已提交ieee医学成像汇刊



1

生成器vs细分器:伪健康

合成

张云龙，林欣，庄毅宏，孙丽艳，黄玥，丁兴浩，王桂生，于益洲，Fellow, IEEE, Yang Lin

***摘要-从病理图像中合成出特定学科的病理图像对算法的开发和临床实践都有价值。*近年来，基于生成式对抗网络(GAN)的方法被提出并取得了良好的效果。在本文中，我们发现GAN中的鉴别器不能准确地识别图像中的病变，进一步阻碍了生成良好的伪健康图像。为了解决这一问题，我们创造性地引入分段器作为鉴别器。能够准确定位病变，有助于提高伪健康图像的视觉质量。然后，我们提出了一种图像增强技术，将原始图像和合成图像之间的残差添加到原始图像中，并利用该技术来解决医学图像分割中存在的低对比度问题。此外，针对缺乏合适的度量标准来衡量合成图像的健康程度的问题，我们提出了一个稳定可靠的度量标准，充分利用标签噪声的两个属性。在公共数据集BraTS上的实验表明，该方法的性能明显优于现有方法。特别是，仅利用30%的训练数据，我们的方法就达到了与现有方法相当的性能。此外，我们还验证了该方法在数据集lt上的有效性。**

***索引术语-医学图像合成，医学图像分割，对抗训练，图像增强，标签噪声***

我的介绍。

伪健康合成被定义为从一个病理图像[3]，[4]合成一个特定学科的无病理图像。生成这样的图像已经被证明对各种医学图像分析任务很有价值，

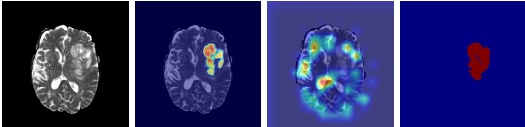
工作是支持中国国家重点研究和发展项目的一部分(2019号yfc0118104),在浙江省重点科研发展计划的一部分(2020号c03073),部分由中国国家自然科学基金资助81671766,61971369,U19B2031, U1605252, 61671309,部分开放基金的科学和技术在自动目标识别实验室6142503190202,部分基础研究基金为中央大学20720180059,20720190116, 20720200003，部分由腾讯开放式基金。

张云龙、林欣、庄奕红、孙立艳、黄玥、丁兴浩，厦门大学信息学院，厦门361005 (e-mail: dxh@xmu.edu.cn)。

王桂生，中国人民解放军总医院第三医学中心放射科(e-mail: wanggs1996@tom.com)

Yu Yizhou就职于Deepwise人工智能实验室，北京100125 (e-mail: yizhouy@acm.org)

杨琳(音译)，杭州西湖大学工程学院，浙江杭州310012 (e-mail: yanglin@westlake.edu.cn)

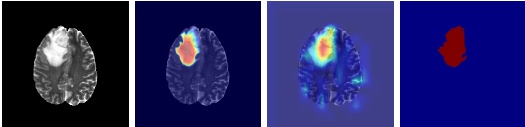


(一)输入

(b) Segmentor

(c)分类器

(d)标签



(e)输入

(f) Segmentor

(g)分类器

(h)标签

图1:分类器和分段器的“视觉解释”。(a-d)代表一个例子。它们分别是一个输入，由Grad-CAM[1]和Seg- Grad-CAM[2]生成的类激活图，以及肿瘤注释。(e-h)是另一个例子的对应图像。

如分割[3]、[5]-[8]、检测[9]，并为病理分析[7]、[10]提供额外的诊断信息。在临床应用中，一个完美的伪健康图像应该同时保持健康(即不健康)。在合成图像中，病理区域与健康区域是和谐的;，属于与输入相同的主题)。注意这两者都是必不可少的。前者的重要性不言而喻，而后者也是相当重要的，因为创造另一个健康的形象是没有意义的。

近年来提出了几种基于gan的方法[4]、[7]、[10]来解决伪健康合成问题，并取得了良好的效果。这些方法的关键部件是一个发生器和一个鉴别器。前者被训练成将病理图像转换成相应的健康图像，而后者与前者竞争，目的是通过双向分类器区分合成图像和健康图像。但是，选择分类器作为鉴别器存在不能准确定位病变的缺点，具体描述如下。首先，分类器不仅基于病理区域，而且基于异质性健康区域(例如，图1(c)中的“肿瘤”解释同时突出显示了健康区域中的病理区域和异质性像素)做出决策，这强制生成器修改突出显示的区域

已提交ieee医学成像汇刊

2

确认分类决策的区域。此外，由于整个合成图像与原始图像有偏差，主体身份将被删除。其次，分类器基于最明显的病理信号像素做出决策。也就是说，只有部分具有最明显信号的病理像素才足以保证正确的分类(例如，图1(g)中的“肿瘤”解释突出了病理区域的核心像素)。因此，剩余的病态信号较弱的像素没有向健康的外观改变，最终损害了合成图像的健康。

为了克服上述缺点，我们选择了隔离指导者作为鉴别器。与分类器相比，分段器能更准确地识别病理区域(图1(b)和1(f)中的“tumor”解释准确地突出了肿瘤区域)。因此，通过分别保留健康像素和转化病理像素，可以同时保持主体的身份和健康状态。前者通过视觉残差实现，后者通过生成器和分段器之间的对抗性训练实现。

在对抗性训练过程中，合成图像中的病态像素逐渐向健康像素靠拢。这些健康的像素应该被标记为“no tumor”，但在训练片段器时被错误地标记为“tumor”，最终导致片段器泛化性能较差。为了解决这个问题，我们提出了一种差分感知损失，它通过对那些看起来令人困惑的健康像素进行静音来提高分段器的泛化能力。

除了提供主观的无病理结果，我们的方法生成的合成图像也有助于分割任务。众所周知，低对比度是医学图像的一个讨厌的本质，这可能导致难以区分不同的解剖结构。因为我们的方法可以有效地解决病理图像空间和健康的信号,我们设计一个图像增强技术的突出病变只需添加病理信号提取到原始图像,然后运用这种技术来提高病变的对比和正常组织,进一步提高分割性能。

伪健康综合的另一个主要障碍是缺乏衡量“健康”的良好定量指标。最近，Xia等人[4]提出了一种度量方法，即对病理图像进行预训练分割器，然后对合成输出的剩余病理信号进行估计。剩余的病理信号间接反映健康状态，越低表示越健康。然而，模型选择对最终结果的影响很大，因此可重复性和再现性较低。更糟糕的是，当在分布空间中出现远离病理性和健康区域的伪影时，它是脆弱和不可靠的。另一方面，主观评价被视为最终决定健康[4]的金标准。然而，它是费时和昂贵的，同时受到观察者之间和观察者内部的可变性，因此也偏离了再现性。受标签研究的启发

拟合错误标签需要更多的时间[11]，较高的学习率阻碍了错误标签的记忆，我们通过估计收敛速度来衡量健康度，提出了一个稳定可靠的度量指标。

为了为这些说法提供证据，我们对BraTS数据集[12]-[14]进行了广泛的评估。此外，我们还在LiTS[15]上展示了部分结果，表明我们的方法有潜力扩展到不同的模式和器官。

