

医学图像压缩与视觉任务联合优化方法

姚超¹, 高梓轩², 陈俊如³, 卢奕鹏⁴

(1. 北京科技大学 计算机与通信工程学院, 北京 100083; 2. 北京交通大学 计算机科学与技术学院, 北京 100044;
3. 无锡学院 集成电路科学与工程学院, 江苏 无锡 214105; 4. 北京大学 集成电路学院, 北京 100871)

摘要: 针对医学图像处理中依赖独立编码组件无法实现数据压缩与机器视觉任务联合优化的问题, 本文构建了一种端到端的机器视觉任务驱动的医学图像压缩网络(machine vision task-driven medical image compression network, MVMICNet)模型, 端到端地实现数据压缩与医学图像分析的和谐统一。为了保持医学图像压缩前后机器视觉任务的性能, 设计了任务感知的改进码率-准确率损失函数, 通过引入任务相关的损失项, 在优化过程中动态平衡码率、重建图像失真与机器视觉任务精度三者之间的关系; 同时, MVMICNet模型采用分阶段训练的模式, 针对机器视觉任务的不同特性进行特定的优化, 确保了模型能够精准捕获对诊断至关重要的特征信息, 实现了压缩效率与任务性能的同步提升, 从而在复杂的医学应用场景中展现出更优越的鲁棒性; 最终, 本文在语义分割和目标检测任务中验证了该框架的有效性。

关键词: 医学图像压缩; 语义分割; 目标检测; 卷积神经网络(CNN); 任务驱动优化

中图分类号: TP 391 文献标志码: A 文章编号: 1005-3026(2026)01-0011-10

Joint Optimization Approach for Medical Image Compression and Vision Tasks

YAO Chao¹, GAO Zi-xuan², CHEN Jun-ru³, LU Yi-peng⁴

(1. School of Computer & Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China; 2. School of Computer Science & Technology, Beijing Jiaotong University, Beijing 100044, China; 3. School of Integrated Circuit Science and Engineering, Wuxi University, Wuxi 214105, China; 4. School of Integrated Circuits, Peking University, Beijing 100871, China. Corresponding author: YAO Chao, E-mail: yaochao@ustb.edu.cn)

Abstract: In medical image processing, the reliance on independent encoding components makes it impossible to achieve joint optimization of data compression and machine vision tasks. To address this issue, an end-to-end machine vision task-driven medical image compression network (MVMICNet) was proposed, achieving harmonious unification of data compression and medical image analysis in an end-to-end manner. To maintain the performance of machine vision tasks before and after medical image compression, a task-aware improved code rate-accuracy loss function was designed. By introducing task-related loss terms, it dynamically balanced the relationship among code rate, reconstructed image distortion, and machine vision task accuracy during the optimization process. Furthermore, the MVMICNet model adopted a stage-wise training approach, specifically optimizing for the different characteristics of machine vision tasks to ensure that the model can accurately capture the feature information crucial for diagnosis. This has achieved a simultaneous improvement in compression efficiency and task performance, thus demonstrating superior robustness in complex medical application scenarios. Finally, the effectiveness of the framework was verified in semantic segmentation and object detection tasks.

Key words: medical image compression; semantic segmentation; object detection; convolutional neural network (CNN); task-driven optimization

收稿日期: 2025-06-06

基金项目: 国家自然科学基金资助项目(62372036, 62120106009, 62332017, U24B20179, U22A2022).

作者简介: 姚超(1986—), 男, 黑龙江绥化人, 北京科技大学副教授。

通信作者: 姚超, E-mail: yaochao@ustb.edu.cn.

随着现代医学影像技术的持续发展,高精度数字影像设备单次扫描将产生数百至数千张图像,如核磁共振成像(magnetic resonance imaging, MRI)、计算机断层扫描(computed tomography, CT)图像、X射线图像等,导致医疗数据量激增,存储与传输成本面临严峻挑战.在此背景下,医学图像压缩的必要性愈发凸显:一方面,高效的压缩技术可大幅降低数据体量,缓解医疗系统的资源压力;另一方面,压缩过程必须谨慎权衡,因为过度压缩可能破坏图像的关键诊断特征,进而对后续的智能分析任务产生不可逆的负面影响.目前,图像压缩算法已经从传统的压缩标准如 JPEG (joint photographic experts group)^[1], JPEG2000^[2], BPG (better portable graphics)^[3]发展到基于深度学习的压缩算法^[4-6].虽然上述方法提供了较高的压缩比,却忽略了医学图像特征复杂多样的本质.无论是复杂的人体解剖结构(如骨骼、肌肉、器官和血管等),还是多模态医学成像技术(如X光、CT、MRI、超声等)提供的不同类型的体内视图,都对压缩算法提出了极高的要求.上述这些方法在压缩过程中很难完全保留这些复杂多样的特征,无法保证压缩后诊断信息的完整性.

医学图像压缩的核心矛盾在于压缩效率与诊断信息完整性的平衡.视觉任务(如医学图像分类^[7-8]、医学图像分割^[9-10]和医学图像识别^[11-12])依赖医学图像中细微的纹理、边缘及对比度特征,而过度压缩可能导致伪影或细节丢失,直接影响诊断准确性.例如,在创伤急救中,CT图像的快速压缩与传输需保留骨骼断裂或出血区域的清晰边界;而在肿瘤检测中,MRI图像的低对比度病灶可能因量化误差而被忽略.因此,医学图像的压缩重建不仅需保证符合人眼感知质量,还应保留足够的语义与结构信息,以满足机器视觉系统在诊断、检测等任务中的准确性要求^[13].因此,解决这一矛盾需要突破传统“压缩-分析”分离的范式,探索压缩算法与视觉任务的联合优化机制,从而在医疗资源的有限性与诊断的精准性之间实现平衡.

