通过基于诊断性CT体积的真实感超声合成，在超声引导脊柱手术中实现增强分割与配准

Ang Li†、Jiayi Han†、Yongjian Zhao、Keyu Li、Li Liu‡

arXiv:2301.01940v1 [eess.IV] 5 Jan 2023

摘要——**本文旨在解决临床超声数据及有意义的标注不可获取或不足的问题，以支持超声引导脊柱手术中的骨分割与配准。尽管超声并非脊柱手术的标准范式，但术中临床超声数据的稀缺是训练神经网络难以逾越的瓶颈。此外，由于超声成像的特点，骨表面难以清晰标注，导致训练后的神经网络无法关注细节。因此，我们提出了一种**计算机模拟**骨超声模拟框架，可从诊断性CT体积合成真实感超声图像。随后，我们利用这些模拟骨超声训练轻量级视觉Transformer模型，使其能够在脊柱超声检查中实现准确且实时的骨分割。在验证实验中，我们通过诊断性脊柱CT体积生成真实感超声模拟数据，以支持无辐射的超声引导椎弓根螺钉置入。将其用于骨分割任务训练时，倒角距离为0.599 mm；用于CT-超声配准时，骨分割准确率达到Dice系数0.93，基于分割点云的配准准确率为0.13∼3.37 mm，且不会引发并发症。骨超声图像在介质界面处会呈现强回声，仅依靠较小邻域的信息，可能导致模型无法区分薄界面与骨表面。为克服这些不足，我们提出采用长程对比学习模块（LCLM），充分挖掘候选区域与其周围像素之间的长程对比关系。消融实验验证了所提出的长程对比学习模块有助于精确定位超声感兴趣区域。此外，训练数据完全由我们提出的超声模拟框架生成，未基于真实临床数据进行微调，这证明了该骨超声真实感模拟框架的有效性。**

给从业者的说明——**本文旨在解决骨超声图像和标注标签不可用或数量不足的问题。我们采用这种数据增强技术，生成与相应 CT 体积相关联的真实感模拟骨超声及标注。当前超声数据增强方法的主要问题是无法生成连续且上下文准确的数据（基于神经网络的算法），以及缺乏实时物理仿真能力（传统的基于声学算法）。**

在本文中，我们通过加入挤压仿真、移动空间仿真和图像增强等技术，改进了基于CT图像的超声仿真，以生成质量更高的图像，并将这些图像直接应用于目标任务，即训练神经网络来引导超声引导脊柱手术。此外，我们通过消融实验展示了真实感超声模拟框架及神经网络各个新模块的有效性。可以得出结论，使用超声模拟框架生成的数据训练的神经网络，有望用于超声引导下的椎弓根螺钉置入手术。未来，我们将把体内 **骨骼超声仿真用作强化学习环境，并部署训练好的智能体，直接引导超声引导手术。**

索引词——**真实感超声模拟、视觉Transformer、骨表面分割、CT-超声配准、超声引导脊柱导航**

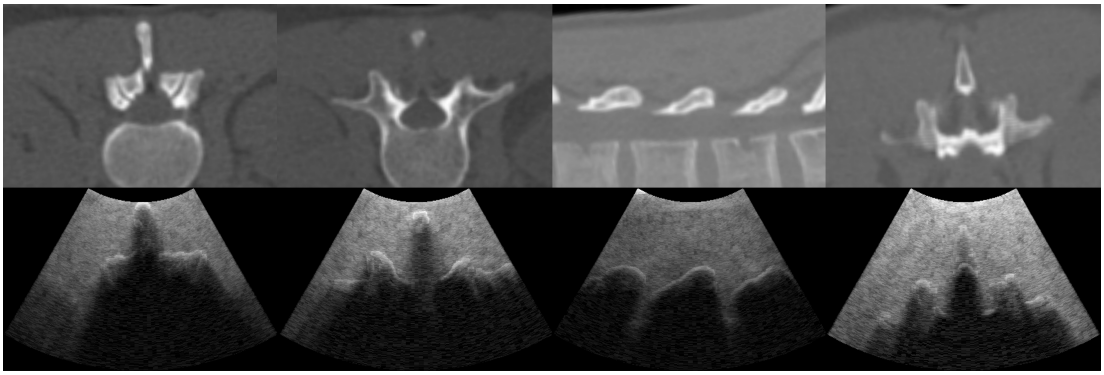
I. 引言

从术中超声数据中分割骨表面，随后进行 CT-超声配准，是超声引导脊柱手术的两个关键步骤。近期研究聚焦于采用深度学习方法，实现骨表面的精准、稳健、实时分割与配准。然而，由于缺乏标准化数据以及患者隐私方面的顾虑，数据量匮乏成为在术中成像领域应用深度学习方法的一大挑战。具体而言，超声并非脊柱相关手术中的标准成像模态，且超声引导脊柱手术并不常见，因此这一问题尤为突出；即使脊柱超声检查可用于临床前采集流程，海量数据的标注仍是一项严峻挑战。另一个限制因素是手动数据采集流程：超声换能器相对于成像脊柱解剖结构的方向欠佳，会导致用户依赖性较强地采集到低质量骨骼扫描图像。

通过数据增强扩大现有数据集规模，以提升模型性能，这一做法已得到广泛研究。其中，获取大型标注数据集的一种基础方法是计算机模拟的脊柱超声图像真实感模拟。超声模拟方法大致可分为三类：基于声学模型的超声模拟、基于图像的超声模拟，以及利用生成对抗网络（GAN）生成虚拟图像。基于声学模型的超声生成算法通常速度很慢 [1]。利用 GAN 生成超声图像可能需要针对不同器官训练不同的网络，同时，难以获取在同一空间中对齐的相应超声体数据和 CT 体积，因此，利用 GAN 生成大规模合成超声图像颇具挑战。基于图像的方法尝试利用模拟的

† 这些作者作出了同等贡献

‡ 表示通讯作者。电子邮箱：lliu@ee.cuhk.edu.hk。AngLi、Yongjian Zhao、Keyu Li 和 Li Liu 均任职于香港中文大学。Jiayi Han 任职于复旦大学。



