面向基础超声模型的基于纹理的框架

Tal Grutman、Carmel Shinar 和 Tali Ilovitsh，IEEE 高级会员

摘要—**超声是应用最广泛的医学成像模态，但它生成的图像具有根本的独特性：组织依赖性散射、反射和声速变化共同形成了一组有限的特征纹理，与自然图像统计特性显著不同。这些由声学驱动的模式使最初为自然图像设计的算法难以应用于超声领域。为弥合这一差距，该领域日益转向基础模型，希望借助其泛化能力。然而，这些模型在超声应用中往往表现不佳，因为它们并非针对超声物理设计，而只是使用超声数据进行训练。因此，将超声领域知识融入既有学习框架至关重要。为此，我们将自监督学习重新表述为纹理分析问题，并提出纹理超声语义分析（TUSA）。借助 TUSA，模型能够学习利用高度可扩展的对比方法，直接从简单的 B 型超声图像中提取真正的领域特定表征。我们使用开源数据、模拟数据和体内数据的组合训练了 TUSA 模型。我们将其潜在空间与若干更大型的基础模型进行比较，结果表明，在独特的在线数据集以及本研究收集的临床眼部数据集上，对于具有挑战性的下游任务，我们的方法使 TUSA 模型具备更强的泛化能力。我们的模型在检测 COVID（70%）、脊髓血肿（100%）和玻璃体出血（97%）方面准确率更高，并且与肝脂肪变性（r = 0.83）、射血分数（r = 0.63）和血氧饱和度（r = 0.38）等定量参数的相关性更强。我们已开源模型权重和训练脚本：https://github.com/talg2324/tusa。**

索引词—**对比学习、超声基础模型、深度学习、超声。**

本研究部分由以色列科学基金会（资助编号：192/22）、ERC StG（资助编号：101041118，NanoBubbleBrain）、以色列癌症研究基金（资助编号：1286686），以及特拉维夫大学 Nicholas and Elizabeth Slezak 心脏研究与生物医学工程超级中心资助。Tal Grutman 和 Tali Ilovitsh 任职于以色列特拉维夫 6997801 特拉维夫大学生物医学工程学院（电子邮箱：ilovitsh@tauex.tau.ac.il）。Carmel Shinar 任职于以色列特拉维夫 6997801 特拉维夫苏拉斯基医疗中心及特拉维夫大学医学院。

I. 引言

DEEP 学习彻底改变了计算机视觉，并已在现实世界和医学影像中得到实际应用。超声从深度学习模型根据数据进行自适应的能力中获益良多，而经典视觉算法往往难以应对其较低的分辨率、对比度和信噪比。近期许多超声研究利用深度学习开展各种应用，包括全息成像[1], 波束形成[2], 体积表示 [3], 等。尽管如此，前沿深度学习算法通常使用的数据规模在超声领域却更难获取，因为监督标签往往需要经过多位专家的严格审查 [4] ，这是由于临床成像中常见的病理检测一旦出错，后果要严重得多[5], [6]。

超声成像的图像生成方式与其他医学成像模态有着根本区别，这给基于自然图像的算法带来了挑战。在典型的超声系统中，换能器阵列向目标介质发射声波。声波传播过程中，会因目标解剖结构的几何形状和组成而发生散射和反射。换能器元件记录返回的回波，随后根据每个回波返回换能器所需的飞行时间及回波幅度重建超声图像，并将其映射为灰度图像。声波在介质中的传播轨迹主要由局部声速以及组织的散射系数和反射系数决定。这些生理参数在局部存在差异。图像中的灰度级形成纹理模式，反映了这些声学特性的空间分布。尽管不同个体、器官和病理状况的声学参数有所差异，但它们在人体内的生理范围有限。因此，尽管人体组织在生物学上具有多样性，超声图像所包含的特征纹理种类仍然有限。器官边界和骨骼通常呈白色，软组织呈现颗粒状散斑模式，而液体则呈黑色。这说明解读图像需要结合额外的背景信息：心脏中的暗色无回声区代表血液，而在产科成像中，同样的影像表现则代表羊水。此外，为了适应弱反射与强散射之间巨大的动态范围，检测到的声波信号会按对数刻度编码为图像像素。这些现象共同作用，使超声图像与经典自然图像及其他放射学成像模态的图像大不相同。

尽管存在这些差异，许多医学成像模态中的视觉内容仍局限于一组有限且明确界定的人体组织。这一约束推动了模态无关模型的发展，例如RadImageNet[7] 和 MedSAM[8]，，它们利用各成像模态间共享的解剖结构来学习可迁移表征。近来，该领域已转向“基础模型”，即采用无监督或自监督方法[9]、在海量数据集上训练的大型模型。随着数据集与模型规模的增长，基础模型学会以超出其训练容量的方式进行泛化

数据。此类模型虽然训练成本高昂，但在下游应用中非常有用。基础模型可用于多种应用，包括在专用数据集上的开箱即用（零样本）使用、为下游分析提取有用特征，或针对更具体的任务进行微调。

这些应用场景在医学影像领域，尤其是超声领域，极具实用价值，因为高质量的标注数据集十分稀缺。目前已有多个基础模型面向一般医学诊断 [8], [10] 以及超声领域专门开发并开放共享 [11], [12], [13]。这些模型被视为自动超声分析的“民主化”，因为它们让普通用户也能使用一个模型，而该模型已具备从可能极其昂贵的数据集和 GPU 运算时长中习得的知识。然而，这些模型在不同超声领域中的泛化能力仍有很大提升空间，因为它们并非专为超声图像设计，而只是使用超声数据进行训练。例如，UltraSam[12] 难以适应使用训练数据中未曾涵盖的特殊探头采集的数据集。这凸显了超声优先设计的必要性：设计应基于底层成像原理，而非隐式地从数据集中提取信息。尽管如此，基础模型仍继续采用通用框架进行训练，因为专业知识难以规模化应用。

因此，将超声领域知识融入经过充分研究的学习框架至关重要，例如掩码自编码 [14] 和 SimCLR [15] 。这些框架可以利用数据增强，无中生有地构造学习信号：将一幅未标注图像与其增强版本进行比较。此类技术已显著提升了自然图像任务的性能，即使在基础模型范畴之外，它们仍是模型预训练中广泛采用的方式[16], [17]。这些技术已开始应用于超声项目，例如超声基础模型（USFM），该模型采用自监督学习，在掩码自编码框架中使用 B 型超声图像进行训练[13]。在本研究中，我们将灰度纹理模式作为经典自监督学习与受超声启发的深度学习架构之间的中介。通过将超声专用纹理分析融入自编码框架，我们提出这样的假设：超声特有的直觉可以嵌入深度学习模型，从而利用人体不同部位共有的纹理信息；其他成像模态也在采用这一方法 [18]。将自监督学习表述为纹理分析问题后，我们便可利用高度可扩展且已在主流应用中取得成功的对比学习框架，从简单的 B 型超声图像中获得真正的领域特定知识。我们的框架基于这样一种直觉：专业人员评估超声图像时，会先识别灰度图像中的宏观结构——这些结构通常由其纹理联系起来——然后再处理眼前的任务。

