# MedSRGAN: 使用生成对抗网络的医学图像超分辨率

# 摘要

由于需要使用有限的辐射剂量获取高质量图像，例如低剂量计算机断层扫描（CT）和低场磁共振成像（MRI），医学成像中的超分辨率（SR）成为医学成像中的一个新兴应用。然而，由于医学图像的复杂性和更高的视觉要求，SR在医学成像中仍然是一个具有挑战性的任务。在本研究中，我们开发了一种基于深度学习的方法，称为使用生成对抗网络的医学图像SR（MedSRGAN），用于医学成像中的SR。我们开发了一种新颖的卷积神经网络，称为残差全图注意力网络（RWMAN），作为我们MedSRGAN的生成器网络，通过不同通道提取有用信息，并更加关注有意义的区域。此外，在MedSRGAN训练期间，我们将内容损失、对抗损失和对抗特征损失的加权和融合成一个多任务损失函数。我们收集了242个胸部CT扫描和110个脑部MRI扫描用于训练和评估MedSRGAN。结果表明，MedSRGAN不仅保留了更多的纹理细节，而且在重建的SR图像上生成了更真实的模式。对由五位经验丰富的放射科医生评分的CT切片进行的平均意见分数（MOS）测试证明了我们方法的有效性。

关键词 医学图像 · 超分辨率（SR） · 深度学习 · 生成对抗网络（GAN）

# 1 引言

医学图像，包括计算机断层扫描（CT）、磁共振成像（MRI）、正电子发射断层扫描（PET），广泛用于临床应用，如非侵入性疾病诊断、解剖成像和治疗计划[2, 7, 17, 24, 25, 32, 37, 38, 43, 47]。然而，这些成像技术存在一些局限性。例如，CT扫描期间不可避免地会发生辐射损伤。低剂量CT（LDCT）是一种临床推荐的减少患者辐射损伤的技术，但代价是图像分辨率低和噪声污染。MRI以其提供人体各部位的解剖、代谢和功能信息的能力而闻名。然而，MRI的主要缺点之一是采集时间长，这使得图像质量容易受到患者运动的影响。为了获得更高的信噪比（SNR），其空间分辨率往往比CT图像更粗糙。此外，低磁场MRI扫描仪也限制了MRI的空间分辨率。PET能够通过成像体内不同部位标记化合物的不同吸收能力，准确检测早期肿瘤。PET的成像原理是基于检测放射性同位素湮灭反应产生的光子对的数量。由于光子对的稀缺和噪声干扰，低分辨率问题也是PET成像的一个挑战。

为了在低辐射剂量和技术限制下提高图像质量，一种解决方案是基于低分辨率（LR）图像检索图像细节以重建高分辨率（HR）图像[12, 29]，这一过程被称为超分辨率（SR）。尽管临床获得的LR医学图像包含真实和基本信息，但它们失去了许多细节，包括高频成分、某些模式（如组织背景和纹理），这些细节对于重建HR图像极为必要。因此，医学图像的SR仍然是一个重大挑战，并且在实践中很少使用。在这项工作中，我们提出了一种基于深度学习的新颖医学图像SR框架，并展示了其有效性。

## 1.1 相关工作

近年来，提出了大量基于神经网络的算法用于自然图像的SR。根据不同的目的，这些方法主要可以分为两大类：高峰值信噪比（PSNR）/结构相似性（SSIM）和高感知质量。具有更高PSNR/SSIM的方法主要通过网络结构创新实现，例如更深的网络、各种特征连接和网络内部特征流动策略[8, 26, 27, 33, 41, 49]。借助均方误差（MSE）损失，它们在PSNR/SSIM指标上取得了显著成果，但SR图像仍然与人类观察相矛盾，因为人眼擅长捕捉高频成分，而不是像素级差异。另一方面，具有更高感知质量的方法[30, 45]主要通过感知损失[22]和GAN[11]实现。这些方法在感知指标上表现良好，但在PSNR/SSIM指标上有所妥协。

基于深度学习的医学图像恢复研究近年来逐渐受到医学成像研究社区的关注，这些研究包括不同成像系统之间的图像转换[1, 19]、超分辨率[3, 4, 31, 34, 35, 39, 48]和去噪[1, 10, 20]。

### 1.1.1 自然图像的高PSNR/SSIM

Dong等人[8]提出了用于单图像SR（SISR）的超分辨率卷积神经网络（SRCNN），这是SISR问题中第一个卷积神经网络（CNN）研究。SRCNN首先使用双三次插值将输入图像放大到目标尺寸，然后将其输入到一个具有三个卷积层的简单网络中。该方法已经能够实现比任何其他传统计算机视觉算法更高的PSNR/SSIM。这项研究表明，端到端学习方法在放大图像时可以有效学习像素级的必要非线性变换。从那时起，深度学习成为自然图像SISR的主流。许多尝试在这一领域进行，以在PSNR/SSIM上实现更高的性能，例如使用递归卷积网络进行特征提取[26]，堆叠更多卷积滤波器并添加长残差连接[27]，以及去除批量归一化层以保留网络的范围灵活性[33]。Shi等人[41]提出了一种用于SR的高效上采样模块，称为像素洗牌模块，它简单地将图像张量重新排列为，以放大图像特征的高度和宽度。此过程不包含可学习参数，并减少了特征通道。研究人员通常在上采样模块之前添加一个额外的卷积层以增加特征通道，这使得上采样过程可学习并保留特征通道。基于这些思想，Zhang等人[49]设计了一个非常深的CNN，带有堆叠的通道注意力模块，称为残差通道注意力网络（RCAN），并在PSNR/SSIM上达到了最先进的水平。

所有这些方法都被认为是面向PSNR的方法，因为它们在网络训练期间仅将均方误差（MSE）或平均绝对误差（MAE）作为损失函数，这是一个直观的想法，以增加PSNR指标。然而，生成的图像仍然会变得过度平滑[30, 45]。由于PSNR通常与人类观察者的视觉评估不一致，这样的结果在医学成像应用中是不可容忍的，可能会阻碍相关临床应用，如疾病诊断和病变检测。

