi', \theta') = (\psi, \theta) - \alpha \nabla_{\psi, \theta} \mathcal{L}_{meta-train}(\mathcal{D}_{tr}; \psi, \theta).$$

最终的目标函数：

$$\operatorname{argmin}_{\psi, \theta} \mathcal{L}_{meta-train}(\mathcal{D}_{tr}; \psi, \theta) + \mathcal{L}_{meta-test}(\mathcal{D}_{te}; \psi', \theta').$$

3.2.3 学习解耦表示

单独编码特定域的图像特征作为补充的基于vector的变分表征，因此，模型致力于学习两独立的向量表征：s用来捕捉域之间的共同图像特征，d用来捕捉每个域的特有特征。此外，将空间解剖学信息编码在一个单独的特征z中用来完成对s和d的解耦。

首先，输入图像X首先被encode成为s和d，接着有一个shallow的分类器 $T_C(d)$ 来预测X的源域(c)，接着用解码器DE结合提取的特征Z以及表征s和d，重铸输入图像。

$$\hat{\mathbf{X}} = DE(\mathbf{Z}, \mathbf{s}, \mathbf{d})$$

DE使用AdaIN layers，AdaIN可以提升解耦效果并且促进Z更好的encode信息。

为了实现这种三重解耦，我们考虑了几种Loss：1. $\mathcal{L}_{KL}(\mathbf{s}, N(0, 1)), \mathcal{L}_{KL}(\mathbf{d}, N(0, 1))$ ，其中 $N(0, 1)$ 为标准正态分布（高斯分布）。2. Hilbert-Schmidt Independence Criterion Loss，要求s和d相互独立。3. 分类loss： $\mathcal{L}_{cls}(\mathbf{c}, \hat{\mathbf{c}})$ 使得d和域相关信息高度相关。4. 重建loss $\mathcal{L}_{rec}(\mathbf{X}, \hat{\mathbf{X}})$ ，定义为二者的l1距离，来学习无监督表征。

进一步，又使用了rank regularisation来进一步实现解耦。

$$\begin{aligned}\mathcal{L}_{DT} = & \lambda_{rank} \mathcal{L}_{rank}(\mathbf{Z}) + \lambda_{KL}(\mathcal{L}_{KL}(\mathbf{s}, N(0, 1)) + \mathcal{L}_{KL}(\mathbf{d}, N(0, 1))) \\ & + \lambda_{rec} \mathcal{L}_{rec}(\mathbf{X}, \hat{\mathbf{X}}) + \lambda_{HSIC} \mathcal{L}_{HSIC}(\mathbf{s}, \mathbf{d}) + \lambda_{cls} \mathcal{L}_{cls}(\mathbf{c}, \hat{\mathbf{c}}),\end{aligned}$$

c代表域label。经过预训练已经提前训练好了超参数 $\lambda_{rank} = 0.1$, $\lambda_{KL} = 0.1$, $\lambda_{rec} = 1$ and $\lambda_{cls} = 1$.

3.2.4 元学习/元测试目标

meta-train的目标包括两部分：

$$\mathcal{L}_{meta-train} = \lambda_{Dice} \mathcal{L}_{Dice}(\mathbf{Y}, \hat{\mathbf{Y}}) + \mathcal{L}_{DT}$$

在meta-test步骤，模型需要1.准确预测图像分割mask。2.解耦Z, s以及d到和meta-train集一样的level。一个naive的策略是依然使用 \mathcal{L}_{DT} 。但是根据前人研究的分析，meta-test阶段训练不稳定。

在考虑了固定的学习以及设计biases，解耦度可以被定义为重建质量以及域分类精确度。

最终的loss：

$$\mathcal{L}_{meta-test} = \lambda_{Dice} \mathcal{L}_{Dice}(\mathbf{Y}, \hat{\mathbf{Y}}) + \lambda_{rec} \mathcal{L}_{rec}(\mathbf{X}, \hat{\mathbf{X}}) + \lambda_{cls} \mathcal{L}_{cls}(\mathbf{c}, \hat{\mathbf{c}}).$$

4.实验结果

4.1 实验设置

两个标准公开数据集： M&M (Multi-centre, multi-vendor & multi-disease cardiac image segmentation dataset) 以及SGGM (Spinal cord gray matter segmentation dataset)：