本文的其余部分组织如下。在第二部分，我们回顾了伪健康合成的相关文献和一些有代表性的对抗方法。在第三节中，我们提出了提议的全球资源vs。在第四部分，我们提出了一个新的度量，a - dice，来衡量“健康”，并通过实验和分析突出其优势。在第五节中，我们进行了广泛的实验来验证我们提出的方法的有效性。第六节总结全文，并进一步讨论了我们的方法的局限性和未来的工作。

2相关的工作

*答:Pseudo-healthy合成*

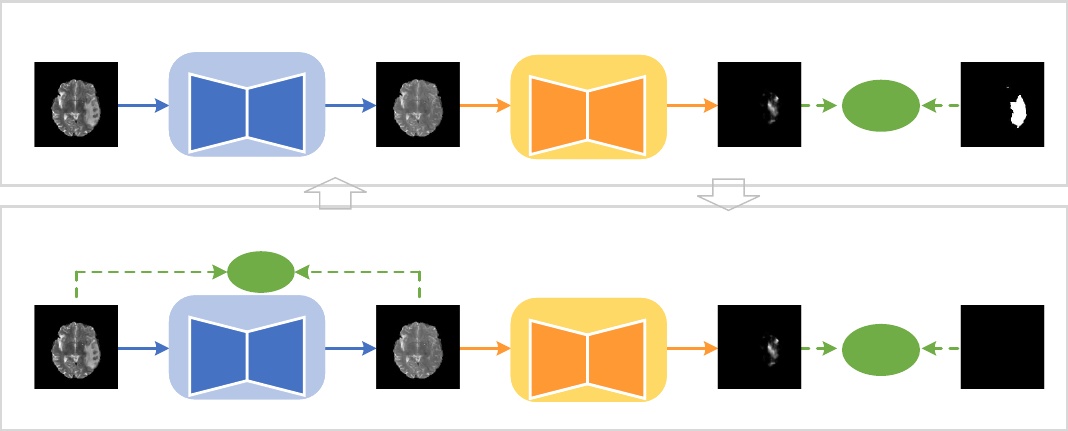
近年来，伪健康合成因其下游任务[4]-[10]，[16]的潜力而受到医学图像分析界的广泛关注。我们把相关的工作分为两大类:病理缺陷(即病理缺陷)。[16] -[18]和基于病理充分性的方法(即，仅在训练阶段包含健康图像)。，在训练阶段拥有大量的病理和健康图像)[4]，[7]，[10]。

基于病理缺陷的[16]、[19]-[25]方法总是与无监督异常检测/分割[26]密切相关，目的是通过学习在训练阶段压缩和恢复健康解剖来学习规范分布。在随后的测试阶段，首先将病理图像压缩到潜在空间。这些方法假定获得的潜在表现接近于伪健康图像的对等物。基于这一假设，从潜在表征中重构出虚假健康图像。实际上，这些方法的假设都超过了理想的[16]，[18]。实际上，由于无法获得伪健康图像对应的最优潜在表征，健康度和主体身份并没有得到保证。

基于病理充分性的方法[4]，[7]，[10]从另一种方式处理伪健康合成。这些方法在训练阶段除了引入相应的图像级[10]或像素级[4]、[7]的病理标注外，还引入了病理图像。Baumgartner等人提出了一种基于gan的基本方案，该方案由一个生成器和一个鉴别器组成。生成器被训练来欺骗鉴别器，并保持重建图像接近原始图像，鉴别器的目的是区分合成图像从未配对的健康图像。需要注意的是，该方法仅使用了图像级的标注，不能准确地翻译出病态像素并保持健康像素。为了缓解这个问题，PHS-GAN[4]和ANT-GAN[7]引入了像素级注释。同时，他们两个

作者张等:生成器与分段器:伪健康合成

3.



步骤A:修复发电机，更新分段器

G

年代

𝐿𝑠1

xp

xs

S (x)s）

欧美

步骤B:更新生成器，修复segmentor

𝐿𝑅G

年代

𝐿𝑠2

xp

xs

S (x)

本产品

图2:培训流程。我们对生成器G采用了编码器-解码器架构，对分段S采用U-Net架构。模型通过迭代交替步骤a和步骤b进行优化。在步骤a中，我们固定生成器G，并使用L更新分段Ss1．在步骤B中，我们修复了分段器S，并用L更新了生成器Gs2 + LR．

为Cycle-GAN[27]的变体。具体来说，PHS-GAN考虑了一对多问题，将病理信息从看似健康的信息中分离出来，并使用像素级标签提取病理的位置和形状。ANT-GAN在将Cycle-GAN应用于伪健康合成的过程中提出了两个改进。第一个是简化优化的捷径，第二个是屏蔽L2损耗，更好的保留正常区域。

我们的方法与Sun et al.[7]和Xia et al.[4]在实验设置下的工作相同，但我们的动机与他们有明显的不同。他们试图从大量健康的图片中学习，将病态的图片转换成健康的外观。相比之下，我们的方法进一步明确利用健康区域和病理区域的外观差异所包含的差异信息，试图通过弥补这些差异来生成无病理图像，直到二者达到和谐。

*b .对抗训练*

采用分段器作为鉴别器的想法是受到对抗训练[28]广泛应用的启发。原始GAN的鉴别器用于鉴别真假图像。在域自适应中，DANN[29]采用了一个鉴别器来区分源域和目标域。在对抗性攻击的研究中，鉴别器是一个分类器，用来将一个例子分类到相应的类[30]，[31]。在图像翻译任务中，鉴别器也是一个分类器，用于检测翻译和标签[27]，[32]之间的高级结构差异。最近Naveen等人将对抗性训练应用到自我监督学习中，并采用一个分类器作为判别器来预测借口标签。并与使用不同分类器作为判别器的相关研究进行了比较

进一步发展对抗性训练范式，并将鉴别器扩展到像素级密集预测任务。

3方法

本节介绍了基于像素级标记的病理-充分性伪健康合成的GVS方法。假设一组病理图像{xp}与它们的像素级病变注释{yt}。我们的目标是训练一个生成器G可以翻译病理图像xp 对应的合成xs 具有优越的健康性和主体同一性。

