

· 论著 ·

## 基于 TransUNet 改进的多任务模型在直肠癌新辅助治疗中的应用\*

尹淑文<sup>1</sup> 丁志鹏<sup>2</sup> 李岩<sup>1</sup> 高国闯<sup>1</sup> 程丽霞<sup>1</sup> 陈英东<sup>1</sup> 王文晗<sup>1</sup> 张秋菊<sup>1△</sup> 周洋<sup>2△</sup>

**【摘要】目的** 本研究基于深度学习方法对分割模型 TransUNet 做部分改进,构建既可分割又可预测分类的多任务模型,并应用于直肠癌新辅助治疗患者敏感人群识别的核磁图像中。**方法** 多任务模型在 TransUNet 基础上添加分类结构,具体为全连接层(输入为 512,输出为 256)、ReLU 激活函数、全连接层(输入为 256,输出为 3),实现三分类结局的预测,即疾病稳定(stable disease, SD),部分反应(partial response, PR)和完全反应(complete response, CR)。以 2015—2017 年哈尔滨医科大学附属肿瘤医院收治的 71 名直肠癌患者新辅助化疗前的 3D 核磁图像提取为 2D 图像作为研究数据,采用 Dice 系数和豪斯多夫距离评价分割效果,准确率、micro-precision、micro-recall 和 micro-F1 score 评价分类效果。**结果** 模型训练了 100 个轮次,分割任务在测试集上的平均 Dice 系数和平均豪斯多夫距离分别为 0.851 和 10.806,分类任务从切片角度,测试集上的准确率、micro-precision、micro-recall 和 micro-F1 score 均为 0.615,从患者角度,四个指标均为 0.857。**结论** 分割任务效果良好,分类任务以切片为单位,效果一般,以患者为单位,考虑到结局为三分类,尚且可以接受。该多任务模型有潜力用于临床进行辅助诊断。

**【关键词】** 深度学习 多任务学习 结直肠癌 新辅助治疗

**【中图分类号】** R195.1 **【文献标识码】** A **DOI** 10.11783/j.issn.1002-3674.2025.01.001

## Improved Multitask Model based on TransUNet in the Neoadjuvant Therapy for Rectal Cancer

Yin Shuwen, Ding Zhipeng, Li Yan, et al (Department of Health Statistics, School of Public Health, Harbin Medical University, Harbin 150081)

**【Abstract】Objective** We improved the segmentation model TransUNet based on deep learning methods to construct a multitask model that can both segment regions of interest and predict classification, which made it possible to identify sensitive populations of patients undergoing neoadjuvant therapy for rectal cancer. **Methods** The multitask model added classification structure on the basis of TransUNet, including the fully connected layer (input of 512, output of 256), the ReLU activation function, and the fully connected layer (input of 256, output of 3), to achieve the prediction of triple classification outcomes (stable disease (SD), partial response (PD), and complete response (CR)). The 3D MRI of 71 rectal cancer patients before neoadjuvant chemotherapy admitted to the Harbin Medical University Cancer Hospital from 2015 to 2017 were extracted into 2D images as data for the study. Dice coefficient and Hausdorff distance were used to evaluate thesegmentation performance, and accuracy, micro-precision, micro-recall, and micro-F1 score were used for classification performance. **Results** The model was trained for 100 epochs, and the average Dice coefficient and average Hausdorff distance for the segmentation task on the test set were 0.851 and 10.806, respectively. For the classification task, the accuracy, micro-precision, micro-recall, and micro-F1 score on the test set were all 0.651 from the slicing perspective, and all four metrics were 0.857 from the patient perspective. **Conclusion** Our model works well in the segmentation task. Although the model performs poorly on the classification task at the slice level, the performance was acceptable at the patient level taking into account the tripartite classification results. The multitask model has the potential to be used in the clinic for assisted diagnosis.

