

学号：2021012992



西北农林科技大学
NORTHWEST A&F UNIVERSITY

2025 届本科生毕业论文（设计）

题目：基于注意力机制的奶牛面部精细分割

学院(系):	信息工程学院
专业年级:	计算机 2102 班
学生姓名:	胡圣权
指导老师:	冯妍
合作指导老师:	
完成日期:	2025 年 6 月

基于注意力机制的奶牛面部精细分割

摘要：在智慧农业蓬勃发展的背景下，奶牛养殖业朝着精准化、智能化方向迈进，奶牛面部分割作为实现个体精细化管理的关键技术，其发展滞后成为制约奶牛养殖智能化升级的瓶颈。传统管理模式存在效率低、误差大等缺陷，深度学习与图像处理技术的发展为解决这一问题提供了新途径。

本研究提出一种基于注意力机制的奶牛面部精细分割模型，我们以 DeeplabV3+ 为框架，在模型中引入了 ECA 通道注意力模块、FPN 特征融合模块和 SSH 语义增强模块。首先我们通过 ECA 模块聚焦关键面部特征，抑制背景噪声。然后再利用 FPN 模块融合深浅层多尺度特征，丰富特征表达。最后借助 SSH 模块优化特征还原能力，提升分割精度。为了完成实验，本研究构建了包含不同光照、角度和个体差异的奶牛面部图像数据集。在经过数据预处理和标注后，将其划分为训练集、验证集和测试集。

实验结果表明，本研究提出一种基于注意力机制的奶牛面部精细分割模型在 mIOU、mPA 和 ACC 三项指标上的性能分别达到 0.6684、0.7797 和 0.9793，显著优于 Unet、PSPnet 等主流算法。消融实验的结果验证了 ECA、FPN 和 SSH 模块能够有效提升模型性能，时间性能分析探究了我们提出的算法与主流算法时间差异。此外，算法对背景表现出优秀的分割性能，对奶牛其他面部关键部位表现出近似的分割性能。综上，本研究为奶牛个体识别、健康监测等智能化应用提供了基础数据支持。未来可探索混合注意力机制融合、跨模态数据集成及数据集扩展等方向，进一步提升模型性能。

关键词：注意力机制；DeeplabV3+；特征融合；语义分割

Accurate Face Segmentation in Dairy Cows Based on Attention Mechanism

Against the backdrop of the vigorous development of smart agriculture, the dairy cattle breeding industry is moving towards precision and intelligence. However, the lagging development of cow face segmentation, a key technology for achieving individual fine-grained management, has become a bottleneck restricting the intelligent upgrading of dairy cattle breeding. Traditional management models suffer from defects such as inefficiency and large errors, while the development of deep learning and image processing technologies has provided new avenues to address this issue.

This study proposes a fine-grained cow face segmentation model based on attention mechanisms. Using DeeplabV3+ as the framework, the model incorporates an ECA channel attention module, an FPN feature fusion module, and an SSH semantic enhancement module. First, the ECA module focuses on key facial features while suppressing background noise. The FPN module then fuses multi-scale features from shallow and deep layers to enrich feature representation. Finally, the SSH module optimizes feature restoration capabilities to improve segmentation accuracy. To conduct the experiments, this study constructed a cow face image dataset containing different lighting conditions, shooting angles, and individual differences. After data preprocessing and annotation, the dataset was divided into training, validation, and test sets.

Experimental results show that the proposed attention-based cow face fine-grained segmentation model achieved performance metrics of 0.6684 in mIOU, 0.7797 in mPA, and 0.9793 in ACC, significantly outperforming mainstream algorithms such as Unet and PSPnet. Ablation experiments verified that the ECA, FPN, and SSH modules effectively enhance model performance, while time performance analysis explored the timing differences between the proposed algorithm and mainstream algorithms. Additionally, the algorithm demonstrated excellent segmentation performance on the background and comparable performance on other key facial regions of cows. In summary, this study provides fundamental data support for intelligent applications such as cow individual recognition and health monitoring. Future research can explore directions such as the integration of hybrid attention mechanisms, cross-modal data integration, and dataset expansion to further improve model performance.

Keywords: Attention mechanism; DeeplabV3+; Feature fusion; Semantic segmentation

目录

第 1 章	引言	1
1.1	研究的目标和意义	1
1.2	研究背景	1
1.3	相关领域研究进展	2
1.4	研究内容	3
1.5	内容安排	3
1.6	本章小结	4
第 2 章	相关理论概述与数据集制作	5
2.1	经典语义分割算法	5
2.1.1	U-Net	5
2.1.2	PSPNet	6
2.1.3	SegNet	6
2.1.4	FCN	6
2.2	注意力机制算法概述	7
2.3	评价指标	7
2.4	数据集制作	8
2.4.1	数据采集	8
2.4.2	数据预处理	8
2.4.3	数据标注	9
2.4.4	制作奶牛面部图像数据集	10
2.5	本章小结	10
第 3 章	基于注意力机制的奶牛面部精细分割模型的建立	11
3.1	基于注意力改进的奶牛面部精细化分割模型搭建	11
3.1.1	Backbone 主干网络	12
3.1.2	ECA 通道注意力模块	13
3.1.3	FPN 特征融合模块	14
3.1.4	SSH 语义增强模块	15
3.2	实验设计	15
3.3	本章小结	16
第 4 章	实验结果与分析	17
4.1	参数确定	17
4.2	消融实验	17

4.3	性能对比与分析	18
4.4	算法在奶牛各部位的性能分析	19
4.5	算法效率分析	21
4.6	本章小结	21
第 5 章	结论与展望	23
5.1	结论	23
5.2	展望	23
参考文献	25
致谢	29

第1章 引言

1.1 研究的目标和意义

随着智慧农业和精准养殖业的发展，奶牛养殖正逐步变得精准化和智能化。作为实现奶牛个体精细化管理的关键技术，奶牛面部分割一直在奶牛现代化养殖中扮演着重要角色。奶牛面部分割通过计算机视觉算法来区分奶牛面部和背景并分离奶牛面部的不同区域。然而，目前鲜有学者对奶牛面部分割领域的算法作进一步的研究。

传统的奶牛养殖管理主要依赖人工观察。这种方法在个体识别、健康监测等环节存在效率低、误差大、实时性差等问题。例如，通过耳标或颈圈进行个体识别易受环境干扰，而基于体表特征和奶牛体况量表的健康评估方法主观性过强，不同的兽医基于自身的经验会产生不同的判断。此外，兽医所有奶牛捕捉面部细微生理变化（如眼部分泌物异常、口鼻溃疡等早期疾病信号）需要大量的人力劳动实现，这使得传统兽医观察的方法在大规模牧场难以实现。随着深度学习与图像处理技术发展，利用计算机替代人工观察奶牛成为新的热点。奶牛面部分割作为其中的一种重要方法，这使得近年来奶牛面部分割技术逐渐被人们关注。

因此，本研究聚焦于提出一种基于深度学习算法对复杂养殖环境下的奶牛图像实现面部分割。首先我们将制作包含不同光照条件、拍摄角度、个体差异的奶牛面部图像数据集。其次，我们将搭建语义分割模型实现奶牛面部的精细分割。最终通过消融实验和对比分析主流分割算法的性能表现对我们提出的算法进行评价。该研究成果将为奶牛个体识别、健康状态评估、行为模式分析等智能化应用提供基础数据和技术支持，有助于实现养殖过程的自动化监控与精准化管理。同时有效提升奶牛养殖的生产效率与经济效益，推动智慧畜牧业的可持续发展。

1.2 研究背景

动物面部分割一直以来都是智慧畜牧业的发展的核心技术。近年来，随着计算机视觉与深度学习的发展。动物面部分割越来越受到学者们的重视。动物面部分割凭借其高效性和实时性为动物的个体识别、健康监测及行为分析提供重要的技术支持。这也使得动物面部分割成为提升养殖效率与动物福利的关键突破口。

近年来，国内外学者在研究动物面部分割的方法中取得了显著进展。李向宇和李慧盈（李向宇等，2024）提出了一种结合形变卷积网络（DCN）和调优卷积神经网络（TCNN）的猪脸特征点检测方法。在收集一个多场景下猪脸的数据集基础上，研究团队利用深度学习算法有效解决了猪脸图像变形的的问题。该方法实现了 5.60% 的低错误率，为牧场猪群的个体识别和健康状态评估提供了基础。何屿彤等（何屿彤等，2024）提出了 YOLOv3-DBSP 模型，该模型在 YOLOv3 中引入了密集块（DenseBlock）和空间金字塔池化（SPP）模块。该模型在羊脸检测任务中达到了 90.18% 的平均精度，比原始

YOLOv3 提高了 9.87%。Chen (Chen., 2024) 开发了一种用于猪面部分割多模态深度学习框架。该研究结合了热成像和可见光图像构建新的数据集。通过卷积神经网络 (CNN) 处理这两种模态数据,并设计特征融合模块来整合信息。实验结果表明,该模型在低光照条件下的分割准确率比单一模态提高了接近 15%。Li 等 (Li., 2024) 提出了一种轻量化的 Transformer 模型。该方法结合边缘计算设备,专门用于处理视频数据。该模型首先通过多尺度可变形注意力模块减少计算负担,其次引入统一查询表示模块,旨在实现并行处理分类、定位和掩码预测任务。实验结果表明该研究实现了每秒 30 帧的实时动物面部分割,准确率达 89%。郭松涛 (郭松涛等, 2022) 开发了一种名为 Tri-AI 的面部识别方法。该方法基于深度神经网络模型实现了野生川金丝猴的高精度识别。实验结果表明,该方法的平均准确率达 94%,其中秦岭种群的识别成功率达 95.6%。Peng 等 (Peng., 2023) 利用合成动物数据集 (SAP) 训练部位分割模型,提出了一种类平衡傅里叶数据混合 (CB-FDM) 方法,有效解决了合成数据与真实数据的分布差异问题。实验表明,该方法在老虎、马等四足动物的部位分割任务中性能显著提升,该方法为小样本学习提供了新思路。Chen (Chen., 2024) 采用 YOLOv4 算法构建了鸡部位检测系统,实现了鸡腿、鸡翅和鸡胸的实时目标检测。在定位鸡的关键部位贡献巨大。实验结果表明,该方法的平均精度 (mAP) 达到 96.8%。