我们最初提出了这一训练框架，并将其称为纹理超声语义分析（TUSA） [19], 。在本文中，我们利用更多数据和稳健的验证研究对其进行了大幅完善和改进。本文将介绍如何将纹理分析融入自监督学习，然后分析 TUSA 模型的潜在空间，并将其与 MedSAM[8],、UltraSam[12], 和 USFM[13], 等多个知名基础模型进行比较，重点阐明可扩展的基于超声的信息在从 B模式图像中提取特征时的重要性

II. 方法

我们的 TUSA 模型通过将图像自编码任务拆分为两个部分来学习纹理。第一项任务是分割问题。给定 K 个纹理通道，使用所选的编码器-解码器网络，将每个 B 模式像素分割到 K 个纹理通道中的相应通道。在第二阶段，每个通道分别与其自身可学习的卷积核进行卷积，然后将各通道相加，以重建原始图像强度。这种任务分离本质上将自编码重新定义为形状与纹理的分离，并以一种能够将二者卷积结合、从而复现原始图像的方式实现（图 1）。

在训练过程中，TUSA 学习依次完成这两项任务。受限数量的纹理核迫使 TUSA 识别出在各种器官之间保持一致的纹理，并利用 B 型超声中反复出现的强度模式，学习分割不同的器官。这种特定于领域的解决方案从根本上仍然是一个图像自编码问题，可以在没有标签的情况下进行训练，因此具有很强的可扩展性。

训练完成后，TUSA 可用于前文所述的任何用途。编码器可以从解码器中分离出来，对 B 模式图像进行编码，以获得基于纹理的潜在表示，从而深入了解图像内容。分割后的纹理图像可以经过滤波，以分离特定纹理，也可以针对特定分割任务进行微调。本文将重点关注由基础模型编码的潜在空间知识。我们将通过三种方法探究基础模型的特征提取能力：可视化、软二元分类和下游回归。

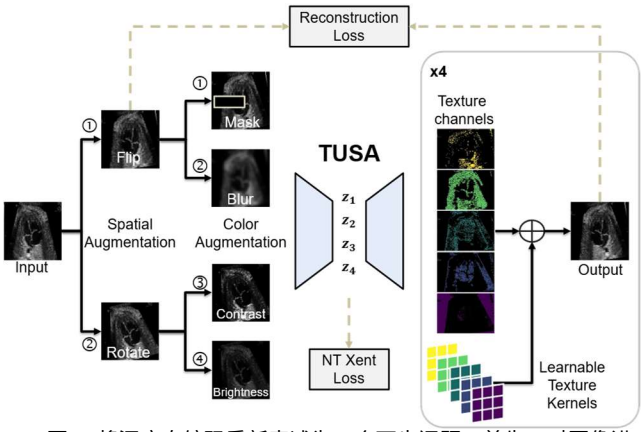

图 1. 将深度自编码重新表述为一个两步问题。首先，对图像进行分割，以生成 K 个纹理通道。第二阶段中，每个通道都与一个独特的可学习卷积核进行卷积，以重建输入强度。

A. 实施与训练 1) 模型实施

我们使用滑动窗口 U-Net Transformer (SwinUNETR)

[20]，一种近期提出的分割架构，融合卷积信息与基于注意力的信息，以提取全局和局部特征。在重建步骤之前，该模型的逻辑值经过 Sparsemax[21] 激活函数处理，促使 TUSA 为图像中的每个像素选择特定的纹理通道。在重建阶段，逐通道可分离卷积使用各不相同的可学习卷积核对每个通道进行卷积，然后通过带有 tanh 激活函数的 1x1 卷积核压缩通道维度，以重建 B 模式强度。我们使用了K=5 个纹理核，因为经验观察发现，在分割阶段结束时，增加纹理核会产生冗余信息。

TUSA 按照 SimCLR 框架进行训练，采用了多种成熟的损失函数，重点关注重建，同时还加入自监督惩罚项和正则化。对输入图像进行两次随机空间增强。随后，对每种空间增强结果分别进行随机颜色增强（擦除、亮度、对比度、高斯模糊）。在模型对这四种增强结果进行前向传播时，我们计算这四幅图像的编码与批次中其余图像编码之间的归一化温度交叉熵（NT-Xent）损失 [15] 。由于模型旨在识别宏观纹理模式，我们期望它能够适应简单的加性和乘性强度变化，因此会计算同一种空间增强下、不同颜色增强所产生逻辑值之间的均方误差。获得该图像的四个输出重建结果后，我们计算重建损失；该损失由 L1 损失、结构相似性（SSIM） [22] 以及采用 RadImageNet 主干网络 [7] 的学习感知图像块相似度（LPIPS） [23] 组成。图 1 展示了这一训练流程。

最后，我们通过在损失函数中弱惩罚模型输出的负逐通道熵，鼓励模型尽可能充分地利用各种纹理通道。这促使 TUSA 最大限度地利用批次中的不同纹理，并有助于防止其学习冗余特征。

TUSA 在远程 Kubernetes 集群上进行训练，该集群使用 Run:ai（特拉维夫，以色列）资源管理器，并配备 NVIDIA RTX A5000 GPU（NVIDIA 公司，美国加利福尼亚州圣克拉拉）。我们使用 PyTorch [24] 框架和 MONAI[25] 项目实现模型并计算 LPIPS 损失。模型采用混合精度训练，共训练 1k 个 epoch，每个 epoch 使用 10k 张图像。我们使用 AdamW [26] 优化器，批次大小为 32，初始学习率为 10^{-4}，并在训练过程中通过余弦退火逐步降至 10^{-6}。我们对身体的主要区域进行均匀采样，以免占比过高的器官（尤其是超声心动图数据集中的器官）主导学习到的纹理（图 2）。USFM [13] , 的训练也采用了类似的采样技术；这对于区分特定纹理的显著程度与其在开源数据集中的可获得性至关重要。我们在验证集上对 TUSA 候选图像进行了人工审核。该验证集由训练数据中涵盖的各个身体区域分别抽取的一帧图像组成。

*2) 训练与评估数据*我们的模型经过训练和评估

基于多种类型的数据集进行训练。首先，我们使用公开可用的、涵盖不同器官和硬件的超声数据集进行训练和评估。训练数据还补充了 k-wave [27],[28] 仿体、计算机模拟仿体，以及先前研究中采集的小鼠活体乳腺癌肿瘤数据 [3]。为补充评估数据，我们加入了在诊所采集的眼部图像临床数据集。训练数据总计包含来自身体 9 个区域的 100k 张 B 型超声图像，以及一小部分实验室先前研究采集的补充数据（小鼠乳腺癌肿瘤和模拟仿体）。各训练数据集列于表 I。