*GVS基本流程图*

所提出的GVS训练流程如图2所示。生成器通过迭代交替步骤A和步骤b，逐步合成伪健康图像。具体步骤如下。

**如图2中的步骤A所示，我们修复生成器G，更新分割器S，对合成图像中的病变进行分割。**损失定义为:



(1）

L在哪里ce 交叉熵损失。

**步骤b。在这一步中，我们修复了分割器S，更新了生成器G，目的是去除病变，保持病理图像的主体身份。**一方面，期望生成器G能够合成不含病变的伪健康图像;因此，一种对抗性损失被利用了，

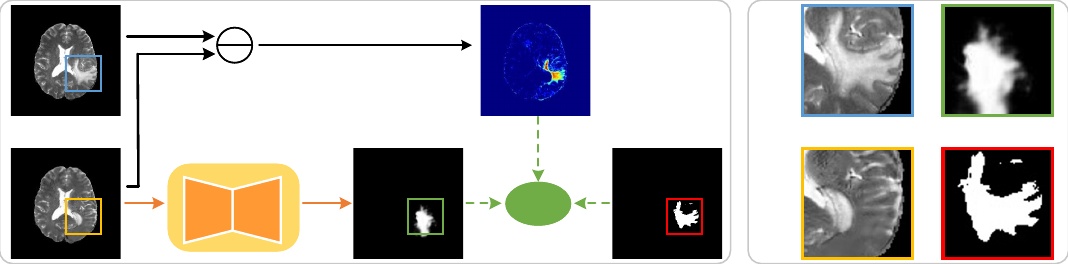


(2）

在yh 表示与y大小相同的零矩阵t．为了欺骗节段者，产生器进一步补偿病理和健康之间的分布差异

已提交ieee医学成像汇刊

4



(a) (b)

年代

𝐿𝑠1

图3:(a)像素级加权交叉熵损失框架。(b)蓝、黄、绿、红框分别表示病理图像、合成图像、分割预测和病变标注。

地区。另一方面，合成图像在视觉上应与病理图像保持一致。因此，使用现有伪健康综合方法[7]中提出的剩余损失:



(3）

L在哪里mse 表示pixel-wise L2 的损失。G的总训练损失为:

（4)

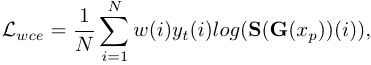


其中λ表示一个权衡健康与身份的超参数，λ > 0。

在迭代训练过程中，分割器S和生成器G相互竞争。片段S试图检测正常和病理区域之间的差异，而生成器G试图弥补它们。最终，G协调病理和正常区域，合成健康样图像。

*B.训练具有较强泛化能力的细分器*

进一步考虑了分割器的泛化能力。在训练过程中，病理区域逐渐向健康区域转化。如图2(b)所示，病变区域的主要部分已经得到很好的转化，所以这些像素应该被标记为“无肿瘤”。但是，基本的GVS在训练分段器时仍然将该区域内的所有像素视为“肿瘤”，这误导了分段器，严重损害了它[11]的泛化能力。如图2(b)所示，我们观察到segmenter的预测严重偏离标签，说明segmenter泛化程度较差。为了应对这一挑战，我们采用了对经过良好转换的像素进行静音的策略。我们首先发现输入和合成输出之间的不同映射在很大程度上反映了像素转换的好坏。也就是说，实质性的差异表示良好的转变，较小的差异表示差的转变(rf。图3 (B))。因此，我们利用差分图作为衡量转换程度的指标，提出了一种差分感知损失(DWL)来分割病变。



(5）

增强图像差异地图地面真相

输入

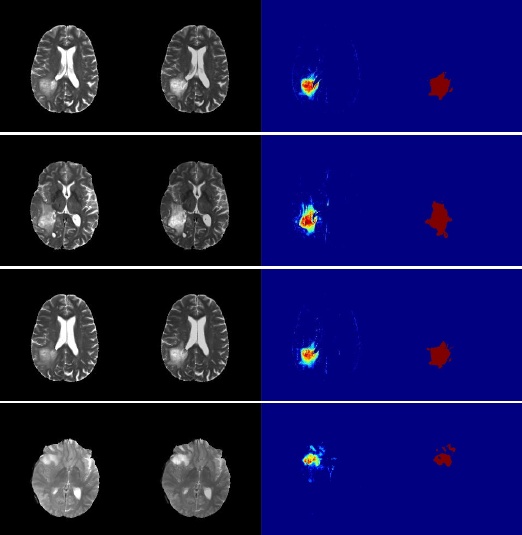


图4:BraTS数据集增强结果示例。从左到右的图像分别代表输入、增强图像、差异图和地面真相。当α = 1.0时得到这些结果。

其中N为像素的个数。定义与差分映射相关联的权值w为:

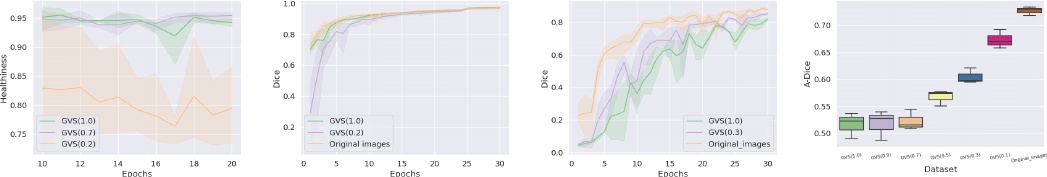


(6）

m =标准化(xp −G (xp)表示归一化差分映射。在这项工作中，w[w < 0.1] = 0.1，因为最小值并不代表完美的转换，有必要保持一个微妙的惩罚。通过升级分割损失L，提出完整的GVSs1 从方程2到感知差异的Lwce．

作者张等:生成器与分段器:伪健康合成

5



(一)“Healhiness”[4]

(b)骰子曲线(lr=0.0001)

(c)骰子曲线(lr=0.1)

(d)的骰子

图5:(a)文献[4]在不同时期提出的“Healthiness”。(b)当分段器以lr = 0.0001进行训练时，每个时期的骰子在训练数据上的得分。(c)在lr = 0.1的条件下，分段器在每个epoch的训练数据上的骰子得分。(d) A-Dice对不同合成图像的评价。这里，GVS(1.0)表示GVS生成的合成图像，使用100%的训练数据进行训练。GVS(0.1)、GVS(0.2)、GVS(0.3)、GVS(0.5)、GVS(0.7)也可以由此推导。很明显，当一个> b时，GVS(a)比GVS(b)更健康。

*C.病灶增强对比*

众所周知，低对比度是医学图像的本质，阻碍了组织分割。图像增强作为解决这一问题的一种有效的预处理技术，在医学图像分割中得到了广泛的应用。

最近Hamghalam等人[35]证实了在BraTS数据集中，增加底层组织的对比度可以有效提高分割任务的泛化能力。受此启发，我们将提出的GVS应用于增强肿瘤和正常组织之间的对比度，然后利用增强后的图像提高分割性能。这个过程可以简单地表述为:

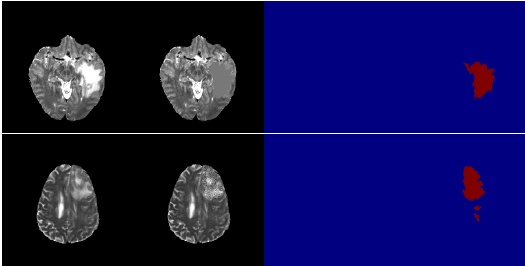


（7)

