实验报告

大作业：基于深度学习的人脸年龄预测

组别：8

姓名	刘畅	学号	202011081053
姓名	姚冠宇	学号	202011260070
姓名	黄梓峻	学号	202011081026
姓名	何佳文	学号	202011081077
姓名	庞志豪	学号	202011150091
姓名	李子豪	学号	202011081088

实验报告内容按以下条目整理：

1. 研究目的
2. 研究背景和现状（简要阐述基于人脸图像进行年龄估计的研究现状）
3. 实验材料（数据集介绍和说明等）
4. 实验方法（包括数据处理、模型构建、训练、预测结果评估等）
5. 结果（文字说明，图表等形式呈现）
6. 讨论（根据自己的结果，也可以结合现有文献的结果，进行解释和讨论）
7. 结论
8. 研究意义
9. 小组分工
10. 参考文献

格式要求：用序号标明提纲的层次目录，标题用四号字、加粗，正文五号字，单倍行距。

1. 研究目的

构建深度学习网络模型，在给定的Adience数据集[1]上完成年龄分类任务

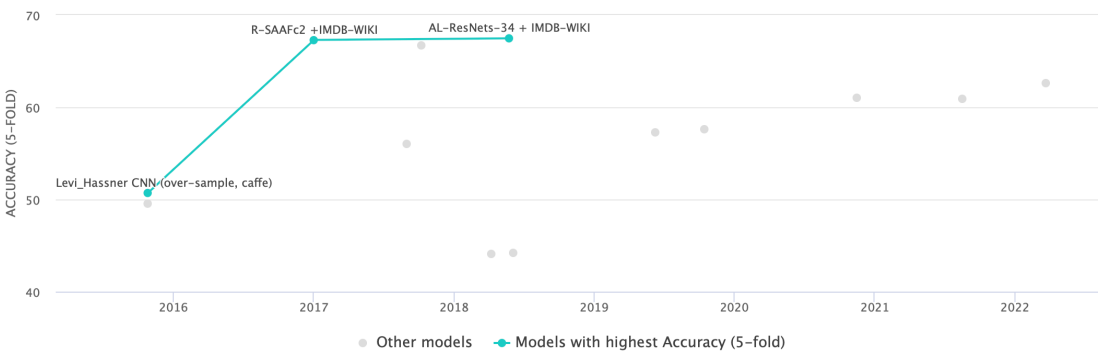
2. 研究背景和现状

在传统机器学习方法中，一般使用专家人工设计特征的方式来对于人脸图像进行年龄预测，然而，受到光照、角度等多种因素的影响，这样设计的模型往往难以达到很好的最终效果。

深度学习使用了端到端的方式，构建了深度神经网络来提取图像特征，从而更好地实现年龄预测。

年龄预测有两种实现方式：分类和回归。如果将每个年龄段看作是一个不同的类，年龄估计可以被看作是一种分类问题；如果将年龄作为一个连续变量，年龄估计也可被视为一种回归问题。针对不同的年龄数据库和不同的年龄特征，分类模式和回归模式具有各自的优越性。

近年来，出现了引入线性代数相关理论的秩不变网络(COnsistent RAnk Logits (CORAL))进行有序回归(ordinal regression)[2]的方法，引入了概率论中后验概率理论辅助识别的算法[3]，使用耳部及轮廓图进行辅助识别的算法[4]等，目前该领域还在不断产生新的方法，效果不断提升。



为了更好的实现分类任务，研究人员提出了众多网络模型，如提出了Relu激活函数、局部响应归一化（LRN）等先进方法的AlexNet[5]，采用了残差结构模块的ResNet[6]等，为更好地实现分类任务提供了众多可供参考的理论方法和网络模型。

