

## 学术前沿专栏:甲骨智能计算

# FEC-PVT: 基于 PVT 架构的甲骨钻凿图像分割网络

刘国奇<sup>a,b</sup>, 李文格<sup>a,b</sup>, 茹琳媛<sup>a,b</sup>, 宋黎明<sup>a,b</sup>, 刘杰<sup>a,b</sup>, 韩燕彪<sup>b,c</sup>

(河南师范大学 a.计算机与信息工程学院; b.甲骨智能计算实验室; c.历史文化学院,河南 新乡 453007)

**摘要:**由于长时间埋藏于地下和风化腐蚀,造成甲骨片破损和甲骨钻凿边界模糊不易分辨,给甲骨钻凿分割带来极大挑战.从甲骨数据库及著录书中系统收集并标注甲骨钻凿图像.基于该数据集,提出一种以 Transformer 为编码器的甲骨钻凿分割网络 FEC-PVT (feature extraction and connection pyramid vision transformer).首先, FEC-PVT 利用 FE\_C 和 FE\_D 模块分别补充低层和高层特征,以获取细节和全局特征;其次, FCOM 模块用交叉注意力让不同层特征交互,获取有效细节;最后, FFDM 模块逐层解码并整合多层次特征,提升解码精度,避免特征丢失.实验验证,所提 FEC-PVT 优于其他的方法,与次优的 DuAT 方法相比,IoU 提高 5.18%.

**关键词:**图像分割;甲骨钻凿;金字塔视觉变换器;卷积神经网络

**中图分类号:**TP391

**文献标志码:**A

**文章编号:**1000-2367(2026)01-0008-09

甲骨作为研究殷商时期的经济、政治、文化发展的重要历史材料<sup>[1]</sup>,不仅其文字内容具有极高的研究价值,甲骨非文字资料同样是研究甲骨学和殷商时期历史的重要内容,其与当时的占卜密不可分.甲骨钻凿是甲骨非文字资料的一个重要组成部分,通过对钻凿形态的判断以及钻凿特点的分析可以帮助研究判断甲骨所属的具体年代和其相关的历史.对于甲骨钻凿形态的正确判断和刻画是研究相关甲骨非文字形态的重要内容,但是目前对于甲骨钻凿形态的判断和刻画大多停留在人工判断阶段,也有少部分学者利用计算机方面的技术来研究.利用计算机相关技术解决甲骨钻凿形态刻画的问题,不仅可以帮助研究人员节约大量人力物力,还对甲骨钻凿边界的描述判断起着指导性的作用.

目前,图像分割任务大多是利用深度学习的方法来确定目标物体的边界.基于编码器-解码器架构的卷积神经网络被广泛应用于图像分割任务中,并取得了显著的成果.2022 年,GAO 等<sup>[2]</sup>提出一种由分割和检测子网组成的双流卷积神经网络(OBM-CNN)来对古甲骨文盾纹进行分割,来界定分辨破损的甲骨文.2022 年, FUJIKAWA 等<sup>[3]</sup>提出利用 YOLO 和 MobileNet 构成一个在线 OBI 识别系统来实现字符的自动检测和识别,第一阶段使用 YOLO 来检测和识别甲骨字符,第二阶段利用 MobileNet 来识别第一阶段未检测到的甲骨文.2023 年,宋传鸣等<sup>[4]</sup>提出一种采用逐步求精的层次分割策略的连通区域拓扑结构约束的甲骨拓片图像分割方法,以达到精细分割结果.2024 年,史小松等<sup>[5]</sup>提出基于 DeepLabV3+ 的甲骨拓片图像文字分割方

**收稿日期:**2024-12-10; **修回日期:**2025-03-17.

**基金项目:**国家社会科学基金重大项目(20&ZD305);国家社会科学基金青年项目(23CZS001);国家自然科学基金青年科学基金项目(C类)(62506121);河南省高等学校智库研究项目(2026ZKYJ38).

**作者简介:**刘国奇(1984—),男,河南新乡人,河南师范大学教授,博士,研究方向为图像分类, E-mail:liuguoqi080408@163.com.

**通信作者:**韩燕彪, E-mail:1430095515@qq.com.

**引用本文:**刘国奇,李文格,茹琳媛,等.FEC-PVT: 基于 PVT 架构的甲骨钻凿图像分割网络[J].河南师范大学学报(自然科学版),2026,54(1):8-16.(Liu Guoqi, Li Wenge, Ru Linyuan, et al. FEC-PVT: An oracle bone drilling image segmentation network based on PVT architecture[J]. Journal of Henan Normal University(Natural Science Edition), 2026, 54(1): 8-16. DOI:10.16366/j.cnki.1000-2367.2024.12.10.0003.)

法,该方法在处理甲骨文拓片文字分割任务有较高的准确率,为甲骨文字的识别、甲骨图片的缀合等提供一定的基础数据。2024年,MENG等<sup>[6]</sup>提出利用YOLOv8实现甲骨文的自动分割,并达到了较好的性能。以上的工作虽然专注于甲骨的分割,但主要集中在对甲骨文的分割,而没有关注到甲骨非文字资料的价值。2024年,LIU等<sup>[7]</sup>提出Gca-pvt-net用于甲骨钻凿的分割,Gca-pvt-net使用群体卷积注意和PVT双分支网络,在甲骨钻凿分割上达到了较好的分类效果。但里面所选取的数据大多是椭圆形长凿与圆孔互包型的钻凿形态,而对其他形态的甲骨数据无过多研究。本文将在数据集上进行扩充,将形态从仅有互包型扩展到3种钻凿形态。

