这个肠道并不存在：用于逼真无线胶囊内镜图像生成的多尺度残差变分自编码器

Dimitrios E. Diamantis、Panagiota Gatoula、Anastasios Koulaouzidis 和 Dimitris K. Iakovidis，电气电子工程师学会高级会员

摘要— **医学图像合成已成为一种有前景的解决方案，可解决基于图像的临床决策支持（CDS）系统中训练机器学习算法所需的医学标注数据有限这一问题。为此，生成对抗网络（GAN）主要用于生成合成图像以进行数据增强，从而支持算法训练过程。然而，在无线胶囊内镜（WCE）领域，现有公开标注数据集的内容多样性和规模有限，对生成对抗网络的训练稳定性和合成性能均产生了不利影响。为寻求一种切实可行的无线胶囊内镜图像合成方案，本文提出一种新颖的变分自编码器架构，名为“这个肠道并不存在”（TIDE）。该架构由多尺度特征提取卷积块和残差连接构成，即使训练图像数量有限，也能生成高质量且多样化的数据集。与旨在扩充现有数据集的当前方法不同，本研究表明，借助 TIDE，可以完全用人工生成的数据集替代真实无线胶囊内镜数据集，同时不影响分类性能。此外，经验丰富的无线胶囊内窥镜专家开展的定性评估和用户评估研究从医学角度证实，TIDE 合成的正常和异常无线胶囊内镜图像均足够逼真。**

索引词— **《内镜》、胃肠道、计算机辅助检测与诊断、多尺度信息融合**

我们感谢“Smart Tourist”项目（MIS 5047243）对本研究的支持。该项目由“强化研究与创新基础设施”行动实施，该行动由“竞争力、创业与创新”运营计划（NSRF 2014-2020）资助，并由希腊和欧盟（欧洲区域发展基金）共同出资。

本研究涉及人类受试者或动物。所有伦理和实验程序及方案均已获得色萨利大学伦理委员会的批准。

D.E. Diamantis、P. Gatoula 和 D.K. Iakovidis 任职于色萨利大学计算机科学与生物医学信息学系，该校位于希腊拉米亚 35131（电子邮箱：{didiamantis, pgatoula, diakovidis}@uth.gr）。A. Koulaouzidis 任职于丹麦欧登塞大学医院和斯文堡医院外科，以及丹麦南丹麦大学临床研究系（anastasios.koulaouzidis@rsyd.dk）
fComputer Science and Biomedical Informatics, University of Thessaly,35131 Lamia, Greece (email: {didiamantis, pgatoula, diakovidis}@uth.gr)A. Koulaouzidis is with the Dept. of Surgery, Odense UniversityHospital and Svendborg Hospital, Denmark, and the Dept. of ClinicalResearch, University of Southern Denmark, Denmark(anastasios.koulaouzidis@rsyd.dk)

I. 引言

胃肠道疾病是导致死亡率和发病率上升的重要原因，并给医疗卫生系统带来不利的经济影响 [1]。早期发现并准确诊断炎症、血管病变或息肉样病变等病理状况，对预防此类疾病至关重要。在有助于胃肠道筛查的方法中，无线胶囊内镜（WCE）[2] 因其无创特性而成为一种重要选择。与软式内镜（FE）等传统技术不同，无线胶囊内镜通过吞服一枚装有微型摄像头、大小如药丸的胶囊来进行检查。胶囊穿行于整个胃肠道，同时录制RGB视频，随后由专业内镜医师进行审阅。此类视频通常包含超过 60,000 帧，因此评估起来颇具挑战。评估录制的无线胶囊内镜视频通常需要 45-90 分钟，即使是经验丰富的临床医生也难免出现人为错误 [2]。为降低此类错误的风险，研究人员提出了多种基于图像的临床决策支持（CDS）系统 [3], [4] 。计算机辅助诊断系统的一项重要性能指标是其泛化能力。训练数据的可用性和多样性会直接影响此类系统的泛化能力。尽管有若干面向非医学应用的公开标注数据集，其中许多规模相当大，例如 ImageNet [5]，但在医学影像领域，通用数据保护条例（GDPR）[6], 等隐私法规使医学数据采集变得困难，即使数据用于研究也不例外。此外，医学数据标注所需的时间和成本也不利于数据集的获取。对于无线胶囊内镜而言，现有开放标注数据集仍然有限，规模通常小于其他数据集，且多样性往往较低，因为其中包含许多相似图像，异常类型的范围狭窄；在大多数情况下，这些数据集还存在严重的类别不平衡 [7], [8]。因此，使用此类数据集训练当前的深度学习计算机辅助诊断系统，会限制其检测和表征异常的效果。

为了解决这些问题，传统的图像增强技术通过旋转、平移等方式丰富数据集，并

训练图像的缩放版本已被用于提升计算机辅助诊断系统的泛化能力。近年来，人们开始研究依赖深度神经网络的方法，通过生成合成图像进一步增加训练图像的数量。生成对抗网络（GAN）[9] 被视为图像合成任务的标准选择。在自然图像或人像合成等数据充足的应用中，生成对抗网络的表现十分出色 [10–12]。然而，在许多数据有限的情况下，使用生成对抗网络会引发各种训练问题，包括模式崩溃、不收敛和不稳定性 [13]。

当前研究 [14–17], 表明，内窥镜图像合成这一问题仍远未得到解决。在大多数情况下，合成图像会因训练数据不平衡或不足而包含伪影，而生成的合成图像是否可信仍是主要挑战。

为解决这些问题，本研究提出了一种基于变分自编码器（VAE）概念的全新无线胶囊内镜图像生成方法 [18]。本文的贡献概括如下：

- 本文提出了一种名为“这个肠道并不存在”（TIDE）的全新变分自编码器架构，该架构结合了多尺度特征提取与残差学习，以捕获输入体积中富含特征的表征，从而能够使用少量样本进行训练。
- 本文将 TIDE 应用于无线胶囊内镜图像合成，旨在完全以合成训练集取代真实训练集。 其他研究者对无线胶囊内镜图像合成的探索仅限于数据增强，即训练集只有一部分由合成图像构成。
- 该研究开展了一项全面的实验评估，涵盖定量和定性方面，其中包括由无线胶囊内窥镜专家进行的用户评估。评估结果证实，合成数据集与真实数据集极难区分。

在这项研究的基础上，我们创建了一个演示网站[1], ，旨在公开展示 TIDE 的性能，并打造首个面向公众开放的实时肠道数据集生成平台。值得注意的是，该平台采用了我们在 [19] 中提出的可扩展机器学习的算法无关架构（ASML）开发。

本文其余部分分为四个章节。第二节概述生成模型在医学影像领域的应用，重点介绍内镜图像的合成。第三节介绍本文提出的用于图像生成的变分自编码器架构。第四节描述评估方法，并列出所开展实验的对比结果。第五节讨论本研究的见解，最后一节则总结全文并提出未来研究方向。