式中α为增强程度。我们在图4中报告了部分增强的结果。从直观上看，增强后更容易区分肿瘤。接下来，我们利用增强后的图像来训练分割器，这与现有的训练管道保持一致。

四、测量健康

在提出我们的度量之前，我们首先分析了Xia等人提出的“健康度”度量。为了评估健康程度，他们首先训练了一个分割器来从图像中评估病理。然后他们使用这个分割器从合成图像中评估病理，并检查估计的病理区域有多大。要了解更多的实现细节，我们建议阅读文献[4]中的第4.4节。这个过程听起来很合理，但仍然存在以下不足之处。首先，在分布空间中，远离病理和健康区域的伪影对预先训练的分割器是脆弱的。因此，在这些情况下，该指标无法准确评估合成图像的健康程度。例如，图6(b)中的伪造图像明显异常。然而，预先训练的片段不能完全识别异常，并进一步导致虚假的高表现。其次，该指标不能定量甚至定性地反映合成图像的健康程度。我们反复pre-trained



(一)输入

(b)伪造图像

(c)预测

(d)标签

图6:预处理后的分割器对两种不同类型的伪造图像的评价结果。从左到右的图像分别代表输入、伪造图像、伪造图像的预测和病变标注。两行分别注入两个不同的工件。第一行的假图像是用正常组织的平均值填充病理区域生成的。第二行在病理区域加入均值为零、协方差为0.2的高斯噪声。

三个分割器，然后用它们来计算相同合成图像的指标。如图5(a)所示，不同时期和运行时的结果波动很大，特别是在GVS(0.2)的情况下。另一个现象是，GVS(0.7)在某些时期的表现可能会超过GVS(1.0)，这是不合理的，因为在更多数据上训练的模型肯定要优于在更少数据上训练的模型。这些现象说明模型的选择极大地影响了定量和定性结果。

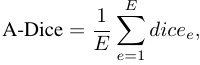
我们的度量是受标签噪声研究的启发。Zhang et al.[11]揭示了一个有趣的现象，即带有假/噪声标签的训练数据需要更多的时间被网络拟合。与此类似，对齐转换良好的像素(即。，这些像素可视为健康像素)，原来的病变注释是反事实的，并妨碍拟合/收敛。为了验证这一点，我们使用健康或syn-

已提交ieee医学成像汇刊

6

(例如，健康图像，GVS(1.0)和GVS(0.2)生成的合成图像)训练一个分段器，然后通过绘制不同时期的训练数据上的骰子分数来评估收敛速度。结果如图5(b)所示，我们发现三种模型最终得到相似的骰子值，但收敛速度不同。在健康图像上训练的模型收敛速度最快，其次是GVS(0.2)和GVS(1.0)。

我们进一步仔细观察了图5(b)中的骰子曲线，发现了定量分析健康度的两个缺点。1)在GVS(0.2)、GVS(1.0)和原始图像上训练的模型表现相似，说明该曲线对不同健康程度的图像具有较低的可判别性。2)初始阶段的性能是不稳定的，可能会误导最终结果。为了解决第一个缺点，我们进一步利用了假/噪声标签的另一个特性。也就是说，较高的学习率会抑制DNN的记忆能力，使其无法拟合[36]标签。因此，简单地提高学习速度(即。，从0.0001到0.1)可以得到如图5(c)所示的结果。我们发现，三种图像的收敛速度差异明显放大。针对第二个不足，我们设计了集成策略来提高稳定性。综上所述，我们提出了一种新的度量来评估收敛速度，即在多时期(a - dice)平均骰子，该度量可以用



（8）

其中E表示总时代，骰子e 表示对e-th时代后的训练数据进行评估的骰子。请注意，较低的A-Dice表示更快的收敛速度，进一步表示更健康的外观。

我们进一步强调了为什么A-Dice可以缓解“健康”中存在的缺陷。首先，任何伪影都是不同于正常组织的分布，可以被分割器快速拟合。因此，由于收敛速度快，将被判定为不健康。第二，尽管每个时代的骰子可能是不稳定的，A-Dice可以通过在多个时代平均骰子来减少偏差。

为了验证A-Dice的稳定性和可靠性，我们在图5(d)中给出了GVS(1.0)、GVS(0.9)、GVS(0.7)、GVS(0.5)、GVS(0.3)、GVS(0.1)七种图像和原始图像的A-Dice，并对其进行定性分析。首先，对于同一幅图像，A-Dice的波动范围约为0.05，小于“Hwealthiness”[4]和主观评价(rf。第二，A-Dice准确地反映了不同合成图像的正确相对顺序。也就是说，当b为>时，GVS(a)的a - dice小于GVS(b)，并且我们发现a - dice在一定程度上与主观评价一致。也就是说，PHS-GAN和ANT-GAN在A-Dice上取得了相近的结果(0.607和0.618,rf。表一)，这也发生在主观度量，"健康程度"(3.800和3.803,rf。这也证实了A-Dice的可靠性。

诉实验

*答:数据集*

我们在两个广泛使用的公共数据集上验证了我们的方法:多模式脑肿瘤分割挑战2019数据集(BraTS19)[12]，[13]和肝肿瘤分割挑战数据集(LiTS)[15]。

**吵闹。**我们使用BraTS的训练集，该训练集包含259 GBM(即。(胶质母细胞瘤)和76个LGG(即。(如低级别神经胶质瘤)的体积已被剥去颅骨，插值到各向同性间隔为1毫米3 并且注册在同一个解剖模板下。每卷包括4种形式(即:， T1, T2, T1c, Flair)，切片大小为240 × 240。我们利用GBM的T2，将它们分成训练集(234卷)和测试集(25卷)。对于每个体积，我们将强度剪辑为[0,V99.5]，其中V99.5是相应体积的99.5%最大像素值。

**分割。**我们使用LiTS的训练数据集，其中包含来自7个不同临床机构的131张肝脏CT扫描。该切片的分辨率为512×512。我们将数据集分为训练集(118次扫描)和测试集(13次扫描)。我们将所有扫描的图像强度值截断到[−200,250]，以去除不相关的细节[37]。

*b .实现细节*

发生器和分段器分别采用编解码器[38]和U-Net[39]架构。我们基于Pytorch实现了该方法。我们采用poly学习速率策略，初始学习速率在每个epoch后乘以(1−totalepochepoch)0.8。基地

学习率设置为0.001,E设置为20。优化器是β的Adam1 = 0.9,β2 = 0.99。Batchsize设置为8为BraTS, 4为LiTS。λ平衡健康和身份设置为10.0。该训练是在NVIDIA TITAN XP GPU上实现的。

*c .基线*

我们将我们的方法与以下三种方法进行比较:

**VA-GAN。**VA-GAN[10]的体系结构包含一个生成器和鉴别器，它们交替更新以生成伪健康图像。此外，VA-GAN还引入了L1 减少合成图像和病理图像之间的差异。我们使用官方代码1 用我们的数据集训练它。

**ANT-GAN。**ANT-GAN[7]是Cycle- GAN的变体。与Cycle-GAN相比，它增加了L2 病理图像与合成图像之间的损失，维持了正常组织，在病理图像转化为健康图像的过程中增加了一条捷径，简化了任务。**PHS-GAN。**PHS-GAN[7]也是Cycle- GAN的变体。为了解决Cycle-GAN的一对多问题，PHS-GAN使用分割网络从病理图像估算出疾病地图，然后使用该地图提供关于疾病位置的信息。我们还使用官方实现2．