以往图像分割依赖于人们的手动标注和识别,费时费力。后来随着计算机技术的发展,研究人员利用机器学习的方法来进行图像分割,但是使用机器学习的方法进行图像分割,仍需要手动设计和选择特征。目前针对图像分割任务的方法大多属于深度学习的范畴,深度学习可以通过深度神经网络提取到数据的不同特征,关注到人眼无法关注到的细微特点,帮助人们辅助处理图像数据。通过深度学习来处理图像数据,比传统的机器学习更为方便,不需要像传统机器学习一样需要手动设计和选择特征,深度学习的方法可以达到自动特征提取。目前常见的深度学习方法处理图像分割任务有U-Net<sup>[8]</sup>、DuAT<sup>[9]</sup>、PVTv2<sup>[10]</sup>、TransFuse<sup>[11]</sup>、AC-SNet<sup>[12]</sup>、CarpNet<sup>[13]</sup>、SegNet<sup>[14]</sup>等。这些网络在他们对应的任务中,有着不错的分割效果,但是其应用到本文所使用的甲骨钻凿分割数据集上时,效果并不理想。

由于甲骨大多长时间埋藏于地下,在挖掘出土之后,甲骨片大多有所破损,其钻凿边界也大多变得模糊不清,不能很好地刻画出钻凿的形态。针对目前所存在的分类网络大多在甲骨钻凿分割数据集上的分割效果不好,本文提出FEC-PVT方法利用PVTv2作为骨干网络,同时增加针对于甲骨钻凿特点添加的3个模块FE\_C、FE\_D、FCOM以及FFDM,帮助分割出边缘信息模糊不清的甲骨钻凿图像。本文的贡献包括以下4个部分:

1)结合现有的甲骨钻凿的数据集和学者对于甲骨钻凿的形态划分方法:小屯甲骨钻凿形态,制作了包含椭圆形长凿与圆孔互包型、单独椭圆型以及圆孔型的甲骨钻凿数据集。

2)设计了特征提取模块FE\_C和FE\_D模块,用于针对PVTv2的底层输出和高层输出进行特征再提取,突出细节特征和全局特征,帮助分辨出钻凿的边界和识别有破损部分的甲骨钻凿。

3)利用特征交互模块FCOM将不同层特征进行连接融合,FCOM模块利用交叉注意力来获取不同层次特征之间的相关性,可以实现特征的有效交互,进而得到更利于分割的特征信息。

4)设计特征解码模块FFDM,针对提取的特征信息,利用逐层递进的解码操作补偿高级语义特征的空间边界信息缺失和低级语义特征的语义信息缺失。

## 1 数据来源及数据集构建

### 1.1 数据来源

本研究以权威甲骨数据库与最新考古著录为核心数据来源,依托“殷契文渊”“甲骨文拓片”“甲骨世界”等数字化平台,系统收集《故宫博物院藏殷墟甲骨文》《中国国家博物馆馆藏文物研究丛书:甲骨卷》《复旦大学藏甲骨集》《吉林大学藏甲骨集》《河南藏甲骨集成》等近年出版著录书的高清影像,构建覆盖不同材质(龟甲/兽骨)、组类的钻凿图像数据集。

### 1.2 数据集构建

数据集中关于甲骨钻凿的有3种形态,分别是单独椭圆型如图1(a)、圆孔型如图1(b)以及椭圆形长凿与圆孔互包型如图1(c)。在这3种形态中,椭圆型长凿与圆孔互包型较为常见,且大多数较完整的甲骨片上均为互包型。相较之下,圆孔型和单独椭圆型的钻凿形态则较少,且多分布于残缺的甲骨片上,或者只在一个较完整的甲骨片上有零散的几个圆孔型或者单独椭圆型的钻凿,其余均为互包型。由此可知,在钻凿形态上存在着严重的类别不平衡问题,椭圆型长凿与圆孔互包型较多,而单独椭圆型和圆孔型较少。对于我们所搜集到的甲骨钻凿图像,在甲骨专家的指导下,根据甲骨钻凿的形态,采用Labelme软件进行对图像进行数据标注,创建了甲骨钻凿数据集。



图1 甲骨钻凿形态示例  
Fig.1 Example of oracle bone drills forms

## 2 方法

本文所提 feature extraction and connection pyramid vision transformer(FEC-PVT)的网络结果如附录图 S1 所示,该网络主要由编码器 PVTv2,特征提取模块 FE\_C、FE\_D,特征交互模块 FCOM,以及特征融合解码模块 FFDM.该模型首先利用 PVTv2 进行特征提取,产生 4 个不同阶段的图像特征.其次,分别使用 FE\_C 和 FE\_D 模块作用于 PVTv2 网络第 2 层和第 4 层的输出特征,对底层特征进行特征增强,对高层特征进行补充.然后利用特征交互模块 FCOM,进行不同层次之间特征交互.最后使用解码器模块 FFDM 进行最后的输出.

同时,该模型使用深度监督学习的方法,在若干个中间层引入监督信号,而并不仅依赖于模型最终输出层的监督.深监督不仅有助于加快模型的训练速度,还能够增强中间层特征的表征能力,同时,其优势在于适应多尺度特征的高效学习.实验证明通过该模型可以有效提升甲骨钻凿分割的精度和准确度,同时也可以辅助学者研究甲骨非文字资料.

### 2.1 编码器 PVTv2 架构

PVTv2<sup>[9]</sup>作为 Transformer 的一个创新性的模型,将 Transformer 模型和 CNNs 中的金字塔结构相结合,使得网络的输出并不是简单的固定大小的图像块,而是可以获得不同尺度上的图像特征,从而提升对图像的分割精度.PVTv2 网络的特征层共有 4 层,分别是  $F_1, F_2, F_3, F_4$ .在这 4 层里面,前两层  $F_1, F_2$  作为低级特征,这两层所提取到的特征一般为浅层特征,但包含着许多无关的噪声.后两层  $F_3, F_4$  作为高级特征,这两层所提取的特征一般为全局特征,获得的特征也更加全面,相比前两层而言,少了许多的细节部分.4 层的输出特点不一,可以实现对图像特征的有效获取.

