

## 基于视频多帧融合的医学超声图像超分辨率重建方法

赵冉, 袁家斌, 范利利

引用本文

赵冉, 袁家斌, 范利利. 基于视频多帧融合的医学超声图像超分辨率重建方法[J]. 计算机科学, 2023, 50(7): 143-151.

ZHAO Ran, YUAN Jiabin, FAN Lili. [Medical Ultrasound Image Super-resolution Reconstruction Based on Video Multi-frame Fusion](#) [J]. Computer Science, 2023, 50(7): 143-151.

---

## 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

### [基于图像数据耦合识别的输电线路安全风险评估方法](#)

Image Recognition Method of Transmission Line Safety Risk Assessment Based on Multidimensional Data Coupling

计算机科学, 2023, 50(6A): 220500032-6. <https://doi.org/10.11896/jsjcx.220500032>

### [基于改进CNN-BP的多波束声纳高程数据预测研究](#)

Study on Multibeam Sonar Elevation Data Prediction Based on Improved CNN-BP

计算机科学, 2023, 50(6A): 220100161-4. <https://doi.org/10.11896/jsjcx.220100161>

### [基于多模态特征融合的时间序列异常检测](#)

Anomaly Detection of Time-series Based on Multi-modal Feature Fusion

计算机科学, 2023, 50(6A): 220700094-7. <https://doi.org/10.11896/jsjcx.220700094>

### [注意力特征融合的李生网络目标跟踪方法](#)

Attentional Feature Fusion Approach for Siamese Network Based Object Tracking

计算机科学, 2023, 50(6A): 220300237-9. <https://doi.org/10.11896/jsjcx.220300237>

### [基于改进的ResNeXt网络结构的遥感图像分类](#)

Remote Sensing Image Classification Based on Improved ResNeXt Network Structure

计算机科学, 2023, 50(6A): 220100158-6. <https://doi.org/10.11896/jsjcx.220100158>

# 基于视频多帧融合的医学超声图像超分辨率重建方法

赵冉 袁家斌 范利利

南京航空航天大学计算机科学与技术学院 南京 211106

(zhaoran@nuaa.edu.cn)

**摘要** 医学超声成像是临床诊断中应用最广泛的成像方式之一。目前超声图像普遍存在分辨率和对比度较低的问题,并且成像过程易受噪声污染。图像超分辨率重建技术被广泛用于改善超声图像的质量。然而,已有的研究工作缺乏对超声视频帧之间互补信息的充分利用,因此效果并不理想。针对此问题,提出了一种基于视频多帧融合的医学超声图像超分辨率重建方法。首先,构建了一个基于卷积神经网络的无监督多帧融合模型,该模型通过对连续的多帧图像进行特征融合,得到具有丰富信息的融合特征图像;然后,建立一个基于无数据知识蒸馏的轻量级图像超分辨率重建模型,通过训练融合特征图像得到教师超分辨率网络,利用训练好的教师网络和生成对抗网络获取的训练数据得到轻量级学生超分辨率网络,最终得到高质量的医学超声图像;最后,在大型超声数据集上进行实验,采用两种图像客观评价指标以及图像分类任务进行评估。结果表明,所提方法与8种已有的图像超分辨率重建方法相比,在提高超声图像分辨率的同时,获得了包含更多信息且具有更高对比度的超声图像。此外,所提方法得到的超分辨率图像在分类网络的识别准确率可达到97.30%,明显优于其他方法,可提高临床诊断效率与准确性。

**关键词:** 医学超声图像;多帧融合;超分辨率重建;无数据知识蒸馏;卷积神经网络

中图分类号 TP391

## Medical Ultrasound Image Super-resolution Reconstruction Based on Video Multi-frame Fusion

ZHAO Ran, YUAN Jiabin and FAN Lili

School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

**Abstract** Medical ultrasound imaging is one of the most widely used methods in clinical diagnosis. However, the resolution and contrast of ultrasound images are low, and the noise is serious. Image Super-resolution is widely used to improve the quality of ultrasound images. However, the existing studies lack full use of complementary information in ultrasound video frames, so the results are not ideal. To solve this problem, this paper proposes a super-resolution method of medical ultrasound images based on video multi-frame fusion. Firstly, it builds a multi-frame fusion model based on convolutional neural network for ultrasound images. The model obtains the fused image with rich information by feature fusion of continuous multi-frame images. Then, it builds a lightweight image super-resolution model based on data-free knowledge distillation. The model trains fused feature images to obtain a teacher super-resolution model. To obtain the final high-quality medical ultrasound image, it creates a lightweight student super-resolution model using the trained teacher model and training data from the generative adversarial network. Finally, the proposed method is tested on large ultrasound datasets, using two objective image evaluation indicators and results on image classification to evaluate its performance. The results show that compared with other methods, the proposed method not only improves the resolution of ultrasound images, but also obtains ultrasound images with more information and higher contrast. The recognition accuracy of the super-resolution image obtained by this method in the classification network can reach 97.30%, which is much higher than that of previous approaches and can increase productivity and the precision of clinical diagnoses.

**Keywords** Medical ultrasound image, Multi-frame fusion, Super-resolution reconstruction, Data-free knowledge distillation, Convolutional neural network

到稿日期:2022-07-23 返修日期:2022-11-22

基金项目:国家重点研发计划(2017YFB0802303);国家自然科学基金(62076127)

This work was supported by the National Key Research and Development Program of China(2017YFB0802303) and National Natural Science Foundation of China(62076127).