II. 相关工作

在医学影像领域，合成图像生成引起了广泛的科学关注，研究人员在多种场景下开展了相关研究，其中主要采用生成对抗网络。许多著名的生成对抗网络架构已应用于医学图像合成。例如，DCGAN [20] 已用于生成逼真的脑部MRI图像 [21]。在条件图像生成方面，Mahapatra 等人 [22] 扩展了 Pix2Pix 框架 [12] ，根据手动分割的区域生成逼真的胸部X线图像，其中包含结节。Jin 等人 [23] 扩展了该框架，用于合成 CT图像中的三维结节。另一种广泛应用的对抗框架 PGGAN [24], 经过训练，可生成包含皮肤病变的逼真皮肤镜图像 [25]。[26], 中提出的对抗学习方案 CycleGAN 已用于跨模态医学图像合成；在 [27], 中，该方案用于脑部图像磁共振成像与计算机断层扫描模态之间的非配对图像到图像转换。Cai 等人 [28] 改进了标准 CycleGAN 框架，使其能够在心脏和胰腺图像的磁共振成像与计算机断层扫描模态之间同时进行三维合成和分割，同时保留解剖结构。 [29], 中，研究人员应用无监督图像到图像转换（UNIT）VAE-GAN框架 [30] ，根据分割后的血管树生成眼底图像。Hirte 等人 [31] 应用了另一种混合 VAE-GAN模型 [32], ，以生成多样性更高的逼真脑部磁共振图像。

在内镜领域，生成逼真的图像已被证明是一项更具挑战性的任务 [33], [34]。这是因为内镜图像中不存在特定的固有模式，其内容也无法像计算机断层扫描、磁共振成像或其他医学成像模态那样用定义明确的结构来描述。 [17] 的工作提出了一种以边缘滤波和掩膜输入为条件的生成对抗网络，用于合成图像。该研究聚焦于息肉图像生成，旨在提升结肠镜视频中的息肉检测效果。然而，该方法存在一些局限，包括息肉生成不确定，以及生成的息肉特征在颜色和纹理方面变化不足。为了解决内镜图像分类中的类别不平衡数据集问题，研究者提出了各种对抗框架，通过生成代表不足类别的图像来进行数据增强。在 [15], 中，研究者采用基于图块的方法，将胃癌病灶融入正常胃镜图像中。然而，正如 [14], 中所述，该方法需要手动确定息肉的位置，否则生成结果可能显得不自然，尤其是在颜色和纹理等息肉特征方面。He 等 [14] 提出了一种基于 [35] 的 GAN 模型的数据增强技术，并遵循对抗攻击过程。最近，[36], 介绍的工作对 [37] 提出的生成框架进行了改造，以生成随机息肉掩膜，再将其与正常结肠镜图像结合，构建

[1] https://this-intestine-does-not-exist.com

条件输入。该条件输入用于训练 CycleGAN 模型 [26]，以合成息肉图像。在 [38], 中，研究者提出了一种以息肉掩膜为条件的双重 GAN 框架，用于扩增结肠镜图像中的息肉病灶。然而，[36] 和 [38] 中的合成结果都取决于息肉掩膜的位置，而这些掩膜有时无法与健康的内镜图像自然融合。近期，StyleGANv2（一种最初为人脸合成而提出的 GAN 架构 [11], ）被用于扩充训练数据集，以检测内镜视频中的息肉病变 [34]。尽管该研究在息肉图像合成方面生成了逼真的图像，但其结果难以复现，因为该研究依赖一个包含数千张可供训练图像的私有数据库。

大多数已发表的内窥镜图像合成研究都侧重于生成结肠镜或胃镜图像，包括正常图像和带有息肉样病变的异常图像，而针对无线胶囊内镜图像生成的研究则较少。在无线胶囊内镜检查中，图像分辨率较低，异常图像的数量通常也较少，因为与柔性内镜检查不同，内镜医师无法控制胶囊内镜对发现的病变连续拍摄多帧。此外，无线胶囊内镜更常用于检查小肠，而柔性内镜很难到达小肠；它对于评估炎症性肠病（IBD），尤其是克罗恩病（CD），具有不可替代的价值 [39]。此外，尽管过去几十年来小肠恶性肿瘤/肿瘤的发病率有所上升，但仍显著低于结直肠或胃部肿瘤的发病率 [40]。

在无线胶囊内镜图像生成方面，Ahn 等人. [41] 对 [42], 中最初提出的混合 VAE-GAN 框架进行了改造，以扩充现有的无线胶囊内镜数据集，从而提升基于图像的计算机辅助诊断系统在小肠异常检测方面的泛化性能。然而，这些结果仅具有指示性，并未明确目标病理状况；合成图像还存在模糊问题，因此很容易与真实图像区分开来。

在合成包含多种炎症性疾病的小肠无线胶囊内镜图像这一更复杂的任务中，[43]提出了一种非平稳纹理合成生成对抗网络（TS-GAN）。然而，生成的图像存在伪影，降低了图像合成的质量。这些缺陷部分归因于训练样本数量有限，因为生成对抗网络模型的性能通常取决于训练数据的数量和多样性。为了解决这一问题，[44] 提出了一种用于生成无线胶囊内镜图像的传统变分自编码器，名为 EndoVAE。该模型由采用单尺度滤波器的卷积层按顺序排列构成。

本文提出了一种用于无线胶囊内镜图像生成的不同变分自编码器架构。与以往架构不同，该架构融入了多尺度特征提取方案和残差连接，旨在生成质量更高的图像。据我们所知，此前尚未提出过这种变分自编码器架构。

III. 方法

本研究提出的新型变分自编码器架构名为 TIDE，它采用一系列多尺度模块（MSB），在多个尺度上提取特征，旨在捕获输入的丰富特征表示。该架构还结合了残差连接，以进一步增强特征提取方案，并解决深度学习模型中遇到的梯度消失问题。图 1 展示了一个 MSB。这些模块有两个目的。首先，它们旨在从不同尺度的输入体积中抽象出图像信息。其次，它们能够对提取的信息进行可学习的融合，从而捕获输入体积更多样化的表示。每个模块接收一个特征图体积作为输入，并将其传递给三个并行卷积层。每个卷积层的卷积核数量相同，并通过使用 3×3、 5×5 和 7×7 卷积核尺寸进行卷积运算，从不同尺度提取特征。因此，可以从输入体积中提取小、中、大尺度的特征。这些卷积层的并行输出沿深度方向拼接，然后传递给另外两个连续的卷积层，从而形成输入体积的丰富特征表示。第一个卷积层的卷积核数量等于拼接后特征图中卷积核数量之和，并执行逐点卷积运算，从而有效地促进拼接输出的聚合。下一个卷积层的卷积核数量与并行卷积层相同。上述所有层的激活函数均为修正线性单元（ReLU）。

TIDE 模型的整体架构如图 2 所示。该模型由两部分组成：编码网络和解码网络。编码器接收由 RGB 内镜图像构成的输入体积，可以是正常图像，也可以是异常图像，记为 x。所提出的编码器由卷积层组成，用于执行池化操作并提取多尺度特征。具体而言，编码器的架构依次包括一个具有 16 个卷积核的卷积层，后接四个 MSB。第一个 MSB 包含 32 个卷积核，之后每个模块的卷积核数量依次加倍。在相邻的 MSB 之间插入卷积层，通过将中间特征体积的大小缩减为原来的一半来执行池化操作。这些层分别由 64、128 和 256 个卷积核组成，并使用大小为 3×3 的卷积核执行卷积运算。