### 2.2 特征提取模块 FE\_C、FE\_D

针对 PVTv2 网络所捕获的低层与高层特征,仍然存在进一步增强和优化的空间,本文设计了特征提取模块(feature extraction\_conv module,FE\_C 和 feature extraction\_dilated\_conv module,FE\_D)分别作用于  $F_2$  层和  $F_4$  层的输出.FE\_C 和 FE\_D 模块分别利用普通卷积和空洞卷积来帮助捕获更多有效的特征,实现对图像特征更加有效的提取.

针对浅层特征,本文选择对  $F_2$  层使用 FE\_C 模块.因为  $F_1$  层所包含的细节特征较多,包含的噪声也相较更多,FE\_C 作用于  $F_1$  层会造成产生更多冗余无关的噪声,不利于获取对图像分割有效的图像特征.考虑到浅层特征所获得的大多是细节特征,而普通卷积专注于提取局部特征,能够有效地帮助模型捕获更丰富的细节表达.因此 FE\_C 模块主要由普通卷积组成,FE\_C 模块如图 2(1)(a) 所示.其整体流程可用式(1)表示.

$$\left\{ \begin{array}{l} x_1 = \bigoplus_{i=1}^2 \text{Con } v_3^i(x), \\ x_2 = \bigoplus_{s \in \{3,5,7\}} \text{Con } v_s(\text{Con } v_5(\text{Con } v_3(\text{Con } v_3(x)))), \\ x_3 = \bigoplus_{s \in \{3,7\}} \text{Con } v_s(\text{Con } v_3(x)) \oplus \text{Con } v_3 D_5(\text{Con } v_5(\text{Con } v_3(x))), \\ x_{\text{out}} = \text{SE}(\text{BN}(\sum_{i \in \{1,2,3\}} x_i)) + \text{CoordAttention}(x), \end{array} \right. \quad (1)$$

其中,  $x$  表示输入的特征,  $\oplus$  表示 cat 操作,  $\text{Con } v_i$  中的  $i$  表示的是卷积核的大小,  $\text{Con } v_3 D_5$  表示卷积核为 3, 空洞率为 5 的空洞卷积。需要注意的是,  $\text{Con } v_3$  由两个卷积组成的, 一个卷积核为  $3 \times 1$ , 另一个卷积核为  $1 \times 3$ 。利用两个卷积核为  $3 \times 1$  和  $1 \times 3$  的卷积替代卷积核为  $3 \times 3$  的卷积, 不仅可以减少参数量和计算量, 还可以捕获不同方向的特征。

针对高级特征, 本文选择对  $F_4$  层使用 FE\_D 模块。因为对于  $F_4$  层的输出结果而言, 包含着全局性的信息, 但是也在获取这些特征的同时, 丢失了许多特征信息。通过使用 FE\_D 模块, 可以捕获更加精细、更加全面、更深层次的特征, 有助于提升图像分割的精度。由于  $F_4$  的输出属于高级特征, 如果使用普通卷积会造成特征更加集中于局部的特征, 而作为高层特征输出应当获取全局性特征。因此对于 FE\_D 模块, 本文主要使用空洞卷积来帮助扩大感受野, 得到更加全局性的特征, FE\_D 模块如图 2(1)(b) 所示。

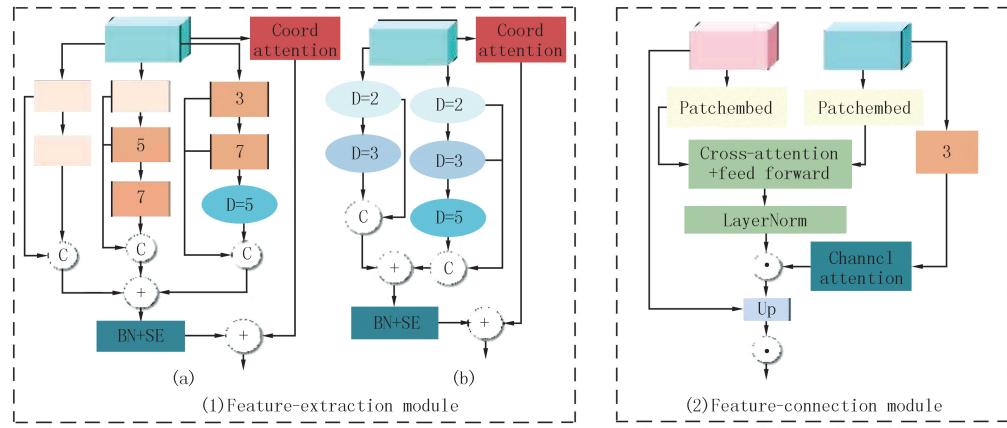


图2 主要模块示意图

Fig. 2 Schematic diagram of main modules

其整体流程可用公式(2)表示, 其中  $x$  表示输入的特征,  $\oplus$  表示 cat 操作,  $\text{Con } v_3 D_i$  表示的是卷积核大小为  $3 \times 3$ , 空洞率为  $i$  的空洞卷积。

$$\begin{cases} x_1 = \bigoplus_{d \in \{2,3\}} \text{Con } v_3 D_d(x), \\ x_2 = \bigoplus_{d \in \{2,3,5\}} \text{Con } v_3 D_d(x), \\ x_{\text{out}} = \text{SE}(\text{BN}(\sum_{i \in \{1,2\}} x_i)) + \text{CoordAttention}(x). \end{cases} \quad (2)$$

通过使用 FE\_C 模块和 FE\_D 模块分别对底层特征和高层特征进行特征补充, 可以帮助获取到更丰富的细节特征和更全面的全局特征。

### 2.3 特征交互模块 FCOM

针对 PVTv2 模块输出的 4 层特征, 本文通过特征交互模块(Feature connection module, FCOM)进行连接融合, 以获得更加全面的特征帮助图像分割。FCOM 模块利用交叉注意力来获取不同层次特征之间的相关性, 可以实现特征的有效交互, 进而得到更利于分割的特征信息。FCOM 模块如图 2(2) 所示。