虽然动物面部分割已经取得了一定的进展,但是鲜有学者研究奶牛面部的分割问题。为解决奶牛面部分割问题,本文提出了一种基于注意力机制的奶牛面部分割算法。本文以 DeeplabV3+ 为算法框架,通过添加 ECA, FPN 和 SSH 模块提升算法性能,实现对奶牛面部的精细分割。

1.3 相关领域研究进展

注意力机制是一种能够增强神经网络特征识别能力的算法。在图像分割领域中,注意力机制往往能够提升神经网络或深度学习模型的性能。早期的注意力机制源于 Bahdanau 等人 (Bahdanau, 2014) 在 NLP 中提出的序列对齐模型。该方法通过动态分配权重聚焦关键语义信息。随着深度学习的发展,注意力机制的发展和应用也逐渐走向成熟。Hu 等人 (Hu, 2018) 提出了 SENet 通过通道注意力模块首次阐述了自适应调整特征通道的重要性。随后,注意力机制井喷式发展,陆续出现了 ECA, CA 等更高效的注意力机制。自 Vaswani 等人 (Vaswani, 2017) 提出 Transformer 架构后。越来越多学者承认自注意力机制在计算机视觉领域中的应用前景。自注意力机制通过计算全局上下文依赖关系,有效捕捉长程语义关联。注意力机制目前已经成为计算机视觉领域必不可少的算法之一。

近年来,越来越多学者尝试将注意力机制应用在图像分割任务中。Oktay 等人 (Oktay, 2018) 提出的注意力 U-Net, 在编码器 - 解码器结构中引入空间注意力模块,旨在增强对目标边界和细节的关注。该方法显著提升医学影像分割精度。Dosovitskiy 等人 (Dosovitskiy, 2020) 的 ViT 和 Liu 等人 (Liu, 2021) 的 Swin Transformer 都通过自注意力机制建模全局语义,为注意力机制的应用提出了新的思路 and 方案。Wang 等人 (Wang, 2023) 提出了一种新的方法名为 SegViT, 该方法通过可学习的 class tokens 与空间特征交

互生成掩码，旨在增强普通 Vit 算法上的自注意力机制模块里的高效语义分割。此外，注意力机制与实时分割模型的融合成为新的热点，Bolya 等提出的 YOLACT (Bolya, 2019) 方法能够通过轻量化注意力设计。以视频为 i 数据集，实现了在移动端实现每秒 30 帧的实例分割。

1.4 研究内容

本研究首先收集奶牛面部图像数据，首先我们采集奶牛分娩疼痛视频并进行抽帧处理，随后我们对数据进行预处理，确保数据质量足以进行语义分割操作。数据预处理完成后，我们搭建基于注意力机制改进的语义分割模型框架，并在自制的奶牛面部数据集上训练语义分割模型，生成语义分割掩码和奶牛面部分割结果。在训练模型的过程中，需要进行模型的参数设置。通过优化模型参数，以获得模型的最佳性能。模型训练完成后，先进行消融实验，探究各个模块对 DeeplabV3+ 性能的提升程度和有效性。随后对比模型分割结果，分析模型在奶牛各个部分的分割效果，并对模型时间性能进行分析。本文以 7:3 的比例划分训练集和测试集，用以评估模型泛化能力。实验时我们采用 mIOU、mPA 和 Accuracy 等评价指标对模型进行性能评估。

本文具体的流程如图 1-1 所示

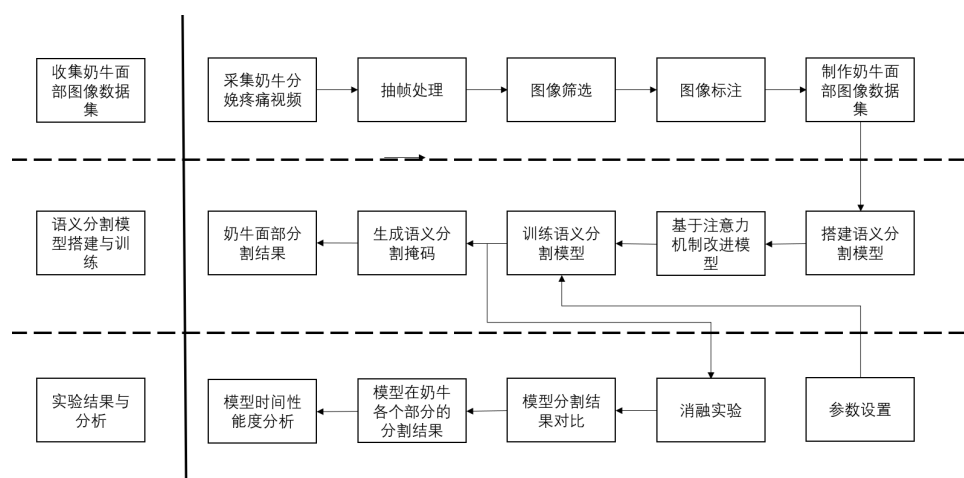


图 1-1 奶牛面部精细分割实验流程

1.5 内容安排

本文一共 5 个部分，每部分的具体内容安排如下：第 1 章引言，在引言中我们介绍研究背景、目标与意义，同时分析传统奶牛养殖管理瓶颈，阐述奶牛面部分割在智慧农业中的重要性，介绍研究方法与论文结构；第 2 章相关理论概述。该部分我们概述经典语义分割算法、注意力机制分类及应用。同时介绍了 mIOU、mPA、ACC 等评价指标；第 3 章基于注意力机制的奶牛面部精细分割模型建立，该部分详述了基于 DeeplabV3+ 框架的改进模型，包括 ECA 通道注意力、FPN 特征融合、SSH 语义增强模块的技术实

现及实验设计；第 4 章结果与分析，该部分介绍了奶牛面部图像数据集制作流程，展示了消融实验、主流算法性能对比及各部位分割性能分析等实现的实验结果；第 5 章结论与展望，该部分着重总结了研究成果，验证算法分割精度优势，提出混合注意力机制融合、跨模态数据集成等未来研究方向，展望算法在智慧畜牧中的应用前景。

1.6 本章小结

这一章主要讲述了是奶牛面部分割研究的相关情况。首先从智慧农业希望奶牛养殖更智能化的需求介绍，分析了以前靠人工管理奶牛的问题，如在识别每头牛个体以及监测它们健康状况时，效率不高还容易出错，从而说明奶牛面部分割作为实现精准化管理的关键技术很重要。其次介绍了研究内容，首先是打造涵盖多种场景的奶牛面部图像数据集，其次是设计一种基于注意力机制的改进分割算法。还介绍了引入的 ECA 通道注意力、FPN 特征融合、SSH 语义增强等模块，这些模块能解决传统算法存在的浅层特征不够、特征还原能力不强的问题。同时，梳理了动物面部分割领域的发展情况，指出在奶牛面部分割研究方面还存在空白。综上，本章明确了研究方向、规划技术路线以及分析相关领域的现状，为后面进行数据集构建、算法实现和实验验证提供了理论依据和思路基础。

第 2 章 相关理论概述与数据集制作

本章主要介绍了经典语义分割算法和注意力机制的发展情况，以及本文涉及的评价指。首先本章介绍了本文实验涉及到的经典的语义分割算法，包括 U-Net、PSPNet、SegNet 和 FCN。U-Net，依靠 U 型的对称结构，再加上跳跃连接的设计，解决了医学图像分割时细节容易丢失的问题。PSPNet 利用金字塔池化模块突破了复杂场景里全局上下文信息建模的瓶颈。SegNet 通过高效的解码设计，让模型变得更轻便，适合轻量级分割任务。FCN 作为第一个全卷积网络，开创了端到端直接做像素级分割的先河，是语义分割算法的开创性工作。随后本章简单介绍了注意力机制。注意力机制能够让模型在处理图像时，能更精准的聚焦到模型的关键区域，抓住重要信息，这样可以实现模型的性能提升。随后，本章详细介绍了评估算法性能的几个核心指标，包括平均交并比 (mIOU)、平均像素准确率 (mPA) 和整体准确率 (ACC)。mIOU 主要衡量分割精度，mPA 关注类别之间的均衡性，ACC 则反映模型整体的预测效果，它们互相补充，能全面评估模型的表现。最后，本章介绍了数据集如何制作，首先进行数据采集，随后依次进行预处理，数据标注和数据集制作。旨在制作出有效的奶牛面部图像数据集为后续的实验铺路。本章通过将这些算法的原理和评价体系进行介绍，展示了语义分割领域的技术发展脉络和实际应用框架。

2.1 经典语义分割算法

本文使用 U-Net, PSPNet, SegNet 和 FCN 四个经典的语义分割模型作为对照组，在后续的实验中与我们提出的模型进行对比实验。旨在评估各个模型在进行奶牛面部分割问题上的性能优劣。

2.1.1 U-Net

U-Net (Olaf, 2015) 是经典语义分割算法，其通过跳跃连接融合不同层次的特征信息。起初用于医学图像分割等对细节敏感的任务。编码器通过卷积和池化操作逐步降低图像分辨率并提取抽象语义特征，解码器则通过反卷积和上采样恢复空间分辨率，同时利用跳跃连接将编码器对应层级的高分辨率特征直接传递至解码器，这种策略可以有效保留边缘、纹理等细节信息。这种 U 型结构使模型能够在较少训练样本下实现高精度分割

U-Net 的编码器通过池化操作把图像尺寸缩小、通道数增加，逐级提取更抽象的深层特征；解码器则反向通过反卷积或上采样实现还原图像尺寸。同时靠跳跃连接”拉上编码器对应层级的浅层细节特征和自己的深层语义信息拼在一起，最后用 1×1 卷积给每个像素完成细化分类。这种设计既解决了传统全卷积网络细节缺失的问题，又通过跳跃连接给梯度回传简化，让深层网络更好训练，现在在农业图像分割、遥感分析等领域