表 I. 训练数据集与器官组分配

器官组	数据集	数量 图像
腹部	AbdominalUS [29]	615
腹部	LiverTumors [30]	635
腹部	OpenKidney [6]	535
乳腺	BrEaST-Lesions_USG [31]	270
乳腺	BUS-UCLM [32]	683
乳腺	BUSBRA [33]	1,873
乳腺	BUSI [34]	780
心脏	CAMUS [35]	21,029
心脏	TED [36]	4,524
心脏	UnityImaging [37]	7,494
胎儿	胎儿腹部结构 分割数据集 使用超声图像 [38]	1,587
胎儿	胎儿切面 [39]	12,336
胎儿	PSFHS [40]	4,064
卵泡	OTU_2d [41]	1,469
Knee	JoCoHS [42]	9,575
Lung	新冠肺炎超声 [43]	19,955
肌肉骨骼	deepMTJ [44]	2,683
肌肉骨骼	FALLMUD [45]	504
Neck	颈总动脉 超声图像 [46]	1,100
Neck	DDTI [47]	480
Neck	甲状腺数据集 [48]	7,056
Lab	小鼠乳腺癌 [3]	720
Lab	k-Wave仿体 [27], [28]	432

所有图像均转换为灰度图，调整至统一分辨率 128^2, , 并映射至 [-1, 1] 范围。我们在较新且更具挑战性的数据集上评估模型，这些数据集来自通常在开源数据中代表性不足的身体部位，其中包括我们专有的眼部超声数据集。我们使用这些具有挑战性的数据集，考察模型在训练分布之外的泛化能力。评估数据集见表 II。

表 II. 评估数据集和任务。

器官组	数据集	任务
Lung	COVID-BLUeS [49]	可视化，二分类 分类， 下游 回归
脊髓	超声脊髓 损伤 [5]	可视化，二元 分类
Eye	玻璃体出血	可视化，二分类 分类
心脏	EchoNet-Dynamic[50]	下游 回归
腹部	肝脂肪变性[51]	下游 回归

*3）人体眼部病理数据集*采集了美国 B 型超声图像，使用

使用配备 15 MHz 和 20 MHz 探头、经美国食品药品监督管理局批准的市售眼部超声系统。所有成像均在实际眼科诊疗机构中，按照常规临床条件，使用 ABSOLU 超声系统（Quantel Medical，法国 Cournon-d'Auvergne）完成。本研究纳入了 54 名患者的眼睛，这些患者曾多次就诊；每次就诊时，每只眼睛均对应一个临床病例。共采集 446 张图像，对应 118 个临床病例，其中 29 个病例含有玻璃体出血。出血的病因各不相同，并与多种潜在眼部病变相关，严重程度差异很大，从遮挡解剖标志的致密混浊到轻微或亚临床表现均有。临床评估通过标准眼科检查进行，并将检查结果与超声检查结果进行比较。在某些病例中，出血仅通过临床检查才得以确认且无法通过超声成像显示，反之亦然。未出现玻璃体病理性混浊的眼睛作为对照组。

所有成像操作均遵循特拉维夫苏拉斯基医疗中心和特拉维夫大学机构审查委员会批准的方案（TLV-0769-25）。

*4）下游任务*我们比较了 TUSA 的潜在空

间与 MedSAM、UltraSam 和 USFM 等成熟的专业基础模型的潜在空间。需要注意的是，这些模型均包含一个具有 8600 万参数的 vit-b 编码器，而 TUSA 使用的是具有 900 万参数的 SwinViT。所有测试均使用相同的图像，但会根据各模型的规格对图像进行调整尺寸和缩放。

评估了三项下游任务（表 II）。为聚焦于各模型提取的信息，我们选择了模型编码器被冻结且未进行微调的任务。首先，使用统一流形逼近与投影（UMAP） [52]进行潜在空间可视化。从每个器官组中随机选取 200 张图像。每个模型将可视化数据集编码为潜在向量，并在此基础上采用无监督 UMAP 算法生成二维可视化结果。接下来，使用线性支持向量分类器（SVC）拟合软潜在空间分类器，以分别对肺部超声中的 COVID-19、脊髓图像中的脊髓血肿以及眼部超声中的玻璃体出血进行分类。为评估是否存在足够的信息以轻松区分病理类别，还展示了各模型针对这些数据集的 UMAP 投影。最后，我们在具有挑战性的回归任务上测试了这些模型。使用冻结的模型编码器对来自单个患者的一系列图像进行编码。一个由

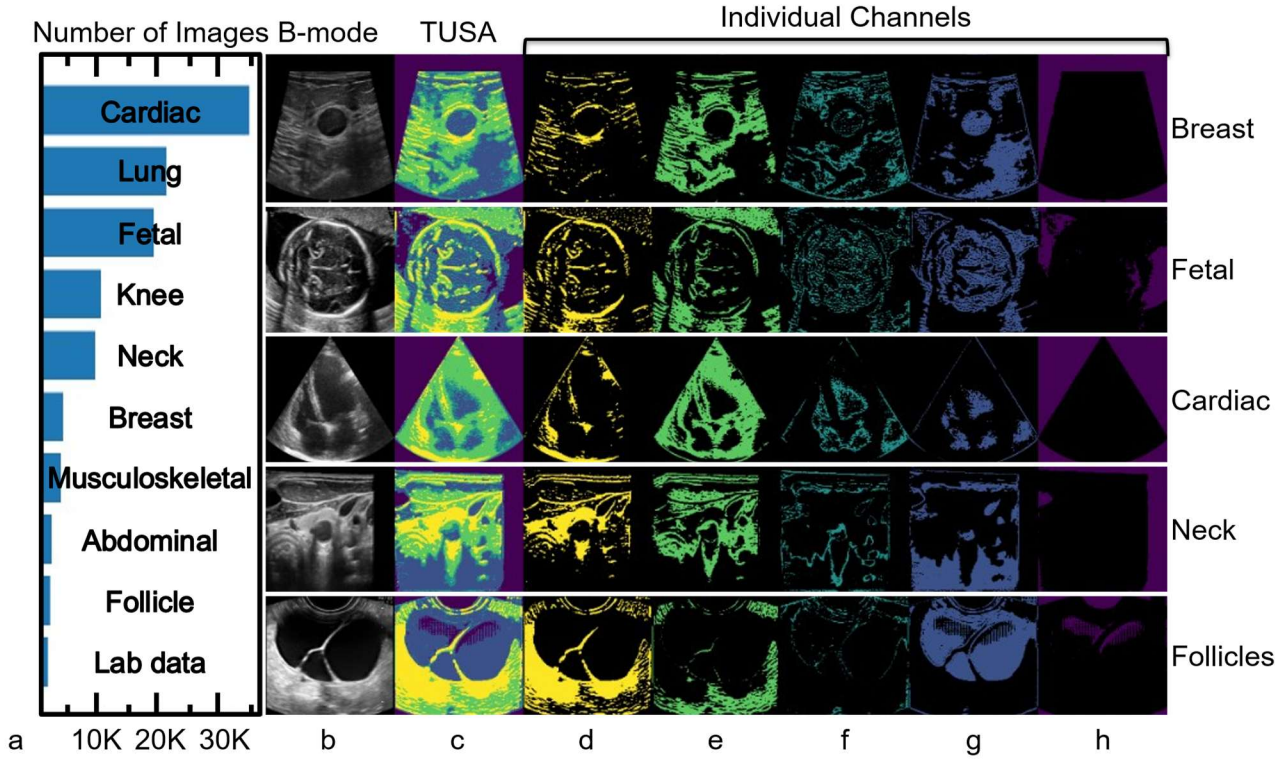

图 2. 按显著程度排序的训练数据器官分布 (a)。样本图像以标准 B 型超声 (b) 和 TUSA纹理图 (c) 的形式呈现。各通道 (d-h) 按回声强度从左（回声最强）到右（无回声）递减排列。