首先, 将 FE\_D 模块的输出结果和 PVTv2 的  $F_3$  层输出结果利用 FCOM 模块进行处理, 得到  $\text{FCOM}_3$ ; 再将  $\text{FCOM}_3$  和 FE\_C 模块的输出结果通过 FCOM 模块进行特征融合, 得到  $\text{FCOM}_2$ ; 最后将  $\text{FCOM}_2$  和 PVTv2 的  $F_1$  层输出结果利用 FCOM 模块进行特征融合, 得到  $\text{FCOM}_1$ 。 $\text{FCOM}_i, i \in \{1, 2, 3\}$  将在解码阶段提供更多丰富的特征信息。

在 FCOM 模块内部, 首先通过块嵌入(patch embedding)操作将两个输入特征转换为高维向量表示, 并对 FE\_D 模块的输出和  $\text{FCOM}_i, i \in \{2, 3\}$  进行卷积和注意力的作用, 得到  $\text{F}_C$ 。接着, 将经过 Patch Embedding 的结果一同输入交叉注意力机制, 来获得高层特征和底层特征之间的关联性。并将交叉注意力获得的结果进行归一化处理得到  $\text{FF}_P$ 。归一化后的  $\text{FF}_P$  和  $\text{F}_C$  再通过卷积融合, 得到初步融合的特征。然后, 对其进行上采样, 并与 PVTv2 网络的  $F_1$  层、 $F_3$  层以及 FE\_C 模块的输出进行卷积融合, 最终得到 FCOM 模块

的输出.

## 2.4 特征融合解码模块 FFDM

对于从 FCOM 模块获得的 3 个尺寸不同的输出特征  $FCOM_i, i \in \{1, 2, 3\}$ , 使用特征融合解码模块 (feature fusion decoding module, FFDM) 进行特征的解码, 来还原详细的特征信息.

首先, 将  $FCOM_3$  进行上采样, 并且使用卷积核为 1 的卷积来调整  $FCOM_3$  的尺寸和  $FCOM_2$  一致, 随后  $FCOM_3$  和  $FCOM_2$  通过 cat 操作组合成  $FCOM_{2\_3}$ , 并将  $FCOM_{2\_3}$  的大小调整至与  $FCOM_2$  相同. 再将  $FCOM_{2\_3}$  与低层特征  $FCOM_1$  进行相应的特征融合, 并将最后的尺寸调整为  $FCOM_1$  的尺寸.

特征解码模块 (FFDM), 针对提取的特征信息进行逐级递进的解码操作, 通过层次化的解码策略逐层恢复细节, 提升对边缘特征的捕获, 补偿高级语义特征的空间边界信息缺失和低级语义特征的语义信息缺失, 以提高分割性能.

综上所述, 本文所提的 FEC-PVT 模型通过使用 PVTv2 提取到 4 个不同尺度大小的特征, 并使用 FE\_C 和 FE\_D 模块对第 2 层和第 4 层的输出特征进行特征再补充; 然后使用 FCOM 模块将所捕获的低级特征和高级特征进行融合; 最后使用 FFDM 模块对特征进行层层递进的解码, 恢复到原始的图像大小.

## 3 实验

本节主要介绍所用数据集、评估指标、超参数设置以及实验运行环境. 同时, 给出与其他模型的对比结果和相关的消融实验, 来评估每个模块的有效性.

### 3.1 实验细节

开发环境为 Ubuntu 20.04、python 3.11 和 PyTorch 2.3, 单个 NVIDIA L40S GPU. 该模型共训练了 150 个 epoch, 使用 AdamW 优化器, 其学习率和权重衰减参数均设置为 0.000 1. 输入图像尺寸为  $544 \times 320$ , 并采用数据增强方法, 包括随机  $90^\circ$  旋转、随机水平翻转 ( $p=0.5$ )、随机垂直翻转 ( $p=0.5$ ), 以提高模型的鲁棒性. 同时, 采用标准化处理, 以匹配预训练模型的输入分布.

为验证所提 FEC-PVT 模型的有效性, 本文使用 Dice 系数<sup>[15]</sup>、IoU<sup>[16]</sup>、Se、Acc<sup>[17]</sup>、HD<sup>[18]</sup> 等 5 个指标来衡量实验结果.

### 3.2 数据集

本文所使用的数据集为依据专家指导所标注的甲骨钻凿数据集, 整个数据集共 133 张图像. 数据大小被调整为  $544 \times 320$ , 并大致以 8 : 1 : 1 的比例划分数据集, 其中训练数据集占 80%.

### 3.3 对比实验

为了证明所提模型的有效性, 本文与一些常见的分割网络: Gca-pvt-net<sup>[7]</sup>、U-Net<sup>[8]</sup>、DuAT<sup>[9]</sup>、TransFuse<sup>[10]</sup>、CaraNet<sup>[13]</sup> 等进行比较. 为保证实验的公平性, 所有实验使用相同的数据集和实验环境. 实验数据结果如表 1 所示.

表 1 所提方法和其他分割方法的分割结果

Tab. 1 Segmentation results of the proposed method and other segmentation methods

Methods	Pub. Year	Dice	IoU	Se	Acc	HD95
ACSNNet <sup>[12]</sup>	2020	0.770 4	0.638 1	0.785 6	0.977 3	23.206 6
CaraNet <sup>[13]</sup>	2021	0.808 8	0.685 4	0.815 5	<u>0.978 8</u>	17.253 8
DuAT <sup>[9]</sup>	2022	<u>0.809 9</u>	<u>0.685 5</u>	0.773 1	0.978 1	15.683 8
Polyp-PVT <sup>[19]</sup>	2021	0.799 7	0.671 1	0.775 4	0.977 4	18.455 4
SegNet <sup>[14]</sup>	2017	0.629 0	0.472 8	0.527 4	0.967 1	28.205 0
TransFuse <sup>[11]</sup>	2021	0.755 4	0.612 2	<u>0.828 4</u>	0.971 4	<u>10.697 6</u>
U-Net <sup>[8]</sup>	2015	0.752 3	0.614 0	0.672 4	0.977 2	21.998 6
U-Net++ <sup>[20]</sup>	2017	0.707 7	0.562 2	0.616 5	0.973 0	22.107 9
Gca-pvt-net <sup>[7]</sup>	2024	0.790 6	0.680 7	0.742 2	0.949 3	17.335 6
Ours	2026	<b>0.847 2</b>	<b>0.737 3</b>	<b>0.873 8</b>	<b>0.982 0</b>	<b>9.092 9</b>

