

**2021年《机器学习》**

**工程报告**



**课 程：** 机器学习

**姓 名：**

**学 号：**

**完成时间：**

一． 工程摘要与每人贡献

摘要：总体介绍本工程的工作。

每人分工：

|  |  |  |  |
| --- | --- | --- | --- |
| 姓名 | 角色  （组长，组员，独自完成） | 工作量  比例 | 负责内容 |
| 张三 | 组长 | 60% | 收集\*数据集；\*算法代码实现；报告的第三/第五部分撰写..... |
|  |  |  |  |
|  |  |  |  |

二、研究背景与意义

说明本工程所涉及的研究背景研究的意义，包括主要涉及领域，主要研究方法，主要存在问题，现有解决方案等等，可以自行补充其他内容。

#### 背景与意义

人脸识别(Face Recognition)是一种依据人的面部特征(如统计或几何特征等)，自动进行身份识别的一种生物识别技术，又称为面像识别、人像识别、相貌识别、面孔识别、面部识别等。通常我们所说的人脸识别是基于光学人脸图像的身份识别与验证的简称。

人脸识别利用摄像机或摄像头采集含有人脸的图像或视频流，并自动在图像中检测和跟踪人脸，进而对检测到的人脸图像进行一系列的相关应用操作。技术上包括图像采集、特征定位、身份的确认和查找等等。简单来说，就是从照片中提取人脸中的特征，比如眉毛高度、嘴角等等，再通过特征的对比输出结果。

情感是任何人际关系中不可避免的一部分沟通。它们可以用许多不同的形式表达出一个人内心的想法和生理状态。因此，使用正确的工具进行表情的识别和追踪是一个重要的课题。人们越来越需要检测一个人的身体状况以及过去几年的情绪。人类情感识别在各个领域都有着广泛的应用。目前主要的应用领域包括人机交互、安全、机器人制造、医疗、通信和汽车领域等。

普遍认为人类主要有六种基本情感：

愤怒（anger）、高兴（happiness）、悲伤（sadness）、惊讶（surprise）、厌恶（disgust）、恐惧（fear）。而大多数表情识别是基于这六种情感及其拓展情绪实现的。

随着人脸的计算机处理技术（包括人脸检测和人脸识别）不断完善，利用计算机进行面部表情分析也就成为可能。总体而言，表情分析是一个非常困难的研究方向，主要体现在表情特征提取的准确性和有效性上。尤其是后者，因为各种表情本身体现在各个特征点运动上的差别就不是很大，例如：嘴巴张开并不代表就是笑，也有可能是哭和惊讶等。

#### 面临的困难

主要困难点是：

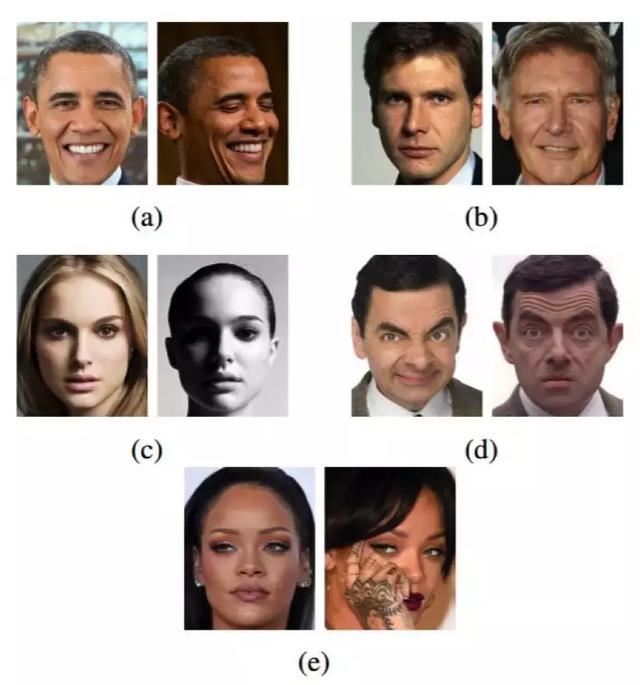
1）表情的精细化程度划分：每种情绪最微弱的表现是否需要被分类。分类的界限需要产品给出评估规则。