都实现了广泛应用。

2.1.2 PSPNet

PSPNet (Zhao, 2017) 的核心创新在于金字塔池化模块 (Pyramid Pooling Module), 通过多尺度上下文信息的融合显著提升复杂场景下的分割精度。该模型采用基于 ResNet 实现架构, 在编码器中通过 (dilated convolution) 替代传统池化操作, 在不降低分辨率的前提下扩大感受野, 保留更多空间细节。金字塔池化模块进一步对深层特征图进行全局平均池化和不同尺度的区域池化将不同层级的上下文信息通过上采样和特征拼接整合, 有效解决了早期模型对全局场景理解不足的问题。

PSPNet 的训练分主要分为两步走, 首先模型通常用 ResNet-101 作为主干网络, 训练时阶段与图像增强同时进行, 这些策略可提升模型的泛化能力。让模型适应更多实际场景。

2.1.3 SegNet

SegNet (Vijay, 2015) 也是一种常用的语义分割算法。其核心是一个可训练的分割引擎, 包含一个编码器网络、一个对应的解码器网络以及一个像素级分类层。编码器网络的架构与 VGG16 网络的前 13 个卷积层相同, 而解码器网络的作用是将低分辨率的编码器特征图映射回全输入分辨率的特征图, 以进行像素级分类。

SegNet 在解码器处理低分辨率输入特征图的上采样方式上做了关键的创新。首先解码器编码器在最大池化步骤里算出来的池化索引, 靠这些索引直接做非线性上采样。上采样完得到的是稀疏特征图, 这时候再让它和可训练的滤波器做卷积操作, 把稀疏的特征填满”变成密集特征图。这套设计有俩好处: 一是让边界划分的准确性显著提升, 二是减少了网络需要训练的参数数量, 还能支持端到端的训练, 整体更高效了。

SegNet 最初是为了满足道路场景理解等应用的需求而设计的, 因此它在内存和计算时间方面都非常高效。后来人们发现 SegNet 与其他语义分割架构相比在参数数量上显著减少, 且在推理时更为高效。

2.1.4 FCN

FCN (Long, 2015) 是最早用深度学习做图像分割的算法。FCN 的核心是把传统卷积神经网络里限制输入尺寸的全连接层, 全换成了全卷积层, 这样不管输入图片是大是小, 都能直接输出每个像素的分类结果。传统方法得拆成好几个阶段, 但是 FCN 直接一步到位, 从原图里学特征和分类器。

FCN 的结构主要有三部分: 卷积层负责提取图像里的各种特征; 上采样层通过反卷积或转置卷积, 把卷积后缩小的特征图放大回原图尺寸; 连接层则靠跳跃连接, 把浅层学的细节信息和深层学的语义信息拼在一起, 让模型既知道像素的位置和分类, 该方法可以提升分割精度。这种端到端的设计, 让模型直接从像素级输入学到像素级输出, 又

快又准。

2.2 注意力机制算法概述

本文对使用注意力机制对传统的语义分割模型 DeeplabV3+ 进行了改进，注意力机制能增强模型分割奶牛面部的能力，为了方便读者对后续的实验的理解，本节对注意力机制进行简单概述。

注意力机制的灵感最早来源于认知科学对人类视觉注意力的研究，简单说它的核心思路是让模型学会自动调整不同特征的重要程度，让模型在处理数据时能集中在关键信息区域，同时过滤掉无关的干扰信息，这样就能提高处理复杂任务的效率。在深度学习里，注意力机制主要通过设计可以学习的权重矩阵，让模型根据输入内容自适应地选择关键特征，现在已经成了提升模型性能的重要方法。根据作用域和计算方式，注意力机制可分为三大类：

(1) 空间注意力：关注特征图中空间位置的重要性，通过生成空间权重矩阵突出关键区域（如目标边界、纹理细节），典型方法包括基于卷积的空间注意力模块（SAM）和基于自注意力的位置编码。

(2) 通道注意力：聚焦特征通道间的语义相关性，通过全局池化或自相关计算通道权重，增强高语义通道的表达能力（如分类相关特征），代表模型如 Squeeze-and-Excitation Networks（SENet）。

(3) 自注意力：捕捉特征内部的长距离依赖关系，通过计算特征向量间的相似性矩阵，实现同一特征图内不同位置的信息交互，有效解决传统卷积神经网络（CNN）感受野受限的问题，典型应用于 Transformer 架构。

语义分割需要精确对齐像素级语义标签与空间位置信息，而传统 CNN 存在长距离依赖建模不足、细节特征丢失等问题。注意力机制的引入通过动态特征筛选与上下文聚合，显著提升了复杂场景下的分割精度

2.3 评价指标

在本文中，我们使用了三种评价指标来评估语义分割算法的性能：Mean Intersection over Union (mIOU)、Mean Pixel Accuracy (mPA) 和 Accuracy (ACC)。这些指标能够有效地反映模型在分割任务中的表现。

Intersection over Union (IOU) 是一个常用的评价指标，用于衡量模型预测结果与真实标签之间的重叠程度。具体来说，IOU 计算的是预测区域与真实区域的交集与并集之比。对于每个类别 i ，IOU 可以表示为：

$$\text{IOU}_i = \frac{P_i \cap G_i}{P_i \cup G_i} \quad (2-1)$$

其中， P_i 是模型预测的第 i 类区域， G_i 是真实的第 i 类区域。为了得到所有类别的平均 IOU，我们计算每个类别的 IOU 并取其平均值，得到 Mean Intersection over Union

(mIOU):

$$\text{mIOU} = \frac{1}{C} \sum_{i=1}^C \text{IOU}_i \quad (2-2)$$

这里, C 是类别的总数。

Pixel Accuracy (PA) 是指正确预测的像素数占全部像素数的比例。对于每个类别 i , PA 可以表示为:

$$\text{PA}_i = \frac{P_i \cap G_i}{G_i} \quad (2-3)$$

同样地, 为了得到所有类别的平均像素准确率, 我们计算每个类别的 PA 并取其平均值, 得到 Mean Pixel Accuracy (mPA):

$$\text{mPA} = \frac{1}{C} \sum_{i=1}^C \text{PA}_i \quad (2-4)$$

Accuracy (ACC) 是指正确预测的像素数占总像素数的比例。它可以表示为:

$$\text{ACC} = \frac{\sum_{i=1}^C (P_i \cap G_i)}{\sum_{i=1}^C G_i} \quad (2-5)$$

2.4 数据集制作

2.4.1 数据采集

数据集收集于 2024 年 12 月至 2025 年 1 月, 收集工作在陕西宝鸡的澳华现代牧业有限公司和宁夏石嘴山的正元生态有限公司进行。总共拍摄了 42 头奶牛视频, 时长共计 1652 分钟, 以 1920×1080 像素的分辨率录制, 帧率为 60Hz。视频录制设备为 SONY HRD-PJ675。拍摄的时候尽可能选择复杂和多样的环境下拍摄, 包括晴天, 夜晚, 阴天和雪天等。图 2-1 展示了拍摄时的场景。

2.4.2 数据预处理

在奶牛养殖智能化研究中, 原始视频数据常因复杂环境存在质量问题, 如奶牛面部被饲草或设备遮挡、头部未完整入镜、摄像机抖动导致图像模糊 (峰值信噪比 $< 20\text{dB}$) 等, 此类无效数据若直接用于模型训练, 易造成特征提取偏差与预测精度下降。为此, 研究构建了系统性的数据预处理流程, 通过多维度处理提升数据质量, 具体步骤如下:

(1) 首先进行视频抽帧处理, 将连续视频流转化为图像序列。针对养殖场监控设备采集的 1920×1080 分辨率、30fps 帧率视频, 采用固定间隔降采样策略, 每 15 帧抽取 1 帧 (等效 2fps), 在避免相邻帧信息冗余 (相似度 $> 95\%$) 的同时保留动态场景特征, 共获得 32556 张初始图像, 覆盖修蹄作业全流程及日常养殖场景, 为后续处理提供基础数据。



图 2-1 数据集获取场景

(2) 其次开展多轮人工筛选，剔除低质量样本。借助 OpenCV 计算图像拉普拉斯方差（阈值设为 150）自动标记模糊图像，结合人工复核剔除运动模糊（模糊核半径 > 5 像素）与对焦失败样本；依据眼、鼻、口周等面部关键区域可见性制定遮挡标准，保留关键区域遮挡 $< 30\%$ 且面部轮廓完整度 $> 70\%$ 的有效样本，剔除单侧眼完全遮挡、口鼻覆盖或头部入镜 $< 50\%$ 的无效样本；同时排除非修蹄场景干扰，确保应激与放松状态样本的场景独立性。经筛选，最终保留 3065 张高质量图像。

(3) 最后进行图像标准化预处理，适配深度学习模型输入要求。采用双三次插值算法将图像统一缩放至 512×512 像素，通过 ROI 定位优先保留面部区域，平衡细节保留与计算效率（显存占用降低 40%）；将像素值从 0-255 灰度范围归一化至 0-1 区间（逐像素除以 255），消除设备亮度差异（如红外图像灰度偏移），稳定神经网络激活函数响应，加速模型收敛（训练周期缩短 20%）。经处理，图像平均峰值信噪比提升至 35dB，关键区域完整度达 92% 以上，为后续模型训练奠定了高可靠性的数据基础，有效增强了复杂环境下的特征表征能力与模型泛化性能。

2.4.3 数据标注

在研究奶牛应激反应时，精准分割它们的面部区域是量化分析情绪状态的关键前提。因为奶牛在应激（紧张、焦虑等状态）时，面部微表情变化主要集中在几个关键生理区域，所以本研究专门针对眼睛、鼻子、嘴巴、眉毛、耳朵，以及面部轮廓和肌肉纹理这 6 个特征部位进行精细标注。这些部位不仅是兽医临床判断奶牛是否应激的重要观察点，比如通过观察眼睛是否睁大、耳朵姿势变化等判断情绪。