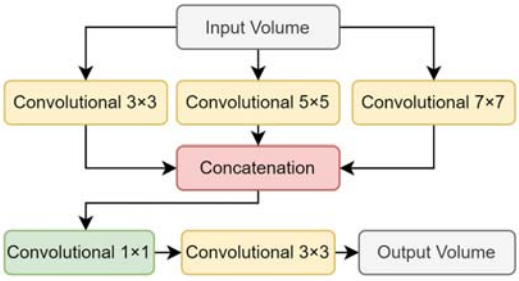


图 1. 多尺度特征提取模块。

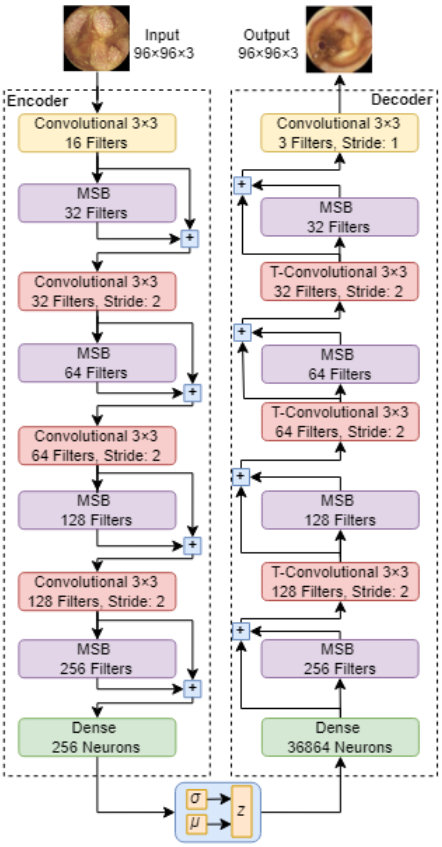


图 2. TIDE 架构。

采用残差连接，以保留从较浅层特征提取模块中提取的特征。因此，每个模块的输入特征图体积都要执行卷积运算，使用 3×3 大小的卷积核，并根据该模块卷积层所采用的卷积核数量进行设置。所得特征图通过加法算子与每个特征提取模块的输出体积进行聚合。最后，将残差连接的输出传递至逐点卷积层，同时保留前一卷积层的卷积核数量。编码器中执行的所有卷积运算均使用 ReLU 作为激活函数。

编码器网络的任务是将输入体积，即内镜图像压缩为两个不同的潜向量，分别对应高斯分布的统计参数：均值 μ 和标准差 σ。因此，编码器卷积部分的输出体积会被展平，并直接输入一个包含 256 个神经元的全连接层；随后连接两个彼此分开的全连接层。每个全连接层均包含 6 个神经元且不使用激活函数，用于估计潜在空间分布的参数 $\boldsymbol{\mu}$，$\boldsymbol{\sigma}$ 。

随后，解码器网络从编码器近似得到的分布中随机采样一个六维向量 $\boldsymbol{z}$ 。因此，解码器根据潜在表示 $\boldsymbol{z}$重建输入体积。解码器架构的顶部设有一个包含 36,864 个神经元的全连接层。接下来，解码器采用编码器的架构，但 MSB 的顺序相反；在解码器中，这些 MSB 由转置卷积层隔开，用于对中间特征体积进行上采样。解码器末端设置了一个转置卷积层，使用 3 个卷积核来预测重建后的 RGB 输入体积。解码器输出的空间维度与内镜图像初始体积 $\boldsymbol{x}$ 的空间维度相对应。因此，所提出的变分自编码器架构合成的图像分辨率与输入图像相同。所有转置卷积运算均采用大小为 3×3 的卷积核，并使用 ReLU 函数作为神经元激活函数；预测层则采用对数 Sigmoid 激活函数。

根据 [18]，用于训练 VAE 模型的反向传播总损失可表示如下：

$$\mathcal{L}(\vartheta, \varphi;\, \boldsymbol{x}_i) = \mathbb{E}_{q(\boldsymbol{z}\,|\boldsymbol{x}_i;\,\varphi)}\, \log p(\boldsymbol{x}_i|\boldsymbol{z}\,;\vartheta) - KL\big(q(\boldsymbol{z}\,|\boldsymbol{x}_i;\varphi)||\, p(\boldsymbol{z};\,\vartheta)\big) \tag{1}$$

其中，第一项对应解码器的重构误差，第二项则近似于 Kullback-Leibler（KL）散度。采用 KL 散度是为了确保编码器将输入体积压缩为服从先验分布 $p(\boldsymbol{z};\ \vartheta)$ 的潜在表示。先验分布 $p(\boldsymbol{z};\ \vartheta)$ 表示为多元高斯分布 $\mathcal{N}(\boldsymbol{z};\ \mathbf{0}, \mathbf{I})$。我们将难以计算的真实后验分布 $p(\boldsymbol{z}|\boldsymbol{x}_i;\ \vartheta)$ 近似为协方差近似呈对角形式的高斯分布，其参数根据公式（2）估计：

$$\log q(\boldsymbol{z}\,|\boldsymbol{x}_i\,;\varphi) = \log \mathrm{N}(\boldsymbol{z}\,;\boldsymbol{\mu}_i, \boldsymbol{\sigma}_i^2\,\mathbf{I}) \tag{2}$$

其中，$\boldsymbol{\mu}_i$ 和$\boldsymbol{\sigma}_i^2$ 是变分自编码器编码器部分的输出。因此，式（1）可表示如下：

$$\mathcal{L}(\vartheta, \varphi;\, \boldsymbol{x}_i) \simeq \frac{1}{2}\sum_{j=1}^{J}\big(1 + \log \sigma_{i,j}^2 - \sigma_{i,j}^2 - \mu_{i,j}^2\big) + \frac{1}{L}\sum_{L=1}^{L} \log \mathrm{p}(\boldsymbol{x}_i|\boldsymbol{z}_l\,;\vartheta) \tag{3}$$

其中

$$\boldsymbol{z}_l = \boldsymbol{\mu}_i + \boldsymbol{\sigma}_i \odot \boldsymbol{\epsilon}_l \ \text{ and } \boldsymbol{\epsilon}_l \sim \mathcal{N}(0, \mathbf{I}) \tag{4}$$

J 表示潜在流形的维数，L 表示从编码器的近似后验分布中进行蒙特卡罗方法采样的样本量，φ 表示编码器网络的参数，ϑ 表示解码器网络的参数，并且符号 $\odot$ 表示哈达玛积运算。

IV. 实验与结果

A. 生成模型的数据集与训练

考虑到无线胶囊内镜的临床实用价值及其在评估小肠炎症性疾病方面的重要性，本研究选用两个数据集进行实验，即 KID数据集2 和 Kvasir-Capsule数据集 [45], [46]。据我们所知，这两个数据集是仅有的公开标注无线胶囊内镜数据集，且包含炎性病变，

例如红斑、糜烂和溃疡。KID [45] 是一个用于评估计算机辅助诊断系统的无线胶囊内镜数据库。该数据库包含 728 张正常图像和 227 张小肠炎性病变图像，分辨率为 360×360 像素。这些图像由 Mirocam® (IntroMedic Co., 首尔，韩国) 胶囊内镜采集。Kvasir-Capsule 数据集 [46] 共包含 34,338 张正常图像和 1,519 张炎性病变图像，分辨率为 336×336 像素。这些图像由 Olympus EC-S10 内镜胶囊采集。尽管该数据集规模相对较大，但其中有许多彼此相似的图像。为减少帧冗余，同时保持两个数据集的规模相当，我们使用 [47] 中所述的图像挖掘方法，从整个数据集中抽取了一个子集，其中包括 728 张正常图像和 227 张具有代表性且互不重叠的异常无线胶囊内镜图像（抽取图像的文件名见本文补充材料）。针对 KID 和 Kvasir-Capsule 数据集，分别使用正常和异常子集单独训练不同的 TIDE 模型即，每个数据集均训练一个模型以生成正常图像，另一个模型以生成异常图像。由于这两个数据集采用不同类型的胶囊内镜采集，因此未将其合并使用。在训练过程中，训练集未采用其他数据增强技术。TIDE 采用提前停止进行训练，训练轮次上限为 5,000 轮，每批包含 128 个样本。考虑到 Adam 优化器在相关应用中的有效性，我们选用该优化器 [48] 训练模型，初始学习率设为 0.001。

