毕业设计开题报告

课题名称：基于yolo的手语识别系统设计与实现

姓名： 万旭涛

指导教师： 茹慧英

**一、选题背景及意义：**

**选题背景：**

在如今飞速发展的时代，文法手语(简称手语)是听障人士与他人进行平等交流和获取外界信息的主要途径。手语是一种主要通过肢体动作来表述的语言，通过无声的肢体语言，听障群体能够充分地向他人表达自己的思想意图，手语作为听障群体的视觉语言是帮助听障人士融入社会的重要交流方式。根据世界聋人联合会的数据，全世界有超过300种手语，虽然随着时间的推移和制度的完善，手语逐渐拥有了规范的语法、明确的语义和健全的词汇体系，但是，由于普通大众一般缺乏系统性的手语学习，难以准确地理解手语的意思，同时受到地域文化差异的限制，使得听障群体在商场、机场、民政等公共服务场景中常常面临着巨大的沟通障碍。因此，健听人士能够理解听障人士的手语表达是解决听障群体无障碍沟通的主要需求。但是目前大多数通信技术都是为了支持口语或书面语而开发的，支持手语的通信技术平台少之又少。虽然语音聊天和视频通话等通信技术和工具已经成为我们生活的重要组成部分，但听障人士在使用这些技术方面仍存在许多问题。近年来，在全球残疾人群体中听力障碍者所占比例较高，随着对特殊群体的了解不断深入，越来越多的研究机构、科研人员和社会各界开始关注面向手语识别的无障碍设施智能化方面的研究。手语识别在计算机视觉和图像处理等领域一直是备受关注的研究热点，具备广泛的应用潜力。

**意义：**

第一，手语识别的研究在听障人士和健全人士之间的交流中扮演着极为重要的角色。由于普通大众在与听障人士进行沟通交流时，很难实时获取对方所要表达的关键信息，这导致听障群体在使用公共基础设施与服务等方面面临着巨大的挑战。而通过手语识别技术可以实现聋哑人与健听人之间无障碍的交流，以促进平等的沟通机会，帮助听障人群更好地融入社会，具有显著的社会效益；

第二，在现实生活中，手语的学习往往需要在专业的聋哑学校学习，由专业的手语 老师进行教授，但是由于不同人的教学习惯和教学标准都不同，这可能会造成区域差异性。针对我国聋哑教育学校的师资大量短缺的现状，将手语识别技术运用到教学中，更有利于普及和统一标准化手语，为听障群体营造一个良好的学习环境，给聋哑学校教学带来很大的便利，对听障人群的教学与学习起到了重要作用。

第三，就科学和学术而言，手语识别相关技术有利于学者们进一步研究基于计算机视觉的动作识别方法，许多应用程序将受益于手语识别技术，这一趋势广泛拓宽了手语识别的应用范围。对社会的影响也举足轻重。因此，在现有的基础上，探索如何应用并改进目前的手语识别算法以解决手语识别过程中的问题，对于基于计算机视觉的动作识别技术具有重要意义。

**二、课题的基本内容：**

目标检测算法相关介绍，首先综合介绍了基于深度学习的目标检测算 法，其次根据目标检测算法的实现方式，将其分为两阶段和单阶段目标检测算法，并进行了更详细的介绍。最后，对本文使用的 YOLO 系列网络模型进行介绍，为后续章节改进优化提供基础。

基于注意力机制的 YOLO手语识别研究。首先简述了注意力机制 的概念和分类。介绍了 CBAM 注意力机制的原理，同时分析了其所存在的不足，并改进了 CBAM 注意力机制。最后，通过实验表明将改进后的 CBAM 注意力机制融合 在 YOLO 骨干网络中可以提升网络对于目标特征信息提取的能力。

改进 YOLO 的锚框计算方法与损失函数。首先介绍了 YOLO 中锚框的计算方法并对其进行了优化。其次分别介绍了Cross Entropy Loss损失函数和 Lovasz-Softmax Loss 损失函数，同时研究了将两者加权结合使用的效果。最后，通过对 比实验以验证两部分改进内容的有效性。

实验与分析，首先对实验的数据集、实验环境和评价标准进行了介绍。其次，将上述两章改进内容在 YOLO 网络中进行结合，并在 American Sign Language Letters Dataset 数据集上进行了训练及测试，验证了本文改进的手语识别模型具有良好的检测性能。

1. **主要研究思路和方法**

1**，**基于深度学习的目标检测算法

相比于传统的目标检测算法，基于深度学习的目标检测算法利用卷积神经网络直接提取图像的特征，避免了繁琐的手工特征提取过程，使其特别适用于图像特征提取工作。基于深度学习的目标检测算法不仅能够提高目标检测的准确率和识别速度，还可以处理更加复杂和多样化的图像数据。其中基于深度学 习的卷积神经网络(Convolutional Neural Network, CNN)被广泛应用于图像处理领域。 在CNN中，输入数据被分成小块区域，通过前向传播逐层处理，提取图像数据的显著特征。其不仅考虑了不同层之间的联系，同时加入了空间信息，因此在图像识别领域表现优异。卷积神经网络极大地提升了网络对于目标特征提取的能力，具备了速度快、精度高、可移植、鲁棒性强的特点。

2，YOLO 系列相关算法

YOLO 算法，实现了目标区域和目标类别的联合预测， 将两个任务融合成一个，将目标检测任务转换为采用单个卷积神经网络作用在图像上， 从中提取预测物体边界框和类别概率的回归问题，实现神经网络只需要看一次就可以检 测目标类别和位置的想法。YOLO 算法在不断的迭代更新，其中 YOLOv1 奠定了整个 YOLO 系列的基础，后续的 YOLO 算法是对其的不断改进创新。