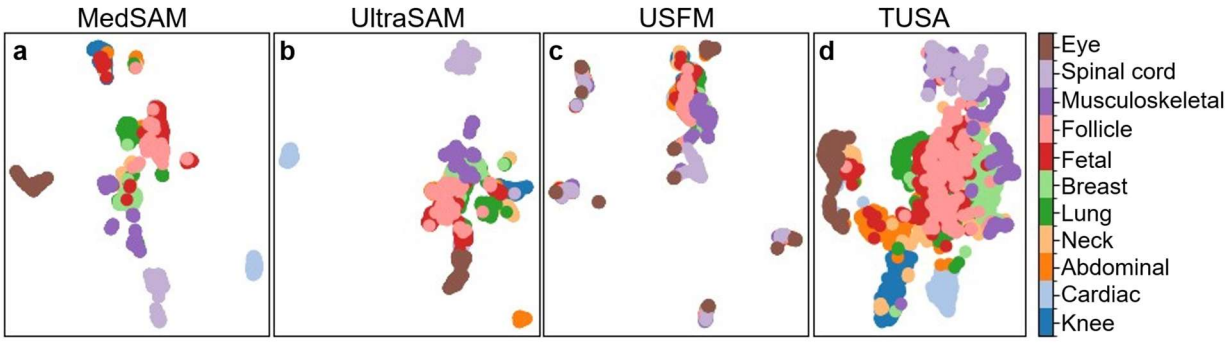

图 3。使用 UMAP 进行潜在空间可视化。从每个器官组中采样 200 张图像，包括独特的眼部数据和脊髓数据，并由各个基础模型以及 TUSA 进行编码。

若干线性层后接时序注意力层
机制组成的小型下游网络经过训练，用于根据心脏嵌入预测射血分数、根据腹部嵌入预测肝脂肪变性，以及根据肺部超声嵌入预测血氧饱和度。

数据集创建、训练和评估的代码以及预训练的 TUSA 模型可在此处获取：https://github.com/talg2324/tusa。

III. 结果

1) TUSA 模型我们训练的 TUS

我们在来自多种器官的超声图像上评估了一个模型（图 2b–h）。这里展示了五种不同器官（乳腺、胎儿心脏、颈部和卵泡）的代表性图像：标准灰阶B模式图像（图 2b）及其 TUSA分解（图 2c），后者是各个纹理通道的叠加（图 2d–h）。分别观察这些通道时，可以发现一致的规律。黄色通道突出显示骨骼、肌纤维和明显的组织边界，这些结构呈现出最强的回声（图 2d）。绿色通道对应大多数软组织（图 2e），而青色通道则捕捉液体中的散斑模式和较弱的组织边界（图 2f）。蓝色通道表示液体，紫色通道反映背景，偶尔也会出现在羊水等非常均匀的液体中（图 2f）。

2) 潜在空间的可视化使用 UMAP算法来获得相

基于模型的映射能力，对其泛化能力形成定性直觉

表 III。评估数据集上的软二分类器结果。

Dataset	Model	Accuracy	F1	Specificity	Sensitivity	AUC
COVID	MedSAM	**69.8**	**70.8**	**70**	**69.7**	71
	UltraSam	60.3	61.5	60	60.6	56.2
	USFM	62	64.7	56.7	66.7	65.4
	TUSA	**69.8**	**70.8**	**70**	**69.7**	**75.1**
Spine	MedSAM	**99.8**	**99.9**	**99.4**	**100**	**100**
	UltraSam	85.7	89.5	**99.4**	81.2	99.1
	USFM	98.3	98.9	95.8	99.2	99.7
	TUSA	**99.8**	**99.9**	**99.4**	**100**	**100**
Eye	MedSAM	89.8	81.8	88.8	93.1	93.4
	UltraSam	93.2	87.5	92.1	96.6	97.8
	USFM	90.7	82.5	91	89.7	92.4
	TUSA	**95.8**	**92.1**	**94.4**	**100**	**99.5**

潜在空间中相似的数据点。绘制了每个器官组中随机抽取的 200 张图像的 UMAP 投影（图 3）。MedSAM 和 UltraSam（图 3a、b）在潜在空间中区分心脏数据方面表现尤其出色，但难以区分其他器官。USFM 在我们的数据上未能提供有用的潜在空间分离效果（图 3c）。我们的 TUSA 模型（图 3d）提取出了高度结构化的潜在空间，在将来自同一区域的图像聚类到一起的同时，保持了身体不同区域之间的分离。

针对评估数据，在 UMAP 投影后分别分析二分类问题（图 4）。对于我们数据集中的任务，USFM 的投影并不能提供特别有用的信息，而 TUSA 和 MedSAM 提供了最佳的类别可分性。这与我们的假设一致：泛化能力通常通过

更大的模型配合更多数据，或具备可扩展领域知识的专用模型来实现。所有模型在 COVID-BLUeS 数据集上的表现都不佳，这与该数据作者的报告一致，即从肺部超声中进行 COVID 检测是一项极具挑战性的任务。对于我们的眼部数据集，所有模型都能在一定程度上区分健康眼睛和玻璃体出血，但只有我们的模型能够形成紧密且可分离的簇，这表明其在潜在空间中具有很高的类别可分性。这与基于纹理的玻璃体出血诊断相一致。

3) 软二元分类为了提供更定量的

衡量每个模型所编码的潜在空间信息，我们针对图 4 中的每个数据集，拟合了一个按类别平衡加权的线性支持向量分类器模型。对于每个模型，我们通过网格搜索独立优化 SVC 参数，以最大化准确率。在 COVID 数据集中，我们从每位患者中随机采样 20 帧，并计算每位患者的平均潜在向量；随后在患者级别进行 5 次交叉验证划分，每次使用被排除的划分评估 SVC 模型。对于脊髓血肿数据，这一步并非必要，因为作者已经将其划分为训练集和测试集。在我们的眼部数据集中，我们同样计算了每个病例的平均潜在向量，并在患者级别进行了 15 次交叉验证划分。

各数据集的准确率、F1 分数、特异度、敏感度和曲线下面积（AUC）见表 III。在 COVID 和脊柱数据集中，TUSA 和 MedSAM 模型的结果非常接近，并且都能很好地泛化到新数据。不过，在眼部出血检测中，TUSA 的表现最佳。