### 1.1.2 自然图像的高感知质量

为了在SISR中恢复可行的高频成分并获得更好的视觉结果，研究人员应用了特征级损失[22]并使用GAN[11]使图像尽可能真实。特征级损失，也称为感知损失，广泛用于许多图像重建任务，包括像素到像素的图像转换（pix2pix）[18]、图像风格迁移[9, 21]和超分辨率[30, 45]。它使用预训练的神经网络分别提取图像输出和地面真值目标的特征，然后在这些特征上计算损失，以指导网络从目标图像中学习语义特征。GAN最初用于假图像合成[11]。它通常包含两个具有不同参数的独立神经网络，一个用于图像生成，另一个用于真实/假图像的判别。判别网络的目标是判断图像是否真实或由生成网络生成，即假的，而生成网络的目标是判别网络目标的对抗形式，因此它使假图像接近真实图像。在训练过程中，生成网络逐渐缩小生成图像与真实图像之间的差距，直到判别网络无法区分差异。

Ledig等人[30]首次提出了基于GAN的SR框架（SRGAN），该框架包含一个用于高分辨率图像生成的生成器网络，一个用于识别生成图像与真实世界图像的判别网络，以及包含感知损失和GAN损失的损失函数。SRGAN在视觉结果上比之前的面向PSNR的方法取得了更好的效果，尽管它在PSNR/SSIM值上有所妥协。在Wang等人的工作ESRGAN（增强SRGAN）[45]中，他们提出了一种密集连接的[16]残差密集块（RRDB）网络用于生成器，并使用相对判别器[23]来确定图像是否比其他图像更真实，并在PIRM-SR挑战赛（ECCV2018）的高感知质量组中获胜。

### 1.1.3 医学图像的重建工作

上述研究在SISR任务中取得了巨大进展，然而，这些研究都没有考虑医学图像的特性，也没有评估其在医学图像上的性能。由于数据分布、模式和纹理的完全不同，直接将这些模型应用于医学图像SR是不可行的。

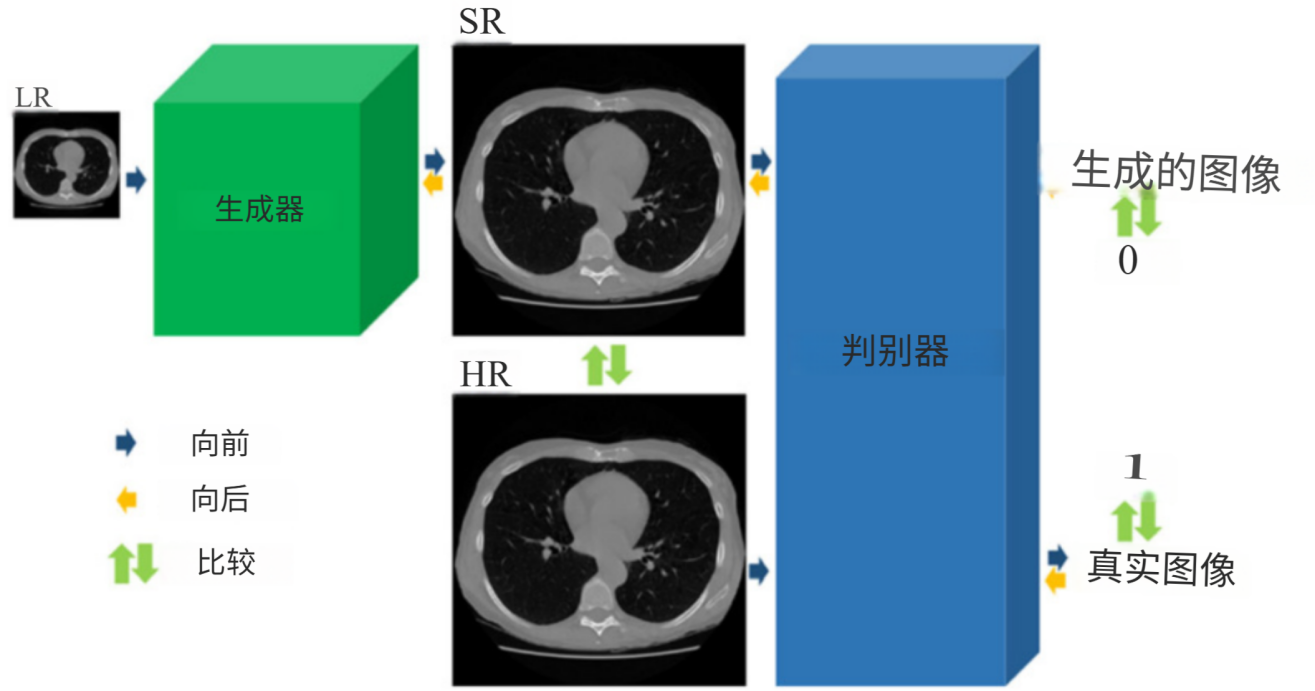
对于医学图像重建工作，Armanoius等人[1]从pix2pix研究中借鉴了思想，并提出了一个基于GAN的框架，名为MedGAN，用于PET-CT转换、MRI运动校正和PET去噪。You等人[48]提出了一个复杂的基于GAN的框架，借鉴了循环GAN[50]的思想，用于单切片CT SR，并且需要独立的LR和HR数据集。由于在同一患者上收集更多CT扫描会造成不必要的辐射损伤，他们仅使用了一个来自新鲜冷冻尸体踝样本的胫骨数据集和一个包含少量扫描的公共腹部数据集进行实验。此外，他们仅关注非常小的局部补丁而不是整个CT扫描，因此这种方法在临床应用中的实用性有限。对于MRI SR，Chen等人[3, 4]提出了3D密集连接超分辨率网络（DCSRN）来恢复结构脑MRI的HR特征，并进一步开发了一个GAN框架来指导DCSRN训练，以进一步提高SR质量。Mardani等人[35]提出了用于压缩感知的GAN（GANCS），以检索具有改进精细纹理细节的高质量图像。尽管在这些方法中观察到了视觉上可靠的结果，但他们没有评估这些生成的图像是否会被放射科医生接受，也不会影响临床诊断。

## 1.2 我们的贡献

在这项工作中，我们开发了一个使用生成对抗网络（MedSRGAN）的医学图像SR框架，用于重建可靠和视觉上真实的SR医学图像，该框架将LR医学图像作为输入，生成SR图像。在MedSRGAN中，我们采用了一种改进的医学图像生成网络，残差全图注意力网络（RWMAN）作为SR图像生成器，一个新的成对判别器来区分HR/SR和LR图像对，以及一个新颖的多任务损失函数，结合内容损失、对抗损失和对抗特征损失，用于指导SR图像获得更多的可靠性和可行性。我们的方法在242个胸部CT扫描和110个脑部MRI扫描上进行了评估，结果表明我们的方法对疾病诊断的影响很小，这表明基于SR的医学成像系统有可能在实践中应用。