通信作者:袁家斌(jbyuan@nuaa.edu.cn)

## 1 引言

随着超声成像技术的进步,医学超声成像已成为临床诊断中应用最广泛的成像方式之一。超声成像由于具有安全、无创、便捷、实时等优点,优于计算机断层扫描(CT)、X射线、核磁共振(MRI)等其他医学成像技术,几乎是产科、妇科、腹部、心血管系统、泌尿系统等成像的首选方法。但是医学超声成像是一个复杂的过程,超声图像的质量受到超声波频率等因素的影响。根据超声成像的基本原理,检测人体组织时需要低频回声信号,以此让回声信号在人体内穿透得更深,而要想得到较高分辨率的超声图像则需要高频回声信号。由此可见,成像设备发射信号的频率和成像质量相互矛盾,若要达到较好的诊断效果,超声图像的分辨率和对比度就会降低,从而限制其临床应用。因此,准确有效的医学超声图像质量增强方法已成为近期的研究热点,具有重要的临床应用价值。

为了提高医学超声图像的质量,可以从两个方面考虑。一是从硬件方面改进成像设备,但是这种方式成本较高,并且随着硬件水平的提高,其技术瓶颈很难突破,实际应用中存在很多局限性。目前广泛利用数字计算机技术提高超声图像的质量,将自然图像质量增强方法迁移到超声图像上,然而该方式没有充分考虑超声图像的特殊性,重建出来的超声图像质量并不理想。因此,目前急需设计高效的医学超声图像增强算法,以生成高清晰度的医学超声图像,为医生针对病症实施精准的治疗措施提供帮助。

目前的研究工作中存在两个主要问题:1)超声图像增强算法缺乏对超声视频帧之间互补信息的充分利用;2)超声设备实时成像对时间有较高的要求<sup>[1-2]</sup>。针对上述问题,本文提出了一种基于视频多帧融合以及无数据知识蒸馏的医学超声图像超分辨率重建方法。

本文的主要贡献包括:

(1)针对单张超声图像存在的信息不足的问题,提出了一种基于无监督的多帧融合模型。利用超声视频数据集来解决超声图像超分辨率重建问题,将超声视频截取的相邻连续多帧作为输入得到单张高分辨率超声图像。

(2)在超声图像超分辨率重建中引入无数据知识蒸馏,构建一个轻量级学生超分网络,以此来适应超声设备实时成像对时间的要求。

(3)在超声视频数据集上进行实验,使用两个图像客观评价指标以及图像质量对预训练图像分类器的识别精度,来评估所提方法对提高超声图像质量的有效性,结果表明,本文方法明显优于目前已有的图像超分辨率的重建方法。

## 2 相关工作

### 2.1 医学超声图像超分辨重建

医学超声图像质量增强技术,主要分为两大类:预处理技术和后处理技术<sup>[3]</sup>。预处理技术即对超声成像的硬件设备进行改进,与所涉信号的物理特性相关。后处理增强技术使用信号处理算法来提高超声图像的质量。后处理增强技术主要包括图像去噪、图像去模糊、图像对比度增强、图像超分辨率重建(SR)等图像增强技术<sup>[4]</sup>。其中,图像SR是计算机

视觉中的一项基本任务,随着深度学习的发展,其被广泛应用于医学图像、人脸识别、卫星云图等场景中,取得了显著的成效<sup>[5]</sup>。图像SR是一种基于算法的图像增强技术,旨在超越硬件限制,利用低分辨率图像(LR)来提高成像系统的空间分辨率、图像的对比度和解决图像的噪声问题。

目前,图像SR可分为两种:1)单图像SR;2)多图像SR。单图像SR的输入是一张LR图像,即仅利用一张LR图像重建得到高分辨率图像(HR)。多图像SR的输入是多幅LR,即利用多幅LR重建得到HR。多图像SR的重建过程通常如图1所示。



图1 多图像SR的过程

Fig. 1 Process of multiple images SR

对于医学超声图像SR,目前的研究主要集中于基于深度学习的单图像SR。Kim<sup>[6]</sup>将图像SR经典模型VDSR(Very Deep Super Resolution)用于肝脏超声图像。Sawatn等<sup>[7]</sup>将滤波和SRCNN相结合,使用空间滤波器对超声图像去噪,然后使用SRCNN得到高分辨率的超声图像。近年来,基于深度学习的图像SR取得了很大的进步,但是计算成本也越来越高,且图像的退化过程复杂多样,因此目前关于图像SR的研究主要集中于3个方面。1)轻量级的SR模型。Gao等<sup>[8]</sup>提出了一种轻量级的特征蒸馏交互加权网络(FDIWN),可以灵活有效地与不同尺度的特征交互。2)盲SR。为了模拟图像真实的退化过程,Wang等<sup>[9]</sup>提出了Real-ESRGAN模型,引入高阶退化建模过程,对具有复杂退化的图像具有良好的效果。Luo等<sup>[10]</sup>提出了一种概率退化模型(Probabilistic Degradation Model,PDM),将退化过程当作一个随机变量进行研究,相比于其他SR方法,PDM可以建模更多样化的退化过程。3)将非局部注意(NLA)引入深度SR模型。Xia等<sup>[11]</sup>提出了一种非局部注意力方法(Efficient Non-Local Contrastive Network,ENLCN),用于图像SR,通过放大注意力中的query和key来赋予相关特征更大的权重,增强聚合特征的稀疏性。以上方法在大规模自然图像数据集上可以取得良好的效果,但是,单张超声图像存在信息缺乏的问题,通过上述方法无法达到理想的效果。基于此,本文利用超声成像的特点,引用Li等<sup>[12]</sup>的思想,将图像融合和超分辨技术相结合来提高超声图像的质量。

### 2.2 多图像融合的相关研究

计算机视觉中,图像融合是将来自两张或多张图像的相关信息组合成单张图像的过程,融合图像代表原来的多张源图像的所有重要特征,且融合图像的质量相对于源图像有明显的提高。

目前传统的图像融合方法主要包括基于小波变换<sup>[13]</sup>的图像融合、基于主成分分析<sup>[14]</sup>的图像融合、基于离散余弦变换<sup>[15]</sup>的图像融合和基于等高线变换<sup>[16]</sup>的图像融合等。

随着人工智能的发展,基于深度学习的图像融合方法已成为研究热点。基于深度学习的图像融合方法根据学习方式可分为监督方法和无监督方法。在监督方法中,需要良好的预期输出,即参考图像,来监督融合网络的训练,因此参考



模型分为 3 个部分:特征提取、特征融合以及图像重建。

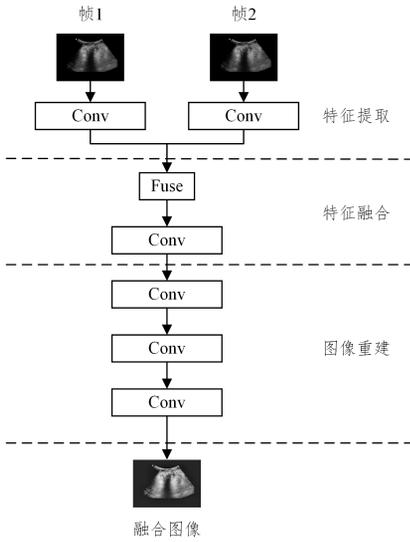


图4 多帧融合模型

Fig 4 Multi-frame fusion model

(1)特征提取模块。本文对 IFCNN 做了改进,采用在大规模数据集 DIV2K 上预训练的 VDSR<sup>[24]</sup> 模型的第 1 个卷积层,其包含 64 个大小为  $3 \times 3$  的卷积核。VDSR 是一种用于超分辨率重建任务的回归模型,它可以广泛地提取图像特征,并且结构相对简单。VDSR 网络结构包含 20 个卷积层,除最后一层外,每一层都包含卷积层和 ReLU 激活函数。本文为了平衡实验复杂性和图像融合的效果,选择了 VDSR 的第 1 个卷积层作为特征提取模块。