1 https://github.com/baumgach/vagan-code

2 https://github.com/xiat0616/pseudo-healthy-synthesis

作者张等:生成器与分段器:伪健康合成

7

*D.其他评估指标*

输入

VA-GAN

ANT-GAN

PHS-GAN

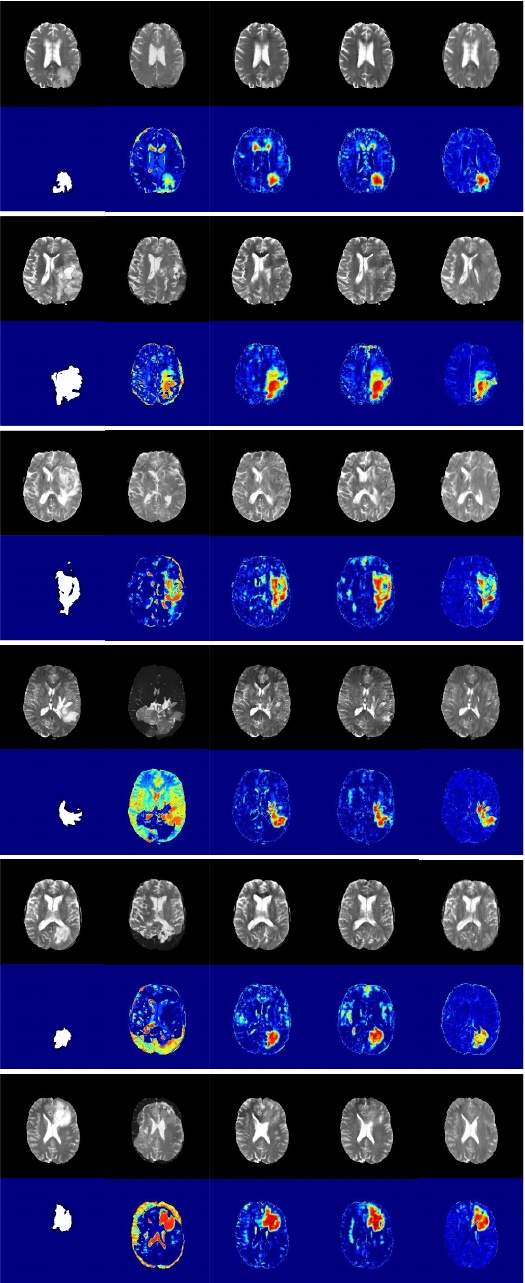
我们的

为了综合评价该方法的有效性，我们分别从健康度和主体身份两个方面对合成图像进行了客观和主观的评价。接下来，我们将分别详细介绍这些指标。

**客观的指标。**伪健康图像的整体质量包括健康度和主体身份。在第四节中，我们描述了我们提出的度量A-Dice，它可以评估伪健康图像的健康程度。在这里，我们介绍了如何测量受试者身份，这可以表示为计算病理和合成图像之间的正常组织的视觉相似性。实现这一目标的常用指标是SSIM和PSNR。这两种方法在实践中都有各自的优点和局限性。

**主观指标。**主观度量是评价伪健康图像质量的金标准。因此，我们采用类似[4]的人工评价方法。与客观指标保持一致，人的评价也由健康性和主体同一性因素组成。主观测试的具体过程如下。

我们随机选取所有比较方法的100个合成输出，排列如下。除病理输入置于第一位置外，所有比较方法的合成输出随机排列在后面四个位置。最后一个位置是病变标注，以便更好地将病理信息传达给评分者。同时，评估者不知道每张图像是由哪种算法生成的。主观评分分为5个等级:5(优秀)、4(好)、3(一般)、2(差)、1(差)。， 5 ~ 1分代表健康程度依次下降;，从5到1的分数表示受试者的身份顺序被抹去)。



*E.与最先进方法的比较*

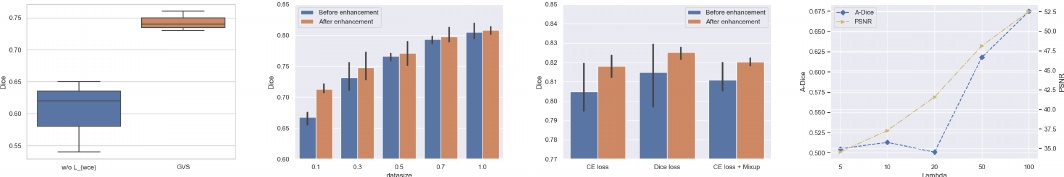
**定性的结果。**定性结果如图7所示。以合成图像的身份和健康度为基础，评价了该方法的有效性。

“健康”可以通过比较病理区域和正常区域是否和谐来判断。如果合成图像中的病理区域与正常区域一致，则该图像为健康图像。相反，在合成图像中，病理区域很容易与正常区域区分开来，这种图像被视为“不健康”。VA-GAN的性能波动很大。部分合成图像具有较好的性能(如图7中的第一和第三个例子)，但由于重建效果较差，大部分图像可以很容易地与健康图像区分开来。PHS-GAN和ANT-GAN可以去除大部分病变，但仍存在一些伪影(rf。最后，提议的GVS移除大部分病灶，并用一个相对健康的区域替换它们。

图7:BraTS数据集定量结果。我们从上到下绘制了六个示例(块)。在每个区块中，第一行显示由VA-GAN、ANT-GAN、PHS-GAN和GVS生成的输入合成图像。第二行为病变标注及对应的差异图。

已提交ieee医学成像汇刊

8



(a) (b) (c) (d)

图8:(a)分段器在GVS中的泛化能力。(b)增强前后分割性能的变化。(c)结合骰子丢失和混淆时分割性能的变化。(d) λ的灵敏度分析。

表一:BraTS数据集的定量结果。我们报告了3次试验的平均值和标准差。

|  |  |  |  |
| --- | --- | --- | --- |
| 方法 | SSIM↑ | PSNR↑ | 掷↓ |
| 原始图像 | - | - | 0.736(±0.310) |
| VA-GAN [10] | 21.89(±1.02) | 0.742(±0.470) | - |
| PHS-GAN [4] | 29.51(±0.47) | 0.966(±0.025) | 0.607(±0.056) |
| ANT-GAN [7] | 28.75(±0.51) | 0.963(±0.018) | 0.618(±0.066) |
| 全球之声(0.3) | **38.79(±0.39)** | **0.995(±0.013)** | 0.615(±0.040) |
| 全球之声w / o Lwce | 37.31(±0.34) | 0.991(±0.012) | 0.549(±0.045) |
| 全球之声(1.0) | 37.31(±0.34) | 0.991(±0.012) | **0.512(±0.035)** |