(1) 首先将预处理后的 512×512 像素图像批量导入 Labelme 平台，建立包含 6 个独立类别的标注标签体系（eye, nose, mouth, eyebrow, ear, facial muscle）。

(2) 由经过兽医行为学培训的标注团队进行首轮逐像素勾勒，对眼睛虹膜与眼白区域、鼻孔轮廓、唇线弧度等细微结构进行亚像素级精度标注

(3) 针对多视角图像中存在的自遮挡问题，通过标注遮罩层记录可见区域边界，确保语义分割的空间完整性。

(4) 采用双人独立标注，对眉毛、面部肌肉等易分歧区域进行二次校准，最终标注精度经 Dice 系数评估均达 92% 以上。每张图像生成对应的 JSON 格式标注文件，包含各部位的多边形坐标序列及语义标签。

2.4.4 制作奶牛面部图像数据集

为构建具有高鲁棒性的深度学习模型，本研究基于 3065 张包含眼睛、鼻子、嘴巴等 6 个关键面部区域标注的高质量图像，按照 6:2:2 的比例严格划分为训练集（1839 张）、验证集（613 张）与测试集（613 张），此举能够为验证集和测试集保留足够独立样本用于客观评估模型性能。

在划分数据时，我们用了分层抽样的方法：先按照拍摄场景（比如不同的光照强度、拍摄角度等）把数据分成多个组，确保每个子集的样本分布和原始数据集几乎一样（类别比例偏差控制在 5% 以内），避免因数据划分不合理导致某些类别样本过多或过少。同时，我们还用了数据增强技术来扩充样本量——通过随机翻转图像、旋转角度、调整亮度等 12 种变换操作，把样本数量从原来的基础上扩充到了 1.8 万张。这样做是为了让模型能更好地适应养殖环境中复杂的光照条件（比如强光、阴影）和多角度拍摄场景（比如正面、侧面拍的奶牛面部）。通过这种有条理的数据分层和增强处理，我们构建的奶牛面部数据集能更全面地覆盖实际场景，为后续分割模型的训练提供高质量数据，确保模型能精准分割奶牛面部的关键区域（如眼睛、鼻子等）。

2.5 本章小结

本文聚焦经典语义分割算法与注意力机制展开研究，指出 FCN 通过全卷积层实现端到端像素级分割的开创性，U-Net 借助跳跃连接融合浅层细节与深层语义以保留分割细节，PSPNet 利用金字塔池化模块整合多尺度上下文信息提升复杂场景处理能力，SegNet 则通过轻量化解码器设计平衡分割精度与计算效率；注意力机制从空间、通道、自注意力三个维度突破传统 CNN 局限，通过动态权重分配让模型聚焦关键特征，在医学图像边缘分割、遥感地物识别等任务中显著提升性能；此外，mIOU、mPA、ACC 等指标分别从类别重叠度、像素准确率、整体预测效果等维度综合评估模型分割表现。本章还介绍了数据集的制作，制作有效的奶牛面部图像数据集有助于后续的建模和实验顺利进展。

第3章 基于注意力机制的奶牛面部精细分割模型的建立

奶牛面部的精细分割是实现奶牛健康监测、行为分析和福利评估的关键技术基础。如 DeeplabV3+ 流行的传统图像分割技术，虽然能获取不同大小的图像信息，但存在两个问题：一是浅层提取的特征里语义信息不够，二是深层特征还原图像具体细节的能力较弱，导致很难准确分割奶牛面部毛发纹理、口鼻眼等复杂区域。为了接近这些难题，本文提出了一种基于注意力机制改进奶牛面部精细分割算法。我们引入 ECA 通道注意力模块，让模型能更有效地判断更重要的特征；同时结合 FPN 多尺度特征融合模块，把不同大小的特征信息有效融合，再用 SSH 模块优化特征还原效果，让模型既能抓住深层语义，又能保留浅层空间信)，构建出一个深浅层特征优势互补的高效分割模型。本章研究围绕主干网络介绍、注意力机制和特征融合模块的融合、以及多个模块协同工作和实验设计展开。

3.1 基于注意力改进的奶牛面部精细化分割模型搭建

DeeplabV3+ (Chen, 2018) 是一种流行的语义分割算法，采用编码器 - 解码器结构，通过改进的空洞空间金字塔池化 (ASPP) 模块捕捉多尺度上下文信息，并引入解码器融合底层高分辨率特征与高层语义特征，显著提升分割边界的细节精度。本文提出的基于注意力机制的奶牛面部精细分割算法解决了 DeeplabV3+ 算法浅层特征不充足和特征还原图像能力差的问题，同时实现了对奶牛图像深层特征的特征增强。算法框架如图 3-1 所示。(1) 奶牛面部的深浅层特征提取：首先将奶牛面部图像作为输入传入到深度卷

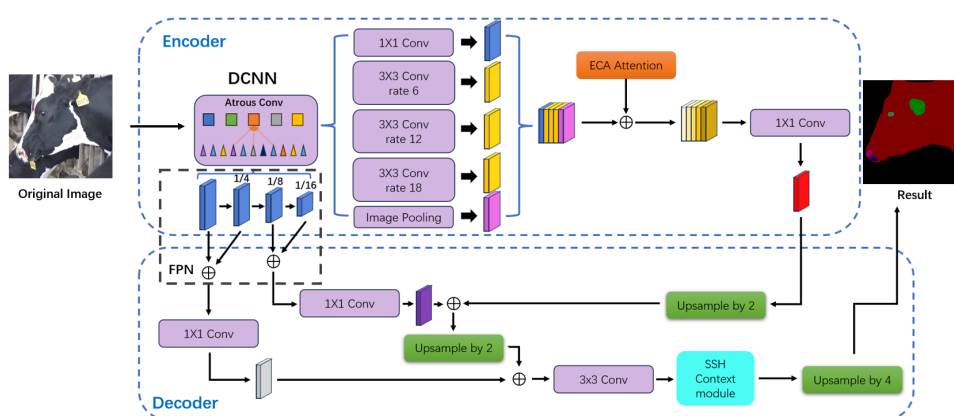


图 3-1 奶牛面部精细分割算法流程

积神经网络 (DCNN) 中提取特征，DCNN 算法通常被称为主干网络 (Backbone)。在编码器部分，Backbone 模块会产生浅层奶牛面部特征 (SFF) 和深层奶牛面部特征 (DFF) 两个输出。

(2) 深层特征的空洞卷积：Backbone 得到的 DFF 会以不同的扩散率进一步进行多

次空洞卷积，这些空洞卷积的扩散率有 6，12，18 三种。随后将 3 个空洞卷积的结果，1x1 卷积的结果和池化结果共 5 个特征进行特征堆叠。这部分又称为 ASPP 模块。

(3) ECA 注意力机制实现特征增强：本文引入了 ECA 通道注意力机制来提高算法对重要的奶牛面部特征的关注度。ECA 通道注意力机制将为来自 ASPP 模块的各个奶牛面部图像特征赋予权重。权重越大的特征越重要。

(4) 深浅层特征的多尺度特征融合：在解码器部分，首先我们基于 FPN 的特征提取方式对 SFF 进行了多尺度特征提取。我们将得到奶牛面部图片大小的 1/4，1/8 和 1/16 三个特征。其次将相对深层的两个特征，将最浅层的特征与 SFF 进行融合。随后将 2 个特征融合的结果进行 1x1 卷积后我们的得到了解码器的深层语义信息 (DDSM) 和解码器的浅层语义信息 (DSSM)。然后特征增强的结果进行 2 倍上采样后与 DDSM 融合，进行 2 倍上采样后与 DSSM 融合。这样处理可以丰富浅层特征的语义信息，同时避免浅层特征和深层特征直接融合带来的误差 (Huang, 2025)。

(5) SSH 增强特征还原能力：由于深度卷积后的图像容易导致大量信息丢失 (Kim, 2016)，为了优化奶牛面部特征的还原结果，我们引入了 SSH 模块。将得到奶牛面部特征层进行 3x3 卷积后输入到 SSH 模块中进行处理，将 SSH 模块的输出结果进行 4 倍的上采样即可得到最终的奶牛面部语义分割结果。奶牛面部疼痛量表需要的 6 个部分在奶牛面部语义分割结果中以不同颜色表现出来。

本文采用的损失函数统一使用交叉熵损失函数，如公式 (3-1) 所示。

$$L(y, \hat{y}) = -\frac{1}{N} \sum_{i=1}^N (y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)) \quad (3-1)$$

其中：L 是损失函数 y 是真实标签 \hat{y} 是预测值 N 是样本数量

3.1.1 Backbone 主干网络

本文以当前最常见的深度卷积神经网络 MobilenetV2 作为 Pain-Deeplab 算法框架中的主干网络。MobileNetV2 (Sandler, 2019) 是一种为移动和嵌入式设备优化的轻量级深度学习模型。它基于倒置残差结构设计。MobilenetV2 在本文中的应用如图 3-2 所示。MobilenetV2 主体主要由三部分组成，首先是一个 1x1 的卷积用于扩展通道数，然后用一个 3x3 的深度可分离卷积提取特征，最后用一个线性 1x1 的卷积用于降。在使用 MobilenetV2 之前，原始的奶牛面部图像需要先进行一个步长为 2 的 3x3 卷积，然后标准化后复制成两路进入 MobileNetV2 网络，如公式 3-2 和公式 3-3 所示。

$$y_{i,j} = \frac{m=0}{2} n = \frac{0}{2} x_{i,m} + \frac{2}{j} n \cdot K_{m,n} \quad (3-2)$$

$$z_{i,j} = \max \left(0, \frac{y_{i,j} - \mu}{\sigma^2 + \epsilon} \cdot \gamma + \beta \right) \quad (3-3)$$