(2)特征融合模块。本文采用元素融合规则,将输入的多个卷积特征直接通过最大元素融合规则进行融合,以此融合不同数量的输入图像。元素融合规则包括 3 类:最大元素融合规则、元素和融合规则以及元素均值融合规则。在医学超声图像中,需要获取各帧图像最清晰的特征(元素最大值),提高融合图像的对比度,因此采用最大元素融合规则。

(3)图像重建模块。此模块包括 3 个卷积层,从融合的卷积特征中重建得到融合图像。

### 3.2.3 融合网络损失函数

IFCNN 中,使用大规模且多样化的多焦点图像数据集来有监督地训练图像融合模型。然而,将多帧融合应用于医学超声图像时,缺乏质量参考图像,因此本文引入无参考图像质量度量作为损失函数。本文受到无参考图像质量评价方法 MEF-SSIM 的启发,引入  $L_{UF}$  损失函数。

$L_{UF}$  从对比度、结构信息和亮度 3 个方面对融合图像的质量进行描述,通过输入图像块的对比度、结构信息和亮度来得到期望的融合图像块。

设  $\{x_k\} = \{x_k | k=1, 2, \dots, n\}$  表示从输入的  $n$  张图像的像素位置  $p$  提取的局部图像块,则:

$$x_k = c_k \cdot s_k + l_k \quad (1)$$

其中,  $c_k$  表示像素位置  $p$  提取的图像块的对比度,  $s_k$  表示图像块的结构信息,  $l_k$  表示图像块的亮度信息。

设期望的融合图像块为  $\hat{x}$ , 对比度为  $\hat{c}$ , 结构信息为  $\hat{s}$ , 由于相邻多帧超声图像之间的亮度变化不显著,因此本文在

设计损失函数  $L_{UF}$  时没有考虑亮度信息。对比度越高意味着图像质量越好,因此将  $c_k$  中的最大值作为融合的期望对比度值,则:

$$\hat{c} = \max\{c_k | k=1, 2, \dots, n\} \quad (2)$$

融合图像期望的结构信息值  $\hat{s}$  通过  $s_k$  的加权和获得,即:

$$\bar{s} = \frac{\sum_{k=1}^n \omega_k s_k}{\sum_{k=1}^n \omega_k}, \quad \hat{s} = \frac{\bar{s}}{\|\bar{s}\|} \quad (3)$$

其中,  $\omega_k$  用于决定每个源图像的局部图像块结构在融合图像局部图像块结构中的贡献,基于源图像块的结构之间的一致性来调整  $\omega_k$ 。由于没有考虑亮度信息,因此  $x_k = c_k \cdot s_k$ , 期望的融合图像块  $\hat{x} = \hat{c} \cdot \hat{s}$ 。使用 SSIM 框架计算出融合图像的像素位置  $p$  处图像块  $y$  的图像质量分数为:

$$Score(p) = \frac{2\sigma_{xy} + C}{\sigma_x^2 + \sigma_y^2 + C} \quad (4)$$

其中,  $\sigma_x^2$  和  $\sigma_y^2$  分别表示  $\hat{x}$  和  $y$  的局部协方差。则融合图像的整体质量度量分数为:

$$Q = \frac{1}{M} \sum_{p \in P} Score(p) \quad (5)$$

其中,  $M$  表示图像块总数,  $P$  代表图像所有像素的集合。最终得到医学超声图像融合框架的损失函数为:

$$L_{UF} = 1 - Q \quad (6)$$

计算出的损失函数  $L_{UF}$  反向传播训练网络。

### 3.3 超分辨率重建网络

医学超声成像具有实时性,超声设备上的探头在扫描身体组织部位的过程中实时生成超声图像,超声成像设备对成像时间有较高的要求,因此部署在超声成像设备上的模型对延迟以及计算资源有着严格的限制。为了达到高效的超声检测,本文在融合图像超分辨率重建中引入无数据知识蒸馏,首先使用融合图像训练一个性能良好但网络结构复杂、参数量较大的教师超分网络,然后利用超分辨率网络的基本特性,通过生成对抗网络(GAN)生成和教师网络原始训练集分布相似的训练数据,最后利用训练好的教师超分网络和生成的训练数据,采用渐进式蒸馏方案<sup>[25]</sup> 训练得到轻量级学生超分网络,并将其作为本文的图像超分辨率重建网络,具体实现过程如图 3 中的“超分辨率模型无数据知识蒸馏过程”部分所示。

#### 3.3.1 教师超分网络

教师网络为带有高效的非局部对比注意模块(Efficient Non-Local Contrastive Attention, ENLCA)的 EDSR<sup>[26]</sup> (Enhanced Deep Super-resolution Network)即高效的非局部对比网络 ENLCN。ENLCN 可以通过非局部注意力、稀疏聚合和对比学习机制,充分利用图像内部相关的非局部特征,来关注图像中信息量最大且密切相关的位置。ENLCN 作为超声图像超分辨率的教师网络能够达到良好的效果,有助于医生对中间病变区域的诊断。将 ENLCA 模块嵌入到深度 SR 网络 EDSR 中,构成 ENLCN 网络作为本文的教师超分网络。该网络包含 EDSR 中的 32 个残差块,并且在每 8 个残差块之后插入一个 ENLCA 模块,残差块的通道数为 256, ENLCA 模块的通道数为 64。非局部注意力具体可表示为:

$$\mathbf{Y}_i = \sum_{j=1}^n \frac{\exp(\mathbf{Q}_i^\top \mathbf{K}_j)}{\sum_{j=1}^n \exp(\mathbf{Q}_i^\top \mathbf{K}_j)} \mathbf{V}_j \quad (7)$$