从表 1 的结果可以看出,本文的所提模型在甲骨钻齿数据集上可以获得更高的分割精度. 其中, 最好的数字用粗体表示, 次好的用下划线表示. 由表 1 分析可知, 本文所提方法的 Dice 和 IoU 指标分别比次优的 DuAT 方法高出 3.73% 和 5.18%; 对于 Se 指标, 本文所提方法相比次优方法 TransFuse 提高 4.54%; ACC 指标, 相比次优的 CaraNet 方法提高了 0.32%; HD 值也相较于 TransFuse 方法的 HD 值减少了 1.604 7.

附录图 S2 为对比实验的分割结果可视化, 其中绿色表示 TP, 黄色表示 FN, 红色表示 FP. 从上到下, 所选甲骨对应的著录编号分别是:《合集》17598 反、《合集》34604 反、《合集》655 反、《合集》6635 反. 分析图 S2 可知, 本文所提出的方法可以分割出大多数的钻齿形态. Gca-pvt-net 模型可以分割出大部分的椭圆型长齿与圆孔互包型钻齿, 且分割准确度较高, 但是对于圆孔型和单独椭圆型却不能很好地分割出来. DuAT 模型和 CaraNet 模型也可以分割出大部分的钻齿形态, 但是也有较多误分区域, 整体精度不高. TransFuse 模型则会受到甲骨片破损的影响, 错将破损部分分割为甲骨钻齿, 而产生误分的钻齿部分.

U-Net 和 SegNet 均主要由 CNNs 组成, 虽能获得大部分的细节特征, 却缺乏全局性的特征, 因此会丢失全局上下文信息. 同时会受到甲骨破损部分的影响, 将非钻齿部分误分为钻齿部分. 除此之外, ACSNet 和 U-Net++ 网络也有类似问题. 而本文所提的方法基于 PVTv2 模型, 不仅有细节特征, 还有全局性特征, 可以帮助更好地去分割甲骨钻齿, 弥补了上述网络中所存在的不足.

DuAT 模型和 CaraNet 模型虽然也有着较高的分割精度, 但是这两种方法的 HD 值均较高, 说明其分割的边界与真实标签之间的边界误差较大. 从相关指标对比结果来看, 本文所提模型在甲骨钻齿分割任务上有良好的性能, 尽管钻齿有 3 种不同形态, 所提模型分割时均能有效分割出目标钻齿. 但是对于少数破损严重的钻齿, 并未达到预期目标.

综合来看, FEC-PVT 模型通过使用 PVTv2 编码器、FE\_C、FE\_D、FCOM 和 FFDM 模块, 不仅边界分割效果良好, 整体上也有着较好的分割性能.

### 3.4 消融实验

为了证明所提模型中各个模块的性能, 本节进行了相关的消融实验, 包括模型中各个模块的消融, 以验证所提模型的有效性和 FE 模块的性能. 模型中各个模块的消融实验结果如表 2 所示, 其中 FE 代表特征提取模块 FE\_C 和 FE\_D, FCOM 代表特征交互模块, FFDM 代表特征融合解码模块. FE\_C 和 FE\_D 模块的消融实验结果如表 3 和表 4 所示, 其中  $FE_{-C_i}$  ( $i=1, 2, 3$ ) 代表  $FE_C$  模块中 3 个不同卷积层,  $C_i$  代表每层卷积核大小或空洞率,  $FE_{-D_i}$  ( $i=1, 2$ ) 代表  $FE_D$  模块中两个不同卷积层,  $C_i$  代表每层卷积核的空洞率. 本节将分析消融实验结果, 证明各个模块的有效性.

表 2 本文所提模块在甲骨钻齿数据集上的消融实验结果

Tab. 2 Ablation experimental results on the oracle bone drill dataset proposed in this paper

Order	Methods	Dice/%	IoU/%	Se/%	Acc/%	HD
a	Baseline(PVTv2)	79.97	67.11	77.54	97.74	18.455 4
b	Baseline+FE	81.84	69.71	86.86	97.92	15.334 7
c	Baseline+FFDM	82.73	70.98	86.51	97.96	13.720 1
d	Baseline+FCOM	82.84	71.03	84.71	98.09	14.278 9
e	Baseline+FE+FFDM	83.06	72.13	84.29	98.10	14.587 5
f	Baseline+FE+FCOM	83.01	71.52	86.17	98.00	14.829 6
g	Baseline+FFDM+FCOM	82.52	70.56	85.96	97.98	15.255 3
h	Ours	<b>84.72</b>	<u>73.73</u>	<b>87.38</b>	<b>98.20</b>	<b>9.092 9</b>

FE 模块的有效性: FEC-PVT 模型利用 FE\_C 和 FE\_D 模块分别作用于浅层特征和高级特征, 以捕获有利于提升分割精度的细节特征和全局特征, 也为后续的解码操作提供更加丰富的信息. 从表 2 中的 a 和 b 的结果对比中可以看出, 加上 FE 模块可以明显提升模型分割的 Dice 和 IoU 指标, 同时也提升了分割的准确性. 对比 g 和 h 结果可以发现, FE 模块的缺少会导致分割的 Dice 指标和 IoU 指标分别降低 2.20% 和 3.23%; 同时, HD 值也增加了 6.162 4. 因此, 使用 FE 模块可获取更加充分的细节特征和全局特征帮助分割图像, 提升分割的性能.