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

图1. 图中展示了生成的超声图像示例，分别对应两块椎骨的关节面、可完整观察椎板的平面、沿棘突扫查的平面，以及任意扫描平面。

超声探头对原始图像（例如 CT 扫描）进行重新采样，然后将图像标量视为器官的声学参数，并根据其声学特性模拟传播过程。然而，与 CT 和 MRI 等其他成像方法不同，超声会显著受到气体的影响。为避免气体的影响，操作者会按压探头，排出探头与皮肤之间的气体，这会导致原始图像中的组织发生畸变。然而，先前的研究忽略了这种畸变。此外，由于不同组织的反射率各异，先前的工作会对图像进行分割，并为每种组织设计特定的传递函数，十分耗时。此外，传统上生成超声图像仅用于辅助二维-三维图像配准 [2] 或培训检查医师 [3] ，并未进一步应用合成超声图像。

本文提出了一种新颖的基于CT的真实感超声合成框架，该框架结合了采用采样方法的自动图像生成，如图3所示。从(a)~(b)中，每列分别表示真实超声图像、对应的CT扫描、反射图和透射图。我们通过对原始图像进行变形来模拟探头按压所产生的畸变，并提出一种可直接应用于整幅图像的自适应传递函数，从而免去针对每种组织设计传递函数的过程，大幅加快超声仿真任务。为了充分发挥所提仿真的优势并实现实时超声图像分割，我们进一步提出了一种配备长程对比学习模块（LCLM）的轻量级视觉Transformer。该模块采用设计的级联空洞卷积层，实现密集的超大感受野，从而提升超声图像分割效果。实验表明，与其他方法相比，所提出的仿真系统达到了最先进的性能，并有助于所提出的视觉Transformer实现真实超声图像分割。

我们的贡献如下：

1）我们提出了一种计算机模拟骨超声模拟框架，可从诊断性CT体积合成逼真的超声图像。

2）我们开发了一种轻量级视觉Transformer模型，可对脊柱超声图像进行精准、实时的骨分割。3）实验表明，与初始CT扫描相比，所提出的计算机模拟骨超声模拟方法显著提升了分割性能，说明该数据增强方法能够为临床真实脊柱超声检查预训练模型。

II. 相关工作

本节讨论真实感超声仿真及相关的骨骼分割工作。

A. 真实感超声模拟

传统的基于声学的超声仿真软件由 Jenson等人在1996年开创。B. T. Cox等人 [4] 于2005年提出了用于快速计算脉冲光声场的K空间方法。随后，Treeby等人 [6] 基于K空间方法优化了超声图像传播模型，并开发了广泛使用的K-wave工具。这种传统的基于声学的算法的优势在于，它能够模拟各种类型的超声成像系统，并尽可能真实地再现图像的物理效应。

由于这些基于声学物理模型的算法在生成大型数据集时效率不高，研究人员开始开发基于光线追踪的方法。Burger 等人[7] 通过将 CT 数据集分割为不同组织，并为每种组织赋予速度、阻抗、散射因子及其他声学特性，开发了一种超声模拟系统；因此，该系统可用于模拟声传播、吸收和散射过程。Cong 等人[8]提出了一种多尺度增强方法，以增强管状结构来模拟血流，并使超声图像更加逼真。Piorkowski 等人 [3] 于 2013年将 Wein [9] 和 Kutter [10] 的算法应用于开发经食管超声心动图（TEE）模拟器，该模拟器对培训医生进行经食管超声心动图检查产生了极其积极的影响。为进一步提高图像生成速度，Wang 等人 [11]使用 NVIDIA 的 Optix 6.0 光线追踪

引擎进行超声蒙特卡罗模拟，取得了良好效果，相较于生成对抗网络和 Field-II [12]表现更佳。

近年来，生成对抗网络（GAN）模型已广泛应用于真实感超声仿真研究。模型通过在超声领域的大型数据集上学习，生成真实感超声图像。Hu 等人 [13]将 GAN 模型用于自由手扫描，以生成超声图像。该模型将空间位置信息作为条件输入，然后输出当前位置的超声图像。这项工作有助于生成相应位置的超声数据，但无法为每位特定患者生成合成超声图像。为解决这一不足，Tom 等人 [14]于 2018 年采用级联生成对抗网络，并以图像分割标签为条件输入，生成更真实的超声图像，从而能够针对不同的分割结果生成合成超声图像。然而，根据他们的报告可以得出结论，即使分割结果准确，基于生成对抗网络的虚拟超声系统仍难以确保边缘强度和形状与真实超声图像一致。

B. 超声分割

早期的一些研究采用手工设计特征进行超声分割，例如主动轮廓 [15], [16]。近年来，UNet 是超声分割中最流行的模型 [17]。许多研究在基于 UNet 的模型中采用了不同的改进方法来进行超声分割。 [18] 提出了一种结合分类和分割任务的多任务 UNet。 [19]提出了一种轻量级 UNet 模型，交替采用3×3 和 1×1 卷积层。他们还引入了一种错误输出抑制机制，将逐块分类结果与分割结果相结合，以消除假阳性。 [20] 在超声图像分割任务中采用了空间注意力。

III. 方法

A. 基于 *CT* 体积的真实感超声模拟

整个超声仿真流程分为几个部分：数据采样、数据筛选、超声发射和图像融合。

a) 成像探头与按压模拟：由于所提出的超声模拟系统旨在训练医生或手术机器人的人工智能代理，因此探头的运动范围不能覆盖整个三维空间。系统中的探头可以在指定的扫描空间内移动，但探头是曲面，即探头表面，而扫描空间无法与其完美贴合。若模拟施加压力，扫描空间会包裹并贴合探头。然而，CT 体积不会随表面发生变形。因此，需要提出一种算法，使 CT 数据能够沿扫描空间随探头表面变形。使用弹簧-质量模型模拟扫描空间内的变形，再利用移动最小二乘算法拟合 CT 变形，当然是可行的，但也有两个缺点。首先，求解弹簧-质量模拟和移动最小二乘方程会消耗大量计算资源。其次，该算法需要额外投入大量工作，在图像上标记锚点。

在这种情况下，我们对 A. Gustafsson 提出的局部平移图像变形算法进行了调整，使其适用于 CT 数据，并根据探头的形状将其应用于图像。该算法只需了解探头的形状以及该切片 CT 数据的 HU 值，即可模拟组织的运动。若要进一步加速，可将 HU参数的权重值设为 1；这样只需计算一次 UV坐标映射，且完全不会影响成像效率。

该公式表示为公式 1

$$\begin{cases} \vec{u} = \vec{x}\frac{r_{max}^2 - |\vec{x}-\vec{c}|^2}{r_{max}^2 - |\vec{x}-\vec{c}|^2 + D}(\vec{m}-\vec{c}) \\ D = \frac{100}{f}\alpha(hu)|\vec{m}-\vec{c}|^2 \end{cases}. \tag{1}$$