其中 $z[i, j]$ 和 $y[i, j]$ 是特征图的第 $[i, j]$ 个元素， μ 是均值， σ^2 是方差， ϵ 用于防止除零错误， γ 和 β 是可学习的参数。第一路执行 MobileNetV2 的三个部分。第二路与第一路

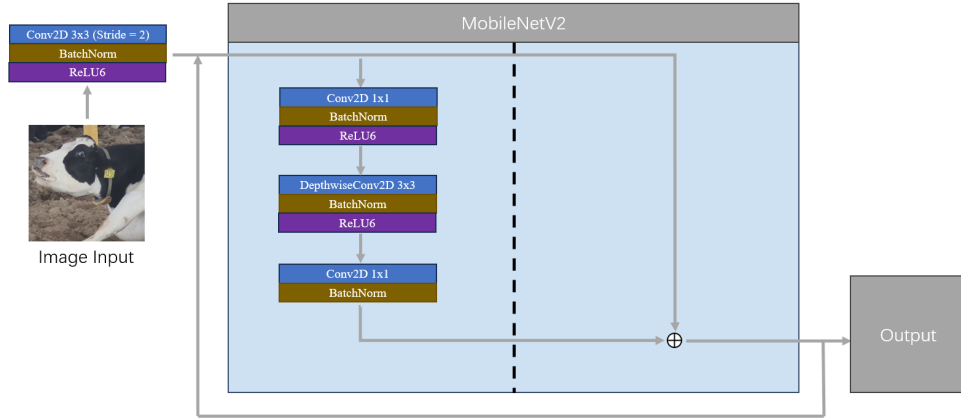


图 3-2 MobilenetV2 算法流程

最终的输出结构做拼接之后即完成一轮 MobileNetV2 处理，如公式 3-4 所示。

$$m = \text{Concat}(\text{Conv3}, \text{DepthwiseConv2}, \text{Conv1}) \cdot z \quad (3-4)$$

其中 Conv_1 、 DepthwiseConv_2 和 Conv_3 分别代表 MobileNetV2 的三个主要操作， z 为输入。主干网络的浅层特征为经历两轮步长为 1 的 MobileNetV2 处理后的输出，深层特征为经历 3 轮步长为 2 和 3 轮步长为 1 的 MobileNetV2 处理后的输出。

3.1.2 ECA 通道注意力模块

ECA (Wang, 2020) 是一种轻量级的通道注意力机制，它的核心思路是通过一维卷积让相邻通道的特征实现局部跨通道交互，而且巧妙地不压缩通道数，这样既能提升模型识别关键特征的能力，又不会增加太多计算量。ECA 能根据输入的特征自动调整一维卷积核的大小。比如遇到简单特征就用小核，遇到复杂特征就用大核，动态控制交流范围，既能准确捕捉通道间的关联，又不会因为强行压缩通道数而丢失信息。和 SE-Net 这类传统通道注意力方法比起来，ECA 几乎不增加参数和计算量，却能让图像分类准确率提升 2% 以上，靠的就是这种精准的“局部交流”设计，既省算力又能增强特征的判别力。目前 ECA 在计算机视觉里应用很广，图像分类、目标检测、分割等任务都能用它来强化模型的特征表达能力。在本文中，我们把 ECA 引入到 ASPP 模块里，给 ASPP 输出的多尺度特征自动分配每个通道的权重，让模型在融合多尺度特征时，能精准挑出更关键的信息，避免杂糅无效特征，从而提升模型对图像上下文信息的利用效率。ECA 在 ASPP 中的应用示意图如图 3-3 所示。其中 K 为 1D 卷积的卷积核大小， σ 为 sigmoid 函数。首先通过 ASPP 模块，我们可以得到堆叠后的奶牛面部特征。ECA 首先对该特征进行全局平均池化 (GAP)，GAP 的公式如公式 3-5 所示。

$$G[c] = \frac{1}{H \times W} \sum_{j=1}^H \sum_{i=1}^W F_{i,j,c} \quad (3-5)$$

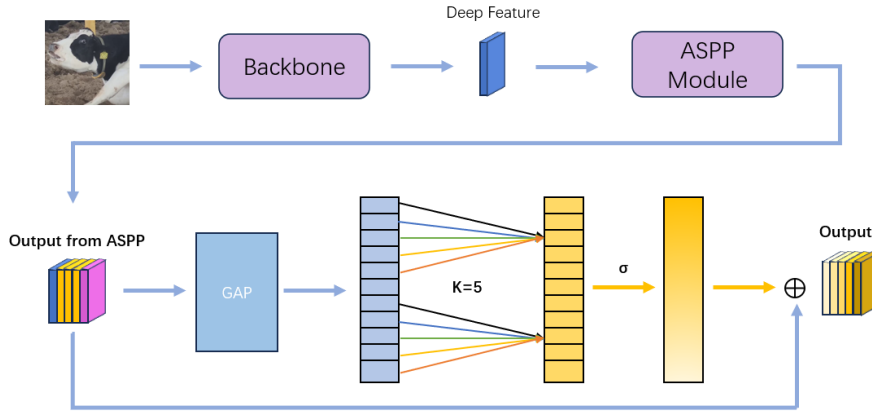


图 3-3 ECA 算法流程

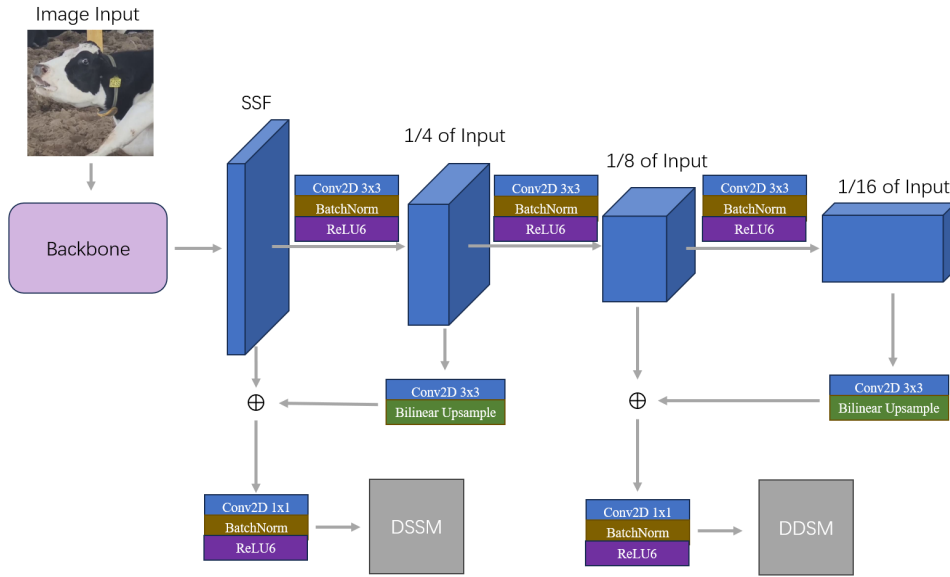
其中 H 是高度， W 是宽度， $F[i, j, c]$ 是输入特征图在位置 (i, j) 和通道 c 的值， $G[c]$ 是输出特征向量的第 c 个元素。随后将 GAP 的结果进行 1D 卷积得到权重向量。然后将得到的权重向量输入到 sigmoid 函数中进行标准化，得到最终的权重向量。

3.1.3 FPN 特征融合模块

FPN (Lin, 2017) 是目标检测里经典的特征融合方法。FPN 核心是通过自顶向下的特征传递和横向连接，把 CNN 不同层次的特征融合在一起。FPN 注意到，高层的特征虽然语义信息强，但分辨率低、细节模糊；低层的特征虽然分辨率高，但语义信息弱。传统方法只用单一层次的特征，导致小目标容易漏检。针对这个问题。FPN 把高层的低分辨率特征通过上采样逐步细化，然后和对应层级的低层高分辨率特征做横向连接。这样每个层级的输出特征就像既有高层的语义信息，又保留了低层的空间细，最终形成一个特征金字塔，让模型能同时处理大、中、小各种尺度的目标。在本文中，我们把 FPN 模块加进 Backbone 网络，专门用来优化多阶段特征的融合。具体来说，它通过自顶向下的路径给各层特征“补语义”，同时用横向连接保留底层的空间分辨率。这样一来，后续的分类和定位任务就能拿到更有效的特征，尤其是对小目标的检测效果提升明显。FPN 在本文中的具体应用如图 3-4 所示。首先将浅层特征经过 3 次步长为 2 的卷积池化操作后得到三个特征层，其大小分别为原图的 $1/4$ ， $1/8$ 和 $1/16$ 。随后将大小为原始奶牛面部图像 $1/4$ 和 $1/16$ 的特征层进行步长为 1 的 3×3 卷积后，用双线性插值法对卷积后的奶牛面部特征层进行上采样。双线性插值的公式如公式 3-6 所示。

$$Q_{x,y} = Q_{11} \cdot (x_2 - x)(y_2 - y) + Q_{21} \cdot (x - x_1)(y_2 - y) + Q_{12} \cdot (x_2 - x)(y - y_1) + Q_{22} \cdot (x - x_1)(y - y_1) \quad (3-6)$$

其中 Q_{11} 是点 (x_1, y_1) 的值， Q_{21} 是点 (x_2, y_1) 的值， Q_{12} 是点 (x_1, y_2) 的值， Q_{22} 是点 (x_2, y_2) 的值。点 (x_1, y_1) 和 (x_2, y_2) 是已知点的坐标。最后，将上采样的结果分别与浅层特征层和原图大小为 $\frac{1}{8}$ 的特征层做特征堆叠。最终能够分别得到 DDSM 和 DSSM。