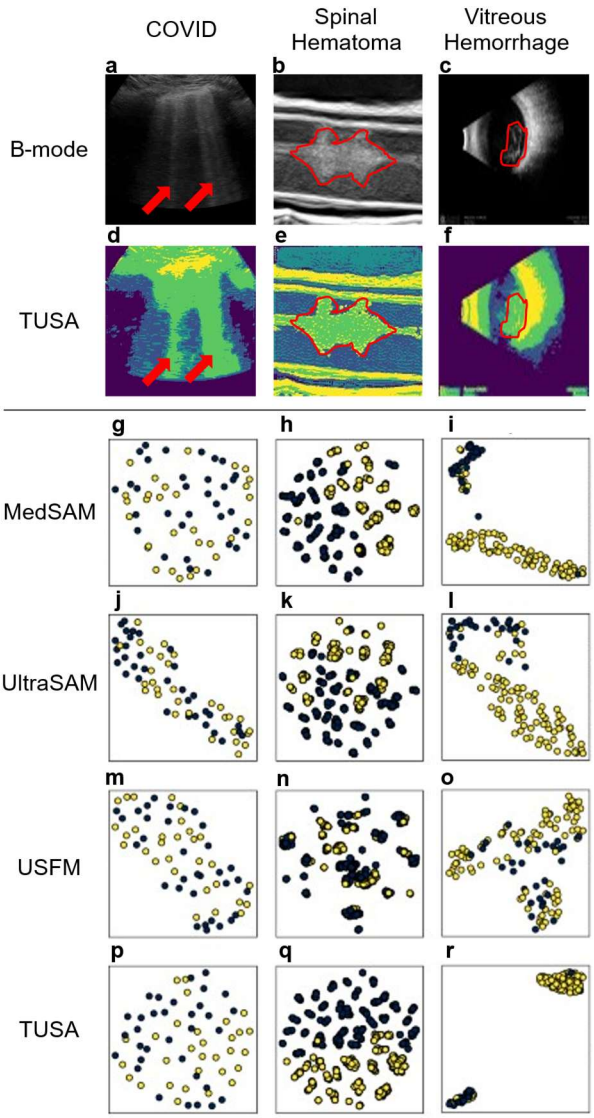

图 4. 使用 UMAP 对二分类问题进行可视化。每种病理状况（a-c）均展示一幅输入 B 型超声图像，并以红色标出病理区域。相应的 TUSA 纹理图供参考（d-f）。分别展示各模型对二分类数据集的 UMAP 投影：MedSAM（g-i）、UltraSam（j-l）、USFM（m-o）和 TUSA（p-r）。具有信息量的潜在空间应能清晰区分正常类别（黄色）与病理类别（蓝色）。

*4）下游回归*我们的最终评估旨在训

在用于预测定量超声指标的下游潜在空间模型中。这尤其困难，因为要量化生理参数，潜在空间必须包含丰富的相关信息，而不能仅仅用于分类。

我们分析了此类下游建模在脂肪性肝病中的表现（图 5）。在健康肝脏中，相较于肝脏，B 型超声图像显示肾脏强回声复合体的回声较强（图 5a）。对应的 TUSA 通道显示，肾脏中散斑较强（黄色通道），而肝脏中呈现规则的软组织纹理（绿色通道）（图 5b）。在脂肪性肝病中，肝脏回声强度显著增高，且由于声波大多被反射，肾脏的回声强度被掩盖（图 5c）。这一现象可在 TUSA 图像中检测到，其中 TUSA 图像此时检测到肝脏中散斑增多（图 5d），并将其分割为黄色通道。基于分割的基础模型（如 MedSAM（图 5e）和 UltraSam（图 5f））无法检测到强度和散斑模式的细微变化，而这些模型编码的潜在向量也难以对肝脂肪变性进行有意义的量化。采用对比学习方案训练的 USFM（图 5g）和 TUSA（图 5h）则能够编码有意义的信息，用于量化肝脏脂肪程度。

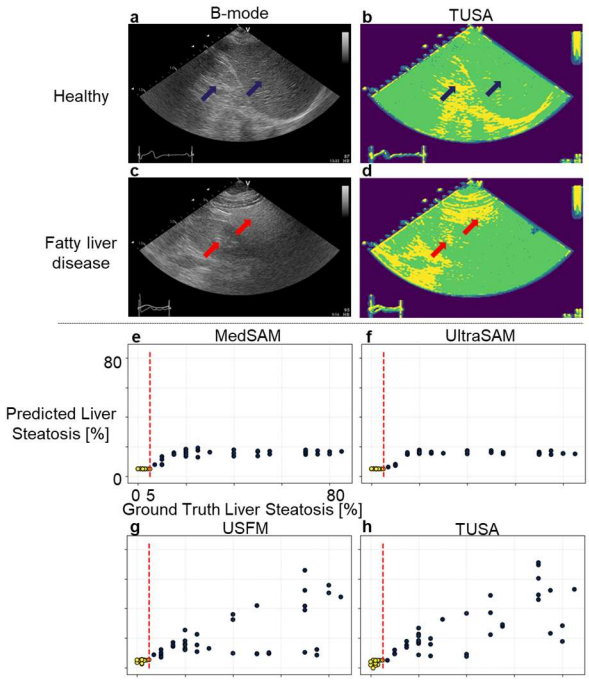

图 5. 脂肪性肝病中的下游回归性能。比较脂肪肝患者（a-b）与非脂肪肝患者（c-d）的 B 型超声图像和 TUSA 通道。蓝色箭头标出了健康的散斑，红色标出了脂肪性病变。回归图中，x 轴为真实标签，y 轴为 MedSAM（e）、UltraSam（f）、USFM（g）和 TUSA（h）的模型预测值。各图的坐标轴相同。在肝脂肪变性数据集中，我们使用了全部 10 帧

按照作者所述，对每位患者采用留一法训练；对于 EchoNet-Dynamic 数据集，我们根据作者的标注，选取收缩期与舒张期之间线性间隔的五帧，并使用作者划分的训练数据和测试数据。在 COVID-BLUeS 中，我们从每位患者的数据中随机抽取 20 帧，并采用 10 组交叉验证划分进行训练，每次在 10% 的数据上评估模型。对于每个数据集，我们报告皮尔逊相关系数 r、平均绝对误差（MAE），以及用于常规回归分类的曲线下面积（表 IV）。在检测肝脂肪变性时，数据集作者将 5% 定义为非酒精性脂肪肝的阈值；在 EchoNet-Dynamic 中，作者则将射血分数 40% 作为低射血分数患者与正常患者的分界阈值。在 COVID-BLUeS 数据集中，我们训练模型预测血氧饱和度，并将 95% 设为健康患者的血氧饱和度阈值，因为 COVID 患者的低血氧可能与 B 型超声伪影有关 [53]。由于 EchoNet-Dynamic 中的左心室分割属于 UltraSam 的训练数据，因此我们在对 EchoNet-Dynamic 进行评估时排除了 UltraSam 模型。每个潜在空间模型均训练十次，

每次在各数据集上使用不同的随机初始化。UltraSam 和 MedSAM 难以从潜在空间中提取有意义的信息，这很可能是因为它们是面向分割的模型，若不进行额外训练，便难以提供与分割无关的信息。USFM 在这些具有挑战性的任务中确实取得了有意义的下游结果，但表现明显逊于 TUSA。