其中,  $\mathbf{Q}_i, \mathbf{K}_j \in \mathbb{R}^c$  和  $\mathbf{V}_j \in \mathbb{R}^{c_{out}}$  分别是特征图  $\mathbf{Q}, \mathbf{K}$  和  $\mathbf{V}$  上位置  $i$  或  $j$  处的像素特征。  $\mathbf{Y}_i \in \mathbb{R}^{c_{out}}$  是位置  $i$  处的输出。为了增加无关特征和相关特征之间的距离,进一步将对比学习引入 ENLCA 模块中。用于训练 ENLCA 的对比学习损失表示为:

$$\begin{aligned} \mathbf{T}_{i,j} &= k \frac{\mathbf{Q}_i^\top \mathbf{K}_j}{\|\mathbf{Q}_i\| \|\mathbf{K}_j\|}, k > 1, \mathbf{T}_{i,j} \in \mathbf{T} \\ \mathbf{T}_i' &= \text{sort}(\mathbf{T}_i, \text{Descending}), \mathbf{T}_i' \in \mathbf{T}', \mathbf{T}_i \in \mathbf{T}_i \\ L_{cl} &= \frac{1}{N} \sum_{i=1}^N -\log \frac{\sum_{j=1}^{n_1} \exp(\mathbf{T}'_{i,j}) / n_1 N}{\sum_{j=n_2}^N \exp(\mathbf{T}'_{i,j} / n_1 N)} + b \end{aligned} \quad (8)$$

其中,  $N$  代表输入大小,  $b$  为常数项,  $n_1$  表示特征图中相关特征和无关特征的百分比,  $n_2$  表示特征图中无关特征百分比的起始索引。  $\mathbf{T}_{i,j}$  通过归一化内积测量  $\mathbf{Q}_i$  和  $\mathbf{K}_j$  之间的相关性。  $\mathbf{T}_i$  和  $\mathbf{T}_i'$  分别表示  $\mathbf{T} \in \mathbb{R}^{N \times N}$  和  $\mathbf{T}' \in \mathbb{R}^{N \times N}$  的第  $i$  行,  $\mathbf{T}_i'$  是  $\mathbf{T}_i$  的降序结果。由此得到 ENLCA 的总损失函数为:

$$\begin{aligned} L_{rec} &= \|\mathbf{I}^{HR} - \mathbf{I}^{SR}\|_1 \\ L &= L_{rec} + \lambda_{cl} L_{cl} \end{aligned} \quad (9)$$

### 3.3.2 学生超分网络

学生网络参照 VDSR。VDSR 原网络采用残差学习,网络结构深度为 20 层,第一层用于处理输入图像,最后一层用于图像重建,中间的每一层都包括卷积层和 ReLU 激活层,通道数为 64,原网络输入经插值放大的低分辨率图像。本文的学生网络将 VDSR 的通道数减半,变为 32,输入为原始的低分辨率图像,在学生网络最后加入子像素卷积层进行上采样,得到放大倍数的输出图像,子像素卷积层解决了 VDSR 输入插值图像导致计算量大的问题。

### 3.3.3 蒸馏方法

考虑到在实际应用中医学超声图像数据集隐私因素的限制,受 DAFL 算法<sup>[27]</sup>的启发,本文采用无数据知识蒸馏方法对轻量级的超分辨率模型进行训练,学生网络通过 GAN 生成一些无标注的训练图片。GAN 的生成器生成和教师网络分布一致的训练数据来训练学生网络。

#### (1) GAN 生成训练数据

本文的无数据知识蒸馏利用图像超分辨率重建的基本特性来帮助 GAN 生成器生成和教师网络训练数据相似的图像。在超分辨率重建任务中,模型将低分辨率图像作为输入,输出对应的高分辨率图像,如果对高分辨率图像进行下采样,使其大小与相应的低分辨率图像大小相同,理论上这两幅图像应该是一样的。利用这个理论,本文假设预训练的教师网络为  $T$ ,待训练的学生网络为  $S$ , $S$  被用于超声设备中,给定一个低分辨率图像  $I_{LR}$ ,经过预训练的教师网络  $T$  后,得到  $T(I_{LR})$ ,然后将  $T(I_{LR})$  下采样得到和  $I_{LR}$  大小相同的图像  $R(T(I_{LR}))$ ,则  $I_{LR}$  和  $R(T(I_{LR}))$  应该是相同的。同样地,在无数据知识蒸馏中,学生网络  $S$  为了从教师网络  $T$  中学到有用的信息,生成器  $G$  生成的训练数据分布应该和教师网络  $T$  的训练数据一致。本文利用了上述  $I_{LR}$  和  $R(T(I_{LR}))$  的一致性。具体地,假设生成器的随机输入为  $z$ ,则生成器生成的图片为  $G(z)$ , $G(z)$  经过教师网络  $T$  后,得到  $T(G(z))$ ,

然后将  $T(G(z))$  重新缩放到低分辨率图像的大小,得到  $R(T(G(z)))$ ,为了和  $T$  的训练数据一致, $G(z)$  和  $R(T(G(z)))$  之间也具有 consistency。即:

$$\begin{aligned} I_{LR} &\Rightarrow T(I_{LR}) \Rightarrow R(T(I_{LR})) \\ G(z) &\Rightarrow T(G(z)) \Rightarrow R(T(G(z))) \end{aligned} \quad (10)$$

用于生成训练数据的 GAN 生成器,包括 3 层网络结构,每层一个卷积层,卷积层后添加批量归一化层(BN)和 ReLU 激活函数。第 1 层和第 2 层卷积有 128 个  $3 \times 3$  的卷积核,第 3 层有 64 个  $3 \times 3$  的卷积核。

#### (2) 渐近式训练策略

在不使用原始数据集的情况下很难通过教师网络提取信息来直接训练学生网络,因此本文采用渐近式训练策略,训练过程中逐渐增加学生超分网络中层的数量,以便从教师网络中提取更多的信息,以更好地训练学生网络。具体来说,首先训练一个和学生网络相似的微型网络,然后逐渐增加微型网络中层的数量,本文使用的学生网络是通道数为 32 的 VD-SR,除了第 1 层和最后 1 层,中间层数从 0 逐渐增加到 18,训练新加的层中的参数,生成器和微型网络的训练交替进行。设  $\{S_k\}$  表示微型学生网络集合,则  $\{S_k\}$  从  $S_0$  逐渐增加到  $S_{18}$ ,每一次迭代时,先固定生成器,训练微型学生网络  $S_k$ ,计算蒸馏损失,反向传播更新  $S_k$  的参数。更新完之后,固定微型网络训练生成器,计算生成器损失优化生成器,当训练的微型网络增加到学生网络之后,则学生网络训练完成。具体做法如算法 1 所示。