B. 定量评估

本实验研究的主要目标是探究，能否仅使用合成图像构成的无线胶囊内镜数据集有效训练分类器，使其准确学会区分真实异常图像和真实正常图像。因此，分类性能可作为定量评估使用 TIDE [49]生成的合成无线胶囊内镜数据集的代表性指标。鉴于无线胶囊内镜数据集类别不平衡，且分类问题为二分类，我们通过考察受试者工作特征（ROC）曲线来量化分类性能。ROC 曲线通过不同决策阈值下真阳性率（TPR）与假阳性率（FPR）之间的权衡，展示分类系统的诊断能力。我们计算了 ROC 曲线下面积（AUC）指标 [50], , 以评估训练所得分类模型的性能；与准确率、敏感度和特异度等仅基于单一决策阈值计算的其他指标不同，该指标不受类别分布不平衡的影响 [50], [51]。

考虑到本研究所需的大量实验工作对计算资源要求较高，我们选择了 LB- FCN 轻量级分类器 [52], , 它是 LB-FCN 的计算效率更高的版本。LB-FCN 是一种用于对内镜图像进行精确分类的先进分类器 [53]。

实验流程概述如下：1）估算 LB-FCN *light* 在每个真实数据集上的基准性能，用于将各数据集中的图像分类为正常和异常（炎症）类别；2）使用训练好的 TIDE 模型随机生成不同组的合成正常和异常无线胶囊内镜图像；3）对于每个数据集，仅使用合成的正常和异常图像训练相同的 LB-FCN *light* 分类器，并且仅在相应的真实图像上进行测试。

表 I 真实图像与合成图像上的分类结果（曲线下面积，%）

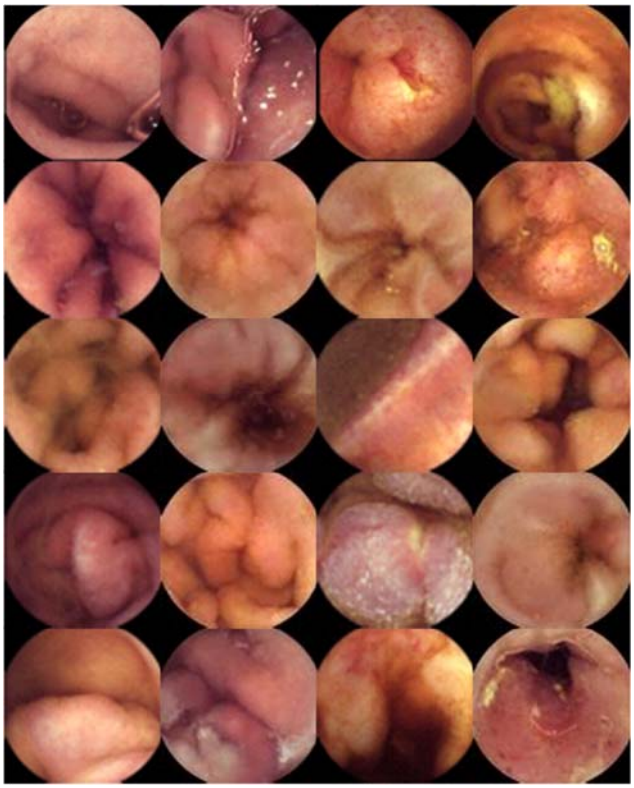
	KID [45]	*Kvasir-Capsule* [46]
Real	90.9 ± 0.8	80.0 ± 0.7
TS-GAN [43]	79.1 ± 0.7	68.8 ± 0.7
CycleGAN [26]	62.9 ± 1.2	72.1 ± 1.1
StyleGANv2 [11]	61.7 ± 1.1	71.8 ± 1.0
EndoVAE [44]	81.9 ± 0.9	71.3 ± 0.3
TIDE	**89.4 ± 1.2**	**80.2 ± 0.6**

为公平比较使用 KID 和 Kvasir- Capsule 数据集获得的分类性能结果，我们采用了数量和比例相同的正常图像与异常图像（即，728 张正常合成图像和 227 张异常合成图像）。所有实验均采用分层 10 折交叉验证方法，以减轻潜在的选择偏倚。

具体而言，合成数据集和包含真实图像的数据集均被划分为十个互不重叠的子集，其中，合成数据集的九个子集用于训练，真实数据集的一个子集用于测试。该过程重复十次，每次选取不同的训练子集和测试子集，直至所有真实数据子集均用于测试。LB-FCN 轻量级的训练设置采用 [44] 中建议的设置。由于合成图像的生成具有随机性，整个实验流程也重复了十次，并记录平均结果及标准差。

将 TIDE 模型定量实验评估的结果与相关最先进模型在 WCE 图像生成方面取得的结果进行了比较即TS-GAN [43], 和 EndoVAE [44]。各模型的分类性能汇总于表 I，其中还列出了使用真实图像在各数据集上获得的参考结果。可以看出，TIDE 的分类性能优于所有参与比较的模型，更重要的是，其性能与参考结果相当。这验证了这样一个假设：真实训练图像可以由合成图像替代，而不会牺牲分类性能。

值得注意的是，我们还尝试了此前用于生成内窥镜图像的方法，包括 CycleGAN [26] 和 StyleGANv2 [11]。训练这些架构颇具挑战，主要是因为可用的无线胶囊内镜训练样本数量较少，由此引发了诸多问题，包括生成的图像质量低、缺乏多样性、存在伪影，以及更重要的模式崩溃 [54]。最初，我们尝试利用预训练模型（在 ImageNet [5] 上训练）进行权重



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

图4. 展示 KID 数据集中小肠组织的真实和合成内窥镜图像。(a) 真实正常图像。(b) TIDE 生成的正常图像。(c) 真实异常图像。(d) TIDE 生成的异常图像。



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

图5. 展示 Capsule-Kvasir 数据集小肠组织的真实和合成内窥镜图像。(a) 真实正常图像。(b) TIDE 生成的正常图像。(c) 真实异常图像。(d) TIDE 生成的异常图像。

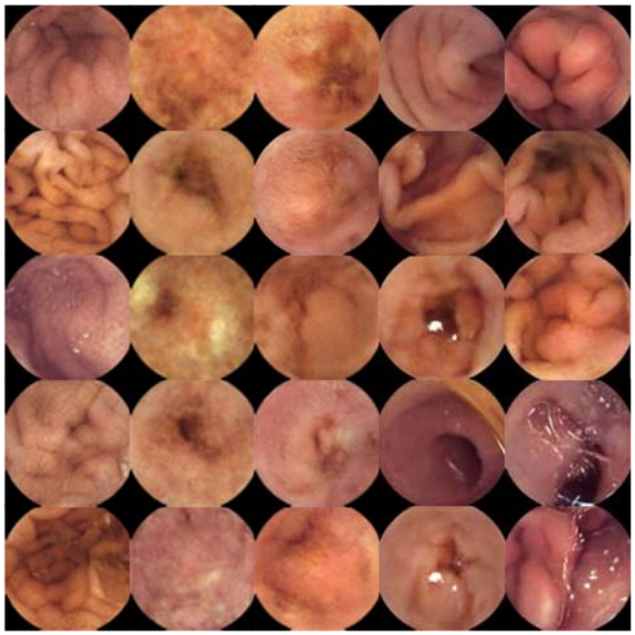
初始化。然而，这两个网络都无法收敛，在训练初期便出现了模式崩溃。从头开始训练 CycleGAN 和 StyleGANv2，生成的图像虽然有些看起来像无线胶囊内镜图像，但大多不自然，且普遍存在噪声伪影。如表 I 所示，使用 CycleGAN 和 StyleGANv2 生成的数据集所获得的分类性能较低，这也反映了这一点。