为了解决上述挑战,本文构建了端到端优化的视觉任务驱动的医学图像压缩网络,通过深度耦合图像压缩与视觉分析任务,实现了二者的协同优化与和谐统一.本文为该网络设计了任务感知损失函数 L_{task} , 缩小原始图像和解压重建图像的视觉任务精度差异,优化码率-准确率.在训练策略上,采用分阶段优化方法,先建立基础的压

缩性能,再针对不同视觉任务特性进行精细化调整.实验验证了 MVMICNet 在保持医学图像压缩性能的情况下,有效提升了语义分割和目标检测等机器视觉任务的精度.

1 相关工作

目前,兼顾信道容量和视觉任务性能的工作范式包含两种:先分析再压缩(analyze then compress, ATC)和先压缩再分析(compress then analyze, CTA)^[14].

ATC 范式首先从原始数据中提取视觉特征,再用适合的编码算法压缩特征,将特征传输到解码端后完成相应的机器视觉任务^[14].提取并压缩语义视觉特征后,再进行特征的传输与重建是一种可行方案^[15].王凯^[16]提出的方法通过双分支特征融合实现高效的医学图像的无损压缩;Herbert 等^[17]提出了一种尺度和旋转不变的兴趣点检测器,其可重复性、独特性和鲁棒性几乎达到最佳性能;Redondi 等^[18]创新性地提出将视觉数据处理前移至感知端,直接对像素域中提取的局部特征进行压缩,跳过传统的图像重建过程,从而在降低传输冗余的同时,显著提升了压缩效率与视觉任务准确率.Liu 等^[19]通过选择性编码特征子集,动态平衡人类视觉与不同机器任务间的比特分配,为机器视觉与人类视觉的联合压缩提供了一种解决方案.在码率受限的情况下,ATC 范式虽然可以获得良好的视觉任务性能,但并没有摆脱手工设计模块的限制,这导致前端提取数据和后端分析数据相互独立,难以实现压缩任务和机器视觉任务的联合优化以同时达到最优.

CTA 范式先压缩图像,以满足有限的网络带宽和存储容量的需求,再进行机器视觉任务的分析.李基臣等^[20]通过对图像进行去噪、增强、复原、分割、提取特征等处理,实现了医学数字图像处理技术在医学影像中的研究与应用.随后,Zabala 等^[21]对重建后的农作物图像进行数字化分类,验证了图像分块操作对视觉任务性能的影响.Chao 等^[22]提出了一种融合尺度不变特征变换的码率失真优化模型,旨在 JPEG 压缩图像重建过程中有效保留图像的关键结构信息及其尺度不变性特征.该方法通过引入尺度不变的特征变换,有助于减小压缩带来的失真对图像多尺度结构的破坏,从而显著提升后续机器视觉任务(如特征匹配与目标识别)的精度与鲁棒性.Shindo

等^[23]将 SAM(segment anything model)的边缘生成能力与压缩结合,提出了一种高效的面向机器视觉任务的图像压缩框架.综上,CTA 范式在高码率下可以获得良好的机器任务分析性能.然而,压缩算法通常是以符合人类视觉特性为目标而设计,并非面向机器视觉任务设计,导致该算法无法满足后续机器视觉任务的性能需求^[24].

2 机器视觉任务驱动的压缩网络

本文提出的机器视觉任务驱动的医学图像压缩网络 MVMICNet 结构如图 1 所示.该网络采用端到端的方式,将图像压缩与机器视觉任务依次串联.在整个框架中,未添加任何增强模块,仅通过调整损失函数和训练策略来实现整体优化

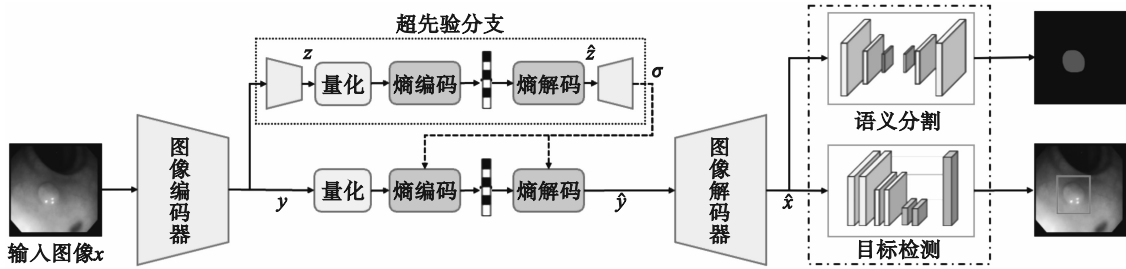


图 1 MVMICNet 框架结构示意图

Fig. 1 Schematic diagram of MVMICNet framework

为减少隐变量的编码冗余,提高熵编码效率, MVMICNet 引入了超先验模型.该分支可以对压缩特征 y 的空间信息进行进一步分析理解,从而推理出 y 的概率分布.图 1 中的超先验分支是一个小型的自编码器,由超先验编码器、量化操作、熵编码操作和超先验解码器组成.图像压缩编码器输出的特征 y , 经过包含 3 个下采样层的超先验编码器, 获得其边信息 z 后, 再进行量化、熵编码与熵解码得到 \hat{z} , 并将其作为超先验解码器的输入来估算 y 的概率分布.超先验解码器的输出结果 y 可以近似为标准差 σ 的高斯分布, 利用 σ 对隐变量特征 \hat{y} 进行建模, \hat{y} 中各值的概率近似地遵循均值为 0, 方差为 σ^2 的高斯分布, 即 $\text{Gaussian}(0, \sigma^2)$, 如式(1)所示:

$$f(x|0, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right). \quad (1)$$

在反向传播的过程中, 边信息 \hat{z} 的均值不断减小, 逐渐趋近于零, 从而实现对码率的优化与压缩效率的提升.同时, 经过量化、熵编码、熵解码后的特征 \hat{y} 输入解码器, 可以重建图像 \hat{x} , 作为

的目标.为统一优化图像编码算法和机器视觉任务的性能, MVMICNet 改进了任务感知损失函数 L_{task} , 并设计了相应的码率-准确率损失函数, 详细内容如下文所示.