在上述公式中， f 控制形变比例。在我们的案例中，我们希望将探头中心线与采样图像上边缘线的交点（如图 4 中的蓝点所示）推至探头顶部（图中红点所示）。根据上述算法，蓝点与红点之间的所有组织都需要向下挤压。假设组织是刚体，那么即使施加非常大的力进行挤压，组织厚度也不会改变。在这种情况下，红点处的原始位置应被推至绿色曲线的顶部。然后可以绘制一条绿色弧线，用来表示组织运动的极限。探头弧线与绿色弧线之间区域内的组织受到挤压，会导致该区域出现强反射或吸收。因此，在模拟传播的过程中，必须使声波在该区域内的衰减较小，否则算法生成的受挤压组织将完全阻挡超声的传播。如图 3 所示，探头曲线下方有几条亮线，它们是受挤压组织的信号，与体内超声图像一致。

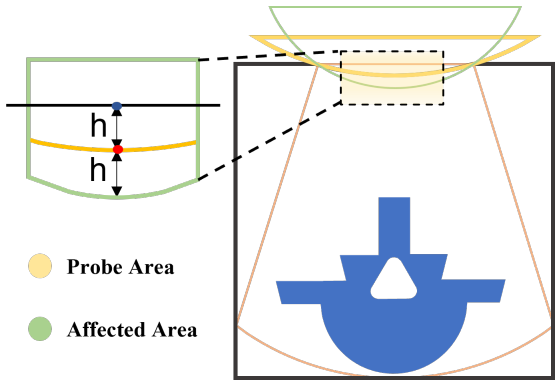
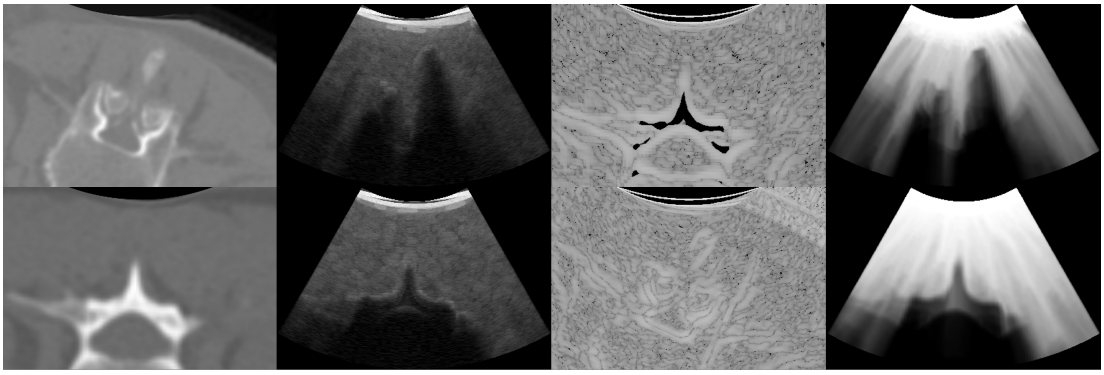


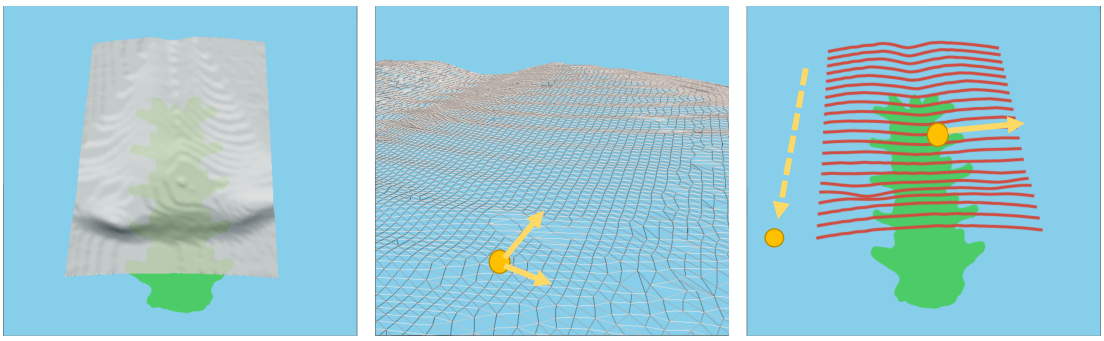
图 2. 该图展示了压力仿真算法。作为图中虚线框区域的放大图，黄色弧线表示探头的形状，黄色弧线与绿色弧线之间的区域是被挤压的目标空间。

b) 探头受限运动空间：如上所述，探头无法在空间中任意移动。本文以脊柱超声为例，设计了两种探头移动约束方式，以实现超声图像采集的自动化。定义 M 为三维空间中的一个网格，并通过扫描目标的边界框截取一个感兴趣的三维子网格 M_r



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

图 3. 图 (a) 显示经过压力仿真处理后的 CT图像，其中皮肤和肌肉被压向探头的曲面。图 (b) 显示通过模拟挤压生成的虚拟超声图像。图 (c) 显示挤压后的反射图。图 (d) 显示处理后的图像不会影响传播。



(a) (b) (c)

图 4. 图中的绿色部分是待扫描对象。图 (a) 显示整个可行空间（皮肤）与扫描对象（脊柱）之间的关系。图 (b) 显示第一种运动方式，黄色圆点表示探头尖端，箭头表示流形中的二自由度运动。图 (c) 显示第二种运动方式，红线表示离散曲面，虚线箭头表示探头可以在不同曲线上切换，箭头表示探头可以沿曲线方向移动。