C. 定性比较

通过分别查看图4和图5中的图像，可以对 KID 和 Kvasir-Capsule 数据集中的真实图像与 TIDE 生成的代表性图像进行定性视觉比较。图中显示，TIDE 生成的内镜图像十分逼真，其多样性也与真实图像相似。尤其值得注意的是，TIDE 生成的合成图像保留了真实组织的可见特征，包括颜色、纹理和形状。对于异常图像，生成的病变不仅看起来与真实异常图像中的病变相似，而且位置自然，与正常组织融为一体。此外，生成的图像还逼真地再现了气泡和碎屑，这些都是真实无线胶囊内镜图像中的常见现象。

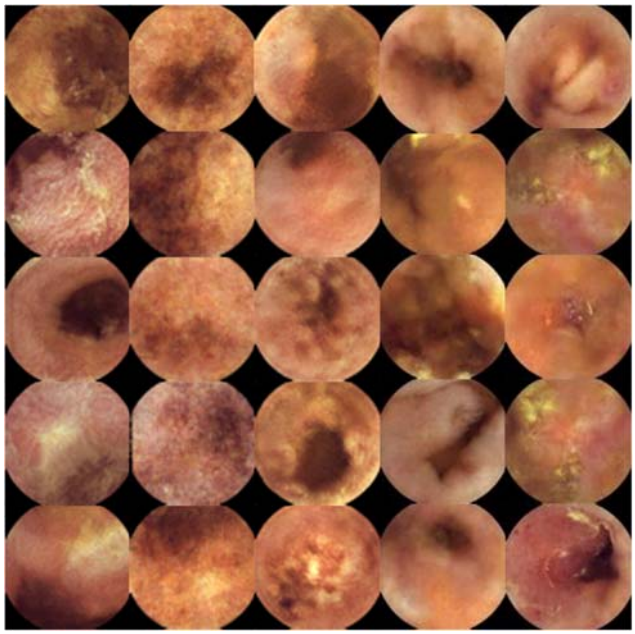
图 6 和图 7 分别对比展示了使用不同先进生成模型生成的正常图像和异常图像，包括 TS-GAN [43], CycleGAN [26], StyleGANv2 [11], 和 EndoVAE [44]。具体而言，对于图 6 中的合成正常图像，从医学角度来看，TS-GAN（图 6(a)）生成的合成图像相当逼真，但图像颗粒感明显，且存在像素化现象，这在常见的小肠胶囊内镜图像中并不多见；CycleGAN（图 6(b)）生成的图像则全部呈现像素化，且整体模糊程度较高，无法据此进行安全可靠的观察；与 CycleGAN 相比，StyleGANv2（图 6(c)）生成的图像有了很大改进，但仍存在明显的图像模糊，整体效果也表明这组正常小肠图像并不真实；EndoVAE（图 6(d) ）生成了一组逼真的正常小肠图像，尽管与先前模型（甚至 TS-GAN）的结果相比有显著改善，但图像模糊程度较高且存在超白伪影，仍暴露出其并非真实数据集；TIDE（图 6(e)）生成的图像比此前的图像更加清晰、分辨率更高，且仅有少量伪影；此外，水/气泡界面也增添了逼真的特征。

关于图 7 中的异常图像，所有基于生成对抗网络的方法（图 7(a-d)）呈现的小肠异常都不够真实。尽管 TS-GAN 是较早用于无线胶囊内镜图像生成的模型，但它生成的图像能较清晰地呈现出某种相关病变过程可能导致的黏膜浸润/硬结，



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

图 6. 不同生成模型生成的合成正常无线胶囊内镜图像。(a) TS-GAN [43]。(b) CycleGAN [26]。(c) StyleGANv2 [11]。(d) EndoVAE [44]。(e) TIDE。



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

图 7. 不同生成模型生成的合成异常无线胶囊内镜图像。(a) TS-GAN [43]。(b) CycleGAN [26]。(c) StyleGANv2 [11]。(d) EndoVAE [44]。(e) TIDE。

然而，这些图像清晰度不足，且存在一些伪影，无法据此得出诊断结论。相比之下，最后一列展示的是黏膜溃疡图像，其纹理逼真，接近真实图像，可用于临床培训及其他用途。

为验证关于生成图像多样性的视觉观察结果，我们开展了一项补充实验研究。我们将核相似度矩阵特征值的香农熵的指数视为一种通用且与领域无关的度量 [55]。设有一组独立样本$x_1, x_2, \ldots, x_n \in \mathcal{X}$，$\mathbf{K} \in \mathrm{R}^{n \times n}$，一个正半定核相似度矩阵，其元素为$\mathrm{K}_{i,i} = 1 \; for \; i \in \{1, \ldots n\}$，以及一个包含$\mathbf{K}/n$特征值的向量$\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n)$。这组样本的多样性可定义为：

$$\delta = exp\left(-\sum_{i=1}^{S} \lambda_i \log \lambda_i\right) \tag{5}$$

可以利用不同的核相似度矩阵来捕捉待评估样本在视觉或语义上的相似性。本研究同时考虑了基于像素和基于特征的相似度核，分别旨在量化图像细节和语义内容方面的多样性。基于像素的相似性通过像素向量之间的余弦相似度来衡量，可捕捉与低级图像特征相关的差异，例如所比较图像的亮度和颜色。基于特征的相似性则通过图像高级特征之间的余弦距离来计算。我们选用在 ImageNet 上训练的 Inception-v3 模型的最后一个池化层作为感知相关的特征提取器 [55], , 该方法也已有效应用于无线胶囊内镜图像表征 [56]。最终，

生成数据集的多样性 δ_g 应大致等于相应真实数据集的多样性 δ_r，即 $\delta_{g=}\delta_r$。我们认为，相对多样性是一个更有意义的度量，其定义为 $\tilde{\delta} = \delta_g/\delta_r$，因为它提供的多样性分数不受用于训练生成模型的真实数据集多样性的影响，从而能够直接比较不同数据集；因此，当 $\delta_{g=}\delta_r$时，该度量达到最大值，即 $\tilde{\delta} = 1$。图 8 展示了本研究生成的所有数据集的相对多样性。根据基于特征的相似性，可以看出，两种基于变分自编码器的模型和 TS-GAN 生成的数据集相对多样性高于 CycleGAN 和 StyleGANv2；而根据基于像素的相似性，除 ST-GAN 生成的异常数据集表现出显著更高的相对多样性外，所有模型的结果彼此相当。不过，在正常数据集所呈现的基于特征的多样性方面，TIDE 优于 EndoVAE。

D. 用户评估

综上，TIDE 生成的数据集所取得的分类性能，与使用相应真实无线胶囊内镜图像时的分类性能相当。此外，与对比生成模型生成的数据集相比，TIDE 数据集中的图像不仅外观更逼真，也更加清晰。为了探究 TIDE 数据集对于无线胶囊内镜领域的内镜医师而言是否也足够逼真，我们开展了一系列视觉图灵测试（VTT）。具体而言，我们使用 TIDE 自动生成的小肠正常图像和炎症性疾病图像进行了三项测试。第一项和第二项测试各包含 55 张图像，其中第一项仅包含