2）表情类别的多样化：是否还需要补充其他类别的情绪，六种情绪在一些场景下远不能变现人类的真实 情绪。因此除了基本表情识别外，还有精细表情识别、混合表情识别、非基本表情识别等细致领域的研究。

3）缺少鲁棒性

4）光照问题。光照问题是机器视觉重的老问题，在人脸识别中的表现尤为明显。由于人脸的3D结构，光照投射出的阴影，会加强或减弱原有的人脸特征。

5）年龄变化，比如说人脸的老化，对于情绪识别来说，人脸老化或许更为重要，因为随着人脸的老化，会出现皱纹等一系列面部特征，这会对模型提取面部特征进行表情分析增加难度。除此之外，人脸图像可变的地方包括头部姿势、遮挡、光照条件和人脸表情等。



图中表示在自然人脸图像中找到的典型变化。（a）头部姿势，（b）年龄，（c）光照，（d）面部表情，（e）遮挡

#### 研究的过程和解决方法

人脸识别涉及到计算机视觉、计算机图形学、图像处理、生理学、心理学等，是一门交叉学科。实际上自七十年代以来，人脸识别已经成为了计算机视觉和生物识别领域被研究最多的主题之一。

对于一些问题，实际上是有一些专门的解决办法的。

对于光照问题，对其进行包括光照强度和方向、人脸反射属性的量化，面部阴影和照度分析等，尝试建立数学模型，以利用这些光照模型，在人脸图像预处理或者归一化阶段尽可能的补偿乃至消除其对识别性能的影响，将固有的人脸属性（反射率属性、3D表面形状属性）和光源、遮挡及高光等非人脸固有属性分离开来。

年龄变化这个问题，我们能做的似乎只有想尽办法提取人的面部特征的数据。因为我们做的是情绪识别，所以，我们的主要目标还是集中在人的面部特征上面，对于像是耳朵等其他特征来说并不是很注重，所以我们可以对影响表情相应的面部特征进行重视。特征提取的时候也是注重诸如嘴巴和面颊肌肉活动的提取。

人的表情是富含多种情绪的，我们需要人为地对训练的一些数据进行分析。如果一种情绪中是包含很多的其他人类情绪的，很多时候我们只能根据这个情绪中我们的划分的情绪所占的百分比来对情绪打标签，或者对此变成一个新的情绪。人情绪的多样性，在我们分类的时候是很大的一个困扰。我们的能做的估计也之后不多扩展需要辨别情绪的，使之不仅仅只有六种基本情绪。

但是，人脸识别技术这些年已经发生了重大的变化。基于人工设计的特征和传统机器学习技术的传统方法近来已被使用非常大型的数据集训练的深度神经网络取代。传统方法依赖于人工设计的特征（比如边和纹理描述量）与机器学习技术（比如主成分分析、线性判别分析或支持向量机）的组合。人工设计在无约束环境中对不同变化情况稳健的特征是很困难的，这使得过去的研究者侧重研究针对每种变化类型的专用方法，比如能应对不同年龄的方法、能应对不同姿势的方法 、能应对不同光照条件的方法等。近段时间，传统的人脸识别方法已经被基于卷积神经网络（CNN）的深度学习方法接替。深度学习方法的主要优势是它们可用非常大型的数据集进行训练，从而学习到表征这些数据的最佳特征。网络上可用的大量自然人脸图像已让研究者可收集到大规模的人脸数据集 ，这些图像包含了真实世界中的各种变化情况。使用这些数据集训练的基于 CNN 的人脸识别方法已经实现了非常高的准确度，因为它们能够学到人脸图像中稳健的特征，从而能够应对在训练过程中使用的人脸图像所呈现出的真实世界变化情况。此外，深度学习方法在计算机视觉方面的不断普及也在加速人脸识别研究的发展，因为 CNN 也正被用于解决许多其它计算机视觉任务，比如目标检测和识别、分割、光学字符识别、面部表情分析、年龄估计等。