#### 算法 1 渐近式知识蒸馏

输入: 预训练的教师超分网络  $T$ ; 微型学生网络层的数量  $P$ ; batchsize 大小  $M$ ; 随机噪声  $\{z\}$

输出: 训练好的学生超分网络  $S$

初始化: 学生微型网络集合  $\{S_i\}_{0 \leq i \leq P}$ , 生成器  $G$ , 微型学生网络初始层数  $k=0$

1.  $S_k \leftarrow S_{\max(k-1, 0)}$ ;
2. 获取生成图像  $\{G(z)\} \leftarrow \{z\}$ ;
3. 得到超分辨率结果  $\{T(G(z))\}$  和  $\{S_k(G(z))\}$ ;
4. 根据式(11)计算  $S_k$  的蒸馏损失函数  $L_{S_k}$ ;
5. 更新  $S_k$  的参数;
6. 固定  $S_k$  的参数,执行步骤 2-步骤 3;
7. 根据式(14)计算生成器  $G$  的损失函数  $L_G$ ;
8. 更新  $G$  的参数;
9. 若达到最大迭代次数,则  $k++$ , 转步骤 1; 否则转步骤 3; 当  $k=P$  时,结束训练。

#### 3.3.4 蒸馏损失函数

教师网络的训练过程和损失函数与 ENLCA 保持一致。用于训练学生网络的损失函数分为两部分,优化学生网络的蒸馏损失函数采用  $L_1$  损失函数。

$$L_{KD} = E_{z \in \rho(z)} [\|T(G(z)) - S(G(z))\|_1] \quad (11)$$

根据 3.3.3 节的生成训练数据的方法,引入生成器的重建损失以及对抗损失,重建损失为:

$$L_R = E_{z \in \rho_z(z)} \left[ \frac{1}{n} \|\mathbf{R}(T(G(z))) - G(z)\|_1 \right] \quad (12)$$

对抗损失为:

$$L_{GEN} = -\log(L_{KD} + 1) \quad (13)$$

因此,生成器的损失函数为:

$$L_G = L_{GEN} + \omega_R L_R \quad (14)$$

其中,  $\omega_R$  为超参数。

## 4 实验

本文的实验包括 5 个部分: 4.1 节介绍了实验的配置环境以及训练细节; 4.2 节收集超声视频数据集用于图像融合和图像超分辨率重建的训练; 4.3 节对图像的客观评价指标进行说明; 4.4 节进行消融实验, 选取图像融合的最佳帧数以及证明图像融合对最终结果的有效性; 4.5 节使用不同的超分辨率重建方法对超声图像进行处理, 从视觉效果、客观效果以及图像分类精度方面对结果进行比较。

### 4.1 实验详情

本文的多帧融合模型和超分辨率模型都在 PyTorch 框架中实现, 在 Tesla K40m 的 GPU 服务器平台上训练和测试, 使用 CUDA 作为网络加速器。实验配置如表 1 所列。多帧融合网络和教师超分网络训练的输入序列的大小均为  $1020 \times 768$ , 学习率和 epoch 设置均相同, 学习率设置为  $10^{-4}$ , 使用 Adam 优化器进行优化。在 200 个 epoch 之后将学习率减半, 在 1000 个 epoch 之后获得最终的多帧融合网络和教师超分网络。多帧融合网络输入的帧数设置为 6。蒸馏损失函数中的  $\omega_R$  设置为 1.0。

表 1 实验配置  
Table 1 Experiment setup

硬件	软件
CPU: 4 Cores	Red Hat 4.8.3-9
RAM: 62GB	CUDA 9.0+CUDNN 7.0.5
GPU: Tesla K40m(45 GB)	Pytorch1.0.0+Python3.6

### 4.2 数据集

超声视频数据集源自一个免费的公开的超声病例库<sup>[28]</sup>, 此超声病例库包含 7672 个病例, 包含超声图像、MRI 图像、CT 图像以及超声视频、MRI 视频和 CT 视频。本文从收集到的 1621 个、119 类不同的超声视频中选取 447 个、37 类质量比较好的超声视频, 截取超声视频的所有帧, 将此数据集称为 US37。从每一类中选取一个超声视频, 将总共 37 个视频的所有帧图像作为测试集, 其余 410 个视频的所有帧图像作为训练集。

教师模型的训练使用训练好的融合网络得到的融合图像。使用双三次插值对图像下采样, 获得相应的缩小 2 倍、3 倍和 4 倍的低分辨率图像。通过随机水平翻转和  $90^\circ$ ,  $180^\circ$  和  $270^\circ$  旋转进行数据增强。

为了进一步证明本文提出的基于多帧融合的超分辨率算法的有效性, 在实验过程中使用不同的超分辨率方法得到超分辨率结果图像, 并在图像分类精度方面对结果进行比较。因此, 在图像分类上, 数据集来自 US37 测试集, 从每一类视频帧中选取 108 张图像, 总共 37 个类别、3996 张图像, 训练集中每一类均有 84 张图像, 测试集中每一类均有 24 张图像, 将此数据集称为 US\_Class。US\_Class 数据集既包含不同组织的超声图像, 也包含同一组织的不同病变的超声图像, 因此增加了分类难度。

### 4.3 评价指标

本文用 2 个图像客观评价指标(IL-NIQE 和信息熵 EN)

和图像分类准确率来客观评估所提方法和对比方法的性能。

(1) IL-NIQE, 一个无参考图像质量评价指标。IL-NIQE 从原始自然图像集合中学习图像块的多元高斯模型, 利用所学习的多元高斯模型, 通过一个类巴氏距离来度量每个图像块的质量, 然后通过平均池化得到总体质量分数。IL-NIQE 值越小, 表明图像的效果越好。