表3 FE\_C 模块在甲骨钻凿数据集上的消融实验结果

Tab. 3 Ablation of FE\_C modules on an oracle bone drill and chisel dataset

Order	Convolutional layer parameter settings	Methods			Dice/%	IoU/%	HD			
		FE_C								
		FE_C <sub>1</sub>	FE_C <sub>2</sub>	FE_C <sub>3</sub>						
a	C <sub>1</sub> :(3×3) * 2	√	√	√	83.13	71.54	14.247 3			
	C <sub>2</sub> :3×3+3+3									
	C <sub>3</sub> :3+3+3									
b	C <sub>1</sub> :3×3+5	√	√	√	83.62	72.09	9.011 6			
	C <sub>2</sub> :3×3+7+7									
	C <sub>3</sub> :3+D5+D5									
c	C <sub>1</sub> :(3×3)×2	√	√	√	83.16	71.63	14.024 5			
	C <sub>2</sub> :3×3+5+5									
	C <sub>3</sub> :3+7+7									
d	C <sub>1</sub> :(3×3)×2	√	×	×	81.71	69.74	19.359 8			
e	C <sub>2</sub> :3+5+7	√	√	×	82.39	70.53	16.921 8			
f	C <sub>3</sub> :3+5+D5	√	√	√	84.72	73.73	9.092 9			

表4 FE\_D 模块在甲骨钻凿数据集上的消融实验结果

Tab. 4 Ablation of FE\_D modules on an oracle bone drill and chisel dataset

Order	Convolutional layer parameter settings	Methods		Dice/%	IoU/%	HD			
		FE_D							
		FE_D <sub>1</sub>	FE_D <sub>2</sub>						
a	D <sub>1</sub> :D <sub>2</sub>	√	√	82.39	70.47	15.720 2			
	D <sub>2</sub> :D <sub>3</sub>								
b	D <sub>1</sub> :D <sub>2</sub>	√	√	83.26	71.59	10.033 5			
	D <sub>2</sub> :D <sub>2</sub> +D <sub>3</sub>								
c	D <sub>1</sub> :D <sub>3</sub> +D <sub>5</sub>	√	√	83.18	71.56	14.401 3			
	D <sub>2</sub> :D <sub>3</sub> +D <sub>5</sub> +D <sub>7</sub>								
d	D <sub>1</sub> :D <sub>2</sub> +D <sub>3</sub>	√	×	80.80	68.40	18.982 9			
	D <sub>2</sub> :D <sub>2</sub> +D <sub>3</sub> +D <sub>5</sub>								
e		×	√	82.47	70.57	14.187 2			
f		√	√	84.72	73.73	9.092 9			

**FE\_C 模块的有效性:** FE\_C 模块作为浅层特征提取模块,帮助捕获有利于提升分割精度的细节特征,为后续的特征融合提供更加丰富的细节特征,表3为 FE\_C 模块的消融实验结果.首先,从 a,b,c 和 f 结果对比中可以看出,相比 a 中的 3 个卷积层均使用 3×3 的卷积核,f 中使用不同尺寸的卷积核会得到更好的效果,IoU 指标提升了 2.19%;相比 b 和 c 的不同卷积核大小,f 有着更好的分割精度,IoU 指标也有所提升.因此,f 中的卷积核大小设置是较为合理的.其次,从 d,e 和 f 的结果对比中可以看出,f 比只有一层和两层卷积层的 d,e,IoU 值分别提高了 3.99% 和 3.20%,同时有着更小的 HD 值.因此,f 中的卷积层数设置是合理的.

**FE\_D 模块的有效性:** FE\_D 模块作为高级特征提取模块,帮助捕获有利于提升分割精度的全局特征,为后续的特征融合提供更加有效的全局特征,表4为 FE\_D 模块的消融实验结果.首先,从 a,b,c 和 f 结果对比中可以看出,相比 a,b 和 c 的不同空洞率大小或卷积核的个数,f 的 IoU 指标提升了 2.14%~3.26%,同时有着较小的 HD 值.因此,f 中的空洞率大小以及卷积核的个数的设置较为合理.其次,从 d,e,f 的结果对比中可以看出,当卷积层为两层时会获得较好的分割结果,f 比只有一层卷积层的 d,e,IoU 值分别提高了 5.33% 和

3.16%, 同时有着更小的 HD 值. 通过 d, e, f 的结果对比可知, f 中的卷积层数设置是合理的.

FCOM 模块的有效性: FCOM 作为融合浅层特征和高层特征的模块, 通过使用交叉注意力来获取不同层次特征之间的关联, 并生成融合后的特征. 对比 a 和 d 的结果可知, 增加 FCOM 模块可以有效提升分割的性能, IoU 提升了 3.92%. 从中可以看出增加对浅层和高层特征的融合, 可以实现局部特征和全局特征的交互, 进而提升分割的精度. 通过对比 e 和 h 的结果, 可以看出 FCOM 对于模型的必要性, 删去 FCOM 之后, 模型的 HD 值增加了 5.494 6, 同时 Dice 和 IoU 的值也均有所降低. 由此可见, 增加 FCOM 模块可以实现局部特征和全局特征的有效交互, 进而模型可更好地分割出目标.

FFDM 模块的有效性: FFDM 作为模型的解码器模块, 通过层层递进的方法进行解码. 相比直接将模型的各层输出变成原始大小再连接的方法而言, FFDM 方法利用层层递进的方法进行解码, 可以使得每一层的特征在解码过程中都尽可能避免特征的丢失. 从 a 和 c 的结果可知, 加上融合模块 FFDM 之后, Dice 增加了 2.76%, IoU 增加了 3.87%. 同时获得了更低的 HD 值, 相比增加 FFDM 模块之前, HD 值降低了 4.735 3. 通过对比 f 和 h 的结果, 可知当去除 FFDM 模块会导致 HD 值的增加和 Dice 以及 IoU 指标的降低, 不利于模型获得良好的分割效果.