**【Key words】** Deep learning; Multitask learning; Colorectal cancer; Neoadjuvant therapy

结直肠癌是常见的消化系统恶性肿瘤,2022 年国家癌症中心发布的报告显示,我国结直肠癌新发病例和死亡病例分别为 40.8 万和 19.56 万,在全部癌种中分别排第二和第四位,且发病率和死亡率呈现上升趋势<sup>[1]</sup>。根据当前美国癌症联合委员会(American joint committee on cancer, AJCC)制定的肿瘤分期手册,结直肠癌分为 I、II、III、IV 期,其中,II~III 期为局部进展期,多数患者确诊时已处于该期,其根治术的预后显

著低于 I 期。目前,对于局部进展期直肠癌(locally advanced rectal cancer, LARC)的标准治疗方法是新辅助放化疗(neoadjuvant chemoradiotherapy, nCRT)加全直肠系膜切除术。新辅助治疗副反应小,患者耐受性好,而且能有效降低临床分期,控制术后转移并提高保肛率<sup>[2]</sup>。有文献报道接受新辅助治疗的患者中大约 20%~30% 治疗后发生临床完全缓解(complete clinical response, cCR)<sup>[3]</sup>, 7%~38% 发生病理完全缓解(pathological complete response, pCR)<sup>[4]</sup>。对于达到 cCR 的患者,建议观察等待策略,可以避免手术创伤,保留肛门直肠功能<sup>[5]</sup>。所以,新辅助治疗前识别出敏感患者非常重要。

有研究从分子标志物角度对新辅助放化疗的效果

\* 基金项目:国家自然科学基金项目(82073666;82473735);黑龙江省自然科学基金(LH2022H067)

1. 哈尔滨医科大学公共卫生学院卫生统计学教研室(150081)

2. 哈尔滨医科大学附属肿瘤医院影像中心

△通信作者:张秋菊, E-mail: qjzh\_81@126.com;周洋, E-mail: zhouyang094@126.com

进行预测,目前还未用于临床<sup>[6]</sup>;也有采用影像组学的方法通过图像预测<sup>[7-8]</sup>,使用预先定义的特征,一定程度上影响了分类的准确性和可靠性。随着深度学习技术的发展,越来越多的研究通过卷积神经网络(convolutional neural network, CNN)提取特征,实现分类或分割任务。卷积神经网络有两个归纳偏置,即局部性和平移等变性,但在学习远程依赖关系时存在局限。近年来,Transformer 模型受到了广泛关注,它基于自注意力机制,能够捕获全局信息,但缺乏卷积神经网络的归纳偏置,且对样本量要求高,数据集较小时不能发挥优势<sup>[9-10]</sup>。TransUNet 模型结合了 U-Net 与 Transformer 的优点,应用于图像分割任务<sup>[11]</sup>。本研究旨在构建一个多任务模型,在 TransUNet 的基础上加入分类结构,进行影像分割的同时,实现患者新辅助化疗三分类结局的预测。

### 材料与与方法

#### 1. 数据来源

本研究数据来源于哈尔滨医科大学附属肿瘤医院,包括 2015—2017 年收治的 71 名直肠癌患者新辅助化疗前的 3D 核磁图像,由具有多年临床经验的放射科医生对肿瘤区域进行标注。