表 IV. 评估数据集上的下游回归结果。

Dataset	Model	r	MAE	AUC
Liver Steatosis	MedSAM	0.72 ± 0.01	17.65 ± 0.08	**1.00 ± 0.00**
	UltraSam	0.72 ± 0.01	17.64 ± 0.08	**1.00 ± 0.00**
	USFM	0.67 ± 0.04	13.50 ± 0.56	**1.00 ± 0.00**
	TUSA	**0.83 ± 0.01**	**10.03 ± 0.19**	0.99 ± 0.00
EchoNet-Dynamic	MedSAM	0.19 ± 0.03	8.62 ± 0.09	0.64 ± 0.02
	USFM	0.57 ± 0.01	7.34 ± 0.06	0.83 ± 0.01
	TUSA	**0.63 ± 0.00**	**6.95 ± 0.08**	**0.87 ± 0.01**
COVID-BLUeS	MedSAM	0.37 ± 0.02	3.76 ± 0.01	0.70 ± 0.02
	UltraSam	0.29 ± 0.01	3.70 ± 0.02	0.70 ± 0.01
	USFM	0.23 ± 0.03	3.83 ± 0.04	0.64 ± 0.02
	TUSA	**0.38 ± 0.03**	**3.65 ± 0.04**	**0.71 ± 0.01**

IV. 讨论

在这项工作中，我们提出了一种新的深度超声模型自监督训练框架，能够利用可扩展的领域知识。我们的 TUSA 框架采用基于 SimCLR 的数据增强方法，在学习分割各纹理核如何描述给定输入图像的同时，学习与常见超声图像纹理相对应的纹理核。该模型实质上将纹理与解剖结构分离，按照我们认为人类会采用的方式分析图像。为了评估这一框架是否有益于超声下游任务，我们评估了采用这种方法训练的模型以及成熟的基础超声模型的泛化能力，旨在揭示更多数据与领域专门化之间的权衡。这一点在两个极端案例中尤为明显：MedSAM 使用了最多的训练数据（包括超声数据，但并非专门针对超声），而我们的 TUSA 模型则是一个规模小得多的专用模型，仅使用 10 万张图像进行训练。在所有潜在空间测试中，这两个模型的表现都优于其他模型：UMAP 投影后，它们对器官组的区分更清晰，二分类结果也更好。在许多示例中，我们的 TUSA 模型表现优于 MedSAM，这表明，将可扩展的专门化与 SimCLR 这类通用框架相结合，能够胜过那些为自然图像而设计、依赖海量数据资源的训练流程。因此，尽管 UltraSam 和 USFM 这类专用模型无法媲美 MedSAM 凭借其庞大数据集获得的泛化能力，我们的模型却通过扩展专业知识实现了这一点。这种做法不一定要取代以数据为先的方法，但可以为基础模型的泛化能力提供一条辅助流程。

从潜在空间可以清楚看出，SAM 模型在训练中大量使用了心脏数据，因为它们能够特别有效地区分心脏数据与其他超声数据；这与相较于其他器官组，超声心动图数据的丰富程度相符。同样，三个基础模型在乳腺肿瘤方面表现良好，尤其是 USFM，其数据集主要由乳腺超声构成。SAM 模型对未见过的脊髓和眼部数据集以及使用较少的膝关节超声数据集均具有良好的泛化能力，而 USFM 难以区分膝关节超声数据集，这表明 USFM 可能无法泛化到未见过的器官组。另一个有趣的现象是，除了能出色地区分各器官组之外，我们的 TUSA 模型还将肌肉骨骼数据集分成两部分，凸显了构成这一器官组的两个数据集之间存在一定差异。评估数据集的潜在空间表明，与基于自然图像的分析相比，基于纹理的分析更能指示这些病理状况（COVID-19、脊髓血肿以及眼部玻璃体出血）（图 4）。这很容易理解，因为这些基础模型的训练数据集并不包含与独特的脊髓和眼部数据集有丝毫相似的数据，这凸显了基础超声模型具备可扩展泛化能力的重要性。我们还使用 SVC 模型（表 III）以更定量的方式验证了这一理论，结果显示，TUSA 编码能够为二分类提供可分性最高的潜在空间。在此我们再次发现，最佳结果来自使用最多数据（SAM 模型）或进行模态专门化（TUSA）。最后，我们证明，在无需大量额外训练的情况下，TUSA 潜在空间中蕴含的纹理知识最适用于真实世界的下游任务（即使使用性能普通的 GPU，采用预编码潜在空间的下游多层感知机也只需不到 10 分钟即可完成训练），尤其是在那些与在线可获取的经典超声数据集不同的困难且独特的任务中（表 IV）。这一点在肝脂肪变性和射血分数预测方面尤为明显，TUSA 的得分远高于其他模型。

我们的 TUSA 模型取得成功具有几项重要意义。首先，我们的可扩展方法可用于继续推动 UltraSam 和 USFM 等大型专科基础模型的发展边界。然而，我们的模型也表明，小型模型能够带来许多相同的益处，并且在域内泛化时往往优于基础模型，具体取决于利用纹理信息能带来多大的优势。

我们的模型是超声优先模型，与规模大得多的 MedSAM 不同，它可能并不适用于放射学中的其他应用。不过，利用模态特异性特征，可学习纹理这一概念或许与模态特异性基础模型相关[54]。我们还可以提出 TUSA 的其他几种应用场景。我们研究了潜在空间中纹理信息的提取，而我们的模型可用作泛化能力良好的小型图像编码器。此外，从 TUSA 的分割步骤中获得的输出通道可视为一种类似于主成分分析的成分分析，从而便于与之交互，或

滤除与特定分析无关的纹理。未来的研究可在尝试估计声学参数时，将图 2、4 和 5 中的 TUSA 图像作为先验。例如，已有研究表明，关于声速的先验知识可显著提高绝对估计的准确性[55]。另外，单个 TUSA 通道可以分离不同类型图像中的相关结构（图 2），下游应用则可将 TUSA 用作形态分割算法的起点。

尽管我们的模型在下游应用中展现出良好前景，但仍有很大的改进空间。虽然我们通过经验选择了K=5 个纹理核，但在更高分辨率下使用多尺度核或许能取得更好的性能。随着模型使用的纹理数量增加，建立一种合并相似纹理核、降低冗余的机制也将大有裨益。我们选择使用规模较小、图像分辨率相对适中的模型，是出于硬件限制；若有更大的数据集和 GPU，也有可能训练出更加通用的模型。此外，我们的模型经过训练，会依赖训练数据所含解剖结构中常见的参考纹理，因此在遇到未包含的纹理时，预计会难以适应，例如使用超声造影剂进行成像时。