2.1 机器视觉任务驱动的医学图像编码框架

MVMICNet 将图像压缩、机器视觉任务两部分串联起来, 实现端到端联合优化的目的.具体地, 图像压缩部分采用自编码器^[4]与超先验模型^[5]相结合的结构.输入图像 x 经过自编码器转换为隐变量特征 y , 随后经过量化、熵编码等操作, 为后续的图像重建与视觉任务分析奠定基础.但要压缩编码过程进行码率优化, 就需要 1 种针对网络参数可微的码流估计方法, 且能对 y 的概率进行准确建模以提高熵编码效率, 从而实现更高的压缩效率.

机器视觉任务网络的输入.

本文将机器视觉任务网络级联在深度图像编码框架之后而不进行网络结构的修改, 因此所提的方法对于不同的下游机器视觉任务是有通用性的.

2.2 任务感知损失函数

图像压缩任务通常被视作一个码率失真优化(rate distortion optimization, RDO)问题, 即 $J = \lambda D + R$. 其中, R 为压缩后的比特率, D 为重建图像与原始图像之间的失真程度, λ 为平衡 R 和 D 权重的拉格朗日乘子, J 为联合优化的目标函数.该公式反映出比特率与失真程度之间的权衡关系, 编码需要的比特数越多, 图像重建的失真程度就越小, 反之亦然.通常, RDO 问题通过在预设的一组 λ 值下进行多次反向传播迭代求解, 从而获得在不同压缩率下的最优平衡.然而, 对原始医学图像进行机器视觉任务分析时, 虽然可以获得高精度的视觉任务性能, 但是原始图像经过下采样、编码、解压重建后带来的失真会降低视觉任务的性能.因此, 这种单一率失真优化方式无法

适用于机器视觉任务驱动的医学图像压缩.

因此,本文引入任务感知损失函数 L_{task} ,在传统率失真损失函数的基础上进行改进,实现对压缩码率与下游视觉任务准确率之间的联合优化,可以表示为

$$L_{\text{task}} = \frac{\sum_{i=1}^m \sum_{j=1}^n (M(x_{i,j}) - M(\hat{x}_{i,j}))^2}{m \times n}. \quad (2)$$

其中: $M(\cdot)$ 表示框架中编解码器下游的机器视觉任务网络的输出; x 和 \hat{x} 分别表示原始图像和重建图像; m 和 n 分别表示图像的长和宽.

改进后的码率-准确率优化损失函数(L_{stage1})可以表示为

$$L_{\text{stage1}} = \lambda_1 D + R + \lambda_2 L_{\text{task}}. \quad (3)$$

其中: λ_1 和 λ_2 分别表示图像失真和下游任务准确率在模型优化中的权重参数.

码率-准确率优化损失函数可以平衡图像压缩率和下游任务准确率之间的关系.这种优化方式有助于根据不同机器视觉任务的需求,有针对性地优化图像压缩算法.

2.3 针对特定任务的分阶段优化模式

为了同时优化图像压缩和机器视觉2种不同类型的任务,该框架采用了分阶段的训练方式.在第一阶段,采用式(3)作为损失函数,优化码率-准确率,获得高质量的重建图像,其对应的机器视觉任务性能逼近原始图像所对应的性能.在训练过程收敛后,MVMICNet保存网络参数,采用有监督的训练方式优化机器视觉任务的性能,结构如图2所示.第二阶段,根据不同机器视觉任务的特点,MVMICNet以图像压缩网络的有损重建图像为输入,加入有监督的标签数据作为引导信息,使用交叉熵损失函数(L_{stage2})进行训练,可以表示为

$$L_{\text{stage2}} = -\frac{1}{N} \sum_k \sum_{c=1}^M s_{kc} \log(p_{kc}). \quad (4)$$

其中: N 表示样本的数量; k 表示第 k 个样本; M 表示类别的数量; c 表示类别,且 $c=1,2,\dots,M$; s_{kc} 为0或1,当 $s_{kc}=1$ 表示第 k 个样本的预测类别与真实标签图像类别 c 相同,否则为0; p_{kc} 表示第 k 个样本属于类别 c 的概率.

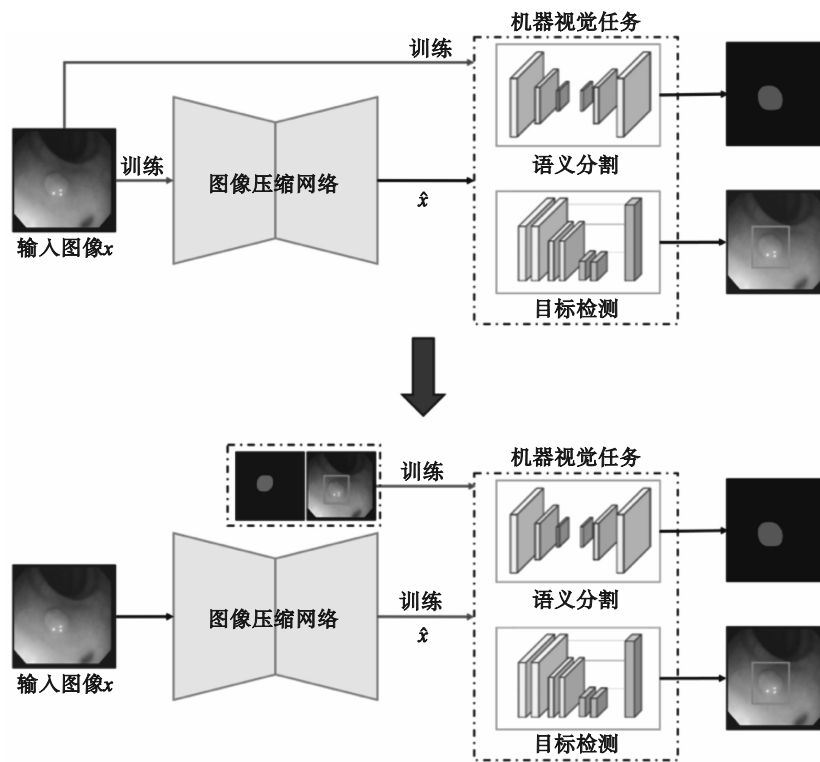


图2 MVMICNet任务特定优化阶段框架结构示意图

Fig. 2 Schematic diagram of MVMICNet framework at task-specific optimization stage