通过将输入图像与合成图像在结构细节、亮度、对比度等方面进行比较来确定身份。由于ANT-GAN、PHS-GAN和所提出的GVS能够准确地重建正常组织，我们绘制了输入图像与合成图像之间的差异图，以放大差异。该方法结合不同的地图和标记，保留了更多的脑组织细节，比其他方法具有更高的重建质量。值得注意的是，脑脊液呈高像素，且靠近病灶。这些区域被其他方法不同程度地削弱，而被我们提出的方法保存完好。总的来说，在某些病例中，VA-GAN不能保持其身份，并丢失部分病变区域。PHS-GAN和ANT-GAN保留了大脑区域，但丢失了一些细节。在所有的方法中，所提方法的性能最好。

**客观的评价。**在BraTS数据集上的定量结果如表i所示。SSIM和PSNR的结果表明，其他方法由于使用了病变标注，相比于VA-GAN，视觉相似性得到了很大程度的提高。与其他方法相比，该方法进一步提高了视觉相似度。对于健康结果，由于VA-GAN不能重建正常组织，如图7)，其A-Dice值没有意义。因此，它不被考虑。S的值dice ANT-GAN和PHS-GAN的值相似，分别为0.618和0.607。与原始图像相比，分别下降了0.118和0.129。建议GVS的A-Dice值为0.512。与现有的方法相比，有了很大的改进。只是从另一个角度

表二:BraTS数据集的主观评价。我们报告了3个评分者的平均值和标准偏差。

|  |  |  |
| --- | --- | --- |
| 方法 | “健康”↑ | “主体身份”↑ |
| VA-GAN [10] | 2.103(±0.414) | 1.943(±0.519) |
| PHS-GAN [4] | 3.667(±0.824) | 3.800(±0.568) |
| ANT-GAN [7] | 3.707(±0.823) | 3.803(±0.518) |
| 我们的问 | **4.457(±0.330)** | **4.390(±0.491)** |

在30%的训练数据下，GVS取得了显著的性能，其中A-Dice接近ANT- GAN和PHS-GAN，而SSIM和PSNR明显超过现有的方法。

**主观评价。**我们在表2中报告了主观结果。主观结果的总体趋势与客观结果相似。也就是说，GVS表现出最好的性能。随后，PHS-GAN和ANT-GAN得到了相似的结果。最后，VA-GAN占据底部位置。特别是不同方法之间的性能差异被放大，我们的GVS在“健康”和“主体认同”方面都比其他方法具有更明显的性能优势。此外，我们的GVS达到了接近理想的性能(“健康”:GVS (4.457) vs上限(5.0);“主体身份”:GVS (4.390) vs上限(5.0)，这表明从专家的角度来看，GVS的表现近乎完美。

*F.感知差异损失的有效性*

在Section III-B中，我们认为由于合成图像与病变标注的不匹配，基本GVS中的分段器泛化能力较差。因此，生成器将被误导，生成的图像视觉质量很差。我们首先评估了添加差分感知损失前后分段器的泛化能力。为此，将病理图像发送给GVS和GVS训练的分割器，并计算相应的骰子分数。结果如图8(a)所示，说明GVS获得了更高的平均骰子分数和更低的方差相比，w/o L的GVSwce，验证了差分感知损失可以提高分段器的泛化能力。

接下来，我们进一步检验了差异意识缺失对健康状态和主体认同的有效性

作者张等:生成器与分段器:伪健康合成

9

原始

增强

原始

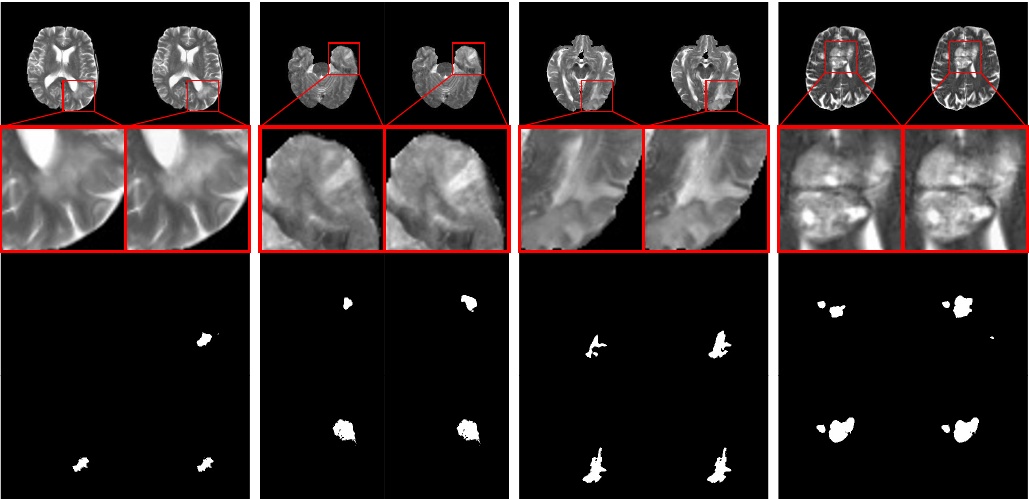
增强

原始

增强

原始

增强



预测放大视图图像

标签

图9:BraTS数据集上增强结果的分割预测示例。注意，这些结果是在α = 0.3时产生的。

表3:不同α情况下的分割性能比较。所有结果在3次试验中取平均值。注意，α = 0.0表示没有增强的图像，括号中的值是通过对比增强前后的分割性能计算出来的。最大改进用粗体表示，第二个用下划线表示。

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 有钱的孩子 | α | 0．0 | 0．3 | 0．5 | 0．7 | 1．0 |
|  | 骰子 | 0.805 | 0.813 (+ 0.008) | 0.812 (+ 0.007) | **0.818 (+ 0.013)** | 0.808 (+ 0.003) |
| 床位数 | α | 0．0 | 0．3 | 0．5 | 0．7 | 1．0 |
|  | 骰子 | 0.643 | 0.655 (+ 0.012) | 0.653 (+ 0.010) | **0.666 (+ 0.023)** | 0.665 (+ 0.022) |

的合成图像。结果如表1所示。我们发现，a - dice有了明显的改善，SSIM略有增加，这说明分割器的泛化能力与合成图像的健康度有密切的关系。

*G.对比增强脑肿瘤分割结果*

增强后的分割性能如表3所示。我们发现在不同的α下，分割性能都有不同程度的提高。接下来，我们将通过定性比较进一步探究增强后可能导致性能改进的原因。我们在图8中展示了四个对比度较低的示例。结果表明，增强突出了病变，最终导致更准确的分割。

然后，我们进行实验来探索在不同数据量下增强带来的改进。结果如图8(b)所示，我们发现数据量越小，改善越明显。我们猜这是因为网络

当数据量很小时，容易过拟合，增强在图像级别明确地规范化了网络，从而更容易找到更好的解决方案。

最后，我们将增强算法与其他能有效提高分割性能的技术(如骰子丢失和混淆)相结合，测试并验证其与其他技术的兼容性。骰子损耗是V-Net[41]中提出的一种损耗函数，是医学图像分割中常用的损耗函数之一。Mixup[42]是一种有效的分类数据扩充技术。最近，它也被引入医学图像分割[43]，并被证明是有效的。结果如图8(c)所示。我们发现这两种技术确实提高了脑肿瘤的分割性能，而我们的增强技术在此基础上进一步提高了分割性能。