# 2 方法

MedSRGAN是一种典型的基于GAN的架构。如图1所示，它由两个神经网络组成：生成器和判别器，以及每个网络的独立损失函数。判别器的目标是判断图像是否为地面真值或生成图像，而生成器的目标来自判别器和图像内容。

 图1 MedSRGAN的基本框架。不同颜色的箭头定义了数据如何向前流动以进行计算，向后流动以更新参数，以及在何处进行比较以计算损失函数

## 2.1 生成器结构

如图2所示，我们将我们的生成器命名为残差全图注意力网络（RWMAN），它是通过将RCAN[49]的残差通道注意力块（RCAB）替换为残差全图注意力块（RWMAB）而修改的。RWMAN在LR尺度上提取丰富的特征并在这些步骤之后进行上采样。它由一个输入卷积、带有长和短残差连接的RWMAB组以及一个包含两个Conv-Pixel Shuffle块的上采样模块组成。

在RCAN[49]中，RCAB首先使用全局平均池化将整个特征图通过所有通道压缩成一个单一值，然后在后续层中使这些值可学习，最后将它们作为先前通道的权重，这一过程被视为通道注意力。注意力机制在许多NLP任务中提高了RNN模型（如LSTM和GRU）的性能[6]，现在广泛用于其他不同领域，包括推荐系统[5, 13]和计算机视觉[15, 44, 46]。一般来说，这种机制通过在当前特征上添加一个子网络来适应性地应用权重，通道注意力通过特征通道学习权重。这一步骤帮助神经网络通过不同通道适应性地学习如何明智地使用信息，但它忽略了图像不同区域的注意力。通常，自然图像的整个地图应该被同等关注以重建更高分辨率的图像，因为所有像素可能反映现实世界中的有意义信息。然而，对于许多医学图像，只有那些包含有用信息的区域才应该被明智地考虑，周围的区域（如空气和空旷区域）是无意义的。我们期望神经网络不仅通过不同通道提取有用信息，而且更加关注有意义的区域。

在RWMAB中，我们使用卷积层获得与输入图像张量形状相同的张量，然后使用sigmoid激活将其范围从0到1形成所有通道的每个像素的权重。这种子网络结构帮助我们的模型适应性地放大或降低每个像素的影响，以使我们的基于GAN的损失函数产生影响。我们堆叠了128个这样的神经网络块，带有长（从开始到结束）和短（每16个块）残差连接，以减少训练难度[14]。

在上采样模块中，我们使用两个Conv-Pixel Shuffle[41]模块生成图像特征，保持特征通道数量，然后通过卷积层减少通道以生成输出图像。

此外，在将LR图像输入神经网络之前，引入了一个均值为零、单位方差的随机高斯噪声作为附加通道进行扰动，如图2所示。这是一个可选操作，在本研究中用’+’表示。例如，带有附加噪声通道的MedSRGAN表示为。将噪声和有意义信息作为输入的GAN框架被视为标准条件GAN[18, 36]。通过这样做，神经网络内部的特征图提供了一些随机性，这可能有助于神经网络在生成同质区域（特别是组织背景）中的更多可行模式时更具适应性，而图像的主要内容仍来自输入LR图像。

为方便起见，由我们的神经网络生成的与相应LR图像对应的HR图像在下文中称为SR（超分辨率）图像：

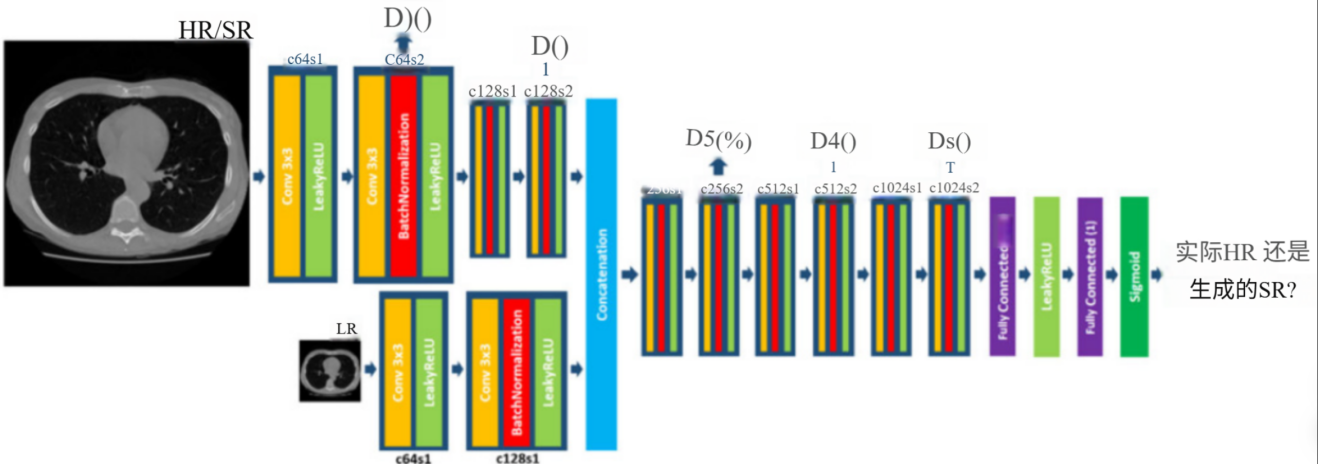
其中是给定生成器的操作。在本研究中，HR表示地面真值图像，而SR表示生成图像。

## 2.2 判别器结构

我们使用图像对（LR，HR/SR）作为判别器的输入，而不是单个HR/SR图像，以区分给定LR图像的SR图像。图3显示了我们的判别器架构。通过这种设计，判别器通过连接从LR和HR路径提取的特征图来学习HR/SR和LR图像对的成对信息，并输出（LR，HR）或（LR，SR）对为真实对的概率。

以CT切片对（LR，HR）为1和（LR，SR）为0，判别器的训练阶段表示为：

其中是判别器的操作。

 图3 判别器结构。对于每个卷积层，通道（c）和步幅（s）在上文指出，所有LeakyReLU在负轴上使用。，表示我们在这些层之后输出特征以计算对抗特征损失

## 2.3 损失函数

在训练期间，生成器和判别器在每次迭代中交替训练不同的数据，以避免模型参数陷入局部最优。在本研究中，使用二元交叉熵（BCE）损失训练判别器：

其中是判别器的操作。

生成器训练的损失函数是内容损失、对抗损失和对抗特征损失的加权和：



其中和是可调的超参数，用于平衡每部分损失的影响。

### 2.3.1 内容损失

这部分损失主要负责图像内容的恢复。通常，为了获得更高的PSNR，HR和SR之间的内容损失可以简单地设置为L1损失或MSE损失：

其中表示图像的大小，假设每个图像为乘。

如等式（5）和（9）所示，最小化L1或MSE是面向PSNR的优化，但它无法重建高频内容并保留视觉真实特征[30, 45]。对于医学图像，纹理特征非常重要，因为这些高频信息在人类视觉判断图像质量时起决定性作用。因此，仅使用面向PSNR的损失函数不足以用于医学图像SR。