(2) 信息熵 (Entropy, EN), 用于评估图像的信息内容, EN 值越高, 意味着图像包含更多信息, EN 值的计算式如下:

$$EN = - \sum_{i=0}^{L-1} p_i \log_2 p_i \quad (15)$$

其中,  $L$  表示灰度总数,  $p_i$  表示图像中相应灰度的归一化直方图。

(3) 图像分类准确率, 图像质量对预训练图像分类器的识别精度有重要的影响, 分类准确率越高, 意味着超分辨率重建得到的图像质量越好。

### 4.4 消融实验

#### 4.4.1 融合帧数选取实验

在多帧融合阶段, 分别对 37 类超声视频中连续的 2 帧、3 帧、6 帧、9 帧和 12 帧进行融合, 通过对得到的融合图像的 IL-NIQE 和 EN 的平均值进行比较, 结果发现选取 6 帧进行融合得到的融合图像的效果最好。其中 4 类超声视频不同帧数融合图像的 IL-NIQE 和 EN 的平均值如表 2 和表 3 所列。可以看出, 对于胃肠道和肱骨头超声, 融合 6 帧得到的融合图像的 IL-NIQE 的平均值是最小的, 虽然腹壁疝和子宫肌瘤融合 6 帧的 IL-NIQE 值不是最小的, 但与最小值相差不大, 并且在熵的度量上, 6 帧融合得到的融合图像的 EN 值较大, 即融合图像包含的信息量较大。因此, 最终选取连续 6 帧融合得到融合图像。

表 2 不同帧数融合的 IL-NIQE 值对比

Table 2 Comparison of different frames fusion on IL-NIQE

	2 帧	3 帧	6 帧	9 帧	12 帧
胃肠道	41.57	41.31	<b>41.01</b>	41.72	42.06
腹壁疝	<b>38.01</b>	38.22	38.02	38.20	39.18
肱骨头	48.37	48.28	<b>47.76</b>	48.97	48.04
子宫肌瘤	37.26	<b>37.01</b>	37.04	37.82	37.53

注: 加粗字体表示最优结果。

表 3 不同帧数融合的 EN 值对比

Table 3 Comparison of different frames fusion on EN

	2 帧	3 帧	6 帧	9 帧	12 帧
胃肠道	5.4469	5.4529	5.4748	5.4924	<b>5.5074</b>
腹壁疝	4.2676	4.2558	4.3468	4.3896	<b>4.4592</b>
肱骨头	4.3571	4.3572	4.3768	4.3907	<b>4.3967</b>
子宫肌瘤	5.6355	5.6363	5.6440	<b>5.6480</b>	5.6473

注: 加粗字体表示最优结果。

#### 4.4.2 多帧融合的有效性

为证明多帧融合的有效性, 本节进行了消融实验, 对以下两种方法进行对比。

(1) 无融合: 将本文方法中的多帧融合部分去掉, 得到一个超分辨率重建模型。

(2) 有融合: 本文提出的基于视频多帧融合的超分辨率重建方法。

放大倍数设置为 2, 在 37 类超声视频上测试, 比较结果图像的 IL-NIQE 和 EN 值, 其中 4 类图像的实验结果如表 4 和表 5 所列。可以看出, 多帧融合能够从源输入图像中提取更多重要

信息,显著提高超声图像的质量,有助于医生的临床诊断。

表4 融合的 IL-NIQE 消融实验结果

Table 4 Results of ablation studies for fusion on IL-NIQE

方法	胃肠道	腹壁疝	股骨头	子宫肌瘤
无融合	64.84	52.64	68.54	56.94
有融合	<b>54.25</b>	<b>48.40</b>	<b>57.30</b>	<b>48.59</b>

注:加粗字体表示最优结果。

表5 融合的 EN 消融实验结果

Table 5 Results of ablation studies for fusion on EN

方法	胃肠道	腹壁疝	股骨头	子宫肌瘤
无融合	4.9192	3.9965	4.2194	5.2835
有融合	<b>5.5917</b>	<b>4.7520</b>	<b>4.9786</b>	<b>5.8397</b>

注:加粗字体表示最优结果。

#### 4.5 方法对比

使用 US37 数据集训练和测试,将本文提出的方法与已有的图像超分辨率方法进行比较,包括 FSRCNN<sup>[29]</sup>, VDSR, EDSR, RCAN<sup>[30]</sup>, ESPCN<sup>[31]</sup>, Real-ESRGAN, ENLNCN 和 FDIWN。由于本文的对比实验既包括单图像超分辨率方法,也包括多图像超分辨率方法,因此选择的比较策略是将连续 6 帧的单图像超分辨率结果的平均值与连续 6 帧得到的多图像超分辨率结果在视觉效果和客观指标上进行对比。在视觉效果和图像分类准确率方面的对比是在放大 2 倍的情况下进行的,Real-ESRGAN 只有放大 4 倍的结果图,因此在 4.5.1 节和 4.5.3 节的对比中没有列出 Real-ESRGAN 的对比结果。

##### 4.5.1 视觉效果

各方法在子宫肌瘤超声上的视觉效果的对比如图 5

所示。图中单图像超分辨率方法的结果图均为 6 帧中的第 3 帧图像,多图像超分辨率的结果图连续 6 帧图像融合后的超分辨率结果。从图中可以看出,单图像超分辨率结果不能充分利用超声图像之间的互补信息,得到的超分辨率图像效果并不明显,而本文采用的基于视频多帧融合的图像超分辨率方法的结果图明显取得了良好的效果,超分辨图像的对比特度明显高于处理之前的图像,可以更清晰地看到病变组织部位。

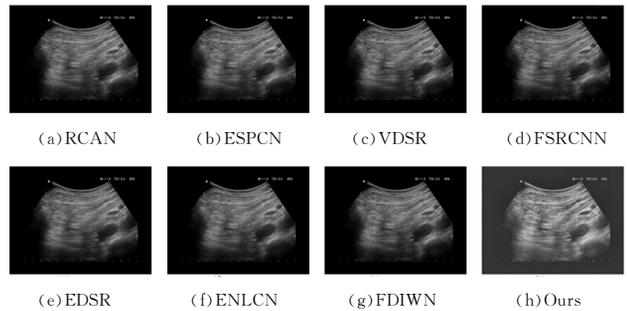


图5 US37 超声数据集上视觉效果的对比如

Fig. 5 Visual results comparison on US37 datasets

##### 4.5.2 客观效果

本文在 US37 测试集上验证了重建效果,本节给出了 4 个具有代表性的组织部位的实验结果,包括胃肠道、腹壁疝、股骨头和子宫肌瘤,它们的 IL-NIQE 和 EN 的对比结果如表 6 和表 7 所列。根据客观对比结果,也可以看出本文方法在已有的超分辨率方法上有了显著提升。