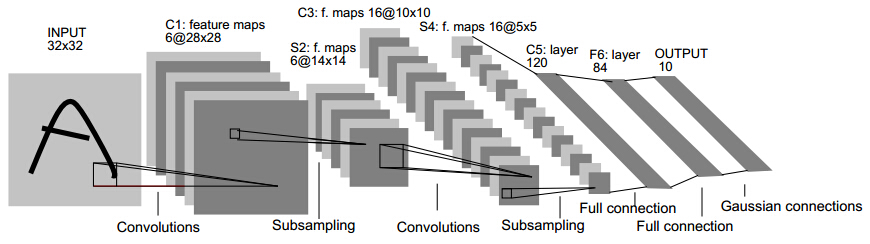
卷积神经网络（CNN）是人脸识别方面最常用的一类深度学习方法。深度学习方法的主要优势是可用大量数据来训练，从而学到对训练数据中出现的变化情况稳健的人脸表征。这种方法不需要设计对不同类型的类内差异（比如光照、姿势、面部表情、年龄等）稳健的特定特征，而是可以从训练数据中学到它们。深度学习方法的主要短板是它们需要使用非常大的数据集来训练，而且这些数据集中需要包含足够的变化，从而可以泛化到未曾见过的样本上。幸运的是，一些包含自然人脸图像的大规模人脸数据集已被公开，可被用来训练 CNN 模型。除了学习判别特征，神经网络还可以降维，并可被训练成分类器或使用度量学习方法。CNN 被认为是端到端可训练的系统，无需与任何其它特定方法结合。

使用神经网络来做人脸识别并不是什么新思想。1997 年就有研究者为人脸检测、眼部定位和人脸识别提出了一种名为「基于概率决策的神经网络（PBDNN）」的早期方法。这种人脸识别 PDBNN 被分成了每一个训练主体一个全连接子网络，以降低隐藏单元的数量和避免过拟合。研究者使用密度和边特征分别训练了两个 PBDNN，然后将它们的输出组合起来得到最终分类决定。另一种早期方法则组合使用了自组织映射（SOM）和卷积神经网络。自组织映射 是一类以无监督方式训练的神经网络，可将输入数据映射到更低维的空间，同时也能保留输入空间的拓扑性质（即在原始空间中相近的输入在输出空间中也相近）。注意，这两种早期方法都不是以端到端的方式训练的，而且提出的神经网络架构也都很浅。

对于基于 CNN 的人脸识别方法，影响准确度的因素主要有三个：训练数据、CNN 架构和损失函数。因为在大多数深度学习应用中，都需要大训练集来防止过拟合。一般而言，为分类任务训练的 CNN 的准确度会随每类的样本数量的增长而提升。这是因为当类内差异更多时，CNN 模型能够学习到更稳健的特征。但是，对于人脸识别，我们感兴趣的是提取出能够泛化到训练集中未曾出现过的主体上的特征。因此，用于人脸识别的数据集还需要包含大量主体，这样模型也能学习到更多类间差异。

用于人脸识别的 CNN 架构从那些在 ImageNet 大规模视觉识别挑战赛（ILSVRC）上表现优异的架构上取得了很多灵感。比如我们耳熟能详的VGG，Googlenet等。

每个模型都有着自己的优势与劣势。LeNet的效果其实并不是很好，而且相对于相同效果的其他模型来说，这个模型对硬件的要求更高。后面几个网络保持的原则基本上就是没有最深，只有更深。模型基本上都是向更深层次发展的。当然Googlenet对于模型广度也进行了非常有意义的探索。这个就对硬件有了更高的要求，对算力有了更高的要求。