一个预训练良好的神经网络，如VGG-16或VGG-19[42]，已被用于在隐藏层提取纹理特征以计算内容损失，在许多任务中包括图像风格迁移[9, 21]、图像转换[18]和图像SR[22, 30, 45]。在本工作中，我们使用预训练的VGG-19来计算内容损失。我们不仅在一个VGG隐藏块上获得损失，而是使用五个隐藏块在每个语义级别上充分提取纹理特征，并结合损失，在较浅的块上使用更多权重，在较深的块上使用较少权重。这是因为较深的特征带来更多的抽象语义信息，而较浅的特征显示相对更多的具体纹理特征信息，如图4所示。用表示VGG-19网络在第i个最大池化层之前的第j个卷积操作，我们的内容损失定义为：

其中是可调参数，用于平衡HR和SR图像之间的直接L1损失，权重和层数用于五个隐藏块，和分别表示两个特征的L1范数（平均绝对误差）和均方误差的计算。根据ESRGAN[45]中VGG特征图的分析，我们在所有隐藏块的ReLU激活之前使用特征图，以避免ReLU激活后的信息丢失。

### 2.3.2 对抗损失

对抗损失旨在帮助生成器使SR图像尽可能接近真实的HR图像，以欺骗判别器。对抗损失具有与相反的形式，旨在使接近1：

其中是判别器的操作。

### 2.3.3 对抗特征损失

与基于VGG的内容损失类似，我们还将我们的判别器视为特征提取器，并计算HR和SR图像以及（HR，LR）和（SR，LR）对之间的隐藏块损失。图像示例如图5所示。最小化对抗特征损失是为了最小化判别器所看到的纹理特征差异，从而帮助生成器从判别器特征中受益，生成更真实的模式。

用表示图3中判别器内部的第个隐藏块，我们的对抗特征损失定义为：

其中权重用于五个隐藏块，表示两个特征的均方误差的计算。

# 3 实验

在本研究中，MedSRGAN和MedSRGAN与双三次插值和基于CNN的SR框架（包括RCAN[49]、ESRGAN[45]和面向PSNR的RWMAN（在实验中表示为RWMAN（P）））进行了比较。RCAN是一种面向PSNR的方法，在几个典型的SISR数据集（如Set5、Set14和Urban100）上实现了最先进的PSNR和SSIM，ESRGAN在PIRM-SR挑战赛（ECCV2018研讨会，一个超分辨率竞赛）的高感知质量组中获胜，这意味着它们在高频成分重建上表现最佳，并获得了最高的感知指数。RWMAN（P）是使用像素级MSE损失训练的RWMAN。

## 3.1 评估

PSNR和SSIM仍用于所有方法的SR框架评估，SR社区缺乏可靠的客观指标来模仿人类观察的判断。PSNR和SSIM计算如下：

其中 是灰度最大值；表示均值和方差，是两个图像的协方差；两个常数，按SSIM的惯例计算。

由于PSNR和SSIM通常与个体视觉判断相矛盾，我们还进行了平均意见分数（MOS）测试，以定性评估SR图像的视觉质量。五位经验丰富的放射科医生被要求对包含100个随机选择的HR CT切片图像和由MedSRGAN和其他参考方法生成的相应SR图像的数据集进行评分，评分范围为1到5，分别表示非常令人烦恼、令人烦恼、稍微令人烦恼、可察觉但不令人烦恼、不可察觉。

## 3.2 CT实验

### 3.2.1 数据和训练细节

我们使用了LUNA 16挑战赛[40]中的242个胸部CT扫描进行实验。随机选择了219个扫描（52,073个切片）用于训练，剩余的23个扫描（5889个切片）用于测试。所有CT切片均为，并用作HR参考。LR切片通过对相应的HR切片应用平均池化获得，大小为。此外，从中国一家医院收集了30个额外的胸部CT扫描（10,732个切片）用于测试MedSRGAN在临床数据上的表现。

所有网络将 CT切片作为输入，并输出 CT切片。我们的网络是完全卷积的，因此它们可以将任意大小的CT切片作为输入并生成切片。LR和HR CT图像的强度被裁剪到 HU（Hounsfield Unit），然后线性缩放到[0, 1]范围。为了训练生成器，我们将，和用于损失函数（4），并使用Adam[28]优化器，和。我们将内容损失的权重设置为比对抗损失大两个数量级，这种设置也应用于基于GAN的方法，如SRGAN和ESRGAN，我们在实验中调整了和，但对生成的图像影响有限。初始学习率设置为，每次迭代减半。这些实验在两块NVIDIA Titan GPU上进行。