Adience数据集先前的5折交叉验证的top-1准确率如上图所示，在50%至70%之间波动。

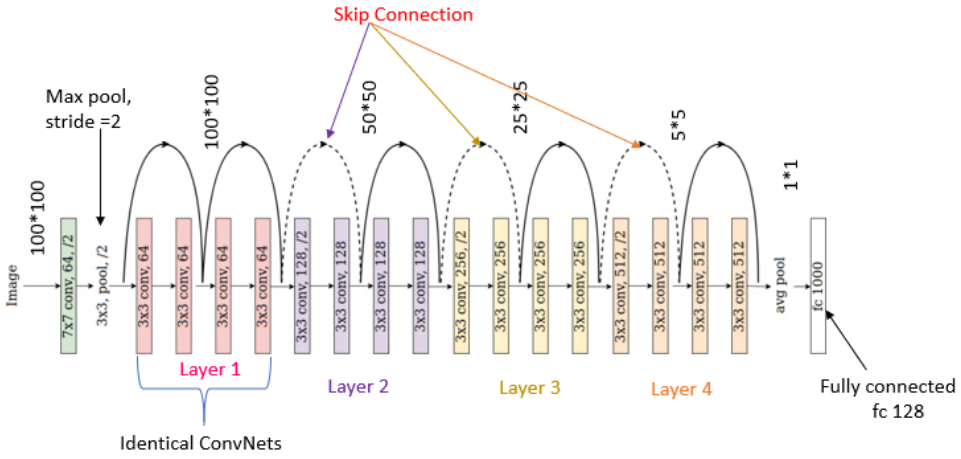
3. 实验材料

Adience数据集对齐后的aligned数据
Label使用原始label

4. 实验方法

0.代码结构:
所需环境在requirements文档中。
训练请运行main.py,并将第112行和114行改为自己的标签数据路径和图片数据路径。
测试模型请运行test_code,并将第42行和44行改为自己的标签数据路径和图片数据路径,第153行改为自己的预训练模型路径。
如果想获取分类错误图片，请在test_code第227行将showerros改为true，并修改第92行为存储错误图片的路径。
分类错误图片示例在error文件夹里。

1.网络结构:
我们采用resnet-18的网络结构，同时使用batch normalization: （具体介绍在讨论部分）



ResNet-18 Architecture

Fruit 360 Input Image size= 100*100 px

同时，为了使梯度能够更好的回传，我们将ReLU激活函数改为了抑制区0.2的LeakyReLU激活函数，为了更好的执行分类任务，我们将最后一层通过连接层后加上了softmax激活函数用来归一化每个类别的概率。

2.数据清洗:
查阅先前论文我们选择将数据集中的图片先变换为224x224分辨率，之后将rgb三个通道的均值和方差分别变为:[0.485,0.456,0.406],[0.229, 0.224, 0.225].
对于标签，我们将将年龄划分为了8个区间：(0-2, 4-6, 8-13, 15-20, 25-32, 38-43, 48-53, 60-)，包括19487张图像。即在此数据集上，将年龄预测转化为8分类问题。

3.数据增强:
为了能更好的利用数据集的数据，我们首先采用了RandomHorizontalFlip（随机水平翻转），RandomRotation(20)（随机旋转20度以内），ColorJitter(brightness=0.1, contrast=0.1,saturation=0.1, hue=0.02)（亮度随机改变原图0.1、对比度随机改变原图0.1、饱和度随机改变原图0.1和色调随机改变原图0.02），尽可能排除了光照和角度等因素的影响。

之后我们对数据进行了gird mask[7]数据增强，具体来说，我们随机（概率为随训练轮次从0.6逐渐涨为0.8）为图片加上如下图所示的mask来让网络进行判断，目标是增强网

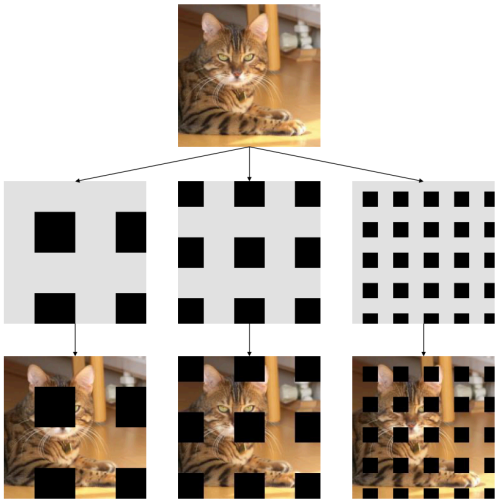


Figure 3. This image shows examples of GridMask. First, we produce a mask according to the given parameters $(r, d, \delta_x, \delta_y)$. Then we multiply it with the input image. The result is shown in the last row. In the mask, gray value is 1, representing the reserved regions; black value is 0, for regions to be deleted.

络的判别能力。

具体执行时，我们先设定网格间隔d和保持率keep ratio r，之后随机初始化最开始的起点的位置$0 < \delta x, \delta y < d$，如下图所示：

对于d和r的选择我们根据原始论文中的结果选择[96, 224]和0.6.