表6 在 US37 数据集上不同超分辨率方法的 IL-NIQE 值

Table 6 IL-NIQE of different SR methods on US37 datasets

	scale	RCAN	ESPCN	VDSR	FSRCNN	EDSR	Real-ESRGAN	ENLNCN	FDIWN	Ours
胃肠道	X2	50.68	51.20	51.67	54.25	<b>48.27</b>	—	50.92	50.71	50.04
	X3	50.30	49.28	50.37	52.25	<b>48.91</b>	—	—	50.32	49.76
	X4	48.69	48.94	50.75	50.79	49.25	<b>44.76</b>	50.55	48.96	48.36
腹壁疝	X2	57.21	57.78	56.33	58.77	57.60	—	56.72	56.55	<b>48.40</b>
	X3	57.26	58.27	56.17	56.21	56.77	—	—	55.77	<b>49.29</b>
	X4	58.97	57.98	57.05	56.08	59.50	59.56	60.62	59.15	<b>55.54</b>
股骨头	X2	72.97	73.21	74.16	73.18	72.95	—	74.06	73.34	<b>57.30</b>
	X3	74.57	73.13	74.81	72.59	73.86	—	—	74.71	<b>64.16</b>
	X4	75.17	72.85	74.17	73.31	73.41	75.30	75.22	75.37	<b>56.30</b>
子宫肌瘤	X2	50.93	51.13	<b>48.35</b>	51.60	51.21	—	51.20	50.97	48.59
	X3	50.72	50.74	51.49	51.12	50.80	—	—	50.01	<b>49.72</b>
	X4	50.19	52.54	52.18	49.27	50.27	<b>45.23</b>	50.20	48.74	50.21

注:加粗字体表示最优结果。

表7 在 US37 数据集上不同超分辨率方法的 EN 值

Table 7 EN of different SR methods on US37 datasets

	scale	RCAN	ESPCN	VDSR	FSRCNN	EDSR	Real-ESRGAN	ENLNCN	FDIWN	Ours
胃肠道	X2	4.2349	4.0619	4.0804	4.0505	4.2106	—	4.2716	4.2772	<b>5.5917</b>
	X3	4.2607	4.1642	4.0856	4.0993	4.2279	—	—	4.2790	<b>6.0451</b>
	X4	4.2281	4.0825	4.0859	4.1191	4.2132	4.1641	4.1816	4.2518	<b>5.8843</b>
腹壁疝	X2	3.0434	2.9164	2.9355	2.9050	3.0295	—	3.0579	3.0809	<b>4.7520</b>
	X3	3.0257	3.0079	2.9359	2.9520	3.0430	—	—	3.0854	<b>5.2443</b>
	X4	3.0460	2.9256	2.9371	2.9730	3.0373	2.8712	3.0052	3.0689	<b>5.0796</b>
股骨头	X2	3.2427	3.1598	3.1529	3.1568	3.2390	—	3.2484	3.2534	<b>4.9786</b>
	X3	3.2443	3.2480	3.1534	3.1818	3.2422	—	—	3.2655	<b>5.4647</b>
	X4	3.2928	3.2211	3.1618	3.2322	3.2871	3.1831	3.2342	3.2819	<b>5.3496</b>
子宫肌瘤	X2	4.5307	4.4057	4.4040	4.4041	4.5306	—	4.5647	4.5530	<b>5.8397</b>
	X3	4.5319	4.5056	4.4053	4.4376	4.5388	—	—	4.5620	<b>6.3287</b>
	X4	4.5321	4.4457	4.4119	4.4345	4.5271	4.6068	4.5024	4.5343	<b>6.1333</b>

注:加粗字体表示最优结果。

### 4.5.3 图像分类准确率比较

本文选择了 Inception-v2, Inception-v3, VGG16, ResNet-50 和 DenseNet-121 作为分类网络。使用 4.2 节描述的 US\_Class 数据集来训练和测试这些分类网络。Adam 被用作优化器,交叉熵被用作损失函数。通过不同超分辨率方法转换的图像的分类精度对比如图 6 所示。

可以看到,与已有的超分辨率重建方法相比,本文方法得到的超分辨率图像具有更高的分类准确率。分类准确率的具体值如表 8 所列。

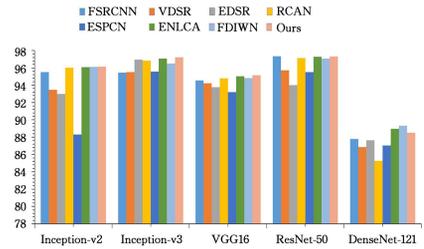


图 6 整体分类准确率

Fig 6 Overall classification accuracy

表 8 分类准确率

Table 8 Classification accuracy

(单位: %)

	FSRCNN	VDSR	EDSR	RCAN	ESPCN	ENLCN	FDIWN	Ours
Inception-v2	95.50	93.47	92.98	96.01	88.29	96.08	96.10	<b>96.14</b>
Inception-v3	95.45	95.51	96.96	96.82	95.57	97.07	96.51	<b>97.22</b>
VGG16	94.55	94.20	93.75	94.77	93.18	95.03	94.82	<b>95.14</b>
ResNet-50	97.30	95.72	93.99	97.14	95.50	97.27	97.07	<b>97.33</b>
DenseNet-121	87.80	86.86	87.64	85.26	87.05	88.96	<b>89.32</b>	88.51

注:加粗字体表示最优结果。

**结束语** 本文提出了一种基于视频多帧融合的医学超声图像超分辨率重建方法,利用超声视频中的相邻连续多帧之间的互补信息,将超声视频中相邻连续多帧融合,结合基于无数据知识蒸馏的超分辨率重建方法,得到包含更多信息且分辨率和对比度更高的医学超声图像。其中,无数据知识蒸馏可以解决医学超声图像的隐私性问题,并且通过无数据知识蒸馏得到的轻量级超分辨率模型更适合部署在实时成像的超声设备上。实验结果表明,本文提出的超声图像超分辨率方法优于已有的其他方法,并且能够有效地辅助医生进行疾病诊断。然而,本文方法中的无监督多帧融合网络还存在一定不足,在多帧融合网络的特征提取模块上使用的 VDSR 模型很难满足医学超声成像实时性要求,后续的工作将对此进行改进,尝试将已有的轻量级超分辨率方法用于多帧融合网络的特征提取模块,设计更加快速高效的多帧融合网络应用于医学超声图像。此外,将着眼于医学超声视频,将知识蒸馏应用于医学超声视频超分辨率重建,解决医学超声视频的质量问题。