### 3.2.2 MOS测试

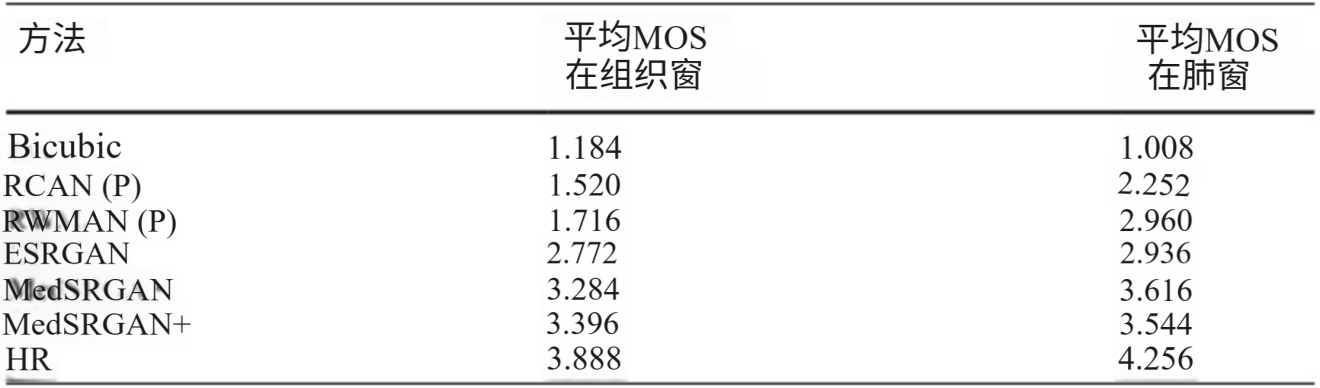
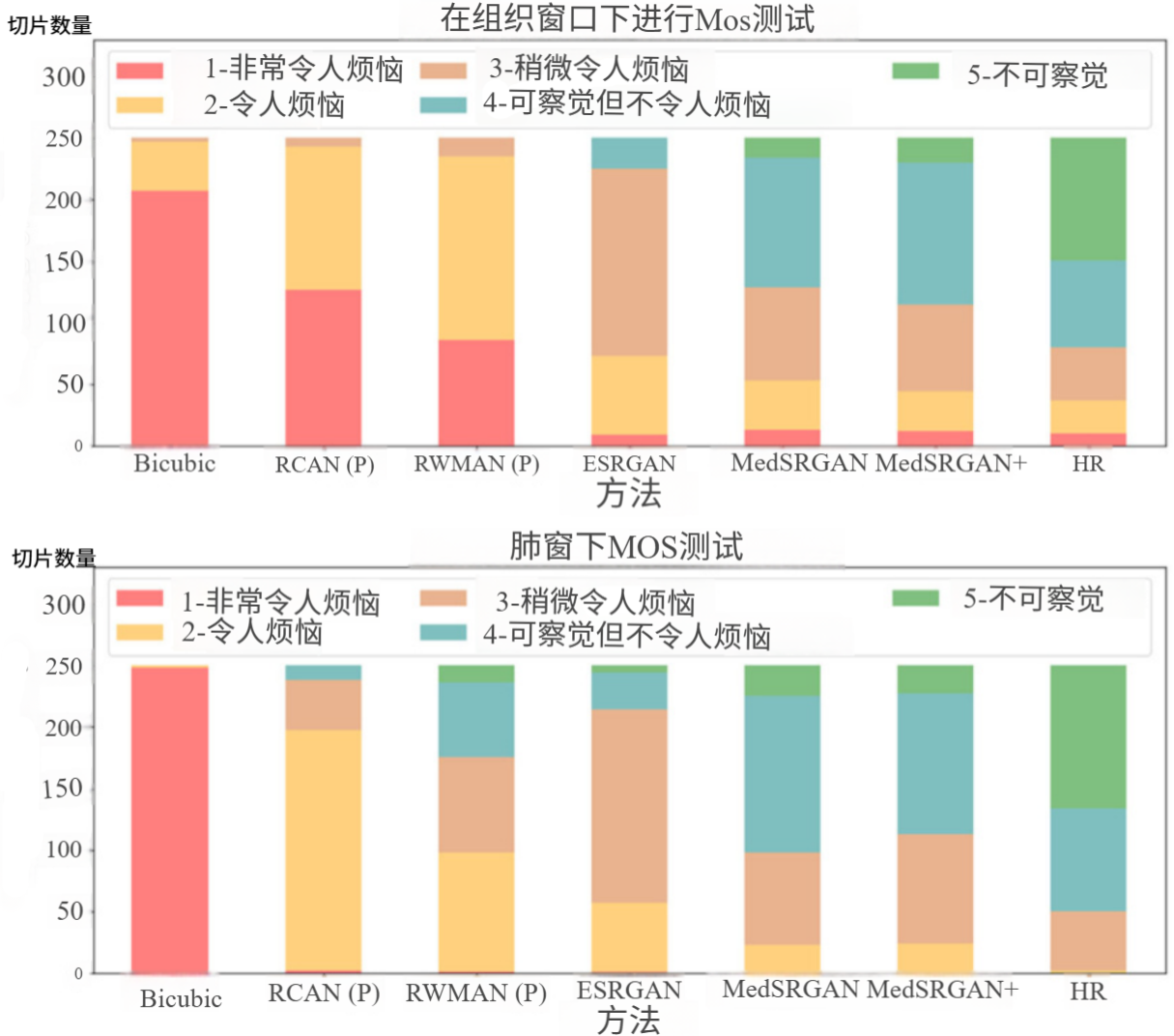
医学图像是放射科医生进行诊断和治疗决策的核心信息部分。医学图像感知的科学致力于理解和改进临床解释过程。在本研究中，我们专注于基于内容的主观视图评估，即我们的MOS测试，来自五位经验丰富的放射科医生在模拟临床环境中的评估。从我们的测试集中随机选择了100个CT切片，形成MOS测试数据集。在这些切片图像中，一半使用 HU显示窗口（组织窗口），另一半使用 HU（肺窗口），这是胸部CT查看的两个重要显示窗口。对于每个切片，HR图像（原始图像）和由包括双三次插值、RCAN（P）、RWMAN（P）、ESRGAN、MedSRGAN和在内的方法获得的SR图像被评分，因此我们的MOS数据集包含个CT切片图像。这些图像被随机打乱，并确保不包含任何可能暗示方法的信息，以避免主观偏见，然后由五位经验丰富的放射科医生评分。结果及其评分分布如表1和图6所示。

### 3.2.3 结果分析

使用不同方法对所有测试图像的平均PNSR/SSIM如表2所示，这些结果在将整个 HU显示窗口缩放到[0, 1]后计算。结果显示，使用基于GAN的方法在这个任务中几乎可以实现与面向PSNR的方法相同的PSNR/SSIM性能。