## 4 总 结

本文主要针对甲骨非文字资料——甲骨钻凿形态进行研究, 通过分析甲骨钻凿数据集的特点及其所存在的甲骨破损, 钻凿边界模糊的问题, 提出了一种基于 PVTv2 编码器的网络 FEC-PVT. 该网络通过使用 FE\_C 和 FE\_D 模块对 PVTv2 提取的浅层和高层特征进行进一步的特征提取, 从而有效增强细节特征的表达能力, 并提升全局特征的捕捉效果, 获得更加丰富且全面的特征表征. 其次, 利用 FCOM 模块对提取到的浅层和高层特征进行融合, 最后再利用 FFDM 模块进行解码操作, 并获取最后的输出结果. 同时, FEC-PVT 模型使用深监督的方法, 在若干个中间层引入监督信号, 而并不仅依赖于模型最终输出层的监督, 帮助训练深层网络, 提高中间层特征的表征能力. 实验证明 FEC-PVT 针对甲骨钻凿分割任务有着良好的性能.

由于甲骨上面的破损与钻凿类似, 仍有区域会被错误分类, 我们未来将尝试抗噪声相关的模型应用于甲骨钻凿分割, 以获得更加准确的分割结果. 目前, 甲骨钻凿数据集的数量较少, 在后续的研究中, 我们将扩充完善数据集, 将模型应用于更加丰富的甲骨钻凿分割数据集上.

附录见电子版(DOI:10.16366/j.cnki.1000-2367.2024.12.10.0003).

## 参 考 文 献

- [1] 李雪山, 韩燕彪.殷商和西周甲骨钻凿异同探析[J].河南师范大学学报(哲学社会科学版), 2017, 44(6): 89-94.  
LI X S, HAN Y B. Similarities and differences between Oracle bone drilling in Yin Shang Dynasty and Western Zhou Dynasty[J]. Journal of Henan Normal University(Philosophy and Social Sciences), 2017, 44(6): 89-94.
- [2] GAO W Z, CHEN S X, ZHANG C S, et al. OBM-CNN: a new double-stream convolutional neural network for shield pattern segmentation in ancient oracle bones[J]. Applied Intelligence, 2022, 52(11): 12241-12257.
- [3] FUJIKAWA Y, LI H Y, YUE X B, et al. Recognition of oracle bone inscriptions by using two deep learning models[J]. International Journal of Digital Humanities, 2023, 5(2): 65-79.
- [4] 宋传鸣, 周雨晴, 张晋豪, 等. 连通区域拓扑结构约束的甲骨拓片图像分割[J]. 闽南师范大学学报(自然科学版), 2023, 36(4): 35-50.  
SONG C M, ZHOU Y Q, ZHANG J H, et al. Image segmentation for oracle bone rubbings with topological structure constraint of connected regions[J]. Journal of Minnan Normal University(Natural Science), 2023, 36(4): 35-50.
- [5] 史小松, 代记圆. 基于 DeepLabV3+ 的甲骨拓片图像文字分割方法[J]. 信息技术与信息化, 2024(4): 216-220.  
SHI X S, DAI J Y. Text segmentation method of Oracle rubbings image based on DeepLabV3+[J]. Information Technology and Informationization, 2024(4): 216-220.
- [6] MENG X Y, PU H T, MENG F. Automatic segmentation of oracle bone inscriptions using YOLOv8[J]. Procedia Computer Science, 2024, 242: 1074-1081.
- [7] LIU G Q, YANG Y P, LI X S, et al. Gca-pvt-net: group convolutional attention and PVT dual-branch network for oracle bone drill chisel segmentation[J]. Heritage Science, 2024, 12: 260.

[8] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation [EB/OL]. [2024-11-16]. <https://arxiv.org/abs/1505.04597>.

[9] TANG F L, XU Z X, HUANG Q M, et al. DuAT: dual-aggregation transformer network for Medical image segmentation [C]// Pattern Recognition and Computer Vision. Singapore: Springer, 2024: 343-356.

[10] WANG W H, XIE E Z, LI X, et al. PVT v2: Improved baselines with Pyramid Vision Transformer [J]. Computational Visual Media, 2022, 8(3): 415-424.

[11] ZHANG Y D, LIU H Y, HU Q. TransFuse: fusing transformers and CNNs for medical image segmentation [C]// Medical Image Computing and Computer Assisted Intervention-MICCAI 2021. Cham: Springer, 2021: 14-24.

[12] ZHANG R F, LI G B, LI Z, et al. Adaptive context selection for polyp segmentation [C]// Medical Image Computing and Computer Assisted Intervention-MICCAI 2020. Cham: Springer, 2020: 253-262.

[13] LOU A G, GUAN S Y, LOEW M. CaraNet: context axial reverse attention network for segmentation of small medical objects [J]. Journal of Medical Imaging, 2023, 10(1): 014005.

[14] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.

[15] MILLETARI F, NAVAB N, AHMADI S A. V-net: fully convolutional neural networks for volumetric medical image segmentation [C]// 2016 Fourth International Conference on 3D Vision (3DV). Stanford: IEEE, 2016: 565-571.

[16] YU J H, JIANG Y N, WANG Z Y, et al. UnitBox: an advanced object detection network [C]// Proceedings of the 24th ACM International Conference on Multimedia. [S.I.]: ACM, 2016: 516-520.

[17] CHANG H H, ZHUANG A H, VALENTINO D J, et al. Performance measure characterization for evaluating neuroimage segmentation algorithms [J]. NeuroImage, 2009, 47(1): 122-135.

[18] HUTTENLOCHER D P, KLANDERMAN G A, RUCKLIDGE W J. Comparing images using the Hausdorff distance [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1993, 15(9): 850-863.