#### 2. 数据预处理

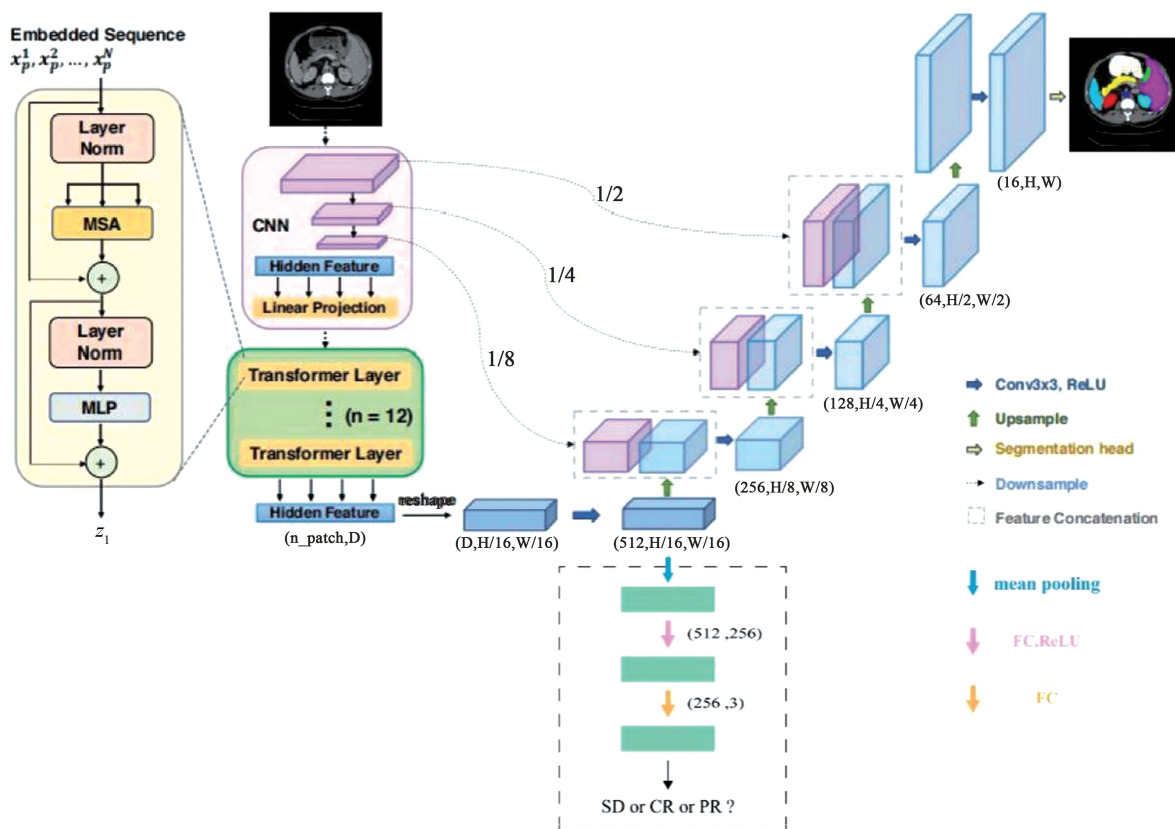
TransUNet 使用 2D 图像训练模型,本研究将每位

患者的 3D 核磁图像提取为 2D 图像,共 636 张。其中的一张图像只包含一个像素点,可用信息有限,将其剔除,最终纳入分析 635 张。由于肿瘤标记区域太小,需要先确定其边界,然后对图像沿边界裁剪,所有图像均统一为 224 \* 224 大小。纳入的 71 名患者中,随机选择 64 位患者的图像用于模型训练,7 位患者的图像用于测试,评价模型效果。

#### 3. 研究方法

本研究以 Chen 等人提出的 TransUNet 模型作为主体结构<sup>[11]</sup>,该模型整体上采用 U-Net 框架,先通过卷积神经网络提取特征,然后将特征图输入标准 Transformer 结构中,两者组合构成编码器。解码器将编码后的特征进行上采样,并与 CNN 提取到的浅层特征进行融合,实现精准定位。TransUNet 模型的任务是分割,本研究在其基础上做部分改进,将单任务模型(分割)改为多任务模型(分割+分类),如图 1 所示。对 TransUNet 底部输出的特征图进行平均池化,然后,添加两个全连接层,两者之间加入 ReLU 激活函数。全连接层的输入分别为 512 和 256,输出分别为 256 和 3,进行分类预测。

分类时,模型可以直接得到单张图像的预测结果。对于患者类别的判定,研究中将每位患者全部图像分类结果的概率分布按列相加求均值,均值最大者对应的类别判定为患者的类。



(注:分割任务参考 Chen et al. 2021<sup>[11]</sup>,虚线框内的结构实现分类任务,为本研究设计)

图 1 多任务模型网络架构

### 4. 模型训练

编码器部分,使用了预训练模型,由 ResNet-50 和 Vision Transformer 组成,在 ImageNet 上完成。训练时,初始学习率设为 0.01,之后根据迭代次数进行衰减,选择 SGD 优化器,每次输入 16 个样本,共训练 100 个轮次。通过损失函数评价模型预测值与真实值的吻合程度,分割部分选择 Dice 损失和交叉熵损失,分类部分选择交叉熵损失,对三者分别加以 0.25, 0.25 和 0.5 的权重求得总损失。

### 5. 评价指标

#### (1) 分割效果评价

采用 Dice 系数和豪斯多夫距离评价分割的效果。Dice 系数通常用于判断两个集合的相似程度,取值位于 [0,1] 之间,其计算公式为:

$$\text{Dice 系数} = \frac{2 \times |X \cap Y|}{|X| + |Y|}, X \text{ 和 } Y \text{ 分别代表真实标签}$$