，探头只有在这个流形中移动才有意义。第一种方式是在网格上自由移动：对于 M_r上的任意点 p，探头 p 的移动方向将沿网格法向量弯曲。在这种情况下，非常小的 dx 或 dy不会造成显著的 dz，因此图像不会发生剧烈变化，也无法采集连续图像。第二种方法是沿前进方向 $M_{r0}...M_{rn}$将 M_r 切分为三维空间中的一系列曲线。在每条折线上，各点只能沿切向方向前进或后退一定距离。这种方法的优点是，探头法向量的各个分量在该曲线上保持一致；它常用于对目标进行自动环形扫描或平面扫描，并确保扫描目标位于一条线或一个点上。此外，这两种移动方式都可由基于强化学习的智能体采用；第一种方式更适用于连续控制，而第二种方式更适用于离散控制。

c) 声反射： 仿真的第一步是将 CT-HU图像映射为声阻抗图像，映射查找表参考 [21]。在映射表中，每个 HU值都会映射到一个声阻抗 Z。随后将处理该声阻抗图像，而非 CT图像。根据声学物理传递特性，我们首先引入菲涅耳方程来计算反射。将声波分为平行于平面的声波和垂直于平面的声波，方程如下：

$$\begin{cases} R_\perp = (\frac{Z_1 cos\theta_i - Z_2 cos\theta_t}{Z_1 cos\theta_i + Z_2 cos\theta_t})^2 \\ R_\parallel = (\frac{Z_2 cos\theta_i - Z_1 cos\theta_t}{Z_2 cos\theta_i + Z_1 cos\theta_t})^2 \\ R = \frac{1}{2}(R_\perp + R_\parallel) \end{cases} . \tag{2}$$

声波从一种介质传播到另一种介质时，不仅会发生反射，还会发生折射。这种折射与光穿过不同介质时的折射具有相同的物理特性，可由斯涅尔定律表示：

$$\frac{sin\theta_1}{sin\theta_2} = \frac{Z_2}{Z_1} \tag{3}$$

事实上，当声波发生折射或弯曲时，其方向会保持不变，直到下一次传播。可以使用

递归光线追踪算法来模拟多级折射和反射值。然而，由于折射声波的振幅较小，对生成图像的质量影响不大，因此所提出的方法不考虑其折射后的方向，而只将其振幅叠加到下一介质中的入射声波上。

辐照度：引入朗伯余弦定理，以考虑声波的辐照度。该公式意味着，组织梯度方向与声波方向越垂直，声反射越强，组织也越明亮。

$$\frac{I_r}{I_i} = cos\theta \tag{4}$$

d) 声衰减：超声在组织中传播时的衰减大致可分为反射、散射和吸收，其中吸收占大多数。论文 [22]指出，牛肝和猪肝等物质对超声的吸收占超声传播衰减的90%，甚至可达100%。即使在牛脑和腿部肌肉等表现出各向异性的组织中，吸收仍是衰减的重要组成部分。因此，将CT扫描映射为吸收图像十分重要。映射表通过对论文 [23]中的数据I进行插值得到。吸收公式由朗伯-比尔定律给出。由于不同组织在相同HU值下的超声吸收能力可能不同，为了使生成的图像效果更好，我们在原始公式中加入了一个可调参数 α 。修改后的公式如下：

$$I_a = I_0 10^{-\alpha * d * f * \frac{1}{10 * \beta}} \tag{5}$$

组织	密度	速度	阻抗	衰减
Skin	1100	1631	1.794	0.22
Fat	916	1435	1.352	0.975
肌肉	1041	1595	1.647	1.47

表 I 不同组织的声学属性：密度（kgm^-3）、速度（ms^-1）、阻抗 10^6（kgm^-2s^-1）、衰减（$dBcm^-1MHz^-n$） [23]

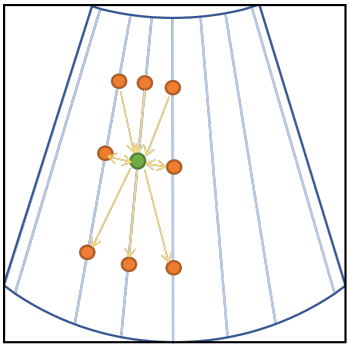


图 5. 传播示意图。

e) 离散场中的传播：超声的传播是生成过程中最耗时的部分。尤其是查找接收组织的步骤，反射和折射需要花费大量时间来检测频繁出现的交点。

与此同时，相邻像素的方向是离散的，而声传播方向是连续的。因此，线段可能相交，而像素却没有交集。八邻域中的每个像素都应视为波传播的障碍物，并将传播方向与像素连线方向的余弦值归一化后作为障碍物权重。也就是说，如果连线方向与传播方向相反，则权重为零；当两个方向完全相同时，障碍物权重最大。

为便于计算反射和折射导致的振幅增强，每个点还会计算其向邻域的波传播。在这种情况下，像素连线方向与传播方向的余弦值为负时，将视为反射。余弦值小于 0.8（一个超参数）时，将视为折射和散射造成的对比度增强。传播示意图如图 1 所示。

f) 反射增强：当换能器接收到来自介质间隙的反射声波时，超声图像中会出现一条亮条纹。由于只有在介质性质发生变化时才会产生反射，因此骨表面与肌肉组织之间应该只有一次反射，但超声图像中经常会在骨骼与肌肉之间显示一条很宽的亮条纹。Hacihaliloglu 等人[24]提出，这条亮条纹是由超声在仰角方向上的厚度造成的。基于这一理论，本文对当前平面前后方的三维梯度纹理进行采样（如图 6 中的蓝线所示）。这些纹理乘以一个与距离成反比的权重值 α，然后累积到当前生成的超声图像中。模拟宽条纹对神经网络训练具有重要意义；如果训练数据中不存在宽条纹，那么模型在真实超声数据上分割骨表面时，输出就会偏离真实标注。

总之，生成的超声图像是将传播图像与反射图像、吸收图像进行计算，然后再将这三幅图像与径向噪声混合得到的。

B. 超声图像分割

在这项工作中，我们提出了一种轻量级超声分割框架，如图7所示。我们采用[25] 作为骨干网络。在每个MobileViT模块之后，我们添加一个LCLM，以进一步进行长程对比学习。随后，我们使用特征金字塔网络（FPN）恢复输入超声图像的初始分辨率。

a) **MobileViT模块：** 为涵盖全局信息，MobileViT模块首先使用 3×3卷积聚合相邻特征。随后将像素分组成图块，每个图块包含 $h \times w$ 个像素，其中 h 和 w分别表示图块的高度和宽度。每个图块的 i_{th} 词元