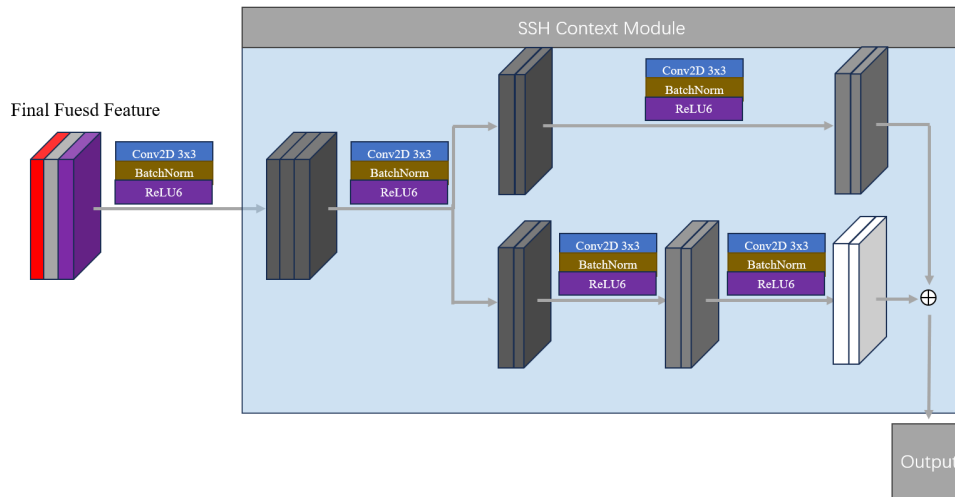


图 3-5 SSH 算法流程

轮训练) 下对比特征提取能力；其次通过递进式消融实验，逐步集成 ECA Attention、FPN、SSH 模块及其组合构建 8 组模型，量化分析单一模块作用与多模块协同效应；再者与 Unet、PSPnet 等四种主流算法在相同条件下开展性能对比，验证所提算法在 mIOU、mPA、ACC 指标上的优势；最后结合可视化图表与量化数据，分析算法对背景及面部关键区域的分割精度、不同复杂度区域处理能力和推理效率，形成涵盖网络架构优化、模块功能验证、算法横向对比及多维度性能评估的完整实验体系。

3.3 本章小结

本文提出的奶牛面部精细分割模型基于 DeeplabV3 + 框架，通过加入 ECA 通道注意力机制、FPN 特征融合模块及 SSH 语义增强模块，很好地解决了传统算法浅层特征不足和特征还原能力差的问题，同时采用 MobilenetV2 作为轻量化主干网络，在保证计算效率的前提下提升了多尺度特征提取能力；实验显示，模型通过空洞卷积与注意力机制增强深层特征，借助多层模块协同工作优化深浅层特征交互，最终利用交叉熵损失函数实现了对奶牛面部 6 个关键区域的精准分割，其输出的高质量语义分割结果可为奶牛面部疼痛量表分析等后续应用提供支持，在畜牧业智能监测领域具备重要实用价值和推广潜力。

第 4 章 实验结果与分析

本章围绕模型训练及性能分析展开研究。实验部分聚焦于不同主干网络、优化器及模块对模型性能的影响，通过对比分析，验证了所提算法在奶牛面部精细化分割任务中的优势

4.1 参数确定

为系统评估不同主干网络对 DeepLabV3+ 模型特征提取能力的影响及优化器对参数优化的作用，本研究设计了多组语义分割对比实验。实验选取 MobileNetV2 (Howard et al., 2017)、Xception (Chollet, 2017) 和 ResNet (He et al., 2016) 作为特征提取主干网络，搭配 Adam (Kingma & Ba, 2015) 和 SGD 两种优化器，总共形成六种模型组合。所有实验在我们获取的奶牛面部图像数据集开展，输入图像统一 resize 至 512×512 像素，采用多尺度数据增强和随机水平翻转预处理，训练过程遵循相同的超参数配置。初始学习率 1e-4，batch size=16，训练周期 50 轮。

主干网络	优化器	mIOU	mPA
MobileNetV2	Adam	0.5554	0.6418
	SGD	0.5312	0.6211
Xception	Adam	0.5117	0.6311
	SGD	0.5263	0.6410
ResNet	Adam	0.5202	0.6210
	SGD	0.5333	0.6318

表 4-1 Apache 选项参数

实验结果表 3-1 所示。从实验结果可以看出，MobileNetV2 凭借轻量级倒置残差块结构与线性瓶颈设计，在语义分割任务中展现出最优的特征表达能力，其与 Adam 优化器的组合取得了 0.5554 的 mIoU 和 0.6418 的 mAP。从优化器对比来看，Adam 优化器通过自适应调整学习率策略，在参数空间搜索中表现出更高效率，验证了自适应优化算法在深层网络训练中的优势。

4.2 消融实验

为了全面分析 ECA Attention、FPN 和 SSH 这三个模块对 DeeplabV3+ 模型效果的提升作用，我们设计了递进式消融实验。具体来说，就是先从最基础的 DeeplabV3+ 模型开始，然后逐步往上叠加模块：先单独加一个模块，再试两个模块组合，最后把三个

模块全加上，这样一共搭出了 8 组对比模型。通过这种方式，我们能清楚分析每个模块在语义分割任务中单独用的效果，以及它们一起用的时候会不会互相冲突。

DeeplabV3+	ECA Attention	FPN	SSH	mIOU	mPA	ACC
√				0.5554	0.6418	0.9760
√	√			0.6045	0.7434	0.9771
√		√		0.6079	0.7576	0.9721
√			√	0.6162	0.7483	0.9731
√	√	√		0.6388	0.7755	0.9741
√	√		√	0.6394	0.7721	0.9723
√		√	√	0.6407	0.7673	0.9739
√	√	√	√	0.6684	0.7797	0.9793

表 4-2 消融实验结果

消融实验结果如表 3-2 所示。√ 代表引入了这个模块，空白表示没有引入这个模块。当单独引入任一模块时，基础模型的 mIOU 指标均获得提升。ECA Attention 使 mIOU 从 55.54% 提升至 60.45%；FPN 的加入验证了多尺度特征融合对复杂场景分割的有效性。SSH 模块则通过边界感知优化，显著增大了算法的 mPA。

此外，模块的两两组合也可以实现对 DeeplabV3+ 算法性能的提升。ECA Attention 与 FPN 的组合较单独 FPN 提升 3.15%，表明通道注意力与多尺度特征的协同优化。FPN+SSH 组合在保持边界分割精度的同时，通过特征金字塔增强了小目标的识别能力。值得注意的是，ECA Attention 与 SSH 的组合在 mPA 指标上达到 77.55%，显示出注意力机制与边界优化的联合优势。

当三个模块完全集成时，我们提出的新的奶牛面部精细化分割算法实现了性能突破。mIOU 达到 66.84%（较基础模型提升 11.3%），mPA 提升至 77.97%，ACC 增至 97.93%。与最佳两两组合相比，三模块联合使 mIOU 进一步提升 2.77%-3.03%，验证了通道注意力、多尺度特征融合与边界优化的协同增效机制。

4.3 性能对比与分析

为系统评估我们提出的奶牛面部精细分割算法的性能，本研究选取 Unet、PSPnet、FCN 和 SegNet 四种主流语义分割算法做对比实验。所有模型采用相同的数据预处理流程（图像归一化至 512×512 像素，随机旋转 ±15° 数据增强）和相同的训练配置（Adam 优化器，学习率 1e-4，训练周期 100 轮），确保实验结果的可比性。表 4-3 展示了对比实验的实验结果。

对比结果如表 3-3 所示，我们的算法在三项核心指标上均展现出显著优势：其 mIOU 达到 66.84%，较次优算法 Unet（63.70%）提升 3.14 个百分点，表明在多类别语义区域

Algorithm	mIOU	mPA	ACC
Unet	0.6370	0.7247	0.9752
PSPnet	0.4318	0.5035	0.9634
FCN	0.5730	0.7759	0.9770
SegNet	0.3570	0.5688	0.9542
New DeeplabV3+	0.6684	0.7797	0.9793

表 4-3 对比实验结果

的整体分割精度上具有明显优势；mPA（平均精度）为 77.97%，超越以边界分割能力著称的 FCN 算法（77.59%），验证了其对疼痛相关细微语义边界的精准捕捉能力；ACC（像素级准确率）达 97.93%，较 PSPnet（96.34%）和 SegNet（95.42%）分别提升 1.59% 和 2.49%，体现出强大的像素级分类可靠性。

分析在 New DeeplabV3+ 在 mIOU，mPA 和 ACC 上性能优秀的原因，本文认为有以下几点。首先相较于传统 Unet 的对称编解码结构，本文算法通过引入更复杂的特征融合模块实现了的多尺度语义特征提取，从而提升了复杂场景下多类别区域的整体分割精度。特别是针对奶牛面部存在的纹理差异、光照变化等干扰因素，算法可能通过增强上下文语义关联，减少了类别混淆问题，进而提高了交并比。其次算法对边界特征有很强的强化处理机制。与 FCN 仅依赖浅层特征进行边界定位不同，本文算法在解码过程中设计了更精细的边界细化模块（SSH），通过融合深层语义信息与浅层边缘特征，实现了对疼痛相关细微结构的准确分割，进一步提升了复杂姿态下面部边界的分割精度。同时，算法得益于其强大的像素分类可靠性。一方面，Adam 优化器结合 $1e-4$ 的学习率提供了稳定的参数更新路径，避免了梯度爆炸或消失问题，使模型能够充分学习到像素级的判别特征；另一方面，算法可能通过更深的网络层次或更高效的特征提取模块，增强了对像素级细节的表征能力，减少了背景与前景像素的误分类现象。特别是在处理奶牛面部与复杂背景的对比度差异时，算法可能通过增强像素级的类别区分性，有效提升了整体分类准确率。此外，算法中的 ASPP 多尺度空洞卷积模块，能够从不同感受野捕捉图像特征，对于奶牛面部不同大小的区域都能精准识别；而 ECA 通道注意力机制则聚焦关键特征通道，抑制冗余信息，使得网络更高效地学习到有助于提升 mIOU、mPA 和 ACC 的核心特征，从而在各项指标上均展现出优异性能。