和分割结果。

豪斯多夫距离用来度量同一样本空间中集合 A 和 B 之间的距离,定义如下:

$$d_H(A, B) = \max(d(A, B), d(B, A))$$

$$d(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\|$$

$$d(B, A) = \max_{b \in B} \min_{a \in A} \|b - a\|$$

$\|a - b\|$  和  $\|b - a\|$  代表集合 A 和集合 B 之间距离的范数,一般采用  $L_2$  范数, A 和 B 分别代表真实标签和分割结果。

#### (2) 预测分类效果评价

采用准确率、micro-precision、micro-recall、micro-F1 score 等指标评价模型分类效果,以表 1 中示意数据为例计算各指标。

表 1 三分类结局混淆矩阵

混淆矩阵	预测值			
	0	1	2	
真实值	0	a	b	c
	1	d	e	f
	2	g	h	i

准确率:预测正确的样本占总样本的百分比,根据表 1,其计算公式为:

$$\text{准确率} = \frac{a+e+i}{a+b+c+d+e+f+g+h+i}$$

micro-precision: micro-precision 常作为多分类问题精确率评价指标,指的是预测为真且真值为真的样本数所占比例,其计算公式为:

$$\text{micro-precision} = \frac{TP_0+TP_1+TP_2}{TP_0+TP_1+TP_2+FP_0+FP_1+FP_2}$$

以 0 分类为例,  $TP_0$  代表真实分类为 0 且预测分类为 0 的真阳性数,即表 1 中 a;  $FP_0$  代表真实分类为非 0 但预测分类为 0 的假阳性数,即表 1 中 (d+g)。其余指标解释类似。

micro-recall: 召回率表示真值为真且预测为真的样本数所占比例, micro-recall 可作为多分类召回率评价指标,其计算公式为:

$$\text{micro-recall} = \frac{TP_0+TP_1+TP_2}{TP_0+TP_1+TP_2+FN_0+FN_1+FN_2}$$

以 0 分类为例,  $TP_0$  代表真实分类为 0 且预测分类为 0 的真阳性数,即表 1 中 a;  $FN_0$  代表真实分类为 0 但预测分类为非 0 的假阴性数,即表 1 中 (b+c)。其余指标解释类似。

micro-F1 score: micro-F1 score 同时考虑了模型的精确率和召回率,取值范围在 0 到 1 之间,其计算公式为:

$$\text{micro-F1 score} = 2 \times \frac{\text{micro-precision} \times \text{micro-recall}}{\text{micro-precision} + \text{micro-recall}}$$

micro-precision、micro-recall、micro-F1 score 和准确率在数值上相等。

## 结 果

### 1. 基线信息

本研究纳入患者 71 名,平均年龄  $55.39 \pm 10.38$  岁,其中男性 52 例 (73.24%),女性 19 例 (26.76%);新辅助治疗后,疾病稳定 28 例、发生部分病理反应 26 例、发生完全病理反应 17 例,三组患者的具体信息如表 2 所示。

表 2 患者基线信息 [n (%)]

变量	SD (N=28)	PR (N=26)	CR (N=17)
性别			
男	23(82.14)	19(73.08)	10(58.82)
女	5(17.86)	7(26.92)	7(41.18)
年龄	54.96±10.71	57.35±10.52	53.12±9.62
吸烟			
是	12(42.86)	11(42.31)	5(29.41)
否	16(57.14)	15(57.69)	12(70.59)
饮酒			
是	9(32.14)	6(23.08)	3(17.65)
否	19(67.86)	20(76.92)	14(82.35)
治疗前 N 分期			
发生转移	21(75.00)	12(46.15)	7(41.18)
未发生转移	7(25.00)	14(53.85)	10(58.82)
治疗前肿瘤直径 (mm)	54.04±12.14	45.04±11.93	45.5±12.18
治疗前肿瘤厚度 (mm)	17.89±3.81	15.92±4.59	15.71±3.67
治疗前肛距 (cm)	5.97±2.53	5.90±1.74	4.59±2.41
病理类型			
腺癌	22(78.57)	22(84.62)	0
粘液癌	4(14.29)	2(7.69)	1(5.88)
未说明	2(7.14)	2(7.69)	16(94.12)
病理分化			
低分化	1(3.57)	0	0
中分化	22(78.57)	21(80.77)	1(5.88)
高分化	3(10.71)	3(11.54)	0(0.00)
未说明	2(7.14)	2(7.69)	16(94.12)