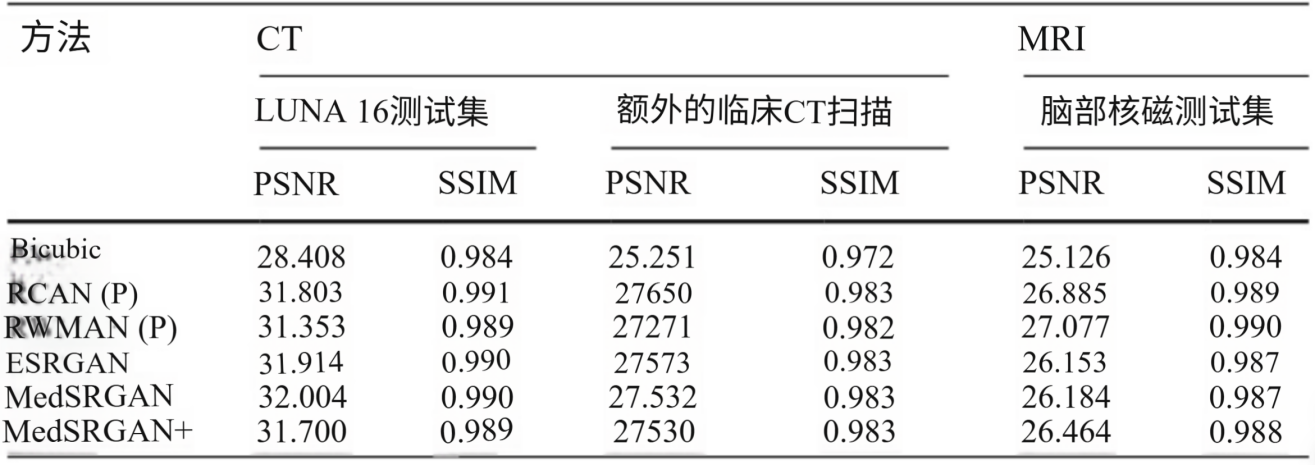
图7显示了图像示例以及PSNR/SSIM值。我们观察到，面向PSNR的RCAN生成的图像虽然具有更高的PSNR/SSIM，但过于平滑，而ESRGAN生成的图像（图7的第一行）在视觉上更不自然，纹理（图7的第二行）不如MedSRGAN生成的图像明显。此外，ESRGAN的图像仍然存在严重的棋盘伪影问题，这表明密集连接的卷积块可能不适用于CT SR。总之，提出的MedSRGAN在重建这些缺失信息方面表现更好，其生成的图像在视觉上更可行。

此外，不同方法在两个显示窗口下的MOS测试结果如表1和图6所示。我们观察到，在组织窗口和肺窗口下，我们的方法在视觉判断上的平均MOS优于其他方法，无论是否使用噪声通道。具体来说，从图6的第一个图表中我们发现，带有噪声通道输入的模型在组织窗口显示时帮助生成器生成更自然的伪影和模式，因此得分更高，但在肺窗口显示时（见图6的第二个图表）视觉质量较低。因此，在使用噪声通道或不使用噪声通道之间可能存在生成自然模式和恢复纹理的权衡。

表1 CT MOS试验结果。方法名称后的(P)表示该模型是在PSNRoriented loss （MSE loss）上训练的，后缀“+”表示包含了噪声信道。黑体字和下划线表示最佳和次佳表现，斜体表示人力资源图像得分   
  
图6组织窗和肺窗的CT MOS检测。方法标题后的(P)表示该模型是在面向psnr的损失（MSE损失）上训练的，后缀 ‘+’ 表示噪声通道被包括在内

具体来说，从图6第一张图的结果来看，我们发现噪声通道输入的模型在组织窗口显示时可以帮助生成器生成更多自然的伪像和图案，从而获得更高的分数，但在肺窗口显示时遇到的视觉质量较差（见图6第二张图）。因此，在生成自然图案和使用噪声通道恢复纹理之间可能存在权衡。

表2 我们实验中所有测试图像的平均PSNR和SSIM。模型在“LUNA 16测试集”和“临床CT数据”中使用LUNA 16训练集进行训练，如第3.2.1节所述。“脑MRI”中的模型使用脑MRI训练数据进行训练，如第3.3.1节所述。（P）表示该模型在面向PSNR的损失（MSE损失）上进行训练，后缀表示包含噪声通道。“PSNR”和“SSIM”列中的粗体和下划线表示在PSNR和SSIM上的最佳和次佳性能



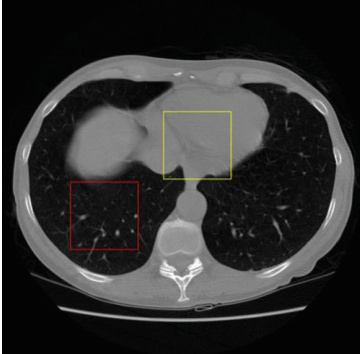
### 3.2.4 消融研究

由于我们的方法以面向PSNR的RCAN作为基线模型，我们修改的有意义的影响值得比较分析。进展如表3所示，简要讨论如下。

替换RCAB为RWMAB 与自然图像相比，医学图像中的一些区域在生成高分辨率图像时可能值得关注。我们使用RWMAB进行全图像特征图注意力，以使网络能够关注那些有意义的区域。为了更好地展示其效果，我们训练了一个面向PSNR的RWMAN进行比较，该模型仅使用像素级MSE损失进行训练。我们观察到，RWMAN（P）在PSNR和SSIM上略有下降，但在MOS测试中有所提高。

使用基于GAN的框架 如前所述，尽管面向PSNR的RWMAN在PSNR和SSIM上表现不如RCAN，但在MOS测试中表现更优。然而，使用面向PSNR的RWMAN生成的图像往往过于平滑，缺乏高频细节。因此，引入了基于GAN的框架和损失函数（4），以允许生成器从对抗学习中获得增强，并生成更真实的结果。

添加输入噪声通道 通过对抗学习，MedSRGAN生成的SR图像更接近真实的HR图像，但图像在模式生成上仍然单调，特别是在组织窗口下显示图像时。为了解决这个问题，我们引入了一个噪声通道作为附加输入，以在网络训练中引入一些随机性。我们观察到，带有噪声输入的MedSRGAN（MedSRGAN+）在组织窗口显示下的MOS测试中略有增加。