3 实验结果与分析

3.1 实验设置

本文第一阶段的训练数据集由 Flickr.com^[6]

提供的 20 745 张自然图像组成,用于预训练图像压缩模块以学习通用的图像表示能力.码率-准确率优化损失函数如式(3)所示,其中, λ_1 用于控制重建图像码率和失真的平衡,分别取 128, 256, 1 024 和 2 048,以训练对应不同码率的压缩模型,

λ_2 设置为 0.000 1. 本文采用 Adam 优化器进行模型训练, 以保证训练过程的稳定性和收敛效率. 第一阶段迭代训练 2×10^6 次, 其中前 1.8×10^6 次迭代的学习率固定为 0.000 1, 以确保模型能够充分学习图像压缩任务的基本特征; 在随后的 2×10^5 次迭代中, 将学习率下调至原先的 1/10, 以实现训练的精细调整和收敛, 提升模型的泛化能力. 本文所有实验均在基于 PyTorch 框架的环境中完成, 硬件平台为配备 11 GB 显存的 NVIDIA GeForce RTX 2080 Ti 显卡. 针对第二阶段的训练, 本文根据不同的视觉任务特点灵活调整参数设置, 结合任务特有的损失函数和训练策略, 进一步提升模型在下游视觉分析任务中的表现, 具体如下文所示.

1) 语义分割任务. 当验证 MVMICNet 在语义分割方面的性能时, 该框架在第二阶段采用结肠镜图像息肉分割数据集 CVC-ColonDB^[25] 的训练集和测试集, 该数据集包含分辨率为 574 像素 \times 500 像素的原始图像以及相应的图像掩模. 训练集和测试集分别包含 303 张和 37 张结肠镜图像. 本文通过数据预处理将原始图像的分辨率统一设置为 352 像素 \times 352 像素. 本文在第二阶段选择 Mask R-CNN^[26] 作为下游语义分割任务的主干模型. 在该阶段训练过程中, 固定图像压缩网络的参数, 仅对语义分割网络进行参数微调. 模型共训练 30 个轮次, 以获得最终优化后的语义分割模型.

2) 目标检测任务. 当验证 MVMICNet 在目标检测任务方面的性能时, 该框架在第二阶段采用胸部 X 光目标检测数据集 ChestX-Det^[27]. 该数据集涉及 13 个类别的像素标注, 分别为 Atelectasis, Calcification, Cardiomegaly, Consolidation, Diffuse Nodule, Effusion, Emphysema, Fibrosis, Fracture, Mass, Nodule, Pleural Thickening 和 Pneumothorax. 整个数据集分成训练集和测试集共 3 500 多张高分辨率胸部 X 光片, 所有图像均来自公立医院临床采集. 本文选用的目标检测模型为 Faster R-CNN^[28]. 同样, 在第二阶段, 固定压缩网络的参数, 开放目标检测网络的参数, 微调 50 个轮次, 以得到最终的目标检测模型.

3.2 性能评估

本文对比算法采用传统 CTA 范式, 输入的原始图像经过 BPG, MBT2018-Mean^[29] 和 Cheng2020-Anchor^[30] 算法压缩后, 再解码重建图像用于分析机器视觉任务的性能. 本文对比算法通

过不同量化系数 q 值来控制码率范围, 而 MVMICNet 通过不同 λ_2 值来控制码率范围. 本文采用峰值信噪比 (peak signal to noise ratio, PSNR) 和多尺度结构相似性 (multi-scale structural similarity, MS-SSIM) 作为评估重建图像的客观指标. PSNR 越高、MS-SSIM 越接近 1, 通常表示重建图像与原始图像之间的失真越小, 重建质量越高. 本文以语义分割和目标检测任务 2 种机器视觉任务为例, 验证 MVMICNet 的性能. 其中, 对于语义分割任务, 本文采用类平均交并比 (mean intersection over union, mIoU) 作为主要评价指标. mIoU 衡量网络预测的分割区域与真实标注区域之间的重叠程度, 定义为二者交集与并集的比值, 并在所有类别上进行平均, 反映整体分割精度. 对于目标检测任务, 本文采用平均精度平均值 (mean average precision, mAP) 作为主要评价指标, 表示所有类别检测精度的平均值, 可全面评估模型的检测性能. 最后, 本文使用每个像素的比特数 (bits per pixel, Bpp) 来评估图像压缩所需的码率.

表 1 展示了 MVMICNet 在 CVC-ColonDB 测试集上的 PSNR, MS-SSIM 和 mIoU 的比较结果. 可以看出, 在相似比特数 (Bpp) 情况下, MVMICNet 比 BPG, MBT2018-Mean 和 Cheng2020-Anchor 算法均有提升. 例如, 在 Bpp=0.065 时, MVMICNet 仅经过 1 个阶段训练获得的语义分割精度 mIoU 就比利用 BPG 在 Bpp=0.132 获得的 mIoU 更高. 图 3 对比了不同算法的码率-准确率曲线. 其中, 虚线部分表示在原始图像 (未经压缩) 上执行语义分割任务所获得的基准精度, 用以作为各压缩方法的参考上限, 而采用 BPG, MBT2018-Mean 和 Cheng2020-Anchor 压缩的 CTA 范式获得的分割精度与原始图像的分割精度差距很大. MVMICNet 不仅可以获得良好的图像压缩性能, 而且可以保持有损重建图像的语义分割精度.