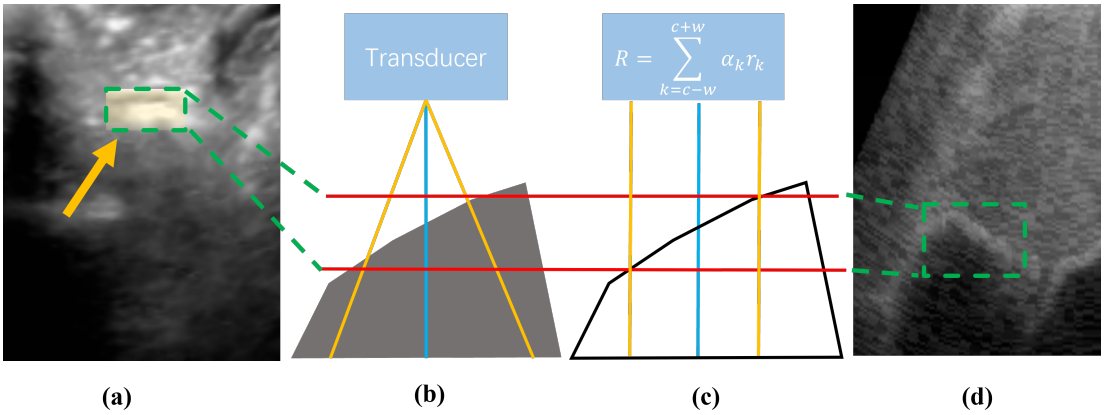


图6。图a)中的箭头指示了粗条纹。图b)展示了超声波的厚度如何导致粗条纹伪影。图c)为三维梯度纹理中采样方式的示意图。蓝线表示成像平面。蓝线与黄线之间的反射值会相乘并累积到成像平面中。图d)展示了系统中模拟出的粗条纹。

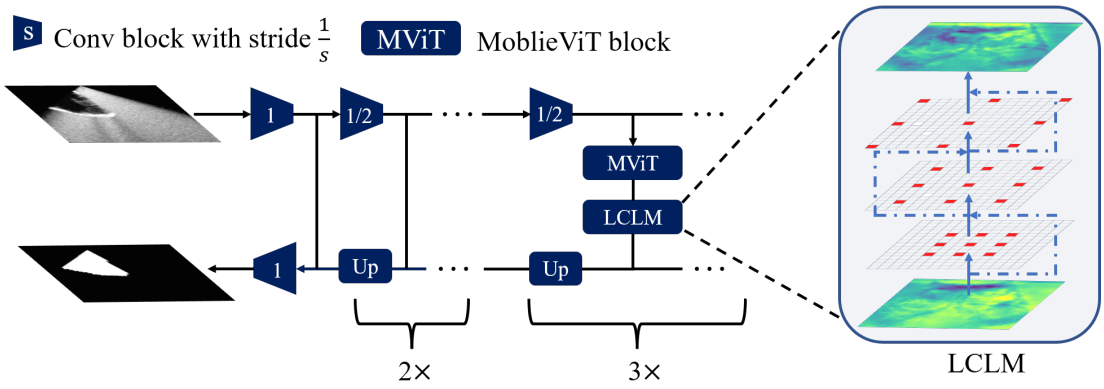


图7。所提出的用于超声分割的轻量级视觉Transformer框架。在前两个阶段，我们仅使用基本卷积层来学习局部表征。在最后三个阶段，我们交替使用MobileViT模块和LCLM，以同时涵盖长程依赖和长程对比。

通过自注意力（SA）彼此传递特征。当 $h <= 3 \wedge w <= 3$时，每个像素都可以访问所有像素的信息。有关 MobileViT模块的详细信息，请参见 [25]。

b) **为什么需要 LCLM：** 我们发现，超声图像中的感兴趣区域（RoI）具有相似的纹理，但相较于其他组织，其振幅较弱。基本的 3×3 卷积能够建模邻近纹理，却无法学习长程纹理。由于待分割区域的局部纹理与其他组织相似，因此我们必须设计一个能够涵盖待分割区域内外纹理的模块。

在视觉Transformer中，如公式6所示，通过对所有词元进行自注意力（SA）来实现长程纹理建模：

$$\widetilde{x} = \text{softmax}(\frac{xW_q(xW_k)^T}{d})(xW_v), \tag{6}$$

其中， $x \in \mathbb{R}^{N\times d}$ 表示维度为 d、 $W_q \in \mathbb{R}^{d\times d_q}$、 $W_k \in \mathbb{R}^{d\times d_q}$ 的 N 输入词元， $W_v \in \mathbb{R}^{d\times d_v}$为可学习参数。我们将自注意力重写为公式如下（见公式 7）：

$$\begin{cases} \widetilde{x}_i = \sum\limits_{j\in\mathcal{A}_i} w_j(x_jW_v) \\ w_j = \frac{exp(x_iW_q(x_jW_k)^T/d)}{\sum_t exp(x_iW_q(x_tW_k)^T/d)} \end{cases}. \tag{7}$$

其中， $\mathcal{A}_i$ 表示词元 x_i可访问的词元集合。由于 W_q 和 W_k 仅用于建模查询之间的关系，因此一旦固定 W_v ， x 所生成的空间便受到限制。同时，如图8(a)所示，SA倾向于从词元中聚合“相似”特征，这有助于模型学习长程依赖。然而，由于超声图像的特性，学习长程对比也很重要，但SA无法做到这一点。

卷积运算可以用与SA相同的形式表示，如公式8所示：

$$\widetilde{x}_i = \sum_{j\in\mathcal{A}_i} \lambda_j(x_jW_v) . \tag{8}$$

在这种情况下，如果 $rank(x_jW_v) = d$，则由 x 生成的空间为 $\mathbb{R}^d$。我们随机选取 1000 组不同的 W_q 和 W_v、1000 组不同的 λ，并固定 W_v。输出如图8所示。如图所示，卷积的输出填满了空间，而自注意力的输出则聚集在一起。

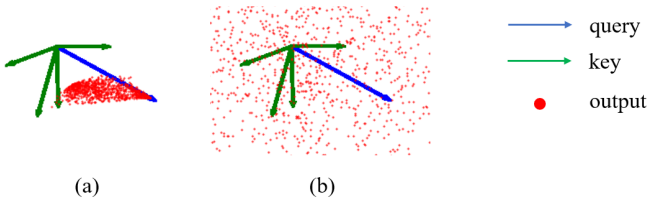


图8. 自注意力（a）和卷积（b）在固定 W_v 时可能产生的输出（红点）。

近期的一项研究 [26] 也表明，采用超大卷积核的卷积神经网络能够从超大感受野（子全局）中学习特征表示，并以更少的参数实现与最先进的Transformer相当的性能。

