AI生成人脸图像鉴别

项目说明文档

为方便评委更好地理解项目实施细节，本参赛小组提供详细说明文档。该说明文档由三部分构成：1**项目背景** 2**模型判别流程** 3**代码测试说明** 以下分别详细说明。

# 项目背景

如今，图像编辑软件的滥用日益严重，导致数字图像的真实性受到质疑。同时，网络社交网络的广泛存在使其成为传播伪造图像以报道假新闻、传播谣言等的主要渠道。此外，人工智能深度学习发展迅猛，尤其以生成对抗网络和扩散模型为主的模型算法发展迅猛。这些生成式的模型能够产生以假乱真的图片，虽然它具有许多新颖应用的潜力，但由于虚假信息的滥用，它也带来了新的风险。例如，一些非专家或媒体出于政治或者经济的目的而滥用生成式技术，导致严重的社会威胁。因此，开发伪造检测技术来帮助人们确定图像真实性极其必要。在这一背景下，本团队开展对AI生成人脸图像鉴别的研究。

# 模型判别流程



**模型网络使用预训练的resnet50，并做了以下优化调整：**

**图像预处理：**

对输入图像做了预处理，如翻转，裁剪，调整亮度对比度和标准化等。该操作能够提高模型的泛化能力，通过上述预处理操作，模型能够学习到更广泛的数据特征，从而在新的、未见过的数据上表现更好。

**MTCNN：**

在训练和测试时加入MTCNN人脸检测提高精度。通过精确的人脸检测，可以减少将非人脸区域错误识别为人脸的情况，降低误检率。同时，MTCNN不仅检测人脸，还对人脸进行关键点定位，有助于实现更准确的人脸对齐，这对于需要严格对齐的应用（如人脸识别）来说尤其重要。此外，MTCNN能够处理不同姿态、不同表情、部分遮挡等情况的人脸，对于本项目契合程度比较高。MTCNN的设计考虑了速度和效率，使其可以在实际应用中实现较为实时的检测。

**Resnet50：**

1. 由于是迁移学习，冻结了模型网络的前三层，仅layer4和fc层更新参数，可以减少过拟合，并让模型在新任务上有更好的适应能力。
2. 添加dropout层和权重衰减来防止过拟合。Dropout通过在训练过程中随机“丢弃”一部分神经元的输出，强迫网络学习更加鲁棒的特征，减少模型对特定训练样本的依赖，同时，权重衰减通过对网络权重施加惩罚，鼓励模型学习较小的权重，从而减少模型复杂度，提高泛化能力。
3. 将全连接层的输出修改为2，用于2分类任务。能够简化决策过程，输出层的每个节点可以表示属于每个类别的概率，使得决策过程更加直接和清晰。
4. 使用正态分布初始化全连接层的权重，以提高训练的稳定性。正态分布初始化可以确保权重不会过大或过小，这有助于避免在训练初期出现梯度消失或梯度爆炸的问题。
5. 使用适用于二分类任务的交叉熵损失（CrossEntropyLoss）。交叉熵损失函数的梯度通常比较平滑，这有助于在训练过程中使用梯度下降算法进行有效学习。
6. 对梯度进行裁剪，避免梯度爆炸，梯度裁剪可以有效地防止在训练过程中出现的梯度爆炸问题，这是深度学习模型训练中常见的问题，特别是当网络非常深或使用长序列数据时，提高训练的稳定性。

**其他优化：**

1. 动态学习率调整，逐步降低学习率以提高训练后期的效果。在训练初期使用较高的学习率可以帮助快速收敛到一个较好的解，而在训练后期降低学习率有助于模型在解空间中更精细地调整，从而提高最终效果。
2. 采用early stopping，如模型在验证集上的性能停止提升，提前终止训练，避免模型在训练集上过拟合。亦可节约计算资源，提前终止训练可以节省时间和计算资源，因为不需要完成所有的训练迭代。

# 代码测试说明

测试请在Ai\_main/CNNDetection目录下运行main.py文件，运行结束后会在Ai\_main目录下生成cla\_pre.csv文件，即模型鉴别结果文件。

**修改main.py中的路径**：

model\_path：模型存放路径（默认即可，找不到路径时复制Ai\_main/CNNDetection/model/中的模型路径）。

dataroot：测试文件夹路径。

**模型训练**：./train-detector.sh {Train-Dir} {Valid-Dir}