图 4 为不同算法的分割图像对应的可视化结果. MVMICNet 获得的分割效果可以清晰地分辨出结肠息肉及具体轮廓, 能够比 BPG 和 Cheng2020-Anchor 压缩算法取得更好的分割效果, 从而验证了本文所提算法的有效性.

图 5 为不同算法在 CVC-ColonDB 数据集上对同 1 张图像的重建结果. 在相近的压缩比下, BPG 和 Cheng2020-Anchor 压缩算法出现了块效应, 而 MVMICNet 获得的重建效果逼近原图, 能够比其他压缩算法取得更好的重建效果.

表 1 CVC-ColonDB 数据集上语义分割精度的对比结果
Table 1 Comparison results of semantic segmentation accuracy on CVC-ColonDB dataset

算法	Bpp	PSNR/dB	MS-SSIM	mIoU	
				第一阶段	第二阶段
BPG	0.090	30.02	0.908	0.498 7	
	0.100	31.40	0.929	0.559 1	
	0.114	33.01	0.944	0.618 3	
	0.132	34.53	0.956	0.668 3	
MBT2018-Mean	0.080	32.55	0.907	0.613 5	
	0.094	33.84	0.938	0.674 5	
	0.111	35.15	0.952	0.722 0	
Cheng2020-Anchor	0.130	35.98	0.969	0.754 8	
	0.068	36.45	0.934	0.697 1	
	0.089	37.98	0.952	0.743 5	
MVMICNet	0.112	39.04	0.972	0.789 8	
	0.137	39.82	0.978	0.819 2	
	0.065	40.38	0.982	0.780 4	0.822 6
MVMICNet	0.083	41.25	0.984	0.825 2	0.831 7
	0.105	42.08	0.987	0.832 6	0.849 0
	0.131	42.83	0.989	0.833 7	0.864 5

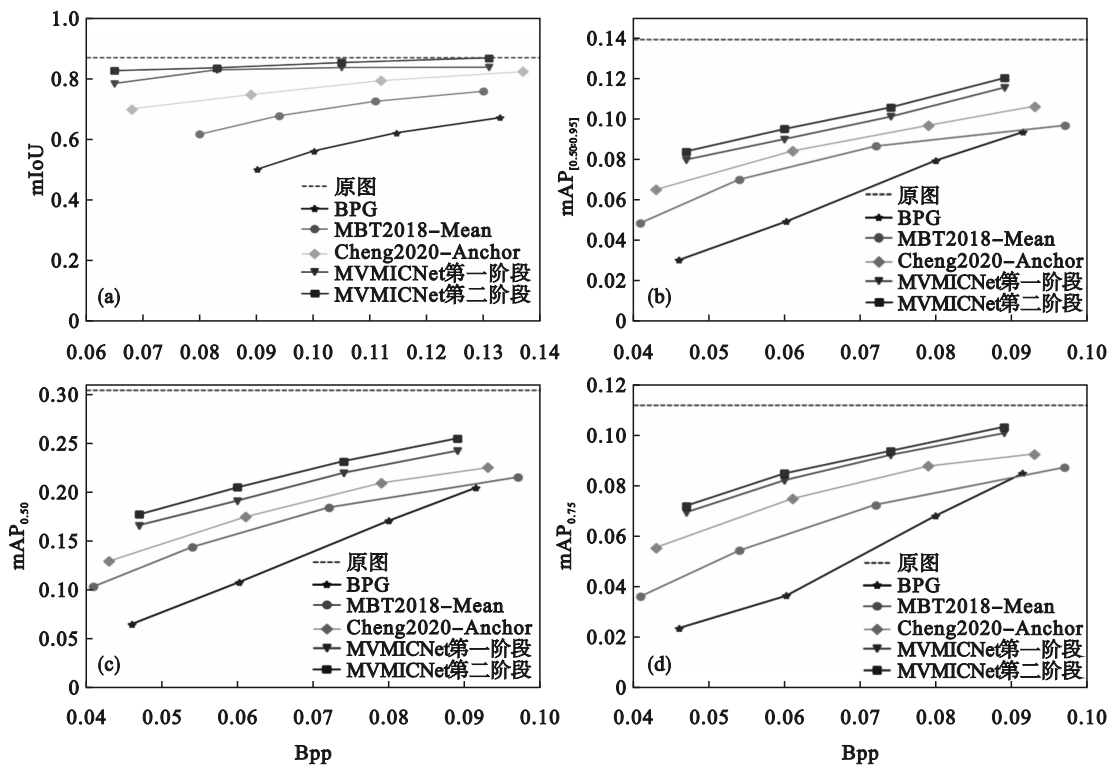


图 3 CVC-ColonDB 和 ChestX-Det 数据集上不同算法的码率-准确率曲线比较结果

Fig. 3 Comparison of code rate-accuracy curves for different algorithms on CVC-ColonDB and ChestX-Det datasets

(a)—CVC-ColonDB; (b)—ChestX-Det; (c)—ChestX-Det; (d)—ChestX-Det.