如前所述，我们需要一个模块来提取长程纹理。由于自注意力无法学习长程对比，而超大的卷积核会带来巨大的计算开销，因此我们需要设计一种轻量且有效的卷积模块来进行长程对比学习，即长程对比学习模块（LCLM）。

c) **通过空洞卷积进行长程对比学习：**为了用更少的参数覆盖长程对比，一种简单的方法是使用空洞卷积。在这项工作中，我们采用了 3 个级联的空洞卷积层，空洞率为 {3,5,11}。结合 MobileViT 模块末尾的一个基本卷积层，LCLM 可密集覆盖多达 41×41 个像素。我们在图9中展示了感受野。每个卷积层后面都接一个归一化层和一个激活层。深度LCLM 仅需 $3 \times 3 \times 3 \times d$ 个参数，甚至少于一个标准卷积层（$3 \times 3 \times d \times d$，$d >> 3$）。

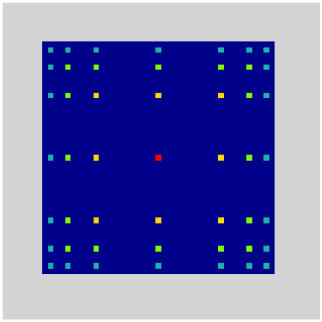


图9. 红色像素的感受野。非灰色像素均被该感受野覆盖。

IV. 实验

A. 基于*CT*体积的真实感超声模拟

为图像生成相关研究提供定量评价标准并非易事。从定性角度来看，与基于深度学习的超声图像生成相比，本文的优势在于能够生成针对患者个体的图像，并更准确地呈现病灶边界，而且与以往基于图像生成的研究相比，我们通过模拟软组织和骨骼中的挤压与散射，使系统生成的图像更加逼真。尽管如此，定量评价也具有重要意义。由于图像生成质量难以进行定量评估，我们根据研究目的来评价图像生成质量。除了用于培训医生进行超声检查外，超声图像生成算法最重要的用途是术中配准和提供深度学习数据集。因此，我们将生成的图像输入各种基于深度学习的分割算法，以验证该图像生成系统对于神经网络预训练的实用性。

B. 基于深度学习的分割

a) **实施细节：**为验证所提出的超声模拟方法和轻量级分割模型，我们使用合成超声图像训练所提出的模型，并在真实超声图像上进行推理。我们将批量大小设为32，使用Adam优化器，并将学习率设为1e-4。

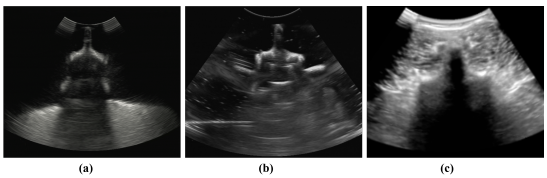


图10. 验证图像示例。图(a)为凝胶模拟物中脊柱体模的超声图像，图(b)为水中的3D打印脊柱的超声图像，图(c)为临床人体的超声图像。

b) **与其他模型的比较：**我们首先在表 II 中将所提出的模型与其他超声分割模型进行比较。表中报告的所有结果均由我们自行复现。与基于 UNet 的模型相比，所提出的模型规模更小且效果更好。与 MViT-FPN 相比，所提出的方法取得了显著更好的性能。

为了使用该模型，在本案例中，分割的目的是实现更好的配准，因此我们去除了存在明显分割错误的案例，得到我们筛选后的结果，并将其用于后续配准。在这种情况下，使用生成数据训练的网络，其效果可以达到甚至优于使用大量真实超声图像和精确标签训练的最先进模型。我们的标签通过 CT分割获得；由于该方法能够标注在某些超声场景中不可见或边界模糊的骨表面，因此对于轻量级神经网络而言，其收敛更具挑战性。

c) 消融研究：为证明合成超声相较于初始 CT 扫描、长程依赖和长程对比的有效性，我们对上述设置进行了验证，结果如表 III 所示。使用 CT 扫描训练模型会导致交并比大幅下降，这表明所提出的方法能够有效合成超声图像。移除 LCLM 后，模型无法学习超声图像的长程对比，导致性能下降。移除 MVIT 后，交并比略有下降，这表明

方法	Dice	CD（TP）	CD（FN）	参数数量	FPS
UNet [17]	0.574	1.085毫米	0.590毫米	131.8M	4.13
MViT-FPN [25]	0.294	4.664毫米	0.824毫米	20.0M	28.1
CNLUnet [19]	0.329	2.868 毫米	1.287 毫米	48.6M	10.9
Ours	0.783	0.599 毫米	1.079 毫米	20.0M	26.3
我们的方法（选定）	0.926	0.227mm	0.184mm	20.0M	26.3

表 II 与其他分割模型的比较。请注意，表中报告的所有性能均由我们自行重新实现。

模型能够从长程依赖中获益，但其作用与长程对比相比并不显著。

方法	基线	合成超声→计算机断层扫描	不含 MViT	不含 LCLM
Dice	0.783	0.040	0.438	0.294
CD(TP)	0.599mm	6.062mm	0.910mm	4.664mm
CD（FN）	1.079mm	3.938mm	0.674mm	0.824mm

表 III 所提模型的消融研究。“SUS”表示合成超声。“SUS→CT”、“不使用 MVIT”和“不使用 LCLM”分别表示使用 CT扫描训练模型、从基线中移除 MVIT，以及从基线中移除 LCLM。

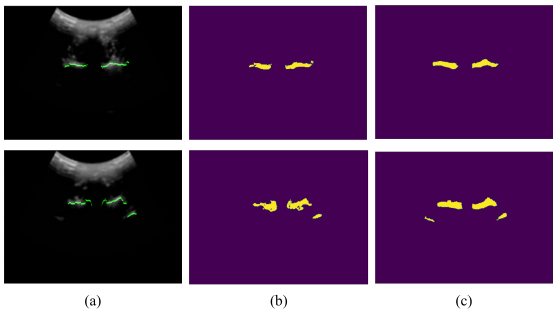


图 11. 分割结果。图(a)、(b)和(c)分别展示了分割标签、网络输出和真实标注。

C. 超声-计算机断层扫描配准的验证

a) 实施细节：提出超声模拟系统，旨在从多种姿态、患者和环境中更好地获取标注数据，从而实现基于超声的手术导航。本文验证了基于合成数据训练的模型输出是否能够将术前CT和规划轨迹对齐至术中阶段，并以椎弓根螺钉置入中的脊柱骨表面配准为实例。