4.4 算法在奶牛各部位的性能分析

为了全面看看我们提出的奶牛面部分割算法在奶牛面部各个部位的效果，图 4-2 到 4-4 给出了在 mIOU、mPA、ACC 三个核心指标上，背景和眼睛、鼻子等六个面部部位共七个区域的量化数据。通过这几个指标综合分析，能清楚看出算法在不同复杂程度区域的分割特点。可见我们的算法对背景有较强的识别和分割能力，模型在面部其他部分的表现较为相近。

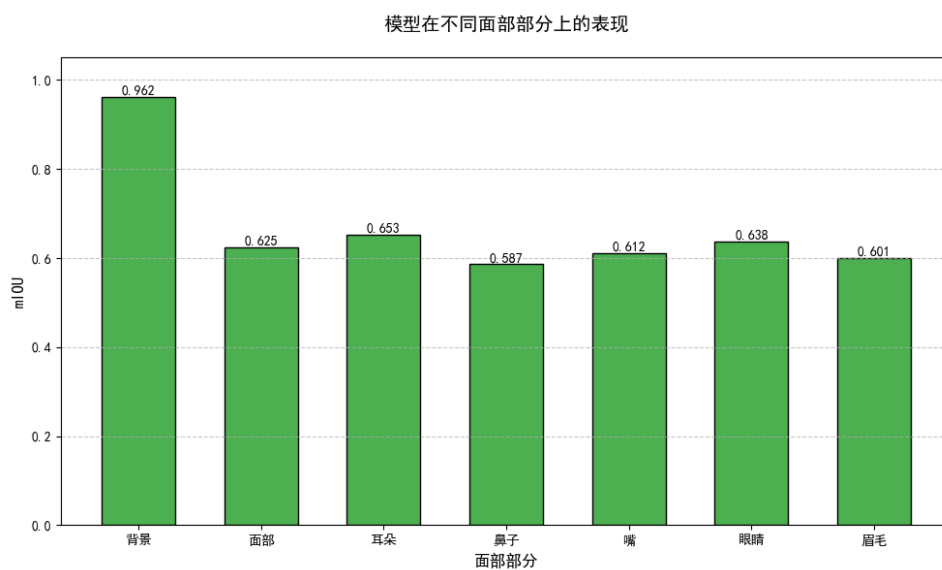


图 4-1 不同部位的 mIOU 表现

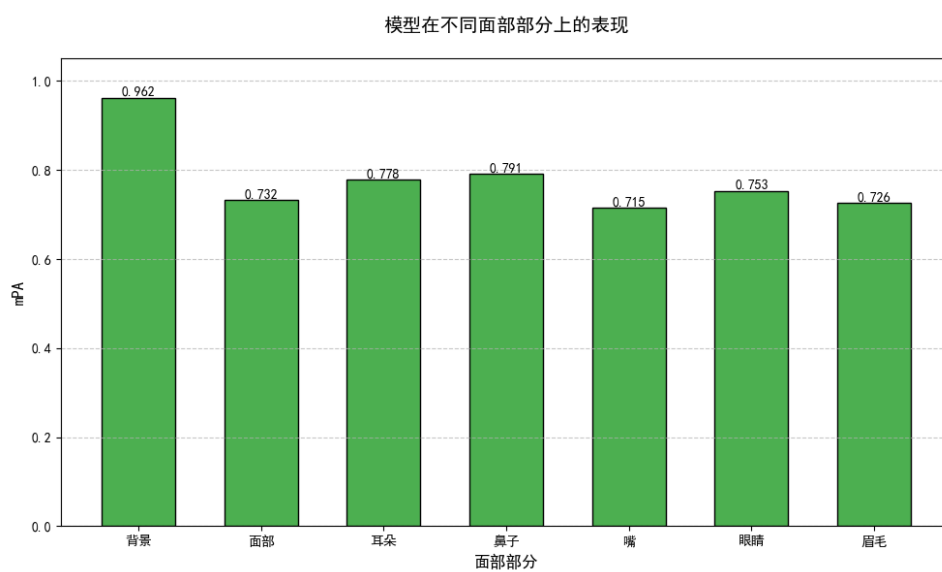


图 4-2 不同部位的 mPA 现

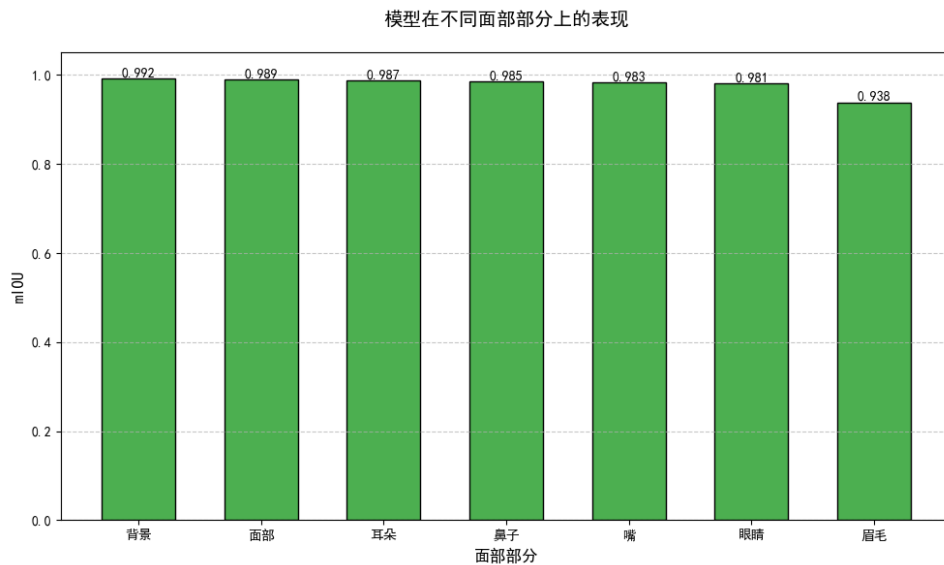


图 4-3 不同部位的 ACC 表现

4.5 算法效率分析

表 3-4 展示了各个模型的时间性能对比。在表中，本研究比较了不同模型的执行时间。从实验结果可知，New DeeplabV3+ 虽凭借 ASPP 多尺度空洞卷积、ECA 通道注意

算法	推理时间 (s)
Unet	54261
PSPnet	62144
FCN	41101
SegNet	58806
New DeeplabV3+	100025

表 4-4 不同算法的时间性能

力、FPN 多级特征融合及 SSH 高倍上采样等复杂模块叠加获得较好语义分割性能，但时间性能显著下降；其他算法中，FCN 作为基础全卷积网络，仅含编解码简单上下采样，无多尺度融合或注意力机制，计算图最简故速度最快，Unet 与 SegNet 通过对称编解码或池化索引恢复特征，结构稍复杂但参数量可控，推理时间中等，而 PSPnet 因通过金字塔池化提取多尺度上下文需拼接与卷积操作，计算量显著增加，运行时间更长，整体呈现出模型结构越复杂，推理速度往往越慢的特点。

4.6 本章小结

本章成功搭建了一套奶牛面部图像数据集，并提出了基于改进 DeepLabV3+ 的精细分割算法。实验发现，用 MobileNetV2 作为主干网络搭配 Adam 优化器，在语义分割

任务里效果最好，而且把 ECA Attention、FPN 和 SSH 模块结合起来，能明显提升模型性能。虽然这个算法在时间性能上比一些轻量级模型稍慢，但它在分割精度和捕捉边界细节方面更有优势，在复杂场景里更实用。

第 5 章 结论与展望

5.1 结论

(1) 本研究提出了一种基于注意力机制的奶牛面部精细分割算法，以 DeeplabV3+ 语义分割算法为框架在其基础上应用了 ECA.FPN 和 SSH 模块增强性能。ECA 注意力机制能让 DeeplabV3+ 算法关注更重要的奶牛面部特征，同时抑制了不重要的特征，这有助于 DeeplabV3+ 算法更关注奶牛面部，而不是背景或噪声。FPN 模块能对 DeeplabV3+ 算法的浅层特征进行扩充，同时实现了深浅层奶牛面部特征的多尺度特征融合，这丰富了可以学习的奶牛面部特征，提升了我们的算法的鲁棒性。SSH 模块提升了算法将奶牛面部特征层还原为奶牛面部语义分割结果的能力，提升了 DeeplabV3+ 对奶牛面部分割的准确性。这些模块的优势均通过消融实验得到了验证。

(2) 通过和传统语义分割算法的对比。实验结果表明，本研究提出的基于注意力机制的奶牛面部精细化分割算法在 mIOU, mPA, ACC 三项指标中都达到了最佳，分别达到了 0.6684, 0.7797 和 0.9793。

(3) 实验结果显示，本研究提出的基于注意力机制的奶牛面部精细化分割算法对背景有较强的分割和识别能力，在 mIOU, mPA, ACC 三项指标分别达到了 0.9620, 0.9620 和 0.9920。这表现出算法能很好的从图片中识别背景和牛脸。针对奶牛面部的 6 个部分，实验结果表明，改进后的 DeeplabV3+ 算法在这 6 个部分上有相近的性能。

本研究提出一种基于注意力机制的奶牛面部精细分割算法，以 DeeplabV3+ 语义分割算法为框架，引入 ECA、FPN 和 SSH 模块提升性能，其中 ECA 通过聚焦关键面部特征、抑制背景及噪声干扰增强算法敏感度，FPN 通过扩充浅层特征并融合深浅层多尺度特征丰富特征维度、提升鲁棒性，SSH 通过优化特征层还原能力提高分割准确性，各模块优势均经消融实验验证；对比实验表明，该算法在 mIOU、mPA、ACC 指标中表现最优（分别达 0.6684、0.7797 和 0.9793），对背景的 mIOU、mPA、ACC 分别达 0.9620、0.9620 和 0.9920，对眼睛、鼻子等 6 个关键部位分割指标表现均衡，充分验证了该语义分割算法在实现奶牛面部精细分割中的潜力。

5.2 展望

本研究通过模块化改进显著提升了 DeeplabV3+ 算法在奶牛面部精细化分割的精度与鲁棒性，为智慧畜牧领域的表型分析提供了关键技术支持。未来可以对我们的算法做出一些新的改进和探索。