在 ChestX-Det 数据集上,目标检测性能的对比结果如表 2 和表 3 所示.图 3 用曲线图展示了当 $\text{IoU}=0.50:0.95$, 0.50 和 0.75 时,对比算法 BPG, MBT2018-Mean, Cheng2020-Anchor, MVMICNet

第一阶段和第二阶段在 ChestX-Det 数据集上的结果.原始图像对应的目标检测精度分别为 $0.139 1$, $0.302 6$ 和 $0.111 8$,如图中虚线所示.无论是在低码率还是高码率情况下, BPG, MBT2018-

Mean 和 Cheng2020-Anchor 压缩算法对应的目标检测精度都明显低于原始图像以及 MVMICNet 第一阶段和第二阶段对应的精度. MVMICNet 在低比特率情况下的目标检测精度低于原始图像, 但随着比特率的升高, 检测精度逐渐上升. 在高比特率下, MVMICNet 获得的检测精度接近于原始图像对应的精度. 图 6 为不同算法在 ChestX-Det 数据集上的 1 张图片目标检测可视化比较结果. 在相近的压缩率下, BPG 和 Cheng2020-Anchor 压缩算法仅检测出了类别 Effusion 并且置信度均低于 MVMICNet 第一阶段和第二阶段. 相较于 MVMICNet 第一阶段, MVMICNet 第二阶段检测出了类别 Effusion 和 Consolidation, 检测框位置几乎与原图一致并且置信度逼近原图, 验证了 MVMICNet 第二阶段训练的必要性. 综上所述, MVMICNet 在保证图像压缩率失真性能的前提下, 能够有效保持医学图像中目标检测任务的高精度表现, 体现了该网络在图像压缩与下游视觉任务间的深度联合优化能力. 通过联合训练和任务感知的损失函数设计, MVMICNet 不仅提升了压缩效率, 还显著增强了重建图像的语义信息保留, 从而满足了医学图像智能分析对图像质量和任务性能的双重需求.

图 7 为不同算法在 ChestX-Det 数据集上对同 1 张图像的重建结果. 在相近的压缩率下, BPG 和 Cheng2020-Anchor 压缩算法出现了较多的模糊伪影, 而 MVMICNet 获得的重建效果逼近原图, 能够比其他压缩算法取得更好的重建效果.

3.3 消融实验

1) 任务感知损失函数权重选择. 为了验证参数 λ_2 对语义分割任务性能的影响, 本节设置了相

应的消融实验, 如表 4 所示. 为了控制变量, 本节固定权重参数 λ_1 为 256, 并分别选择 4 个不同的 $\lambda_2=[0.1, 0.001, 0.000 1, 0.000 01]$ 进行训练.

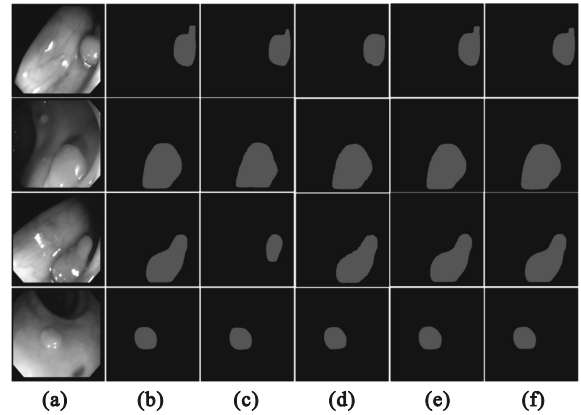


图 4 不同算法在 CVC-ColonDB 数据集上的语义分割可视化结果对比

Fig. 4 Visual result comparison of semantic segmentation by different algorithms on CVC-ColonDB dataset

(a)—原始图像; (b)—原始图像对应的语义分割结果; (c)—BPG 压缩图像的语义分割结果; (d)—Cheng2020-Anchor 压缩图像的语义分割结果; (e)—MVMICNet 的第一阶段压缩图像的语义分割结果; (f)—MVMICNet 的第二阶段语义分割结果.

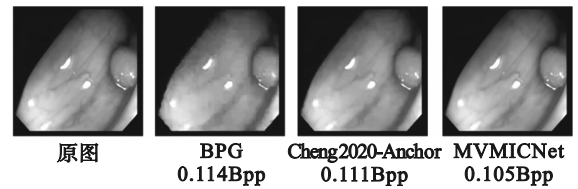


图 5 不同算法在 CVC-ColonDB 数据集上对同 1 张图像的重建结果

Fig. 5 Reconstruction results of the same image from CVC-ColonDB dataset by different algorithms

表 2 ChestX-Det 数据集上目标检测精度的对比结果 1
Table 2 Comparison results 1 of object detection accuracy on ChestX-Det dataset

指标	MVMICNet				BPG				
	$\lambda_2=0.000 01$	$\lambda_2=0.000 1$	$\lambda_2=0.001$	$\lambda_2=0.1$	$q=34$	$q=31$	$q=28$	$q=25$	
Bpp	0.047	0.060	0.074	0.089	0.046	0.060	0.079	0.091	
PSNR/dB	41.14	41.93	42.57	43.10	35.98	37.50	38.76	39.52	
MS-SSIM	0.984 7	0.987 7	0.989 8	0.991 3	0.959 4	0.969 5	0.976 7	0.983 4	
IoU= 0.50:0.95	第一阶段	0.079 7	0.089 9	0.101 0	0.115 4	0.030 0	0.049 1	0.030 0	0.093 4
	第二阶段	0.083 8	0.094 9	0.105 6	0.120 1				
mAP IoU=0.50	第一阶段	0.164 9	0.190 1	0.218 4	0.241 0	0.064 0	0.106 9	0.064 0	0.203 3
	第二阶段	0.175 9	0.203 6	0.230 4	0.253 6				
IoU=0.75	第一阶段	0.069 3	0.082 0	0.092 0	0.100 6	0.023 5	0.036 3	0.023 5	0.084 7
	第二阶段	0.071 9	0.084 7	0.093 6	0.103 1				

表 3 ChestX-Det数据集上目标检测精度的对比结果 2
Table 3 Comparison result 2 of object detection accuracy on the ChestX-Det dataset