验证分为两部分：第一部分验证刚体配准的准确性，即平移和旋转；第二部分验证术前规划轨迹经过上述变换后是否会导致手术并发症，也就是术中轨迹是否会触及重要器官。

实验在三种仿体上进行：水中的脊柱仿体、水中的3D打印人体脊柱，以及琼脂凝胶中的牛脊柱。配准步骤分为粗配准和针对每个节段的单独配准。粗配准使用所有点云来计算大致的位置和方向，最终配准采用去噪迭代最近点（ICP）方法。配准结果如下表所示。

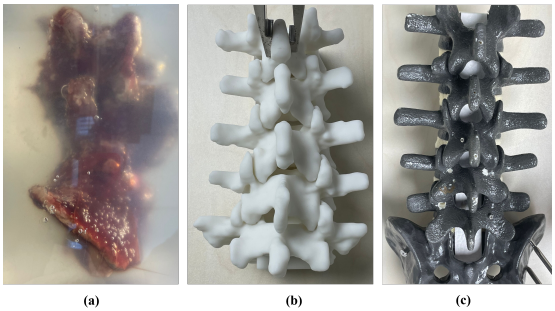


图 12. 用于配准验证的体模模型。

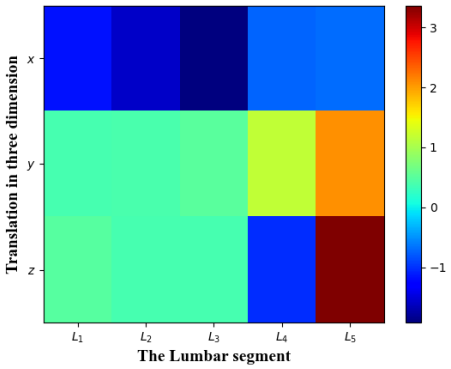


图 13. 三维空间中的配准均方误差损失。

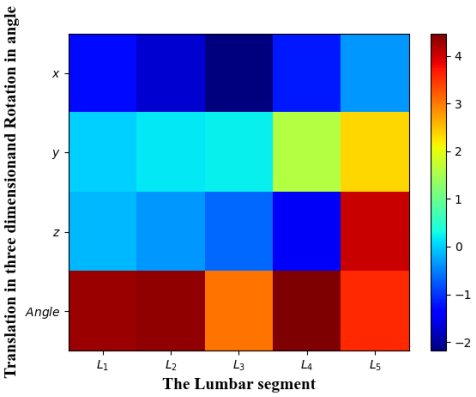


图14. 通过配准矩阵得到的椎弓根螺钉平移和旋转。旋转以角度表示，单位为度。

b) 术中模型配准验证：在配准阶段，模型以超声图像为输入，输出分割标签，并提取其最大连通域的上轮廓，因此可利用超声探头的标定矩阵生成点云。超声标定通过双十字线模体完成 [27]，，均方误差（最小平方损失）为 0.97 mm。由于

各个分段的坐标系并不重合，因此下一部分将通过椎弓根螺钉的旋转来表示角度方面的旋转评估。为评估配准误差，采用点对点均方误差，该误差也是迭代最近点算法的目标函数。配准结果如图15所示，其中白色模型表示术中空间中的真实标注，绿色模型表示超声配准结果。总之，使用生成的超声图像训练的分割算法可用于配准，误差为 $0.133 \sim 3.366$ mm。最大误差出现在第五腰椎的 z轴方向，约为 3 mm，这可能是因为第五腰椎的形状与其他腰椎的形状差异较大。

c) 椎弓根螺钉置入可行性验证：为验证系统在目标任务中的配准效果，我们采用专家规划的椎弓根螺钉置入术中方案，将螺钉尖端的平移和旋转与超声点云配准结果及真实标注进行比较。旋转角度由三维空间中两个椎弓根轴向量之间的夹角确定，并以度为单位表示。如图14所示，螺钉尖端的平移误差介于0.027 mm至4.031 mm之间，其中最大误差仍出现在第五腰椎的z轴上，这与上述评估结果一致。旋转误差介于3.05度至4.47度之间，因此可以得出结论：使用生成的超声图像的预训练模型能够以较小误差完成对齐。椎弓根螺钉置入结果如图16所示，表明该配准不会使患者面临并发症风险。

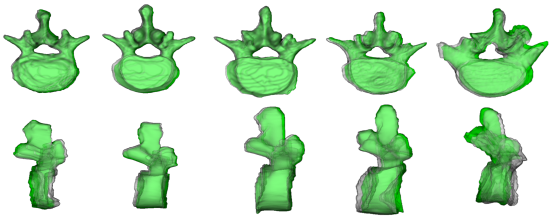
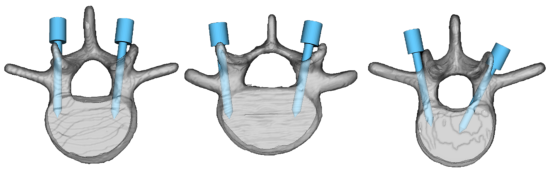


图15. 配准的可视化。图中绿色部分是通过超声分割得到的配准结果，白色部分是真实标注。案例取自3D打印患者脊柱，对应图12中的(b)。



(a) (b) (c)

图16. 基于配准的螺钉置入可视化。图中展示了经验丰富的外科医生在术前CT上规划的标准长度和直径螺钉的姿态，并与通过图像分割获得的配准矩阵对应的术中脊柱真实标注进行比较。案例取自标准脊柱模体，对应图12中的(c)。

V. 结论

本文提出了一种基于CT图像和声学特性的超声图像仿真方法，可根据特定患者的CT 体积自动生成超声仿真环境，从而训练和验证基于超声的深度学习与强化学习算法。此外，我们提出了一种用于超声图像分割的轻量级视觉Transformer，其网络结构具有更高的分割准确率和模态泛化能力。

实验表明，与使用CT图像训练的模型相比，使用由计算机断层扫描系统通过真实感超声仿真生成的超声图像训练的模型具有更高的分割准确率，并且能够在无并发症的情况下执行术中配准流程。与其他分割算法相比，所提出的Transformer具有实时计算效率、更高的分割准确率和更强的泛化能力，这对于超声引导的外科应用尤为重要。