尽管如此，从我们的模型中提取的已学习纹理核，仍提供了有关人体常见纹理的重要信息；我们的流程还能将解剖形状与 B模式纹理天然分离，从而有望根据目标解剖结构的形状生成合成图像。然而，最重要的是，我们的模型仅有 22M 个参数，且专用代码很少，却具备出色的超声泛化能力；它主要依赖 MONAI 库的标准实现，易于即插即用，也便于针对下游及零样本场景进行微调。SAM 和 USFM 模型通常无法在没有高端图形处理器的情况下进行微调，因此无法做到这一点：SAM 和 USFM 模型的编码器基于拥有 86M 个参数的 vit-b，并要求输入图像的分辨率为1024^2 (SAM)或256^2 (USFM)。这意味着，对于下游微调，我们的模型是更易于部署的候选方案：其在较小的128^2 图像上进行反向传播时，通常能轻松适配入门级 GPU，同时下游性能相当或更佳。我们公开了代码和模型权重，以此作为超声领域专用下游应用的起点。

参考文献

[1] D. Schein、T. Grutman 和 T. Ilovitsh，《基于深度学习的超声波束整形，用于生成时空声学全息图》，*IEEE*超声、铁电与频率控制汇刊，第70卷，第6期，第551–561页，2023年6月，doi：10.1109/TUFFC.2023.3264580。[2] H. Cho、S. Park、J. Kang 和 Y. Yoo，《深度相干学习：一种用于医学领域高质量单平面波成像的无监督深度波束形成器》超声”，《超声学》，第143卷，页码107408，2024年9月，doi: 10.1016/J.ULTRAS.2024.107408。[3] T. Grutman、M. Bismuth、B. Glickstein 和 T. Ilovitsh，“基于隐式神经表示，从稀疏超声图像进行可扩展的三维重建”，*npjAcoustics 2025 1:1*，第1卷，第1期，第1–10页，2025年8月，doi: 10.1038/s44384-025-00018-5。[4] A. Vaswani 等，“Attention Is All You Need”，2017年6月。[5] A. Kumar 等，“一个新颖的开源超声数据集，包含用于脊髓损伤定位和解剖结构分割的深度学习基准。”[6] R. Singla 等，“开放肾脏超声数据集”，2023年，第155–164页。doi: 10.1007/978-3-031- 44521-7_15。[7] X. Mei 等，“RadImageNet：用于有效迁移学习的开放放射学深度学习研究数据集”，*Radiol Artif Intell*，第4卷，第5期，2022年9月，doi: 10.1148/ryai.210315。[8] J. Ma、Y. He、F. Li、L. Han、C. You 和 B. Wang，“分割医学图像中的一切”，*NatureCommunications 2024 15:1*，第15卷，第1期，第1–9页，2024年1月，doi: 10.1038/s41467-024-44824-z。[9] M. Paschali 等，“放射学中的基础模型：是什么、如何运用、为何采用，以及为何不采用”，《放射学》，第314卷，第2期，2025年2月，doi: 10.1148/RADIOL.240597/ASSET/IMAGES/LARGE/RADIOL.240597.VA.JPEG。[10] S. Zhang 等，“BiomedCLIP：基于一千五百万组科学图文对预训练的多模态生物医学基础模型”，2025年。 [在线]。获取地址：https://www.ncbi.nlm.nih.gov/pmc/[11] M. Christensen、M. Vukadinovic、N. Yuan 和 D. Ouyang，“用于超声心动图解读的视觉语言基础模型”，*Nature Medicine* |，第30卷，第1481–1488页，2024年，doi: 10.1038/s41591-024-02959-y。[12] A. Meyer、A. Murali 和 N. Padoy，“UltraSam：利用大型开放分割数据集构建的超声基础模型。”[在线]。获取地址：https://github.com/CAMMA-[13] J. Jiao 等，“USFM：一种通用超声基础模型，可推广至不同任务和器官，助力标签高效图像分析”，*Med ImageAnal*，第96卷，页码103202，2024年8月，doi: 10.1016/J.MEDIA.2024.103202。[14] K. He、X. Chen、S. Xie、Y. Li、P. Dollár 和 R. Girshick，“掩码自编码器是可扩展的视觉学习器”，2022年。[15] T. Chen、S. Kornblith、M. Norouzi 和 G. Hinton，“一种用于视觉表征对比学习的简单框架”，2020年，访问日期：2025年1月15日。[在线]。获取地址：https://github.com/google-research/simclr。[16] M. Saeed、R. Muhtaseb 和 M. Yaqub，“有限数据条件下超声心动图分割的对比式预训练”，《计算机科学讲义》（包括《人工智能讲义》和《生物信息学讲义》子系列），第