指标	MBT2018-Mean				Cheng2020-Anchor			
	$q=5$	$q=4$	$q=3$	$q=2$	$q=5$	$q=4$	$q=3$	$q=2$
Bpp	0.041	0.054	0.072	0.097	0.043	0.061	0.079	0.093
PSNR/dB	37.36	38.60	39.98	40.81	39.41	40.63	41.63	42.36
MS-SSIM	0.969 1	0.974 0	0.978 8	0.984 5	0.979 6	0.983 2	0.985 4	0.986 6
IoU=0.50:0.95	0.048 4	0.069 9	0.086 4	0.096 7	0.064 9	0.084 0	0.096 7	0.106 2
mAP								
IoU=0.50	0.102 4	0.142 7	0.183 1	0.214 1	0.128 2	0.173 4	0.208 2	0.223 7
IoU=0.75	0.036 1	0.054 2	0.072 3	0.087 1	0.055 4	0.074 7	0.087 6	0.092 4

本实验使用不同 λ_2 的值通过式 (3) 对 MVMICNet 进行优化,并在训练得到的图像压缩模型上使用 Cityscapes 验证集^[31]进行验证以确定最佳的 λ_2 权重.实验结果如表 4 所示.值得一提的是,本文使用的语义分割模型 ERFNet^[32]利用未经压缩的 Cityscapes 验证集进行语义分割任务时,mIoU 为 72.62%.实验结果验证了不同数量级的 λ_2 对模型在图像压缩精度和语义分割任务精度上均有显著影响,具体表现在图像压缩的率失真性能与语义分割任务的精度之间的权衡关系上.当 λ_2 值较大时,模型训练过程中语义分割任务的梯度权重较高,导致模型更偏重分割任务的

优化,然而这种倾斜会导致图像压缩的率失真性能下降,表现为重建图像的压缩效率降低和失真增加;反之, λ_2 值过小时,模型更侧重于重建图像的质量优化,率失真性能得到提升,但同时难以有效保证语义分割任务的精度.综合考虑,实验验证 λ_2 取 0.000 1 能较好地平衡两者之间的矛盾,使得模型在保证良好图像压缩性能的同时,兼顾了下游语义分割任务的精度需求.此外,第二阶段的有监督训练进一步提升了 MVMICNet 在语义分割任务中的适应性与泛化能力,使其最终表现几乎达到未经压缩的原始图像水平,充分展示了联合优化策略在实际应用中的有效性与优势.

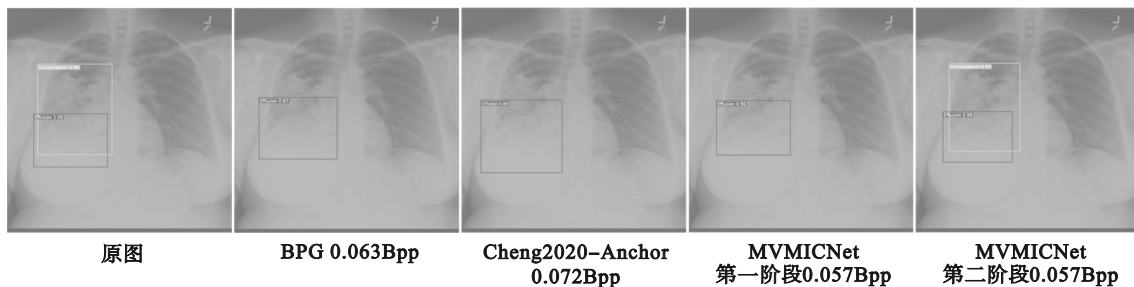


图 6 不同算法在 ChestX-Det数据集上目标检测可视化结果比较

Fig. 6 Visual result comparison of object detection by different algorithm on ChestX-Det dataset

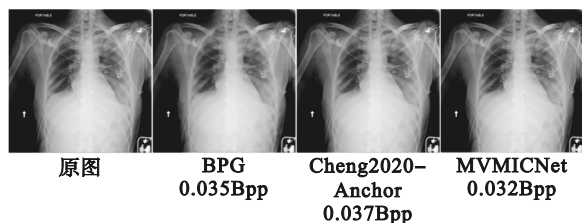


图 7 不同算法在 ChestX-Det数据集上对同 1 张图像的重建结果

Fig. 7 Reconstruction results of the same image from ChestX-Det dataset by different algorithm

2) 针对特定任务的分阶段优化消融实验.在第二阶段训练中,结合机器视觉任务的具体需求,加入有监督的标签数据,利用该任务原本设计的损失函数进行优化,以提升模型在下游任务

中的性能表现.为此,本节的消融实验用于验证分阶段训练策略优化的重要性与可行性.表 1 和表 2 分别展示了 MVMICNet 在 2 个下游机器视觉任务中的性能对比,包括仅进行第一阶段训练策略优化与完成两阶段训练策略优化的结果.实验结果显示,单纯进行第一阶段训练时,模型在保证压缩率和失真之间达到一定平衡的同时,机器视觉任务的性能存在一定程度的下降,说明仅靠率失真优化难以充分兼顾下游视觉任务需求.引入第二阶段训练后,通过结合任务相关的有监督标签信息及其专用损失函数,模型进一步调整压缩特征以更好适应视觉任务,从而显著提升了语义分割和目标检测的精度.

表 4 参数 λ_2 对语义分割任务性能的影响
Table 4 Impact of parameter λ_2 on semantic segmentation performance

λ_2	Bpp	PSNR/dB	mIoU/%	准确率/%
0.1	0.252 30	24.250	33.040	45.5
0.001	0.245 60	31.880	45.780	63.0
0.000 1	0.243 20	35.080	58.030	79.9
0.000 01	0.243 10	34.570	52.710	72.6

4 结 语

本文提出了机器视觉任务驱动的医学图像压缩网络 MVMICNet, 通过构建端到端的联合优化框架, 创新性地实现了图像压缩与视觉分析任务的深度协同与和谐统一, 突破了传统处理流程中压缩与分析相互割裂的局限. 相应地, 本文设计了码率-准确率损失函数, 能够自适应地保留对视觉任务至关重要的诊断特征. 针对不同任务, 本文设计了分阶段优化策略, 提高压缩场景下多个机器视觉任务的性能. 实验结果表明, MVMICNet 在保证压缩率失真性能的前提下, 能够保持甚至提升机器视觉任务的性能. 验证了压缩与视觉任务联合优化的可行性, 为医学图像智能处理提供了新的技术范式.