注:SD 代表疾病稳定;PR 代表部分反应;CR 代表完全反应

## 2. 模型训练

经过 100 个轮次的训练,学习率和总损失的变化趋势如图 2、图 3 所示。随着迭代次数的增加,学习率逐渐下降,最后接近于 0;总损失整体呈波动下降,训练结束时,降至 0.05 以下。

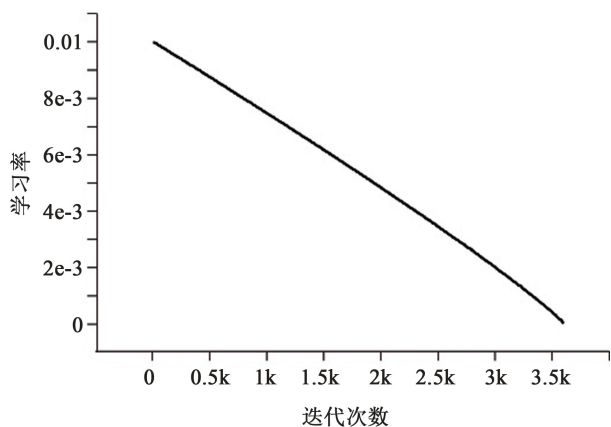


图 2 训练过程学习率变化趋势

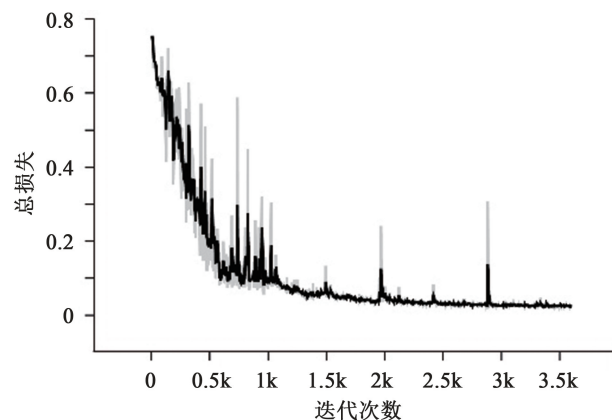


图 3 训练过程总损失变化趋势

## 3. 评价指标

### (1) 分割任务

对于整个测试集,平均 Dice 系数为 0.851,平均豪斯多夫距离为 10.806。每位患者选择 1 张切片,原始图像、真实标签、预测标签、Dice 系数和豪斯多夫距离如表 3 所示。

表 3 部分测试集原始图像、真实标签、预测标签、Dice 系数及豪斯多夫距离

编号	原始图像	真实标签	预测标签	Dice 系数	豪斯多夫距离
1				0.883	3.606
2				0.843	8.944
3				0.942	1.000
4				0.758	21.260
5				0.877	8.000
6				0.812	17.889
7				0.884	5.000

### (2) 预测分类任务

从切片角度, 准确率、micro-precision、micro-recall 和 micro-F1 score 均为 0.615; 从患者角度, 以上四个指标均为 0.857。每位患者新辅助治疗效果的真实分类和预测分类如表 4 所示。

表 4 每位患者的真实分类和预测分类

患者	真实分类	预测分类
1	0	0
2	0	0
3	1	1
4	2	0
5	1	1
6	1	1
7	1	1

注:0 代表疾病稳定,1 代表部分反应,2 代表完全反应

### 讨 论

TransUNet 为单任务模型, 本研究在其基础上加入分类结构, 构建了多任务模型, 同时完成直肠癌患者影像的分割和新辅助治疗效果的分类。研究结果显示, 分割任务在测试集上平均 Dice 系数为 0.851, 效果良好; 分类任务在以切片为单位的测试集上效果一般, 但以患者为单位, 且考虑到结果为三分类, 效果可以接受。