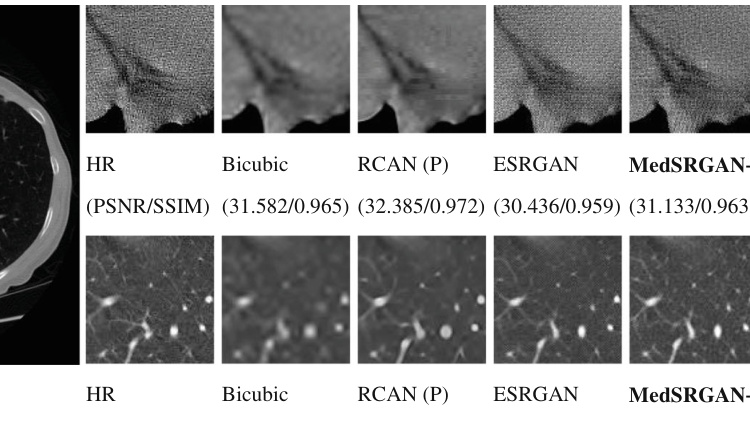
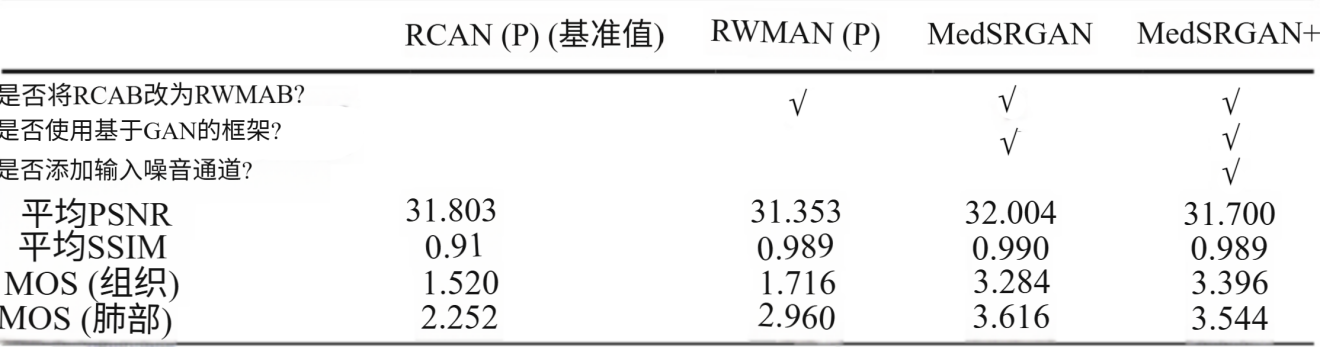
 (PSNR/SSIM) (29.931/0.830) (31.074/0.884) (30.167/0.864) (30.066/0.868)

图7 不同方法在切片上裁剪的图像补丁的结果。（P）表示该模型在面向PSNR的损失（MSE损失）上进行训练，后缀表示包含噪声通道。黄色框中的图像补丁显示在第一行，其显示窗口为，240] HU（组织窗口），红色框中的图像补丁显示在第二行，其显示窗口为，0] HU（肺窗口）。面向PSNR的RCAN生成的图像过于平滑，尽管其PSNR/SSIM较高。对于基于GAN的方法，ESRGAN生成的图像仍然存在棋盘伪影问题，而MedSRGAN+摆脱了这些烦人的伪影，并且在模式和纹理上更接近HR图像。最佳查看方式为屏幕放大

表3 我们实验中的消融研究。（P）表示该模型在面向PSNR的损失（MSE损失）上进行训练，后缀’+’表示包含噪声通道。粗体和下划线表示最佳和次佳性能。此表显示了我们修改在视觉质量上的进展，特别是在MOS测试中



### 3.2.5 结节检测测试

我们认为，重建良好的SR图像不应影响现实世界的诊断。为了更好地展示诊断效果，我们在一个复杂的计算机辅助检测（CAD）系统上对HR CT扫描和MedSRGAN+生成的SR CT扫描进行了结节检测测试。该检测系统经过可靠的肺结节注释训练，并在实践中使用。测试集如第3.2.1节所述，包括23个扫描，其中35个注释的肺结节来自LUNA 16。结节直径的分布和检测结果如图8所示。总的来说，CAD在HR扫描上报告了165个结节，其中31个为真阳性（TP）结节，在SR扫描上报告了156个结节，其中32个为TP，这表明重建的SR图像在这个测试中几乎不影响结节检测。

### 3.2.6 临床数据结果

MedSRGAN框架也在中国一家医院的30个额外临床CT扫描（10,732个切片）上进行了测试，这些数据由TOSHIBA Aquilion CT扫描仪获取，大小为。这些数据可能面临与LUNA16[40]数据集不同的数据分布问题，因为不同的CT扫描仪可能产生不同的噪声模式或其他伪影。然而，我们的方法仍然能够生成具有高视觉质量的图像，如图9所示，其中包含HR参考和不同方法生成的结果。图9的第一行显示在 HU（组织窗口）下，第二行显示在 HU（肺窗口）下。使用不同方法对所有10,732个切片进行SR的平均PNSR/SSIM也收集在表2中。尽管MedSRGAN与其他方法（除了双三次插值）在平均PSNR和SSIM上只有微小差异，但MedSRGAN在重建这些数据的图案和纹理信息方面优于任何其他方法，并且我们的结果在视觉上更接近真实图像，如图9所示。

## 3.3 MRI实验

### 3.3.1 数据和训练细节

为了展示我们方法的灵活性，我们从中国一家医院收集了110个脑部MRI扫描用于MRI实验。所有MRI扫描由PHILIPS Ingenia MR系统获取，大小为，80个扫描（7,661个切片）用于训练，其余30个扫描（2,118个切片）用于测试。

MRI的成像方法与CT有很大不同，每个体素在MRI中的值没有特定的物理意义，因此在训练和测试之前对每个MRI扫描进行了零均值归一化（减去均值并除以标准差）。LR MRI切片通过对原始HR切片进行平均池化获得。我们将 ，和用于损失函数（4），并使用Adam[28]优化器，和。初始学习率设置为，每次迭代减半。

### 3.3.2 结果

使用不同方法对所有脑部MRI测试图像的平均PNSR/SSIM如表2所示。所有指标在裁剪图像上计算，以消除非身体区域的影响，这会使PSNR/SSIM非常高但无意义。定量结果表明，除双三次插值外，所有方法在PSNR/SSIM上表现相似，尽管RWMAN（P）的值最高。

图10显示了从原始MRI图像中裁剪的一些图像补丁及其PSNR和SSIM。我们观察到，面向PSNR的RCAN生成的SR图像虽然具有高PSNR/SSIM，但过于平滑，但在某些补丁中，MedSRGAN和MedSRGAN+的PSNR和SSIM高于RCAN。ESRGAN生成的图像仍然存在较少的可行伪影，而在重建这些细节方面表现更好，结果在视觉上更接近真实的HR图像。

## 3.4 效率和参数计数

我们对这些方法的运行速度进行了实验，结果如表4所示。注意，RCAN[49]在其实验中使用了10个残差组，每个组包含20个RCAB，因此我们在实验中使用相同的配置训练RCAN（P）和RWMAN（P）。参数计数和存储也收集在表4中。