*λ的灵敏度分析*

我们的GVS只包含一个超参数λ，用来平衡视觉残差L的功率R 和对抗性损失Ls2 在训练发电机g的时候

已提交ieee医学成像汇刊

10

合成图像

输入

标签

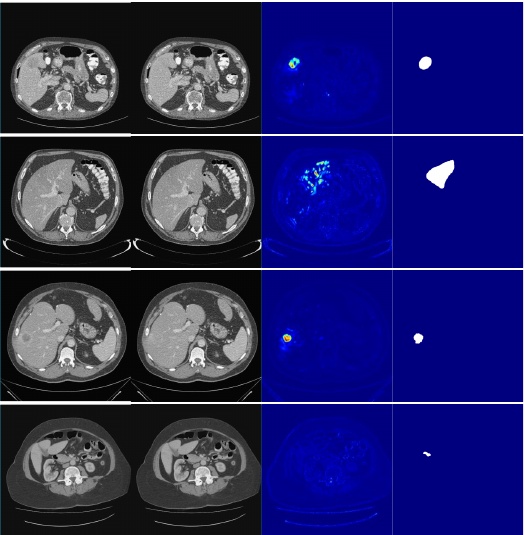


图10:LiTS数据集的定量结果。我们从上到下绘制了四个示例(块)。在每个块中，第一行显示输入、合成图像、差异映射和标签。

我们调查了GVS对不同λ选择的敏感性，并在图9 (e)中报告了结果。我们发现“健康”和“主体认同”占据了相反的位置。良好的健康状态(较低的A-Dice值)意味着较差的主体身份，反之亦然。特别是λ∈[5,20]时，健康度略有变化。

*I.在LiTS数据集上扩展的结果*

提出的GVS也在一个公共CT数据集上进行了评估。结果显示在图10中,表明该方法决定了高对比度图像的身份和转换病变,如图10中第一行所示,但结果对应于图像对比度降低并不令人满意,如底下一行在图10中所示。这是因为LiTS数据集中的图像具有更高的分辨率，更重要的是，病变的体积更小，对比度更低。

我们也将合成的结果应用到对比度增强分割中，结果如图9 (f)所示。我们首先发现性能的提高比BraTS更显著。我们认为这是由于对比度增强在低对比度的情况下更有利于分割。α越高，改进越多的现象也证实了这一观点。

六、讨论与结论

在这项工作中，我们的工作主要包括三个方面。第一个是提出一种对抗框架，包括

生成器和分段器解决了伪健康合成的问题。

实验结果表明，与现有方法相比，我们的算法具有优异的性能。除了V-E小节的结果和分析外，我们进一步强调了健康与主体身份的竞争(如V-H小节所示)。在GAN[44]中也验证了一个相似的属性，即重建与对抗性损失之间的矛盾。因此，对于我们的GVS来说，在健康和主体认同方面取得优异的表现是令人印象深刻的。

在Section III-B中，我们设计了差分感知损失来缓解分割器泛化能力较差的问题，该方法被验证对提高分割器的泛化能力和合成图像(rf。V-F节)。但我们认为这个设计比较粗糙，还有改进的空间。例如，我们不认为最初的合成是不完善的，差异图是没有意义的。因此，如何改进差分感知损失或设计一种新的方案来提高分段器的泛化能力是我们的下一个目标。

在V-F节中，我们揭示了分割器的能力与合成图像的健康度之间的紧密关系。在此基础上，我们进一步推测，加强细分市场有助于进一步提高全球vs的能力。另一方面，3D全卷积网络家族已成为3D医学图像分割领域的主流选择。因此，我们计划对GVS进行3D结构升级，以进一步提高合成性能。此外，所得到的结果是否能够提高三维分割器的分割性能也是可以探索的。

第二部分主要工作是利用病理信息进行伪健康综合，增强正常组织与病变之间的对比。Section V-G和V-I的结果表明，基于GVS的增强技术通过增强对比度有效地提高了当前分割模型的性能。此外，该技术与其他适合分割的技术具有良好的兼容性。因此，它可以作为一个额外的技巧来提高分割性能。最后，但也是最重要的，在较少的训练数据上训练的分段器从增强中获益更多，这是一个有趣和显著的现象。在医学图像分割中，利用小数据集训练模型是一个非常重要和具有挑战性的课题。GVS可能是解决这一问题的有效方法，值得进一步探索。

第三部分提出了一种稳定可靠的衡量合成图像健康度的指标——a - Dice。此外，我们认为这个指标有潜力扩展更多的应用程序。例如，我们计划使用这个度量来测量图像中的病理信息，并探索它与下游任务(如肿瘤分级和生存预测)之间的关系。

综上所述，实验结果积极地证实了我们在本文中的主张。这可能会对伪健康合成的发展产生积极的影响

作者张等:生成器与分段器:伪健康合成

11

1)提高伪健康综合的性能;2)开发一个有意义的下游任务;3)提出一种具有较强重现性和可辨别性的新指标来衡量“健康”。另一方面，我们也认为所提出的GVS和A-Dice可以与其他医学图像分析任务相关联，从另一个角度处理它们。

参考文献

[1] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh，和D. Batra，“Grad-cam:通过基于梯度的定位从深度网络的视觉解释，”在IEEE计算机视觉国际会议学报，2017年，第618-626页。

[2] K. Vinogradova, A. Dibrov，和G. Myers，“通过梯度加权类激活映射实现可解释语义分割”，

[3] C. Bowles等人，“用于白质病变分割的伪健康图像合成”，医学成像模拟与合成国际研讨会。施普林格，2016,pp. 87-96。

[4] T. Xia, A. Chartsias, and S. A. Tsaftaris，“基于病理解纠集和对抗学习的伪健康综合”，医学图像分析，vol. 64, p. 101719, 2020。

[5] C. Bowles等人，“通过图像合成和异常值检测的脑损伤分割”，《神经图像:临床》，第16卷，第643-658页，2017年。

[6] D. H. Ye, D. Zikic, B. Glocker, A. Criminisi, E. Konukoglu，“模态传播:基于数据驱动的正规化的主题特异性扫描的相干合成”，国际医学图像计算和计算机辅助干预会议。施普林格，2013,pp. 606-613。

[7]孙l .，王军，黄勇，丁旭东，H. Greenspan，和J. Paisley，“一种基于对抗学习的医学图像合成方法，”中国生物医学工程学报，2013。

[8] S. Andermatt, A. Horv´ath, S. Pezold和P. Cattin，“使用健康来源图像的分布差异进行病理分割”，国际MICCAI脑损伤研讨会。施普林格，2018,pp. 228-238。

[9] Y. Tsunoda, M. Moribe, H. Orii, H. Kawano, H. Maeda，“用于肺结节检测的胸片数据库伪正常图像合成”，高级智能系统。《科学》，2014，第147 - 155页。

[10] C. F. Baumgartner, L. M. Koch, K. Can Tezcan, J. Xi Ang, and E. Konukoglu，“使用wasserstein gans的视觉特征归类”，《IEEE计算机视觉和模式识别会议论文集》，2018,pp. 8309-8319。