磁共振作为一种常规检查手段, 在结直肠癌的诊断中发挥着重要作用。传统模式下, 临床医生根据经验阅片, 给出诊断结论, 这难免受到主观因素的影响。近年来, 深度学习在计算机视觉领域快速发展, 而且核磁图像获取简单, 二者结合辅助诊断越来越多地应用于实践。这种形式极大增加了图像分割和分类预测的准确性, 也有助于疾病的早期检测, 为进一步治疗提供便利。本研究构建的深度学习模型在训练过程中, 结合了迁移学习以降低样本量小带来的影响<sup>[12]</sup>。最终, 我们得到了尚且满意的结果, 说明该模型在临床实践中具有一定的应用价值。

本研究也存在着一定的不足。首先, 原始数据为 3D 核磁图像, 但样本量仅有 71, 难以满足 3D 模型的训练要求, 我们从中提取了 2D 切片, 既满足了 TransUNet 对输入的要求, 也在一定程度上增加了样本量, 理论上讲, 这样操作损失了一定的信息, 比如, 空间结构。其次, 目标区域占整张图像的比例很小, 我们对图像进行裁剪并统一了大小, 但这样处理使原始图像变得模糊, 质量降低, 可能影响模型的效果。最后, 分类

任务中, 随机选择的 7 名患者只有 1 名发生了完全反应, 从分类结果看, 模型错分了这名患者, 可能是测试集中该类患者数量太少导致的随机结果, 也可能是训练集中发生完全反应的患者数量少于另外两类, 模型没有充分学习这类图像的特征。

对本研究而言, 我们还需继续收集图像, 并尝试对 TransUNet 模型进行升维, 在样本量允许时, 完成 3D 模型的训练。此外, 除单独使用图像, 还可以与临床信息或多种组学比如基因组学、蛋白组学等结合, 对患者的生存结局、疾病分型或分子亚型等进行预测, 从而为患者提供个性化的治疗方案。

### 参 考 文 献

[ 1 ] Zheng R, Zhang S, Zeng H, et al. Cancer incidence and mortality in China, 2016 [ J ]. Journal of the National Cancer Center, 2022, 2 ( 1 ): 1-9.

[ 2 ] 付尚志, 张佳节, 李万平. 直肠癌术前新辅助化疗研究进展 [ J ]. 临床军医杂志, 2015, 43(4): 409-412.

[ 3 ] Glynne-Jones R, Hughes R. Complete Response after Chemoradiotherapy in Rectal Cancer ( Watch-and-Wait ): Have We Cracked the Code [ J ]. Clinical Oncology, 2016, 28(2): 152-160.

[ 4 ] 彭海花, 游凯云, 高远红, 等. 直肠癌新辅助化疗后病理完全缓解的相关研究进展 [ J ]. 中华胃肠外科杂志, 2013, 16(6): 592-596.

[ 5 ] 李心翔, 李大卫. 直肠癌新辅助化疗临床完全缓解后处理的争议与共识 [ J ]. 中国实用外科杂志, 2020, 40(3): 325-329.

[ 6 ] Daye D, Tanaka I, Jain R, et al. Predictive and Prognostic Molecular Biomarkers for Response to Neoadjuvant Chemoradiation in Rectal Cancer [ J ]. International Journal of Molecular Sciences, 2017, 18 ( 3 ): 573.

[ 7 ] Horvat N, Veeraraghavan H, Khan M, et al. MR Imaging of Rectal Cancer: Radiomics Analysis to Assess Treatment Response after Neoadjuvant Therapy [ J ]. Radiology, 2018, 287(3): 833-843.

[ 8 ] Wang F, Tan BF, Poh SS, et al. Predicting outcomes for locally advanced rectal cancer treated with neoadjuvant chemoradiation with CT-based radiomics [ J ]. Scientific Reports, 2022, 12(1): 6167.

[ 9 ] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale [ EB/OL ]. ( 2021-06-03 ) [ 2022-07-14 ]. <https://arxiv.org/abs/2010.11929>.

[ 10 ] Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need [ EB/OL ]. ( 2017-06-12 ) [ 2022-12-05 ]. <https://arxiv.org/abs/1706.03762>.

[ 11 ] Chen J, Lu Y, Yu Q, et al. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation [ EB/OL ]. ( 2021-02-08 ) [ 2022-09-16 ]. <https://arxiv.org/abs/2012.04306>.

[ 12 ] 黎英, 宋佩华. 迁移学习在医学图像分类中的研究进展 [ J ]. 中国图象图形学报, 2022, 27(3): 672-686.

(责任编辑:张悦)