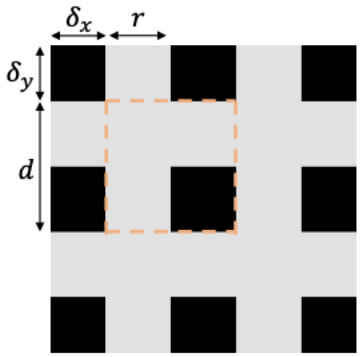


Figure 4. The dotted square shows one unit of the mask.

4. 交叉验证：我们将数据集分为5折（没有划分验证集是因为我们不需要做超参数调整），每次选4折作为训练集，1折为测试集，总共训练5个模型出来，最终以5个模型的平均表现作为我们的评估指标来尽量避免在数据集上的过拟合。

5. 训练方法：我们每折进行100轮训练，共进行5折交叉验证。数据标签采取one-hot编码，对应的损失函数选择交叉熵损失函数。初始学习率设置为0.1。同时，在训练到第20，50，75轮时学习率乘以0.1以随训练轮次调整学习率。

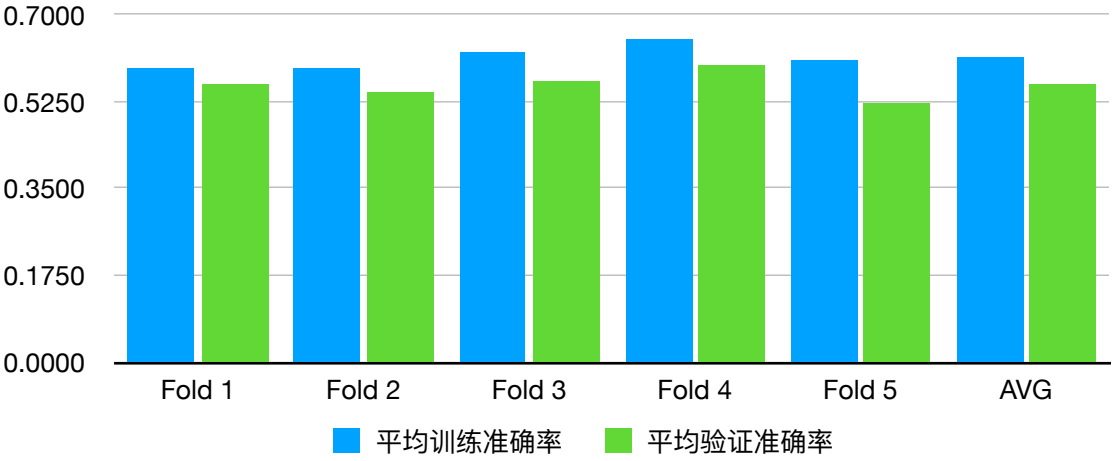
5. 结果

具体的训练输出附在了word文档中可以查看，总共进行了5折交叉验证，每折训练100轮，这里展示5折的结果。同时展示模型最终测试结果

```
our model has size=47.92 MiB
inference time: 216.52788496017456
evalue accuracy:62.06%
one-off accuracy:80.35%
```

五折交叉验证准确率

	平均训练准确率	平均验证准确率
Fold 1	0.5939	0.5598
Fold 2	0.5945	0.5457
Fold 3	0.6258	0.5640
Fold 4	0.6529	0.5994
Fold 5	0.6094	0.5212
AVG	0.6153	0.5580



6. 讨论

1. 模型具体介绍:

图像输入是batchsize*3*224*224，先通过一个7*7*64的卷积，但是步长设置为2，使得图像的大小缩小了一半;

每经过两次卷积我们进行一次残差链接。每过一次卷积我们先进行batch_normalization再经过leakyrelu激活函数。我们定义两次卷积为一个block，每两个block为一个layer，resnet-18总共有4个layer。

在con2_x的刚开始，通过一个最大值池化，步长设置为2，使得图像又缩小了一半;

然后是con2_x、con3_x、con4_x、con5_x一共8个残差块;

在con3_1、con4_1、con5_1都进行了2倍的下采样;