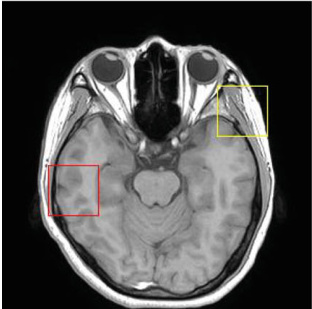
我们发现，1）MedSRGAN框架和ESRGAN在GPU机器上生成图像几乎是实时的；2）RCAN在运行速度上没有优势，这可以用其RCAB结构中的额外通道减少和增加卷积来解释，而我们的网络不包含这样的步骤；3）ESRGAN在速度上略优于MedSRGAN，但参数计数更大。总之，我们的方法不仅在视觉结果上表现突出，而且在速度和参数计数方面也具有优势。

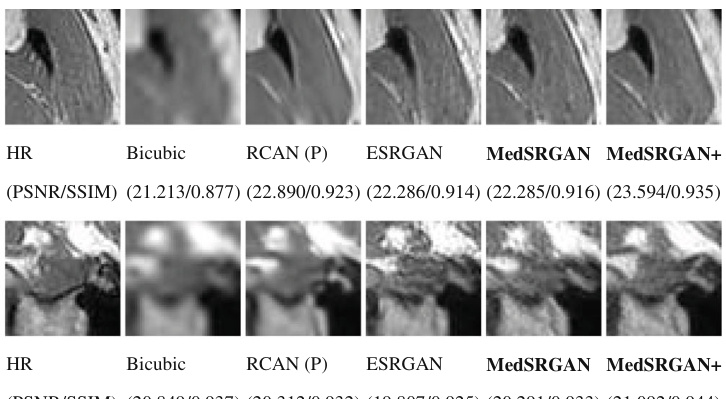
# 4 讨论

在本研究中，我们提出了MedSRGAN作为一个端到端的框架，用于医学图像超分辨率任务。MedSRGAN由一个用于生成超分辨率医学图像的新型神经网络、一个成对的判别器和一个基于GAN的新型损失函数组成。该框架在很大程度上确保了重建图像的内容和细节的充分性以及真实性。

然而，我们想指出的是，MedSRGAN的超分辨率图像仍然不完美，因为仅基于低分辨率图像重建高分辨率图像中的所有信息是不可能的，尽管MedSRGAN已经解决了模糊、过度平滑问题和奇怪的伪影。图像生成的超分辨率实际上是一个欠定问题，这意味着输出超分辨率图像包含比其对应的低分辨率图像更多的信息。为了解释这些额外信息来自何处，我们回顾了整个实验过程。在MedSRGAN框架的训练阶段，每个LR-HR对都被视为一个训练样本，并独立地输入到训练过程中，使用一个损失函数来确保生成的超分辨率图像尽可能接近高分辨率图像。







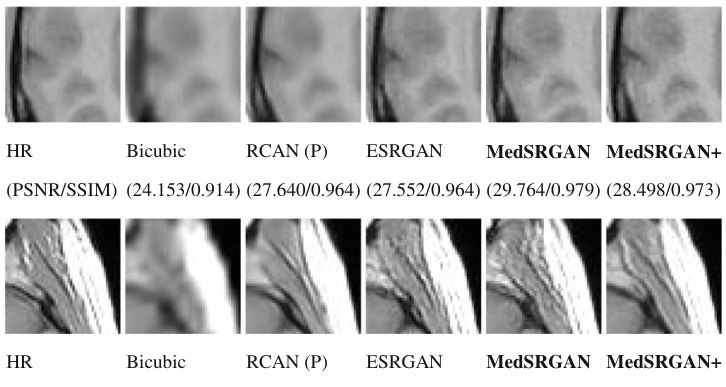
  
(PSNR/SSIM) (18.823/0.931) (22.011/0.968) (21.117/0.961) (20.077/0.950) (21.562/0.965)

图10 不同方法在从原始MRI切片中裁剪的 图像块上的结果。(P)表示该模型在PSNR导向的损失（MSE损失）上进行训练，后缀 表示包含噪声通道。显示窗口被裁剪到[0, 480]。红色框中的图像块显示在第一行，黄色框中的图像块显示在第二行。PSNR导向的RCAN虽然具有高PSNR/SSIM，但带来了过度平滑的图像。对于基于GAN的方法，我们方法生成的图像在图案和纹理上更接近高分辨率图像。这些结果表明，我们的模型在生成可行图案方面表现更好。最佳查看方式是放大屏幕。

因此，一般缺失的信息被编码到神经网络中。假设低分辨率图像包含足够的信息供放射科医生进行诊断，这些隐藏在神经网络中的编码信息可以表现为高频边缘、特定伪影和某种医学图像的图案。在特定的医学图像模态中，这些高频边缘、特定伪影和图案实际上

表4 不同超分辨率方法的速度、参数计数和存储情况。(P)表示RCAN在PSNR导向的损失（MSE损失）上进行训练。RCAN (P)使用了10个残差组，每个组包含20个RCAB，如[49]中所提出的。后缀 ‘+’ 表示包含噪声通道。粗体和下划线表示最佳和次佳性能。  


在自然图像中，这些特征通常是单调且刻板的，这使得在机器学习框架中相对容易学习。但必须指出的是，在极小的焦点小于 像素的高分辨率图像中，没有任何方法能够生成这种图像，因为所有关于这些焦点的信息都已丢失。这种情况可能会在我们进行低分辨率CT扫描时发生。

为了找出超分辨率图像是否影响诊断，我们在高分辨率CT扫描和MedSRGAN生成的超分辨率CT扫描上进行了简单的结节检测测试，如第3.2.5节所示。结果表明，该结节检测系统在重建的超分辨率CT扫描上也表现良好，因为它在不同大小的真阳性结节上获得了几乎相同的检测结果。此外，我们的MedSRGAN生成的超分辨率CT扫描在人类视觉上最接近真实的高分辨率扫描，如MOS测试所示。因此，我们的方法在人类视觉和保留结节信息方面至少是有效和可行的。然而，这些评估可能无法涵盖所有诊断情况，我们相信应该进行更多的实验来评估。

# 5 结论

在本研究中，我们提出了一种基于GAN的通用医学图像超分辨率框架（MedSRGAN），其中包括作为生成器的残差全图注意力网络（RWMAN）、用于图像对的判别器以及用于训练生成器的新型损失函数。结果表明，我们的方法可以直接用于CT和MRI图像的超分辨率重建，并且其结果保留了更多的可行纹理细节，并在高分辨率图像上生成了真实的图案。MedSRGAN的可靠和有效结果表明，使用超分辨率方法在临床获得的低分辨率图像（如LDCT、低磁场MR和MR波谱成像）上检索更多图像细节是可行的。未来，我们将对MedSRGAN进行更深入的研究，例如研究一种更复杂的方法来评估超分辨率重建性能。