(1) 探索混合注意力机制的深度融合是一个比较新颖的思路，ECA 通道注意力机制提出已有一段时间，而且空间注意力机制同样可能具有在奶牛面部分割任务上应用的前景。比如可以将 Vision Transformer 的全局自注意力与 ECA 局部通道注意力相结合，在强化远距离特征语义关联的同时保持计算效率，或者通过稀疏化自注意力掩码降低耳尖

与口鼻等非相邻区域的冗余计算，这些都可以推动算法在牧场边缘计算终端的即插即用。

(2) 考虑到单一视觉数据难以全面表征奶牛的生理状态，未来需构建跨模态智能分析体系。RGB 图像只是一种奶牛面部的体现，我们可以进一步通过融合可见光图像、红外热成像与生物传感器等数据进行研究

(3) 丰富数据集也是很有必要的，可以勇敢构建覆盖多品种（荷斯坦、娟珊牛等）、多生长阶段的奶牛面部图像来丰富数据集。

奶牛面部分割技术对乳业项目意义重大，其可结合时序模型通过捕捉咀嚼频率、眨眼周期等面部肌肉运动轨迹，建立“分割结果-行为模式-健康状态”的多层映射关系，实现奶牛健康全周期监控，还能与奶牛面部应激量表融合、集成多模态设备构建实时监控系统，在修蹄等高应激场景中通过边缘计算快速响应，为精准养殖决策提供支撑并提升健康管理效率与产业效益；该技术不仅是计算机视觉领域的突破，更是智慧畜牧迈向精准化的关键一步，未来需在算法效率与精度平衡、单一模态与多源数据融合、技术创新与标准化建设中拓展应用边界与构建产业生态，随着边缘计算、数字孪生等技术深度融合，其有望成为奶牛健康监测的“数字基础设施”，为保障国家奶业安全与推动畜牧业高质量发展提供持续科技赋能。

参考文献

- CHOLLET F, Xception: Deep learning with depthwise separable convolutions[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. [S.l.]: [s.n.]: 1800–1807.
- LIN T.-Y, DOLLÁR P, GIRSHICK R, et al., Feature Pyramid Networks for Object Detection[C/OL]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). [S.l.]: [s.n.]. <https://arxiv.org/abs/1612.03144>.
- NAJIBI M, NETTO P, CHIERICHETTI V, et al., SSH: Single Stage Headless Face Detector[C/OL]// Proceedings of the IEEE International Conference on Computer Vision (ICCV). [S.l.]: [s.n.]. <https://arxiv.org/abs/1708.03979>.
- VASWANI A, SHAZEER N, PARMAR N, et al., Attention Is All You Need[J/OL]. NeurIPS, <https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91f75e58e5.pdf>.
- 何屿彤, YOLOv3-DBSP模型在羊脸检测中的应用[J]. 农业工程学报, 40(5): 189–196.
- 李向宇, 李慧盈, 基于形变卷积神经网络的猪脸特征点检测方法[J]. 智慧农业资讯, 12(3): 45–56.
- 郭松涛, 基于 Tri-AI 技术的金丝猴个体识别系统[J]. 西北大学学报, 52(4): 589–596.
- BADRINARAYANAN V, KENDALL A, CIPOLLA R, SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation[J]. ArXiv preprint arXiv:1511.00561,
- BAHDANAU D, CHO K, BENGIO Y, Neural Machine Translation by Jointly Learning to Align and Translate[J/OL]. ArXiv preprint arXiv:1409.0473, <https://arxiv.org/pdf/1409.0473>.
- BOLYA D, ZHOU C, XIAO F, et al., YOLACT: Real-Time Instance Segmentation[J/OL]. ICCV, https://openaccess.thecvf.com/content_ICCV_2019/papers/Bolya_YOLACT_Real-Time_Instance_Segmentation_ICCV_2019_paper.pdf.
- CHEN L, Multimodal deep learning for pig facial segmentation[J]. Computers and Electronics in Agriculture, 208(107563): 15–33.
- CHEN L.-C, ZHU Y, PAPANDREOU G, et al., Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation[C/OL]// Proceedings of the European Conference on Computer Vision (ECCV). [S.l.]: [s.n.]. <https://arxiv.org/abs/1802.02611>.
- CHEN Y, Automatic classification of chicken parts using YOLOv4[J]. Journal of Food Science, 89(6): 3456–3464.
- DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al., An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale[J/OL]. ArXiv preprint arXiv:2010.11929, <https://arxiv.org/pdf/2010.11929>.

- HE K, ZHANG X, REN S, et al., Deep residual learning for image recognition[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. [S.l.]: [s.n.]: 770–778.
- HOWARD A G, ZHU M, CHEN B, et al., Mobilenetv2: Inverted residuals and linear bottlenecks[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. [S.l.]: [s.n.]: 4510–4520.
- HU J, SHEN L, SUN G, Squeeze-and-Excitation Networks[J/OL]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(8): 2011–2023. <https://ieeexplore.ieee.org/document/8578390>.
- INSTITUTE Z A R, SegGPT: A General-Purpose Visual Model for Arbitrary Segmentation Tasks[J/OL]. Zhongguancun Forum, <https://cloud.tencent.com/developer/news/1089585>.
- 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. [S.l.]: [s.n.]: 1646–1654.
- KINGMA D P, BA J, Adam: A method for stochastic optimization[C]// International conference on learning representations. [S.l.]: [s.n.].
- LI W, Real-time animal facial segmentation using lightweight Transformer[J]. IEEE Transactions on Intelligent Transportation Systems, 25(3): 2123–2135.
- LIU Z, LIN Y, CAO Y, et al., Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows[J/OL]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(8): 5130–5143. <https://ieeexplore.ieee.org/document/9557403>.
- OKTAY O, SCHLEMPER J, FOLGOC L L, et al., Attention U-Net: Learning Where to Look for the Pancreas[J/OL]. ArXiv preprint arXiv:1804.03999, <https://arxiv.org/pdf/1804.03999>.
- OQUAB M, VO T, HASSANI A, et al., DINOv2: Learning Robust Visual Features without Supervision[J/OL]. ResearchGate, https://www.researchgate.net/publication/370058767_DINOv2_Learning_Robust_Visual_Features_without_Supervision.
- RONNEBERGER O, FISCHER P, BROX T, U-Net: Convolutional Networks for Biomedical Image Segmentation[C]// International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). [S.l.]: [s.n.]: 234–241. DOI: 10.1007/978-3-319-24574-4_28.
- SANDLER M, HOWARD A, ZHU M, et al., MobileNetV2: Inverted Residuals and Linear Bottlenecks[C/OL]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). [S.l.]: [s.n.]. <https://arxiv.org/abs/1801.04381>.
- WANG Q, DONG B, YANG S, ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks[C/OL]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). [S.l.]: [s.n.]. <https://arxiv.org/abs/1910.03151>.

WANG Z, SHEN C, VAN GOOL L, SegViT: Semantic Segmentation with Plain Vision Transformers[J/OL]. ArXiv preprint arXiv:2302.05523, <https://arxiv.org/pdf/2302.05523>.
ZHAO H, SHI J, QI X, et al., Pyramid Scene Parsing Network[C]// IEEE Conference on Computer Vision and Pattern Recognition (CVPR). [S.l.]: [s.n.]: 6230–6239. DOI: 10.1109/CVPR.2017.662.

plain

致谢

当笔尖落在“致谢”二字上，才惊觉大学时光已悄然行至尾声。四载春秋，如白驹过隙，那些课堂上的求知、实验室的探索、操场上的奔跑，此刻都化作记忆长河里闪烁的星子，见证着从青涩到成熟的蜕变。

首先要深深致谢我的导师冯妍老师。从毕设选题时的迷茫困惑，到开题论证的反复打磨，再到论文撰写中的逐字批注，冯老师始终以春风化雨般的耐心指引着我。记得初次汇报时，我对着粗糙的框架忐忑不安，是您用“学术研究就像剥茧抽丝”的比喻鼓励我；当实验数据遇到瓶颈时，您深夜发来的修改建议邮件，字里行间都是严谨的治学态度与温暖的关怀。您不仅教会我如何构建理论体系、优化研究方法，更以谦和务实的学者风范让我懂得，做学问当如匠人雕琢，需怀敬畏之心、持恒久之功。这份师恩，终将成为我学术道路上的灯塔。

感谢我的父母，二十余载的养育之恩，是你们用双肩扛起生活的风雨，让我得以在象牙塔里安心逐梦。电话里那句“别担心，家里都好”，视频时偷偷抹掉的泪光，返程行囊里塞满的家乡味道，都是无声的爱与支持。你们教会我坚韧与善良，让我明白人生的价值不在于成就高低，而在于心怀热忱、步履不停。这份亲情，是我永远的港湾。

还要感谢并肩走过四年的挚友们。是你们在图书馆的深夜伴读、在答辩前夜的模拟演练、在生日时的惊喜相聚，让枯燥的学业充满了温度。我们曾为一道难题争得面红耳赤，也曾为毕业分离偷偷抹泪，那些共同走过的操场夜话、食堂笑谈，都将成为青春最珍贵的注脚。愿我们此去各自精彩，顶峰相见时，仍能忆起西农路 22 号的蝉鸣与夕阳。

最后，想对自己说一声谢谢。谢谢那个在凌晨三点实验室里坚持调试数据的自己，谢谢那个面对修改意见不气馁的自己，谢谢那个始终相信“慢慢来，比较快”的自己。这段旅程有过迷茫与挫败，却也让我学会在风雨中挺直脊梁。此刻收笔，不是终点，而是新的起点。

愿此去历经山河，仍怀赤子之心；愿未来征途漫漫，常遇星辰大海。冯老师的谆谆教诲、家人的温情陪伴、朋友的真挚情谊，都将化作翅膀，助我飞向更广阔的天空。

行文终章，纸短情长，唯愿时光清浅处，一步一安然。