最后一层先经过一个自适应平均池化层，然后一个全连接层映射再经过softmax概率归

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
		3×3 max pool, stride 2				
conv2_x	56×56	[3×3, 64; 3×3, 64] ×2	[3×3, 64; 3×3, 64] ×3	[1×1, 64; 3×3, 64; 1×1, 256] ×3	[1×1, 64; 3×3, 64; 1×1, 256] ×3	[1×1, 64; 3×3, 64; 1×1, 256] ×3
conv3_x	28×28	[3×3, 128; 3×3, 128] ×2	[3×3, 128; 3×3, 128] ×4	[1×1, 128; 3×3, 128; 1×1, 512] ×4	[1×1, 128; 3×3, 128; 1×1, 512] ×4	[1×1, 128; 3×3, 128; 1×1, 512] ×8
conv4_x	14×14	[3×3, 256; 3×3, 256] ×2	[3×3, 256; 3×3, 256] ×6	[1×1, 256; 3×3, 256; 1×1, 1024] ×6	[1×1, 256; 3×3, 256; 1×1, 1024] ×23	[1×1, 256; 3×3, 256; 1×1, 1024] ×36
conv5_x	7×7	[3×3, 512; 3×3, 512] ×2	[3×3, 512; 3×3, 512] ×3	[1×1, 512; 3×3, 512; 1×1, 2048] ×3	[1×1, 512; 3×3, 512; 1×1, 2048] ×3	[1×1, 512; 3×3, 512; 1×1, 2048] ×3
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Table 1. Architectures for ImageNet. Building blocks are shown in brackets (see also Fig. 5), with the numbers of blocks stacked. Down-sampling is performed by conv3_1, conv4_1, and conv5_1 with a stride of 2.

一化到输出。

2. 测试集预处理:

首先对图像裁切到224x224，之后将rgb三个通道的均值和方差分别变为:[0.485,0.456,0.406],[0.229, 0.224, 0.225].

3. 测试机准确率评估方式:

我们采用top-1准确率，具体来说，对网络输出的归一化概率选取概率最大的位置作为网络预测输出与标签进行对比，如果相同则预测正确，以此计算准确率。

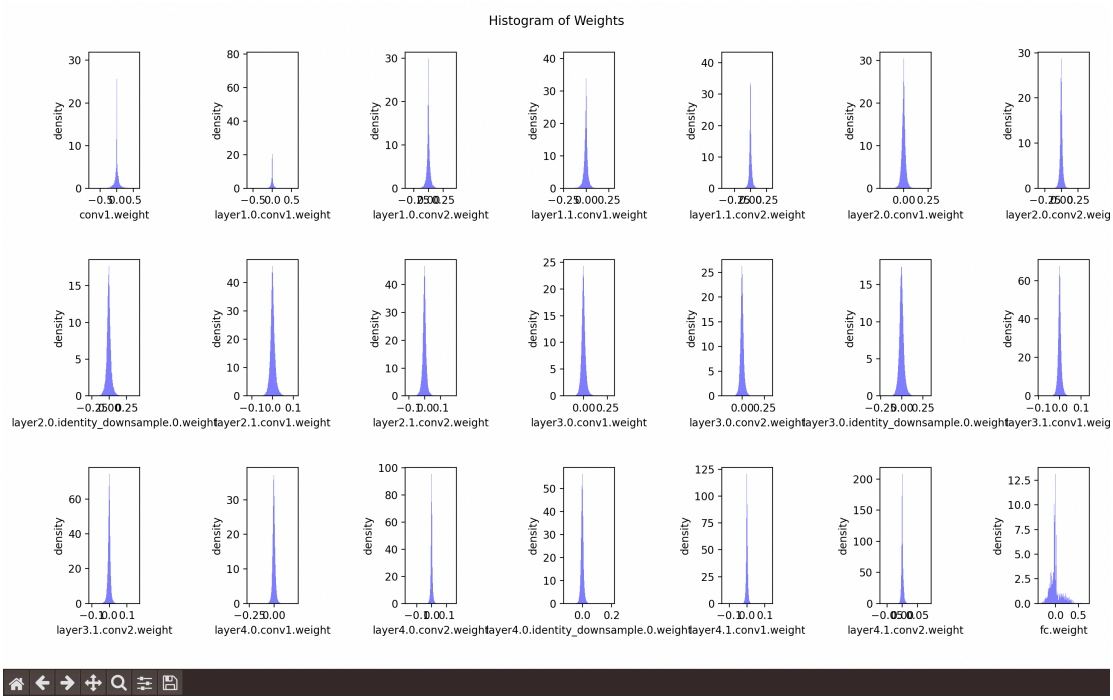
4. 模型大小、泛化能力，与推理时间（采用fold0的网络参数）：

```
our model has size=47.92 MiB
inference time: 216.52788496017456
evalue accuracy:62.06%
one-off accuracy:80.35%
```