### 参考文献

- [1] YAO Z W, YANG F, HUANG J, et al. Improved CycleGANs for Intravascular Ultrasound Image Enhancement [J]. Computer Science, 2019, 46(5): 221-227.
- [2] LIU H, LIU J, HOU S, et al. Perception consistency ultrasound image super-resolution via self-supervised CycleGAN [J]. Neural Computing and Applications, 2021, 33(1): 1-11.
- [3] KUMAR P, SRIVASTAVA S, SAI Y P. High-Performance Medical Image Processing [M]. New York: Apple Academic Press, 2022: 51-61.
- [4] YU Z P, BAI G Z, LIU H Z, et al. Research on Underwater Image Enhancement Based on Histogram Stretching in UCM Algorithm [J]. Journal of Chongqing Technology and Business University: Natural Science Edition, 2022, 39(5): 10-16.
- [5] BASHIR S M A, WANG Y, KHAN M, et al. A comprehensive

review of deep learning-based single image super-resolution [J]. PeerJ Computer Science, 2021, 7(7): 621-677.

- [6] KIM J S. Improved image resolution during zooming in ultrasound image using deep learning technique [C] // 2020 IEEE International Ultrasonics Symposium (IUS). IEEE Press, 2020: 1-3.
- [7] SAWATN A, KULKARNI S. Ultrasound Image Enhancement using Super Resolution [J]. Biomedical Engineering Advances, 2022, 3(1): 1-9.
- [8] GAO G, LI W, LI J, et al. Feature distillation interaction weighting network for lightweight image super-resolution [C] // Proceedings of the AAAI Conference on Artificial Intelligence. AAAI Press, 2022: 661-669.
- [9] WANG X, XIE L, DONG C, et al. Real-esrgan: Training real-world blind super-resolution with pure synthetic data [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. IEEE Press, 2021: 1905-1914.
- [10] LUO Z, HUANG Y, LI S, et al. Learning the degradation distribution for blind image super-resolution [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE Press, 2022: 6063-6072.
- [11] XIA B, HANG Y, TIAN Y, et al. Efficient Non-Local Contrastive Attention for Image Super-Resolution [C] // Proceedings of the AAAI Conference on Artificial Intelligence. AAAI Press, 2022: 6063-6072.
- [12] LI Z, LI S, WANG J M, et al. A Novel Multi-Frame Color Images Super-Resolution Framework based on Deep Convolutional Neural Network [C] // Proceedings of the 2016 5th International Conference on Measurement, Instrumentation and Automation. Atlantis Press, 2016.
- [13] VANMALI A V, KATARIA T, KELKAR S G, et al. Ringing artifacts in wavelet based image fusion: Analysis, measurement and remedies [J]. Information Fusion, 2020, 56(4): 39-69.
- [14] KAUR H, KOUNDAL D, KADYAN V. Image fusion techniques: a survey [J]. Archives of computational methods in Engineering, 2021, 28(7): 4425-4447.

- [15] NOOR A,GAFFAR S,HASSAN M T, et al. Hybrid Image Fusion Method Based On Discrete Wavelet Transform (DWT), Principal Component Analysis(PCA) and Guided Filter [C]// 2020 First International Conference of Smart Systems and Emerging Technologies. IEEE Press,2020:138-143.
- [16] ABBASI AGHAMALEKI J,GHORBANI A. Infrared and visible image fusion based on optimal segmenting and contour extraction [J]. SN Applied Sciences,2021,3(3):1-14.
- [17] PAN Z, YU M,JIANG G, et al. Multi-exposure high dynamic range imaging with informative content enhanced network [J]. Neurocomputing,2020,386(14):147-164.
- [18] LI J, GUO X, LU G, et al. DRPL: Deep regression pair learning for multi-focus image fusion [J]. IEEE Transactions on Image Processing,2020,29(3):4816-4831.
- [19] MA B,ZHU Y, YIN X, et al. SESF-Fuse: an unsupervised deep model for multi-focus image fusion [J]. Neural Computing and Applications,2021,33(11):5793-5804.
- [20] MA J, TANG L, XU M, et al. STDFusionNet: An infrared and visible image fusion network based on salient target detection [J]. IEEE Transactions on Instrumentation and Measurement, 2021,70(4):1-13.
- [21] GOU J P, YU B S, MAYBANK S J. Knowledge distillation: A survey [J]. International Journal of Computer Vision, 2021, 129(6):1789-1819.
- [22] MA K, ZENG K, WANG Z. Perceptual quality assessment for multi-exposure image fusion [J]. IEEE Transactions on Image Processing,2015,24(11):3345-3356.
- [23] ZHANG Y, LIU Y, SUN P, et al. IFCNN: A general image fusion framework based on convolutional neural network [J]. Information Fusion,2020,54(2):99-118.
- [24] KIM J, LEE J K, LEE K M. Accurate image super-resolution using very deep convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE Press,2016:1646-1654.
- [25] ZHANG Y, CHEN H, CHEN X H, et al. Data-free knowledge distillation for image super-resolution[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Computer Vision Foundation,2021:7852-7861.
- [26] LIM B, SON S, KIM H, et al. Enhanced deep residual networks for single image super-resolution[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE Computer Society,2017:1132-1140.
- [27] CHEN H, WANG Y H, XU C, et al. Data-free learning of student networks[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul: IEEE Press,2019: 3513-3521.
- [28] TACO G. Ultrasound Images & Clips[EB/OL]. www. ultrasoundcases. info.
- [29] DONG C, LOY C C, TANG X. Accelerating the super-resolution convolutional neural network [C] // European Conference on Computer Vision. Amsterdam: Springer Press,2016:391-407.
- [30] ZHANG Y, LI K, LI K, et al. Image super-resolution using very deep residual channel attention networks[C]// Proceedings of the European Conference on Computer Vision. Munich: Springer Press,2018:286-301.
- [31] SHI W, ABALLERO J, HUSZAR F, et al. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE Press,2016:1874-1883.



**ZHAO Ran**, born in 1998, postgraduate. Her main research interests include deep learning and medical image processing.



**YUAN Jiabin**, born in 1968, Ph.D, professor, doctoral supervisor, is a member of China Computer Federation. His main research interests include high-performance computing, quantum computing, deep learning, medical image processing, etc.

(责任编辑:喻黎)