参考文献

[1] K. Wang、E. Teoh、J. Jaros 和 B. E. Treeby，“使用 k-wave 工具箱对吸收介质中的非线性超声传播进行建模：实验验证”，载于*2012 年 IEEE 国际超声研讨会*。IEEE，2012 年，第 523–526 页。[2] G. Ning、X. Zhang 和 H. Liao，“基于强化学习的自主机器人超声成像系统”，*IEEE 生物医学工程汇刊*，第 68 卷，第 9 期，第 2787–2797 页，2021 年。[3] A. Pi´orkowski 和 A. Kempny，“基于计算机断层扫描图像的经食管超声心动图模拟器”，*IEEE 生物医学工程汇刊*，第 60 卷，第 2 期，第 292–299 页，2012 年。[4] J. A. Jensen，“Field：用于模拟超声系统的程序”，载于第十届北欧-波罗的海生物医学成像会议，第 *4* 卷，增刊 *1*，第 *1* 部分：*351–353*。Citeseer，1996 年。[5] B. T. Cox 和 P. C. Beard，“使用 k 空间方法快速计算流体中的脉冲光声场”，美国声学学会期刊，第 117 卷，第 6 期，第 3616–3627 页，2005 年。[6] B. E. Treeby、J. Jaros、A. P. Rendell 和 B. Cox，“使用 k 空间伪谱方法对具有幂律吸收的非均匀介质中的非线性超声传播进行建模”，美国声学学会期刊，第 131 卷，第 6 期，第 4324–4336 页，2012 年。[7] B. Burger、C. Abkai 和 J. Hesser，“基于 CT 模型的动态超声模拟及其在医学教育中的应用”，健康技术与信息学研究，第 132 卷，第 56 页，2008 年。[8] W. Cong、J. Yang、Y. Liu 和 Y. Wang，“基于 CT 图像的快速自动超声模拟”，医学中的计算与数学方法，第 2013 卷，2013 年。[9] W. Wein、S. Brunke、A. Khamene、M. R. Callstrom 和 N. Navab，“用于诊断成像和图像引导干预的自动 CT-超声配准”，医学图像分析，第 12 卷，第 5 期，第 577–585 页，2008 年。[10] O. Kutter、R. Shams 和 N. Navab，“基于 CT 图像的医学超声可视化与 GPU 加速模拟”，生物医学中的计算机方法与程序，第 94 卷，第 3 期，第 250–266 页，2009 年。[11] Q. Wang、B. Peng、Z. Cao、X. Huang 和 J. Jiang，“结合 OptiX 引擎与蒙特卡罗路径追踪的实时超声模拟器”，载于 *2020 年 IEEE 系统、人类与控制论国际会议（SMC）*。IEEE，2020 年，第 3661–3666 页。[12] J. A. Jensen，“使用 FIELD II 模拟先进超声系统”，载于*2004 年第二届 IEEE 生物医学成像国际研讨会：从纳米到宏观（IEEE 目录号 04EX821）*。IEEE，2004 年，第 636–639 页。[13] Y. Hu、E. Gibson、L.-L. Lee、W. Xie、D. C. Barratt、T. Vercauteren 和 J. A. Noble，“采用空间条件生成对抗网络的自由手持超声图像模拟”，载于分子成像、运动器官重建与分析，以及卒中成像与治疗。Springer，2017 年，第 105–115 页。[14] F. Tom 和 D. Sheet，“使用具备对抗学习的深度生成网络模拟病理真实感超声图像”，载于*2018 年第 15 届 IEEE 生物医学成像国际研讨会（ISBI 2018）*。IEEE，2018 年，第 1174–1177 页。

[15] B. Liu、H.-D. Cheng、J. Huang、J. Tian、X. Tang 和 J. Liu，“基于概率密度差的超声图像分割主动轮廓”，模式识别，第 43 卷，第 6 期，第 2028–2042 页，2010 年。[16] M. Talebi、A. Ayatollahi 和 A. Kermani，“使用遗传主动轮廓进行医学超声图像分割”，《生物医学科学与工程杂志》，第 4 卷，第 2 期，第 105 页，2011 年。[17] O. Ronneberger、P. Fischer 和 T. Brox，“U-Net：用于生物医学图像分割的卷积网络”，载于医学图像计算与计算机辅助干预国际会议。Springer，2015 年，第 234–241 页。[18] G. Zhang、K. Zhao、Y. Hong、X. Qiu、K. Zhang 和 B. Wei，“Sha-mtl：用于自动乳腺癌超声图像分割与分类的软硬注意力多任务学习”，《国际计算机辅助放射学与外科学杂志》，第 16 卷，第 10 期，第 1719–1725 页，2021 年。[19] M. B. Shuvo、R. Ahommed、S. Reza 和 M. Hashem，“Cnl-unet：一种用于多模态生物医学图像分割并具有错误输出抑制功能的新型轻量级深度学习架构”，《生物医学信号处理与控制》，第 70 卷，第 102959 页，2021 年。[20] H. Pan、Q. Zhou 和 L. J. Latecki，“Sgunet：用于甲状腺结节分割的语义引导 U-Net”，载于*2021*年 *IEEE* 第*18*届生物医学成像国际研讨会（*ISBI*）。IEEE，2021 年，第 630–634 页。[21] K. Raum、J. Reißhauer 和 J. Brandt，“使用时间分辨扫描声学显微镜进行皮质骨各向异性阻抗估计的频率与分辨率依赖性”，《生物医学材料研究杂志*A*辑：生物材料学会、日本生物材料学会、澳大利亚生物材料学会和韩国生物材料学会官方期刊》，第 71 卷，第 3 期，第 430–438 页，2004 年。[22] M. Lyons 和 K. Parker，“软组织中的吸收与衰减：II. 实验结果”，《*IEEE*超声学、铁电学与频率控制汇刊》，第 35 卷，第 4 期，第 511–521 页，1988 年。[23] P. R. Hoskins，“与动脉超声成像和血流速度测量相关的组织物理特性”，《医学与生物学中的超声》，第 33 卷，第 10 期，第 1527–1539 页，2007 年。[24] I. Hacihaliloglu，“骨表面的超声成像与分割：综述”，《技术》，第 5 卷，第 02 期，第 74–80 页，2017 年。[25] S. Mehta 和 M. Rastegari，“MobileViT：轻量级、通用且适用于移动设备的视觉Transformer”，国际学习表征会议，2022 年。[26] X. Ding、X. Zhang、Y. Zhou、J. Han、G. Ding 和 J. Sun，“将卷积核扩展至 31×31：重新审视卷积神经网络中的大卷积核设计”，国际学习表征会议，2022 年。[27] G. Carbajal、A. Lasso、L. G´omez 和 G. Fichtinger，“改进基于 N线模体的自由手持超声校准”，载于计算机辅助放射学与外科学，2013 年。