3，基于注意力机制的YOLO手语识别研究

在 YOLO 算法的 基础上进行改进，将改进后的 CBAM 注意力机制加入到骨干网络中，增强特征空间信息以及通道信息之间的关系，重点关注手势区域特征，以提高手语识别的准确率。

3.1通道注意力机制

通道注意力机制关注输入信号在不同通道之间的相关性，通过学习不同通道的权值，并用权值对不同通道进行多重划分，以增强网络对关键通道域的注意。对于输入的特征图(C H W ) F R × × ∈，基于通道域的注意力通常首先对通道内的信息直接全局平均池化，产生一个通道向量 C FC ∈ R 。

3.2 空间注意力机制

空间注意力可以被视为使神经网络探索输入信息中重要的空间区域的方式。空间注意力机制，通过分配不同的权重来关注关键目标区域的所有空间位置，以获取更多的空间信息。在手语识别任务中，空间注意力机制将网络的关注点聚焦于目标的手部区域，以最小化复杂背景对任务的干扰。

3.3 混合注意力机制

混合注意力机制中分为通道域和空间域，将输入信号同时考虑在两者内，通过通道注意力和空间注意力的串联或并联的方式，形成一个综合的注意力机制。在通道注意力机制的基础上添加空间注意更有助于捕获空间维度的信息，因为其聚焦于特征图上的有效信息的具体位置，能够产生更好的特征描述信息，从而很好地做到补充和完善通道注意力机制的作用。

4. 改进 YOLO 的锚框计算方法与损失函数

首先对锚框(Anchor)计算方法进行改进，使用 K-means++聚类算法提高先验锚盒的尺寸匹配度，使得先验锚框与手语图像匹配地更加精准；其次将 Cross Entropy Loss 损失函数和 Lovasz-Softmax Loss 损失函数加权结合使用，试图让网络在模型训练过程中收敛的更加稳定，以便更好地满足实际应用的需求。

1. **课题的重点拟解决的问题**
2. 基于yolo融合了空间注意力和通道注意力的 CBAM 注意力机制虽然可以提高网络的检测准确率。但由于CBAM中的通道注意对通道特征的空间维度进行了压缩，导致部分通道特征信息的缺失的问题。
3. 为了进一步提高网络的准确性和稳定性，本文首先使用 K-means++聚类算法 代替原始的 K-means 聚类算法，提高了先验锚框的尺寸匹配度，实现先验锚框与目标手语图像的精确匹配。
4. 基于改进 YOLOv5s 的手语识别实验。首先搭建了 YOLO 手语识别任务实验环境。其次将注意力机制、聚类算法和损失函数这三个方面的改进方法进行组合，通过消融实验验证了三个方面的改进方法组合使用的有效性，并与不同网络模型进行了对比实验。
5. **总体安排和进度**

2023.12.01-2023.12.25 选题、课题调研、收集资料、外文翻译、编写开题报告；

2023.12.26-2023.12.29 开题答辩

2024.02.26-2024.03.03 阅读有关书籍或文献、初步构想、论文的初稿撰写；

2024.02.26-2024.03.24 毕业实习

2024.03.25-2024.05.08 详细论证 、中期工作汇报；

2024.05.09-2024.05.31 论文撰写与完善；

2024.06.01-2024.06.04 资料提交上传；

2024.06.05-2024.06.11 指导教师评阅 、评阅教师评阅；

2024.06.12-2024.06.16 毕业论文答辩；

2024.06.17-2024.06.20 毕业设计总结、评优。

1. **主要参考文献**

[1]王琪.基于深度学习的手语识别系统设计与实现[D].西安电子科技大学,2021.DOI:10.27389/d.cnki.gxadu.2021.000324

[2]邢晋超.基于改进YOLOv5s的手语识别算法研究[D].中北大学,2023.DOI:10.27470/d.cnki.ghbgc.2023.001085

[3]邢晋超,潘广贞.改进YOLOv5s的手语识别算法研究[J].计算机工程与应用,2022,58(16):194-203.

[4]蒲俊福,基于深度学习的视频手语识别研究[D]. 中国科学技术大 学,2020.DOI:10.27517/d.cnki.gzkju.2020.000530.

[5]韩凯. 基于深度学习的目标检测研究[D].西南科技大学,2018.

[6]吕情深.基于目标检测的中国手语识别研究[D]. 新疆大 学,2020.DOI:10.27429/d.cnki.gxjdu.2020.000068.

[7]刘建伟,刘俊文,罗雄麟.深度学习中注意力机制研究进展[J].工程科学报,2021,43(11):1499-1511.DOI:10.13374/j.issn2095-9389.2021.01.30.005.

[8]王丽光,张根源,刘子龙.基于单目视觉的实时手语识别系统设计与实现[J].电子科技,2017,30(03):130-133+137.DOI:10.16180/j.cnki.issn1007-7820.2017.03.036

[9]罗文杰.基于深度学习的手语识别研究与应用[D].南昌大学,2023.DOI:10.27232/d.cnki.gnchu.2023.003074

[10]高辉.基于深度学习的手语识别系统设计[D].哈尔滨理工大学,2023.DOI:10.27063/d.cnki.ghlgu.2023.000303

[11]王婧瑶,范飞,刘豪宇等.基于机器视觉的聋哑人手语识别——语音交互系统[J].物联网技术,2021,11(12):3-5.