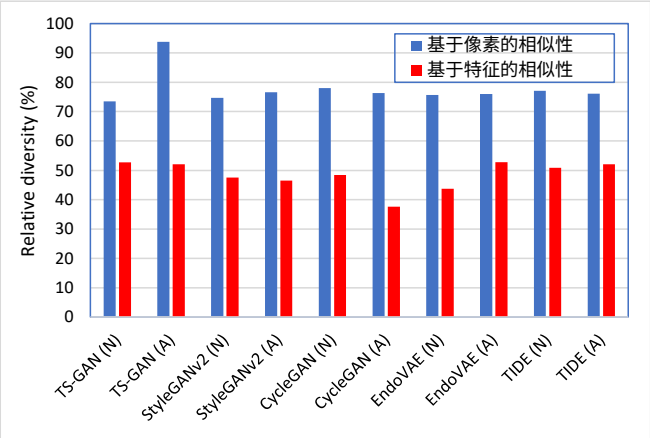


图 8. 基于像素和特征相似性核的相对多样性：不同生成模型生成的所有合成数据集中正常（N）和异常（A）样本的相对多样性。

来自 KID 数据库的真实图像，第二项则仅包含由 TIDE 架构生成的合成图像。最后一项测试将两者结合，共包含 110 张图像。随后，三位拥有 10 至 30 年经验的内镜医师参加了测试，他们需要辨别合成图像和真实图像。值得注意的是，为避免任何选择偏倚，我们没有向参与测试的专家透露各项测试中真实图像与伪造图像的比例。此外，直到本研究结束后，我们才向他们公布每项视觉图灵测试的结果。

表 II 总结了三位内镜医师进行的所有视觉图灵测试结果。在第一个视觉图灵测试中，仅包含由 TIDE 架构生成的人工图像，内镜医师的平均准确率为 46.1±7.3%（范围为 38.2% 至 52.7%）。第二个视觉图灵测试仅包含真实图像，平均准确率为 66.1±8.6%（范围为 56.4% 至 72.7%）。第三个视觉图灵测试包含真实图像和由所提方法合成的人工图像，所得平均准确率为 50.0±1.8%（范围为 48.2% 至 51.8%）。将真实图像视为正类预测、合成图像视为负类预测，平均敏感度和特异度分别为 65.5±10.1%（范围为 56.4% 至 76.4%）和 34.6±6.5%（范围为 27.3% 至 40.0%）（表 III）。以上结果证实，TIDE 生成的内镜图像难以与真实图像区分。

表 II 所有视觉图灵测试结果汇总

	视觉图灵测试		
	1st 视觉图灵测试	2nd 视觉图灵测试	3rd 视觉图灵测试
	准确率 （%）	准确率 （%）	准确率 （%）
内镜医师 *I*	38.2	72.7	50.00
内镜医师 *II*	47.3	69.1	51.8
内镜医师 *III*	52.7	56.4	48.2
Mean	46.1 ± 7.3	66.1 ± 8.6	50.0 ± 1.8

表 III 第 3 次视觉图灵测试中的平均敏感度和特异度

	3rd 视觉图灵测试	
	敏感度 *(%)*	特异度 （%）
内镜医师 *I*	63.6	36.4
内镜医师 *II*	76.4	27.3
内镜医师 *III*	56.4	40.0
Mean	65.5 ± 10.1	34.6 ± 6.5

V. 讨论

医学影像中标注数据集的有限可用性，是开发基于图像的计算机辅助诊断系统的关键进展所面临的障碍。尤其是在无线胶囊内镜领域，由于诊断率仍然较低，且无线胶囊内窥镜专家在检查数千张图像时难以始终保持专注而不受干扰，因此对这类进展的需求十分迫切 [3]。本研究的动机在于需要公开可用的无线胶囊内镜数据集作为基准，从而促成图像分析研究人员之间富有成效的竞争，有效改进其方法，以便应用于临床实践。我们提出了一种能够生成合成无线胶囊内镜图像的多尺度残差变分自编码器架构，并利用公开可用的数据集证明，这类图像可以替代真实图像，用于训练检测异常的机器学习系统。小肠中可能发现的异常种类繁多，而无线胶囊内镜主要适用于小肠检查。作为概念验证，本研究聚焦于炎性病变，这类病变代表了与疾病相关的一系列异常，例如影响全球数百万人的 IBD 和克罗恩病 [1]。

过去十年间，研究人员在医学及非医学领域提出了多种图像生成方法。生成对抗网络及其变体在生成人脸合成图像方面取得了巨大成功，也有多项研究报告了医学图像生成方面的出色成果 [21], [23], [27], [28]。然而，这些研究成功的重要因素包括数量庞大且多样化的训练数据，以及训练图像内容相对对齐；例如，人脸图像会根据面部特征进行对齐，而计算机断层扫描或磁共振成像图像则可根据所呈现的身体结构进行对齐。另一方面，生成胃肠道合成内镜图像更具挑战性，因为其内容更加多样，缺少可用于对齐的特征。相关研究（第 II 节）报告的结果基于规模大得多、通常不公开的训练集。这些研究还指出了应用基于生成对抗网络的生成模型时存在的问题。例如，值得注意的是，[38] 报告称，使用生成对抗网络模型生成的图像训练的当代分类系统，在性能提升到一定程度后便趋于饱和，即使向训练集添加更多合成息肉图像也无济于事。这被归因于生成对抗网络模型无法引入新的、未曾见过的特征。该生成对抗网络只是在操纵训练集中的现有特征，试图重复利用同一组特征来生成外观新颖的合成

息肉。这是基于生成对抗网络的图像生成模型的一项常见局限，也可能解释了我们研究中 TS- GAN 的结果。尽管与其他参与比较的生成对抗网络模型相比，该模型生成的图像更为逼真，且多样性相对较高，但生成的数据集仍不足以达到与相应真实数据集相当的分类性能。

使用 KID 数据库生成的合成数据集进行分类，其分类性能通常高于使用 Kvasir-Capsule数据集生成的合成数据集，无论采用何种生成模型。造成这一现象的原因可能是，与真实 KID 数据集中的图像相比，真实 Kvasir-Capsule 图像中的病变通常较小。

本研究对比的两种变分自编码器架构，即 EndoVAE 和 TIDE，所得结果表明，尽管二者都能生成相当逼真的图像，但 TIDE 生成的图像具有更好的多样性，且其中描绘的结构更加清晰，这对分类性能的提升起到了重要作用。

在实际应用中，TIDE 可基于真实的匿名化数据集生成合成标注无线胶囊内镜数据集，而真实数据集则可安全地保存在医疗服务提供者的场所内。合成数据集可以公开分享给（技术）研究界，而无需担心任何法律或伦理问题，因为这些数据并非真实数据，而是通过统计处理生成的，且无法据此识别任何个人信息。

六、结论

本文提出了 TIDE，这是一种新型变分自编码器架构，在深度学习模型中结合了多尺度特征提取与残差学习，用于生成合成无线胶囊内镜图像。本文开展了一项概念验证案例研究，探讨在基于图像的计算机辅助诊断系统中生成小肠正常图像和异常图像，以检测小肠炎性病变。对 TIDE 进行实验评估后，得出以下关于所提出架构的主要结论：

- 该架构能够生成清晰度和多样性更高的合成图像，可完全替代用于无线胶囊内镜图像分类的真实训练集。• 即使训练样本数量有限，该架构也能有效生成逼真的无线胶囊内镜图像。• 即使是经验丰富的无线胶囊内窥镜专家，也难以辨别 TIDE 生成的合成图像。