参考文献:

- [1] Wallace G K. The JPEG still picture compression standard [J]. *IEEE Transactions on Consumer Electronics*, 1992, 38(1): 18-34.
- [2] Christopoulos C, Skodras A, Ebrahimi T. The JPEG2000 still image coding system: an overview [J]. *IEEE Transactions on Consumer Electronics*, 2000, 46(4): 1103-1127.
- [3] Sullivan G J, Ohm J R, Han W J, et al. Overview of the high efficiency video coding (HEVC) standard [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2012, 22(12): 1649-1668.
- [4] Ballé J, Laparra V, Simoncelli E P. End-to-end optimized image compression [C]// Proceedings of the International Conference on Learning Representations. Toulon, 2017: 1611.01704.
- [5] Ballé J, Minnen D, Singh S, et al. Variational image compression with a scale hyperprior [C]// Proceedings of International Conference on Learning Representations. Vancouver, 2018: 1802.01436.
- [6] Liu J H, Lu G, Hu Z H, et al. A unified end-to-end framework for efficient deep image compression [EB/OL]. (2020-02-09) [2025-05-10]. <https://arxiv.org/abs/2002.03370>.
- [7] 乔思波, 庞善臣, 王敏, 等. 基于残差混合注意力机制的脑部 CT 图像分类卷积神经网络模型 [J]. *电子学报*, 2021, 49(5): 984-991. (Qiao Si-bo, Pang Shan-chen, Wang Min, et al. A convolutional neural network for brain CT image classification based on residual hybrid attention mechanism [J]. *Acta Electronica Sinica*, 2021, 49(5): 984-991.)
- [8] 张诗源, 赵桐溪, 戚飞越, 等. 基于小波变换的智能生物医学图像分类算法 [J]. *应用数学进展*, 2025(3): 16-25. (Zhang Shi-yuan, Zhao Tong-xi, Qi Fei-yue, et al. Intelligent biomedical image classification algorithm based on wavelet transform [J]. *Advances in Applied Mathematics*, 2025(3): 16-25.)
- [9] 江贵平, 秦文健, 周寿军, 等. 医学图像分割及其发展现状 [J]. *计算机学报*, 2015, 38(6): 1222-1242. (Jiang Gui-ping, Qin Wen-jian, Zhou Shou-jun, et al. Medical image segmentation and its development status [J]. *Chinese Journal of Computers*, 2015, 38(6): 1222-1242.)
- [10] 周涛, 董雅丽, 霍兵强, 等. U-Net 网络医学图像分割应用综述 [J]. *中国图象图形学报*, 2021, 26(9): 2058-2077. (Zhou Tao, Dong Ya-li, Huo Bing-qiang, et al. U-Net and its applications in medical image segmentation: a review [J]. *Journal of Image and Graphics*, 2021, 26(9): 2058-2077.)
- [11] 刘飞, 张俊然, 杨豪. 基于深度学习的医学图像识别研究进展 [J]. *中国生物医学工程学报*, 2018, 37(1): 86-94. (Liu Fei, Zhang Jun-ran, Yang Hao. Research progress of medical image recognition based on deep learning [J]. *Chinese Journal of Biomedical Engineering*, 2018, 37(1): 86-94.)
- [12] 苏华强, 雷海军, 雷柏英. 多分支特征融合分类网络用于 CXR 图像识别 [J]. *信号处理*, 2025, 41(2): 253-266. (Su Hua-qiang, Lei Hai-jun, Lei Bai-ying. Multi-branch feature fusion classification network for chest X-ray image recognition [J]. *Journal of Signal Processing*, 2025, 41(2): 253-266.)
- [13] Duan L Y, Liu J Y, Yang W H, et al. Video coding for machines: a paradigm of collaborative compression and intelligent analytics [J]. *IEEE Transactions on Image Processing*, 2020, 29: 8680-8695.
- [14] Wang S R, Wang Z, Wang S Q, et al. End-to-end compression towards machine vision: network architecture design and optimization [J]. *IEEE Open Journal of Circuits and Systems*, 2021, 2: 675-685.
- [15] Girod B, Chandrasekhar V, Chen D M, et al. Mobile visual search [J]. *IEEE Signal Processing Magazine*, 2011, 28(4): 61-76.
- [16] 王凯. 基于双分支特征融合的高动态范围医学影像压缩研究 [D]. 哈尔滨: 哈尔滨工业大学, 2022. (Wang Kai. High dynamic range medical image compression based on two-branch feature fusion [D]. Harbin: Harbin Institute of Technology, 2022.)
- [17] Herbert R, Tuytelaars T, Gool L V. SURF: speeded up robust features [C]// Proceedings of the European Conference on Computer Vision. Graz, 2006: 404-417.
- [18] Redondi A, Cesana M, Tagliasacchi M. Rate-accuracy optimization in visual wireless sensor networks [C]// The 19th IEEE International Conference on Image Processing. Orlando, 2013: 1105-1108.
- [19] Liu L, Chen Z H, Hu Z H, et al. An efficient adaptive compression method for human perception and machine vision tasks [EB/OL]. (2025-01-08) [2025-05-10]. <https://arxiv.org/abs/2501.04329>.
- [20] 李基臣, 亓玉龙, 胡海瑞, 等. 数字图像处理技术在医学影像中的研究与应用 [J]. *电子技术与软件工程*, 2022(9): 194-197. (Li Ji-chen, Qi Yu-long, Hu Hai-rui, et al. Research and application of digital image processing technology in medical images [J]. *Electronic Technology & Software Engineering*, 2022(9): 194-197.)

(下转第 51 页)