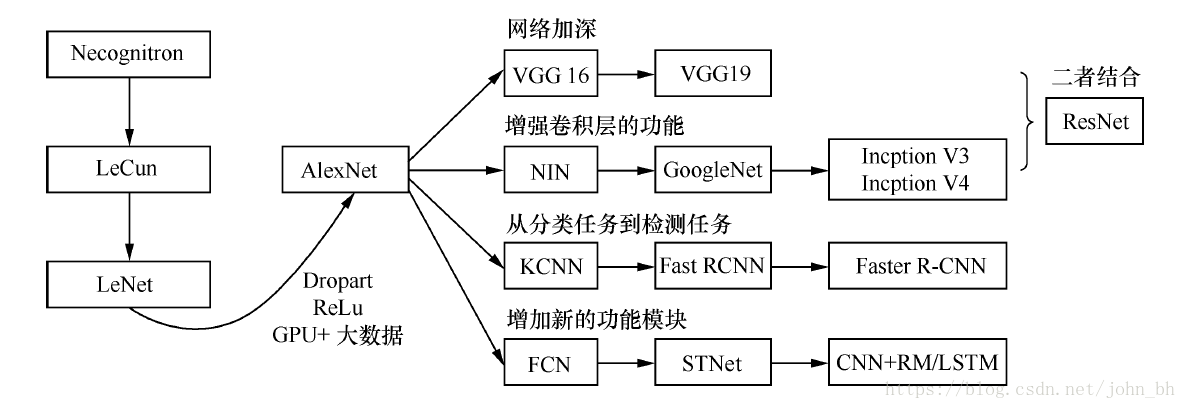
元老级的LeNet图像

Googlenet，14年比赛冠军的model，这个model证明了一件事：用更多的卷积，更深的层次可以得到更好的结构。（当然，它并没有证明浅的层次不能达到这样的效果）

Googlenet图像

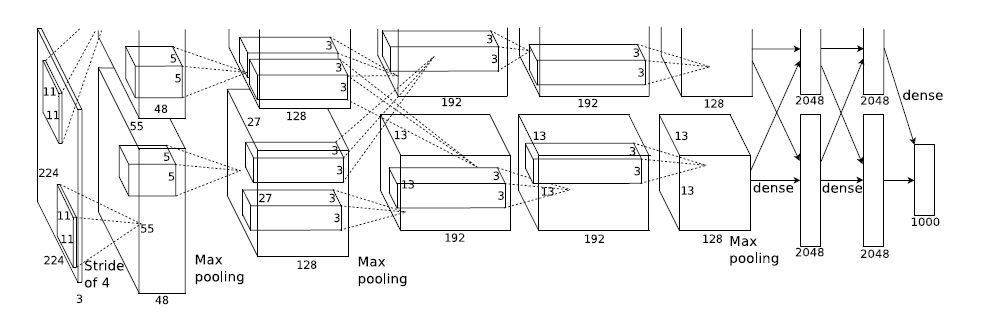
本项目使用的是VGG和Alexnet模型进行训练研究。而VGG更是继承了Alexnet的衣钵，层数更深。这两个模型在最初项目选择的时候就是根据当前数据量以及要求准确率来决定的。