未来的研究方向包括将所提出的框架用于生成涵盖整个胃肠道、具有各种异常和病理状况的图像。此外，所提出方法的通用性使其有望用于生成其他医学成像模态的合成图像。

致谢

我们谨向参与本研究的匿名内镜医师表示感谢。

参考文献

[1] A. D. Sperber、S. I. Bangdiwala、D. A. Drossman、U. C. Ghoshal、M. Simren、J. Tack、W. E. Whitehead、D. L. Dumitrascu、X. Fang、S. Fukudo 等，《全球功能性胃肠道疾病的患病率与疾病负担：罗马基金会全球研究结果》，《胃肠病学》，第160卷，第1期，第99–114页，2021年。[2] A. Riphaus、S. Richter、M. Vonderach 和 T. Wehrmann，《由内镜护士判读胶囊内镜——一项比较试验》，《胃肠病学杂志》，第47卷，第3期，第273–276页，2009年3月。[3] X. Dray、D. Iakovidis、C. Houdeville、R. Jover、D. Diamantis、A. Histace 和 A. Koulaouzidis，《小肠胶囊内镜中的人工智能——现状、挑战与未来前景》，《胃肠病学与肝病学杂志》，第36卷，第1期，第12–19页，2021年。[4] D. Psychogyios、E. B. Mazomenos、F. Vasconcelos 和 D. Stoyanov，《MSDESIS：多任务立体视差估计与手术器械分割》，《*IEEE*医学影像汇刊》，第41卷，第3218–3230页，2022年。[5] J. Deng、W. Dong、R. Socher、L.-J. Li、K. Li 和 L. Fei-Fei，《ImageNet：大型分层图像数据库》，载于*2009*年*IEEE* 计算机视觉与模式识别会议，2009 年，第 248–255 页。[6] P. Voigt 和 A. Von dem Bussche，《欧盟通用数据保护条例（GDPR）》，实用指南，第*1*版，*Cham*：*Springer*国际出版公司，第10卷，第3152676期，第10–5555页，2017年。[7] R. Pannala、K. Krishnan、J. Melson、M. A. Parsi、A. R. Schulman、S. Sullivan、G. Trikudanathan、A. J. Trindade、R. R. Watson、J. T. Maple 和 D. R. Lichtenstein，《胃肠内镜中的人工智能》，*VideoGIE*，第5卷，第12期，第598–613页，2020年11月。[8] G. Pascual、P. Laiz、A. García、H. Wenzek、J. Vitrià 和 S. Seguí，《面向无线胶囊内镜的基于时间的自监督学习》，《生物学与医学中的计算机》，第146卷，文章编号105631，2022年。[9] I. Goodfellow、J. Pouget-Abadie、M. Mirza、B. Xu、D. Warde-Farley、S. Ozair、A. Courville 和 Y. Bengio，《生成对抗网络》，载于神经信息处理系统进展，2014年，第27卷，第2672–2680页。[10] P. Isola、J.-Y. Zhu、T. Zhou 和 A. A. Efros，《利用条件对抗网络进行图像到图像转换》，载于*IEEE*计算机视觉与模式识别会议论文集，2017年，第1125–1134页。[11] T. Karras、S. Laine、M. Aittala、J. Hellsten、J. Lehtinen 和 T. Aila，《分析并改进StyleGAN的图像质量》，载于*IEEE/CVF*计算机视觉与模式识别会议论文集，2020年，第8110–8119页。[12] T.-C. Wang、M.-Y. Liu、J.-Y. Zhu、A. Tao、J. Kautz 和 B. Catanzaro，《利用条件生成对抗网络进行高分辨率图像合成与语义操控》，载于*IEEE*计算机视觉与模式识别会议（*CVPR*）论文集，2018年。[13] D. Saxena 和 J. Cao，《生成对抗网络（GAN）：挑战、解决方案与未来方向》，《*ACM*计算综述》，第54卷，第3期，2021年5月。[14] F. He、S. Chen、S. Li、L. Zhou、H. Zhang、H. Peng 和 X. Huang，《通过生成对抗网络和对抗训练合成结肠镜图像，以增强息肉检测器》，载于*2021*年第*18*届生物医学成像国际研讨会（*ISBI*），2021年，第1887–1891页。[15] T. Kanayama、Y. Kurose、K. Tanaka、K. Aida、S. Satoh、M. Kitsuregawa 和 T. Harada，《利用生成对抗网络合成内镜图像检测胃癌》，2019年，第530–538页。[16] F. Mahmood、R. Chen 和 N. Durr，《通过对抗训练对合成医学图像进行无监督反向域适应》，《*IEEE*医学影像汇刊》，第PP卷，2017年。[17] Y. Shin、H. A. Qadir 和 I. Balasingham，《利用条件对抗网络合成异常结肠息肉图像以提升检测性能》，《*IEEE*开放获取》，第6卷，第56007–56017页，2018年。[18] D. P. Kingma、M. Welling 等，《变分自编码器导论》，《机器学习基础与趋势®》，第12卷，第4期，第307–392页，2019年。[19] D. E. Diamantis 和 D. K. Iakovidis，《ASML：可扩展机器学习的算法无关架构》，《*IEEE*开放获取》，第9卷，第51970–51982页，2021年。[20] A. Radford、L. Metz 和 S. Chintala，《利用深度卷积生成对抗网络进行无监督表征学习》，第*4*届国际学习表征会议（*ICLR 2016*），圣胡安，波多黎各，*2016*年*5*月*2*—*4*日，会议论文集，2016年。