可见，我们的模型在较小的参数量上实现了较短的推理时间和较高的准确率。

5. 模型参数分布：（横坐标代表参数值纵坐标代表处于这个值的参数数量）

可以看出模型参数大多数分布在区间[-0.5,0.5]，呈钟形分布，没有出现极端值。说明我们模型收敛较好。



6. 错误分类图片与分析：

如果想获取分类错误图片，请在test_code第227行将showerros改为true，并修改第92行为存储错误图片的路径。

下面展示一些分类错误图片（色调很奇怪是因为数据预处理时做了三个通道的均值和方差的调整）



可以看到分类错误的图片多为年轻女性，这类图片的年龄区间在现实中也不好判断。有些女士虽然年龄很大但是看着依旧年轻，加上女士的发型多变，也会影响最终的判断。

从下图分类错误样本模型实际的输出也可以看到，对于分类错误图片，模型在中间几类的预测概率十分接近，说明模型也不能在细分的区间上进行准确的判断。

```
tensor([4.5086e-03, 9.7799e-01, 1.0216e-02, 8.9212e-04, 2.9818e-03, 2.0891e-03,
        6.2009e-04, 7.0342e-04])
```

7. 结论

通过实验，我们成功的利用较少的参数量实现了较高准确率的分类效果，这说明了可以利用gridmask、HSL值的随机偏移等方式实现数据增强，并利用LeakyRELU、batch normalization、残差连接等多种方法相结合，调整神经网络结构，从而使神经网络能够更加高效地捕捉数据的内部特征，从而取得更优的识别效果。同时，随着训练轮次调整学习率的方法也可以使网络更好的收敛。我们对网络参数分布的统计方法也可以更好的观察到网络的拟合情况，如果出现非钟形的参数分布或者极端的参数出现大概率说明网络没有很好收敛。

同时，我们的模型也存在着不能在细分的区间上进行准确的判断的问题。还有改进的空间。如在网络结构方面可以引入自注意力机制，模型轻量化方面可以尝试剪枝，训练方法方面可以改进损失函数，数据增强方面可以尝试其他数据增强方法。

由于时间有限，训练次交叉验证的时间也较长，我们并没有做有关消融实验来进一步验证方法的有效性，更多的是直接引用论文的结果，之后可以进行相关实验的补充。

8. 研究意义

通过实验，我们了解了在实现深度网络预测人脸图片年龄的过程中，进行数据处理以及网络超参数调整的方式方法，了解了防止网络过拟合的可用方式，并了解了该领域的最新研究成果，如:

在数据处理方面，我们了解了调整数据归一化的方法以及对图像数据进行的翻转、旋转、HSL值的随机偏移以及最新的gridmask数据增强法。

在网络设计方法方面，我们了解了batch normalization方法以及采用LeakyReLU函数替代ReLU函数的改进方法。

在训练方法方面，我们采用随训练轮次减小学习率的方法来保证收敛。

在更准确评估模型方面，我们了解了k折交叉验证法的实际应用。我们引入了对网络参数分布的统计，通过观察网络的参数分布来判断收敛情况。

在最新研究成果方面，我们了解到了如何巧妙利用数学原理设计神经网络结构，从而更加全面地提取数据特征的方法以及最新的网络结构的原理及其实际应用。

9. 小组分工

姚冠宇：模型设计，训练方式设计，模型评估的代码实现；实验报告撰写。
刘畅：负责相关文献与模型的查找，实验报告撰写
黄梓峻：负责ppt的撰写，数据读取，代码注释
何佳文：代码注释，模型设计，数据增强。
庞志豪：负责模型训练，完善实验报告
李子豪：负责ppt的撰写，查找相关资料图片

10. 参考文献

[1] Levi G , Hassncer T . Age and gender classification using convolutional neural networks[C]// IEEE Conference on Computer Vision & Pattern Recognition Workshops. IEEE Computer Society, 2015:34-42.

[2] Wenzhi Cao, Vahid Mirjalili, Sebastian Raschka. Rank consistent ordinal regression for neural networks with application to age estimation[C]// Pattern Recognition Letters. 2019:140, 325-331

[3] Yunxuan Zhang, Li Liu, Cheng Li, and Chen Change Loy. Quantifying Facial Age by Posterior of Age Comparisons[C]// Proceedings of British Machine Vision Conference. 2017

[4] Dogucan Yaman, Fevziye Irem Eyiokur, Hazım Kemal Ekenel. Multimodal Age and Gender Classification Using Ear and Profile Face Images[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2020.

[5]Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton. ImageNet classification with deep convolutional neural networks[C]// Communications of the ACM. 2017:84-90.

[6]Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE 2016:770-778

[7] https://arxiv.org/abs/2001.04086