[11] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals，“理解深度学习需要重新思考泛化，”arXiv预印本arXiv:1611.03530, 2016。

[12] S. Bakas等人，“利用专家分段标签和放射组学特征推进肿瘤基因组图谱胶质瘤mri收集”，《科学数据》，第4卷，2017年第170117页。

[13] B. H. Menze等人，“多模态脑肿瘤图像分割基准(brats)，”IEEE transactions on medical imaging, vol. 34, no.;10，页1993-2024,2014。

[14] S. Bakas等人，“在brats挑战中识别脑瘤分割、进展评估和总体生存预测的最佳机器学习算法，”arXiv预印本arXiv:1811.02629, 2018。

[15] P. Bilic等人，“肝肿瘤分割基准(lits)，”arXiv预印本arXiv:1901.04056, 2019。

[16] X. Chen和E. Konukoglu，“使用约束性对抗自动编码器对脑mri病变的无监督检测”，

[17] D. Sato等人，“在急诊头部ct卷中使用自动编码器进行无监督异常检测的初步研究”，《医学成像2018:计算机辅助诊断》，第10575卷。国际光学与光子学学会，2018,p. 105751P。

[18] T. Schlegl, P. Seeb¨ock, S. M. Waldstein, G. Langs, and U. Schmidt- Erfurth，“f-anogan:生成对抗网络的快速无监督异常检测，”医学图像分析，vol. 54, pp. 30-44, 2019。

[19] C. Baur, B. Wiestler, S. Albarqouni, and N. Navab，“Deep autoencoding models for unsupervised anomaly segmentation in brain mr images”，国际MICCAI脑损伤研讨会。施普林格，2018,pp. 161 - 169。

[20] D. Zimmerer, F. Isensee, J. Petersen, S. Kohl, K. Maier-Hein，“使用变分自动编码器的无监督异常定位”，发表于医学图像计算和计算机辅助干预国际会议。施普林格，2019,pp. 289-297。

[21] D. Zimmerer, S. A. Kohl, J. Petersen, F. Isensee, K. H. Maier-Hein，“基于上下文编码的无监督异常检测变分自编码器，”arXiv预打印arXiv:1812.05941, 2018。

[22] N. Pawlowski等人，“使用贝叶斯卷积自动编码器的无监督脑ct病变检测”，2018。

[23] C. Baur, B. Wiestler, S. Albarqouni, and N. Navab，“Scale-space autoencoders for unsupervised anomaly segmentation in brain mri，”MICCAI, 2020。

[24] C. Baur, R. Graf, B. Wiestler, S. Albarqouni, and N. Navab，“隐写异常:抑制循环隐写术用于无监督的大脑mri异常检测”，MICCAI, 2020。

[25] B. Nguyen, A. Feldman, S. Bethapudi, A. Jennings, and C. G. Will- cocks，“基于非监督区域的异常检测与图像修复”，ArXiv, vol. abs/2010.01942, 2020。

[26] C. Baur, S. Denner, B. Wiestler, N. Navab, and S. Albarqouni， " Autoen- coders for unsupervised anomaly segmentation in brain mr images: A comparative study， "医学图像分析，第101952页，2021年。

[27] J.-Y。Zhu, T. Park, P. Isola, and A. A. Efros， " Unpaired image to image translation using循环一致的对抗网络，" in Proceedings of the IEEE international conference on computer vision, 2017, pp. 2223-2232。

[28] I. Goodfellow等人，“生成对抗网络”，《神经信息处理系统进展》，2014年，第2672-2680页。

[29] Y. Ganin等人，“神经网络领域对抗训练”，《机器学习研究杂志》，第17卷，第1期。1、2096 - 2030页。

2016.

[30] I. J. Goodfellow, J. Shlens，和C. Szegedy，“解释和利用敌对的例子，”arXiv预印本arXiv:1412.6572, 2014。

[31] T. Miyato, S.-i。Maeda, M. Koyama, S. Ishii，“虚拟对抗性训练:一种监督和半监督学习的正则化方法”，《IEEE模式分析和机器智能汇刊》，第41卷，第2期。8, pp. 1979-1993, 2018。

[32] P. Isola, j - y。Zhu, T. Zhou, and A. A. Efros，“基于条件对抗网络的图像到图像转换”，《计算机视觉与模式识别学报》，2017,pp. 1125 - 1134。

[33] M. Minderer, O. Bachem, N. Houlsby, and M. Tschannen， " Automatic shortcut removal for self-supervised representation learning， " arXiv preprint arXiv:2002.08822, 2020。

[34] P. Vasuki, J. Kanimozhi, M. B. Devi，“医学影像不同领域的图像预处理技术综述”，载于

*2017年IEEE电气、仪表与通信工程国际会议(ICEICE)。*IEEE, 2017，第1-6页。

[35] M. Hamghalam, B. Lei，和T. Wang，“高组织对比mri合成使用多阶段注意gan胶质瘤分割，”预打印arXiv:2006.05030, 2020。

[36] D. Tanaka, D. Ikami, T. Yamasaki，和K. Aizawa，“联合优化框架与噪声标签学习，”2018 IEEE/CVF计算机视觉和模式识别会议，pp. 5552-5560, 2018。

Li X.， Chen H.， Qi X.， Dou Q.， c.w。傅,中国。“从ct体分割肝脏和肿瘤的H-denseunet:混合密集连接unet”，IEEE transactions on medical imaging, vol. 37, no. 1。12, pp. 2663-2674, 2018。

[38] J. Johnson, A. Alahi, and L. Fei-Fei，“知觉损失的实时风格转移和超分辨率，”ArXiv, vol. abs/1603.08155, 2016。

[39] O. Ronneberger, P. Fischer，和T. Brox，“U-net:卷积网络用于生物医学图像分割”，国际医学图像计算和计算机辅助干预会议。施普林格,

2015年,页234 - 241。

[40] A. Hore和D. Ziou，“图像质量度量:Psnr vs. ssim”，2010年第20届模式识别国际会议。IEEE, 2010，第2366-2369页。

F. Milletari, N. Navab和S.-A。“V-net:全卷积神经网络用于体积医学图像分割，”2016第四届3D视觉国际会议，pp. 565-571，

2016.

[42] H. Zhang, M. Cisse, Y. N. Dauphin，和D. Lopez-Paz，“混淆:超越实证风险最小化”，arXiv预印本arXiv:1710.09412, 2017。

[43] Z. Eaton-Rosen, F. Bragman, S. Ourselin和M. J. Cardoso，“改善医学图像分割的数据增强”，2018。

[44] S. Ramasinghe, K. Ranasinghe, S. Khan, N. Barnes, and S. Gould， " Conditional generative modeling via learning the latent space， " arXiv预印本arXiv:2010.03132, 2020。

已提交ieee医学成像汇刊

12

[45] Zhang p, Zhong Y.， Deng Y.， Tang X.， and X. Li， " deep learning based on deep learning in biomedical image analysis， " journal of biomedical image engineering， " journal of biomedical image engineering, 2019. (ei检索)