[21] C. Bermudez、A. J. Plassard、L. T. Davis、A. T. Newton、S. M. Resnick 和 B. A. Landman，《利用深度学习学习脑部 MRI 隐式流形》，载于*Medical Imaging 2018: Image Processing*，2018年，第10574卷，文章编号105741L。[22] D. Mahapatra、B. Bozorgtabar、J.-P. Thiran 和 M. Reyes，《利用样本选择和条件生成对抗网络，高效地对图像分类与分割进行主动学习》，载于医学图像计算与计算机辅助干预国际会议，2018年，第580–588页。[23] D. Jin、Z. Xu、Y. Tang、A. P. Harrison 和 D. J. Mollura，《利用三维条件生成对抗网络生成符合 CT 真实感的肺结节模拟数据，以实现稳健的肺部分割》，载于医学图像计算与计算机辅助干预国际会议，2018年，第732–740页。[24] T. Karras、T. Aila、S. Laine 和 J. Lehtinen，《通过 GAN 渐进式增长提升质量、稳定性与多样性》，国际学习表征会议，2018年。[25] C. Baur、S. Albarqouni 和 N. Navab，《利用生成对抗网络生成高度逼真的皮肤病变图像》，载于*OR 2.0* 上下文感知手术室、计算机辅助机器人内窥镜、基于临床图像的操作与皮肤图像分析，施普林格，2018年，第260–267页。[26] J.-Y. Zhu、T. Park、P. Isola 和 A. A. Efros，《利用循环一致对抗网络进行非配对图像到图像转换》，载于*IEEE*国际计算机视觉会议论文集，2017年，第2223–2232页。[27] H. Yang、J. Sun、A. Carass、C. Zhao、J. Lee、Z. Xu 和 J. Prince，《利用结构约束 CycleGAN进行非配对脑部 MR 到 CT 合成》，载于医学图像分析中的深度学习与临床决策支持多模态学习，施普林格，2018年，第174–182页。[28] J. Cai、Z. Zhang、L. Cui、Y. Zheng 和 L. Yang，《迈向跨模态器官转换与分割：循环一致且形状一致的生成对抗网络》，医学图像分析，第52卷，第174–184页，2019年。[29] P. Costa、A. Galdran、M. I. Meyer、M. Niemeijer、M. Abràmoff、A. M. Mendonça 和 A. Campilho，《端到端对抗性视网膜图像合成》，《*IEEE*医学影像汇刊》，第37卷，第3期，第781–791页，2018年。[30] M.-Y. Liu、T. Breuel 和 J. Kautz，《无监督图像到图像转换网络》，神经信息处理系统进展，第30卷，2017年。[31] A. U. Hirte、M. Platscher、T. Joyce、J. J. Heit、E. Tranvinh 和 C. Federau，《利用深度生成模型逼真地生成弥散加权脑部磁共振图像》，磁共振成像，第81卷，第60–66页，2021年。[32] H. Huang、R. He、Z. Sun、T. Tan 等，《IntroVAE：用于摄影图像合成的内省变分自编码器》，神经信息处理系统进展，第31卷，2018年。[33] D. K. Iakovidis、M. Ooi、Y. C. Kuang、S. Demidenko、A. Shestakov、V. Sinitsin、M. Henry、A. Sciacchitano、S. Discetti、S. Donati 等，《面向下一代测量系统的信号处理路线图》，测量科学与技术，第33卷，第1期，文章编号012002，2021年。[34] D. Yoon、H.-J. Kong、B. S. Kim、W. S. Cho、J. C. Lee、M. Cho、M. H. Lim、S. Y. Yang、S. H. Lim、J. Lee、J. H. Song、G. E. Chung、J. M. Choi、H. Y. Kang、J. H. Bae 和 S. Kim，《利用生成对抗网络合成结肠镜图像，并通过卷积神经网络增强对无蒂锯齿状病变的检测》，《科学报告》，第12卷，第1期，文章编号261，2022年1月。[35] T. R. Shaham、T. Dekel 和 T. Michaeli，《SinGAN：从单幅自然图像中学习生成模型》，*IEEE/CVF* 国际计算机视觉会议论文集，第4570–4580页，2019年。[36] A. Sams 和 H. H. Shomee，《基于 GAN 的逼真胃肠道息肉图像合成》，载于*2022*年第*19*届 *IEEE* 生物医学成像国际研讨会，2022年，第1–4页。[37] T. Karras、M. Aittala、J. Hellsten、S. Laine、J. Lehtinen 和 T. Aila，《利用有限数据训练生成对抗网络》，神经信息处理系统进展，第33卷，12114，2020年12104页。–。[38] H. A. Qadir、I. Balasingham 和 Y. Shin，《基于简易 U-Net 的合成息肉图像生成：息肉到阴性图像及阴性图像到息肉》，《生物医学信号处理与控制》，第74卷，文章编号103491，2022年。[39] M. Pennazio、E. Rondonotti、E. J. Despott、X. Dray、M. Keuchel、T. Moreels、D. S. Sanders、C. Spada、C. Carretero、P. C. Valdivia 等，“用于诊断和治疗小肠疾病的小肠胶囊内镜与器械辅助小肠镜：欧洲胃肠内镜学会（ESGE）指南——2022年更新”，《内镜》，第 55 卷，第 01 期，第 58–95 页，2023年。[40] E. Rondonotti、A. Koulaouzidis、D. E. Yung、S. N. Reddy、J. Georgiou 和 M. Pennazio，“小肠肿瘤性疾病”，《胃肠内镜诊所》，第 27 卷，第 1 期，第 93–112 页，2017年。[41] J. Ahn、H. N. Loc、R. K. Balan、Y. Lee 和 J. Ko，“发现小肠病变：基于内镜图像的学习系统面临的挑战”，《计算机》，第 51 卷，第 05 期，第 68–76 页，2018年5月。[42] A. B. L. Larsen、S. K. Sønderby、H. Larochelle 和 O. Winther，“基于学习得到的相似性度量实现超越像素的自编码”，国际机器学习会议，第 1558–1566 页，2016年。[43] D. E. Diamantis、A. E. Zacharia、D. K. Iakovidis 和 A. Koulaouzidis，“以人工生成的内镜图像替代真实图像进行卷积神经网络训练”，载于*2019 IEEE* 第*19*届生物信息学与生物工程国际会议（*BIBE*），2019年，第 519–524 页。[44] D. E. Diamantis、P. Gatoula 和 D. K. Iakovidis，“EndoVAE：利用变分自编码器生成内镜图像”，载于*2022 IEEE* 第*14*届图像、视频与多维信号处理研讨会（*IVMSP*），2022年，第 1–5 页。[45] A. Koulaouzidis、D. K. Iakovidis、D. E. Yung、E. Rondonotti、U. Kopylov、J. N. Plevris、E. Toth、A. Eliakim、G. W. Johansson、W. Marlicz 等，“KID 项目：用于研究的基于互联网的胶囊内镜数字视频图谱”，《*Endoscopy international open*》，第 5 卷，第 06 期，第 E477–E483 页，2017 年。[46] P. H. Smedsrud 、 V. Thambawita、S. A. Hicks、H. Gjestang、O. O. Nedrejord、E. Næss、H. Borgli、D. Jha、T. J. D. Berstad、S. L. Eskeland、M. Lux、H. Espeland、A. Petlund、D. T. D. Nguyen、E. Garcia-Ceja、D. Johansen、P. T. Schmidt、E. Toth、H. L. Hammer、T. de Lange、M. A. Riegler 和 P. Halvorsen，“Kvasir-Capsule：视频胶囊内镜数据集”，《科学数据》，第 8 卷，第 1 期，第 142 页，2021年5月。[47] D. K. Iakovidis、S. Tsevas 和 A. Polydorou，“通过无监督图像挖掘缩短胶囊内镜阅片时间”，《计算机化医学影像与图形》，第 34 卷，第 6 期，第 471–478 页，2010年。[48] D. P. Kingma 和 J. Ba，“Adam：一种随机优化方法”，*arXiv* 预印本 *arXiv:1412.6980*，2014年。[49] A. Borji，“生成对抗网络评估指标的利弊：最新进展”，《计算机视觉与图像理解》，第 215 卷，第 103329 页，2022年。[50] J. Huang 和 C. X. Ling，“使用曲线下面积和准确率评估学习算法”，《电气电子工程师学会知识与数据工程汇刊》，第 17 卷，第 3 期，第 299–310 页，2005年。[51] P. Branco、L. Torgo 和 R. P. Ribeiro，“类别不平衡领域预测建模综述”，《*ACM*计算综述》（*CSUR*），第 49 卷，第 2 期，第 1–50 页，2016年。[52] D. E. Diamantis、D.-C. C. Koutsiou 和 D. K. Iakovidis，“使用轻量级回看式全卷积神经网络进行阶梯检测”，载于《神经网络工程应用》，J. Macintyre、L. Iliadis、I. Maglogiannis 和 C. Jayne 编，瑞士恰姆：Springer国际出版公司，2019年，第 522–532 页。[53] D. E. Diamantis、D. K. Iakovidis 和 A. Koulaouzidis，“用于计算机辅助内镜的回看式全卷积神经网络”，《生物医学信号处理与控制》，第 49 卷，第 192–201 页，2019年。[54] H. Thanh-Tung 和 T. Tran，“生成对抗网络中的灾难性遗忘与模式崩溃”，载于*2020*年国际神经网络联合会议（*IJCNN*），2020年，第 1–10 页。[55] D. Friedman 和 A. B. Dieng，“Vendi 分数：一种用于机器学习的多样性评估指标”，*arXiv* 预印本 *arXiv:2210.02410*，2022年。[56] X. Li、H. Zhang、X. Zhang、H. Liu 和 G. Xie，“探索迁移学习在小规模类别不平衡内镜图像胃肠道出血检测中的应用”，载于*2017*年 *IEEE* 第*39*届医学与生物学工程学会年度国际会议（*EMBC*），2017年，第 1994–1997 页。