M&M：320个对象，来源于不同国家的不同诊断中心，并且用不同的扫描设备，具体细节略。

SCGM：数据同样收集来源于四个不同的医疗中心，使用不同的MRI系统。每个域都有10个标记对象和未标记对象。

基准模型： nnUnet：一个自适应的基于2D和3D U-Nets的模型，并不特定针对于域泛化。

SDNet+Aug：把输入图像划分为空间解剖学以及非空间形态因素。与我们的方法相比，没有meta-learning的策略，只解耦隐含特征。

LDDG：最新最优的基于域泛化的医学图像分析模型，也使用了rank loss在全监督情形下。

4.2 实现细节及结果

用Adam分类器进行优化。使用一定比例的标记数据，剩下的则为未标记数据。使用Dice和Hausdorff Distance作为评估指标。

Table 1. Dice (%) results and the standard deviations on M&Ms dataset. For “SD-Net+Aug.” and our method, the training data contain all the unlabeled data and 2% or 5% of labeled data from source domains. The other models are trained by 2% or 5% labeled data only. Bold numbers denote the best performance.

Source	Target	nnUNet	SDNet+Aug.	LDDG	SAML	Ours	
2%	B,C,D	A	52.87 ₁₉	54.48 ₁₈	59.47 ₁₂	56.31 ₁₃	66.01₁₂
	A,C,D	B	64.63 ₁₇	67.81 ₁₄	56.16 ₁₄	56.32 ₁₅	72.72₁₀
	A,B,D	C	72.97 ₁₄	76.46 ₁₂	68.21 ₁₁	75.70 _{8.7}	77.54₁₀
	A,B,C	D	73.27 ₁₁	74.35 ₁₁	68.56 ₁₀	69.94 _{9.8}	75.14_{8.4}
	Average		65.94 _{8.3}	68.28 _{8.6}	63.16 _{5.4}	64.57 _{8.5}	72.85_{4.3}
5%	B,C,D	A	65.30 ₁₇	71.21 ₁₃	66.22 _{9.1}	67.11 ₁₀	72.40₁₂
	A,C,D	B	79.73 ₁₀	77.31 ₁₀	69.49 _{8.3}	76.35 _{7.9}	80.30_{9.1}
	A,B,D	C	78.06 ₁₁	81.40 _{8.0}	73.40 _{9.8}	77.43 _{8.3}	82.51_{6.6}
	A,B,C	D	81.25 _{8.3}	79.95 _{7.8}	75.66 _{8.5}	78.64 _{5.8}	83.77_{5.1}
	Average		76.09 _{6.3}	77.47 _{3.9}	71.29 _{3.6}	74.88 _{4.6}	79.75_{4.4}

表1、2、3和4（图见原paper）显示，始终在心脏和灰质分割上取得了最好的泛化性能。特别是在低数据状态下，与最佳表现基线相比，我们在M&Ms上提高了Dice≈5%，在SCGM上提高了≈3%。对于m&m中100%的注释(见附录)，我们的模型仍然优于基线。我们还在附录中展示了定性结果，其中可以直观地观察改进后的性能。

M&M：由于解耦可以使用未标记的数据，从而使模型表现得更好。从而得到比baseline模型更优秀的泛化性能。（还比较了几个baseline的缺点，略）。

SCGM：在SCGM上也取得了一致的改进，在其他器官也得到了应用。模型受益于每个领域额外的10个未标记的主题，从而带来更好的整体性能。

消融实验：测试了解耦过程中的关键变量。这里引用了一个Distance Correlation（DC）来衡量解耦程度。最终结果得到 $\mathcal{L}_{rank:t}$ 、 \mathcal{L}_{cls} and \mathcal{L}_{HSIC} 都促进了解耦，得到了更好的模型泛化效果。

5. 总结

如前面所述（尤其是part I），本文的贡献在于提出了一个新颖的实现域泛化的半监督元学习框架。

使用了重建方法来促进解耦帮助模型使用未标记的数据（这样的脏数据相比于精心标注好的数据可获得性大增）。使用了low-rank的正则化作为学习的bias来提高解耦从而提升泛化性能也是文章的一大亮点。大量的实验结果表明，尤其在标记数据不足的情况下，模型都比当前最优的基线模型优秀。

Expectation：其他域的补充未标记数据能否进一步提升模型性能。