13413 计算机科学讲义，第680–691页，2022年，doi: 10.1007/978-3- 031-12053-4_50。[17] G. Migliorelli、M. C. Fiorentino、M. Di Cosmo、F. P. Villani、A. Mancini 和 S. Moccia，“对比学习在胎儿超声成像标准切面分类中的应用”，《计算机生物学与医学》，第174卷，页码108430，2024年5月，doi: 10.1016/j.compbiomed.2024.108430。[18] G. Wang、H. Lu、S. Ma、N. Li、P. Sang 和 Y. Wang，“纹理知识引导的弱监督脑图像分割”，《生物医学信号处理与控制》，第113卷，页码108948，2026年3月，doi: 10.1016/J.BSPC.2025.108948。[19] T. Grutman 和 T. Ilovitsh，“2025年深度学习医学影像会议：用于超声分割与分析的自监督语义纹理分解”，2025年。[20] A. Hatamizadeh、V. Nath、Y. Tang、D. Yang、H. R. Roth 和 D. Xu，“Swin UNETR：用于MRI图像中脑肿瘤语义分割的Swin Transformer”，《计算机科学讲义》（包括《人工智能讲义》和《生物信息学讲义》子系列），第12962卷计算机科学讲义，第272–284页，2022年，doi: 10.1007/978-3-031-08999- 2_22。[21] A. F. T. Martins 和 R. F. Astudillo，“从Softmax到Sparsemax：稀疏注意力模型与多标签分类”，2016年2月。[22] Z. Wang、A. C. Bovik、H. R. Sheikh 和 E. P. Simoncelli，“图像质量评估：从误差可见性到结构相似性”，《*IEEE*图像处理汇刊》，第13卷，第4期，第600–612页，2004年4月，doi: 10.1109/TIP.2003.819861。[23] R. Zhang、P. Isola、A. A. Efros、E. Shechtman 和 O. Wang，“深层特征作为感知度量的非凡有效性”，访问日期：2025年1月15日。[在线]。获取地址：
htt
ps://www.github.com/richzhang/PerceptualSimilarit y。[24] J. Ansel 等，“PyTorch 2：通过动态Python字节码转换和图编译加速机器学习”，载于第*29*届*ACM*编程语言与操作系统架构支持国际会议，第*2*卷（*ASPLOS '24*），ACM，2024年4月。doi: 10.1145/3620665.3640366。[25] M. Jorge Cardoso 等，“MONAI：用于医疗保健深度学习的开源框架”，2022年。[26] I. Loshchilov 和 F. Hutter，“解耦权重衰减正则化”，访问日期：2025年8月8日。[在线]。获取地址：
https://github.com/loshchil/AdamW-and-SGDW[27] B. E. Treeby 和 B. T. Cox，“k-Wave：用于模拟和重建光声波场的MATLAB工具箱”，*https://doi.org/10.1117/1.3360308*，第15卷，第2期，页码021314，2010年3月，doi: 10.1117/1.3360308。[28] F. Yagubbbayli、D. Sinden 和 W. Simson，“k-Wave-python。”[在线]。获取地址：
https://github.com/waltsims/k-wave-python
[29] “AbdominalUS”，
https://github.com/ftsvd/USAnotAI。[30] R.-F. Lu 等，“人工智能提升肝细胞癌筛查的诊断准确性并减少工作量”，《*NPJ Digit Med*》，第8卷，第1期，第500页，2025年8月，doi：10.1038/S41746-025-01892-9。[31] A. Pawłowska 等，“用于超声乳腺病灶分析的精选基准数据集”，《*Sci Data*》，第11卷，第1期，2024年12月，doi: 10.1038/S41597-024-02984-Z。[32] N. Vallez、G. Bueno、O. Deniz、M. A. Rienda 和 C. Pastor，“BUS-UCLM：乳腺超声病灶分割数据集”，第1卷，2024年，doi：10.17632/7FVGJ4JSP7.1。[33] W. Gómez-Flores、M. J. Gregorio-Calas 和 W. Coelho de Albuquerque Pereira，“BUS-BRA：用于评估计算机辅助诊断系统的乳腺超声数据集”，《*Med Phys*》，第51卷，第4期，第3110–3123页，2024年4月，doi：10.1002/MP.16812。[34] W. Al-Dhabyani、M. Gomaa、H. Khaled 和 A. Fahmy，“乳腺超声图像数据集”，《*Data Brief*》，第28卷，第104863页，2020年2月，doi：10.1016/j.dib.2019.104863。[35] S. Leclerc 等，“利用开放的大规模二维超声心动图数据集进行深度学习分割”，《*IEEE Trans Med Imaging*》，第38卷，第9期，第2198–2210页，2019年9月，doi: 10.1109/TMI.2019.2900516。[36] N. Painchaud、N. Duchateau、O. Bernard 和 P. M. Jodoin，“通过强制时间一致性进行超声心动图分割”，《*IEEE Trans Med I maging*》，第41卷，第10期，第2867–2878页，2022年10月，doi：10.1109/TMI.2022.3173669。[37] G. Duffy 等，“利用心血管深度学习对左心室肥厚进行高通量精准表型分析”，《*JAMA Cardiol*》，第7卷，第4期，第386–395页，2022年4月，doi：10.1001/JAMACARDIO.2021.6059。[38] K. S. Da Correggio 等，“基于超声图像的胎儿腹部结构分割数据集”，第1卷，2023年，doi：10.17632/4GCPM9DSC3.1。[39] X. P. Burgos-Artizzu 等，“评估深度卷积神经网络对常见母胎超声切面的自动分类性能”，《*Sci Rep*》，第10卷，第1期，第10200页，2020年6月，doi: 10.1038/s41598-020-67076-5。[40] Y. Lu 等，“用于耻骨联合-胎头分割的JNU-IFM数据集”，《*Data Brief*》，第41卷，第107904页，2022年4月，doi: 10.1016/j.dib.2022.107904。[41] Q. Zhao 等，“MMOTU：用于无监督跨域语义分割的多模态卵巢肿瘤超声图像数据集”，访问日期：2025年8月8日。[在线]。可访问：https://github.com/cv516Buaa/MMOTU_DS2Net。[42] N. V. Yerich 等，“用于骨关节炎临床研究的标准化、实用型膝关节超声方法：约翰斯顿县骨关节炎项目”，《*ACR Open Rheumatol*》，第2卷，第7期，第438–448页，2020年7月，doi：10.1002/acr2.11159。[43] J. Born 等，“利用可解释的超声图像加速肺部病变检测

分析”，《*Applied Sciences*》*2021*年，第*11*卷，第*672*页，第11卷，第2期，第672页，2021年1月，doi：10.3390/APP11020672。[44] C. Leitner 等，“以人为中心的机器学习方法：超声图像中的肌腱连接处跟踪”，《*IEEE Trans Biomed Eng*》，第69卷，第6期，第1920–1930页，2022年6月，doi：10.1109/TBME.2021.3130548。[45] “FALLMUD：小腿肌肉超声数据集——语言与视觉人工智能。”访问日期：2025年8月8日。[在线]。可访问：https://kalisteo.cea.fr/index.php/fallmud/[46] A. Momot，“颈总动脉超声图像”，第1卷，2022年，doi：10.17632/D4XT63MGJM.1。[47] L. Pedraza、C. Vargas、F. Narváez、O. Durán、E. Muñoz 和 E. Romero，“开放获取的甲状腺超声图像数据库”，*https://doi.org/10.1117/12.2073532*，第9287卷，第188–193页，2015年1月，doi：10.1117/12.2073532。[48] H. Gong 等，“少即是多：甲状腺结节诊断的自适应课程学习”，《计算机科学讲义（含《人工智能讲义》和《生物信息学讲义》子系列）》，第13434卷（计算机科学讲义），第248–257页，2022年，doi：10.1007/978-3-031-16440-8_24。[49] N. Wiedemann 等，“COVID-BLUeS——一项关于人工智能在肺部超声分析中价值的前瞻性研究”，《*IEEE J Biomed Health Inform*》，2025年，doi：10.1109/JBHI.2025.3543686。[50] D. Ouyang 等，“用于逐搏评估心脏功能的基于视频的人工智能”，*252* | 《自然》 |，第580卷，2020年，doi：10.1038/s41586-020-2145-8。[51] M. Byra 等，“利用深度卷积神经网络在超声图像中进行迁移学习，以评估肝脂肪变性”，《国际计算机辅助放射学与外科学杂志》，第13卷，第12期，第1895–1903页，2018年12月，doi：10.1007/s11548-018-1843-2。[52] L. McInnes、J. Healy、N. Saul 和 L. Großberger，“UMAP：统一流形逼近与投影”，《*J Open Source Softw*》，第3卷，第29期，第861页，2018年9月，doi：10.21105/JOSS.00861。[53] T. Omer、C. Cousins、T. Lynch、N.-N. Le、D. Sajed 和 T. Mailhot，“COVID-19肺部超声表现：一项描述性回顾性研究”，《*Cureus*》，第14卷，第3期，2022年3月，doi：10.7759/CUREUS.23375。[54] A. A. Mohammed、W. Tavanapong、C. Fonder 和 D. S. Sakaguchi，“CountXplain：利用基于原型的密度图估计实现可解释细胞计数”，《机器学习研究论文集》，第85卷，第1–18页，2025年，访问日期：2025年11月22日。[在线]。可访问：https://github.com/NRT-D4/CountXplain。[55] P. Stähli、M. Frenz 和 M. Jaeger，“利用脉冲回波超声进行稳健声速重建的贝叶斯方法”，《*IEEE Trans Med Imaging*》，第40卷，第2期，第457页，2021年，doi：10.1109/TMI.2020.3029286。