基于多尺度特征融合与语言建模的手写数学公式识别

一.成果完成人及所在单位：刘翔昊 计算机与信息工程学院 指导教师：王大寒

二.成果简介：

本成果研究了一种基于多尺度特征融合与语言建模的手写数学公式识别方法。本研究基于 Encoder-Decoder 架构，首先使用编码器编码公式图像（见图 1）特征，接着使用 DenseASPP 方法获取输入图像的多尺度特征，然后使用 Non-local Neural Network 获取特征中的长距离依赖，充分利用局部和非局部的图像特征。除此之外，还使用字符分类器以更好地区分和识别相似符号，使用 FastText 语言模型获取数学表达式中隐含的语义信息，最后使用解码器解码特征为 LaTeX 序列。这项工作在多个数据集上验证了所提出方法的有效性和可行性，可以有效提高手写数学公式识别率。

三. 主要研究内容及创新点：

本研究主要设计了一种端到端的手写数学公式识别模型，包括编码器、DenseASPP，Non-local Neural Network、字符分类器、FastText 语言模型和解码器（见图 2、图 3、图 4）。编码器使用 DenseNet 将输入图像编码为高级特征。DenseASPP 融合输入图像的多尺度特征，扩展感受野，从而减轻手写数学公式中长文本和不同大小符号所导致的上下文信息丢失和提取特征中的细粒度细节的影响。Non-local Neural Network 可以捕捉特征中的长距离依赖，充分利用利用局部和非局部的图像特征。字符分类器模块使用了 focal loss，调整了简单样本和难样本的权重，使模型更加关注易被误分类的样本。FastText 作为预训练语言模型，为训练集中的数学表达式生成相应嵌入向量，并学习数学表达式中的语义信息。解码器使用带有注意力机制的 GRU 将高级特征转换为一系列 LaTeX 符号。本研究在 CROHME 2014、CROHME 2016、CROHME 2019 数据集上验证了所提出方法的有效性和可行性。



图 1 手写数学公式图像

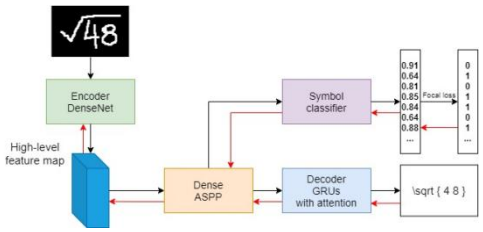


图 2 DenseASPP 模型整体结构图

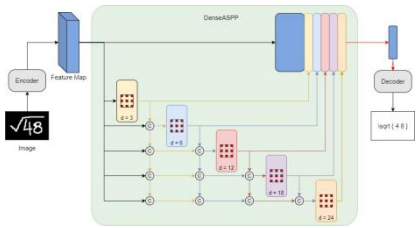


图 3 DenseASPP 结构图

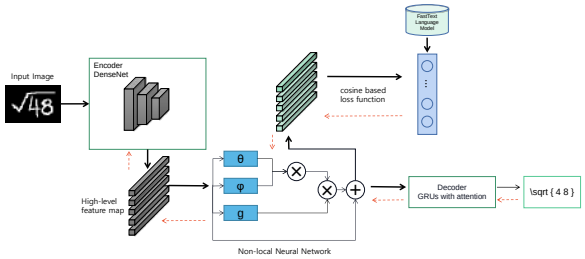


图 4 Non-local Neural Network & Language Model 模型整体结构图

四. 主要研究成果及应用情况

在 ICCPR 2022 发表论文 2 篇。
在 PRCV 2022 发表论文 1 篇。

联系人：刘翔昊，　　电　话：13914567901　　Email: 578312491**@qq.com**