[19] DONG B, WANG W H, FAN D P, et al. Polyp-PVT: polyp segmentation with pyramid vision transformers [J]. CAAI Artificial Intelligence Research, 2023: 9150015.

[20] ZHOU Z W, SIDDIQUEE M M R, TAJBAKHSH N, et al. Unet++: redesigning skip connections to exploit multiscale features in image segmentation [J]. IEEE Transactions on Medical Imaging, 2020, 39(6): 1856-1867.

## FEC-PVT: An oracle bone drilling image segmentation network based on PVT architecture

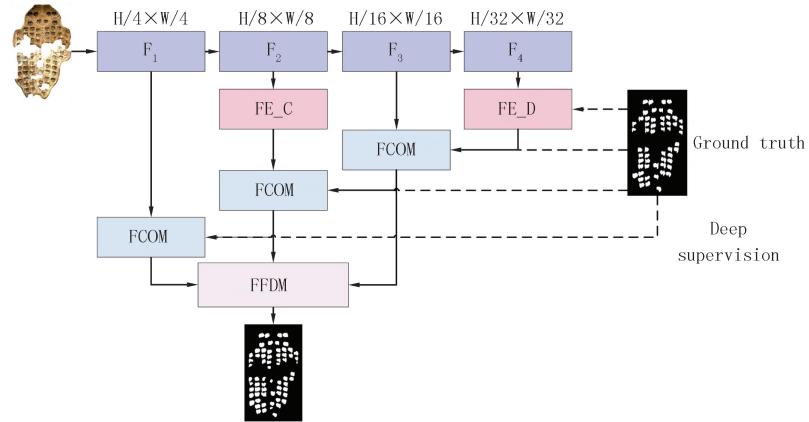
Liu Guoqi<sup>a,b</sup>, Li Wenge<sup>a,b</sup>, Ru Linyuan<sup>a,b</sup>, Song Liming<sup>a,b</sup>, Liu Jie<sup>a,b</sup>, Han Yanbiao<sup>b,c</sup>

(a. College of Computer and Information Engineering; b. Oracle Intelligent Computing Laboratory;  
c. College of History and Culture, Henan Normal University, Xinxiang 453007, China)

**Abstract:** Due to long-term burial in the ground and weathering and corrosion, the oracle bone fragment is damaged and the boundary of the oracle bone drill is blurred and difficult to distinguish, which brings great challenges to the division of the oracle bone drill and chisel. This study systematically collects and annotates oracle bone drill and chisel images from oracle bone databases and cataloged books. Based on this dataset, this paper proposes a feature extraction and connection pyramid vision transformer (FEC-PVT) with Transformer as encoder. FEC-PVT uses FE\_C and FE\_D modules to supplement low-level and high-level features, respectively, to obtain detailed and global features. Secondly, the FCOM module uses cross-attention to allow different layer features to interact and obtain effective details. Finally, the FFDM module decodes layer by layer and integrates multi-level features to improve the decoding accuracy and avoid feature loss. Experimental verification shows that the FEC-PVT proposed in this paper outperforms other methods. Compared with the suboptimal DuAT method, its IoU increases by 5.18%.

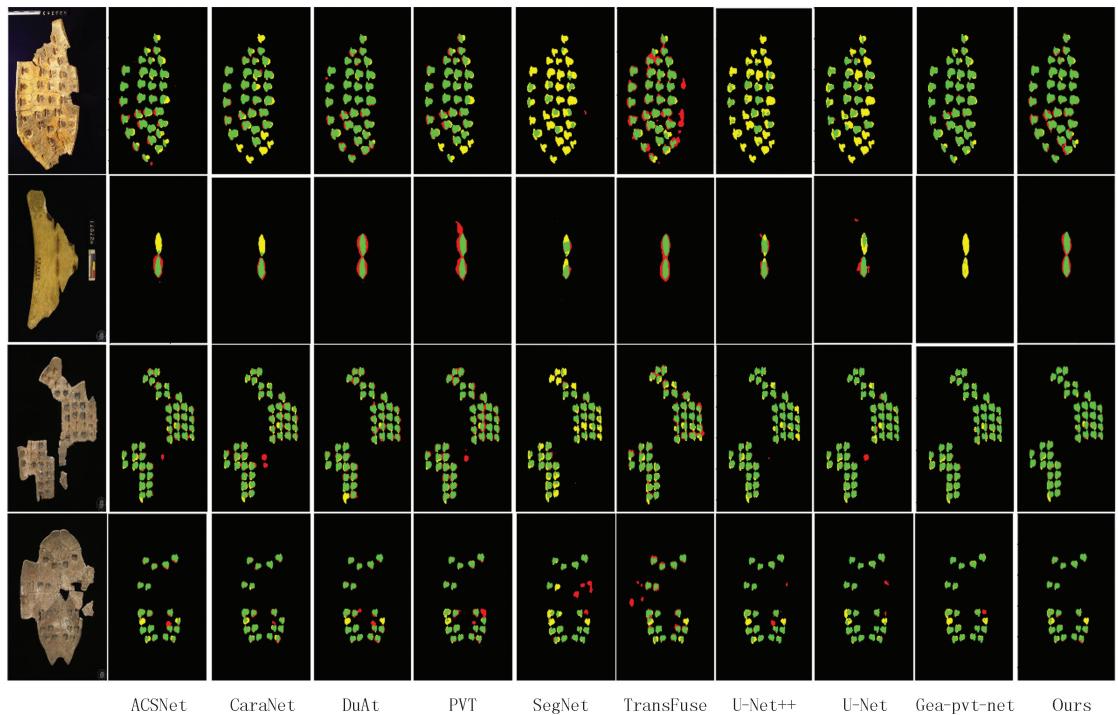
**Keywords:** image segmentation; oracle bone drilling; pyramid vision transformer; convolutional neural networks

## 附 录



图S1 FEC-PVT网络架构

Fig. S1 FEC-PVT network architecture



图S2 分割结果的可视化

Fig. S2 Visualization of segmentation results