卷积神经网络的发展



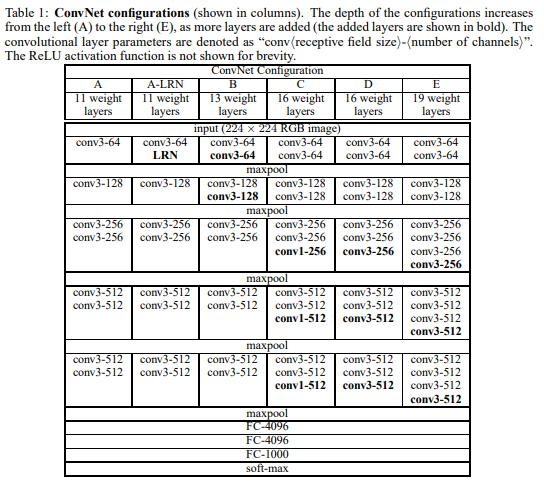
接下来会介绍一些与本项目的相关的网络模型发展。

1. Lenet。以现在的眼光来看，LeNet 绝对是一个小网络，也没什么特点。但是，LeNet 是 CNN 网络结构的开山鼻祖，第一次定义了 CNN 网络结构。其定义了卷积层、采用池化层下采样，用 Tanh 作为非线性激活函数。
2. AlexNet。这个网络在当年的 ILSVRC（ImageNet Large-Scale Visual Recognition Challenge）中以超过第二名10.9个百分点的绝对优势一举夺冠，引起了许多学者对深度学习的研究，可以算是深度学习的热潮的起始标志。相对于传统的CNN，Alexnet有了一些很重要的改动，比如数据增强、Dropout、ReLU激活函数等，对于防止过拟合有很好的作用，同时支持多GPU并行运算，训练速度大大加快。

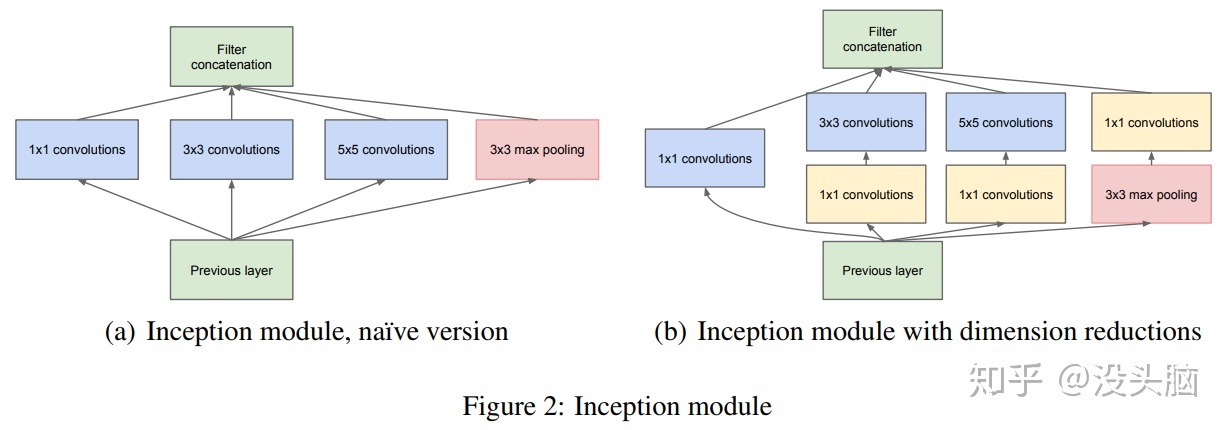


Alexnet模型图像

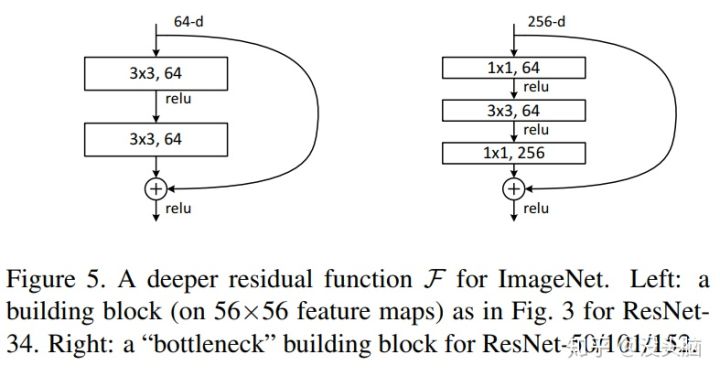
1. VGG。2014 年，Simonyan 和 Zisserman 提出了 VGG 系列模型（包括VGG-11/VGG-13/VGG-16/VGG-19），并在当年的 ImageNet Challenge 上作为分类任务第二名、定位（Localization）任务第一名的基础网络出现。VGG更是继承了Alexnet的衣钵，层数更深。VGG加“深”了前面由五个卷积层与激活函数叠加的部分，使得每部分并不是一个卷积层加一个激活函数组成，而是多个这样的组合组成一部分（有人习惯称这个为 Conv Layer Group），每个部分之间进行池化操作。但是VGG 与当时其他卷积神经网络不同，不采用感受野大的卷积核。

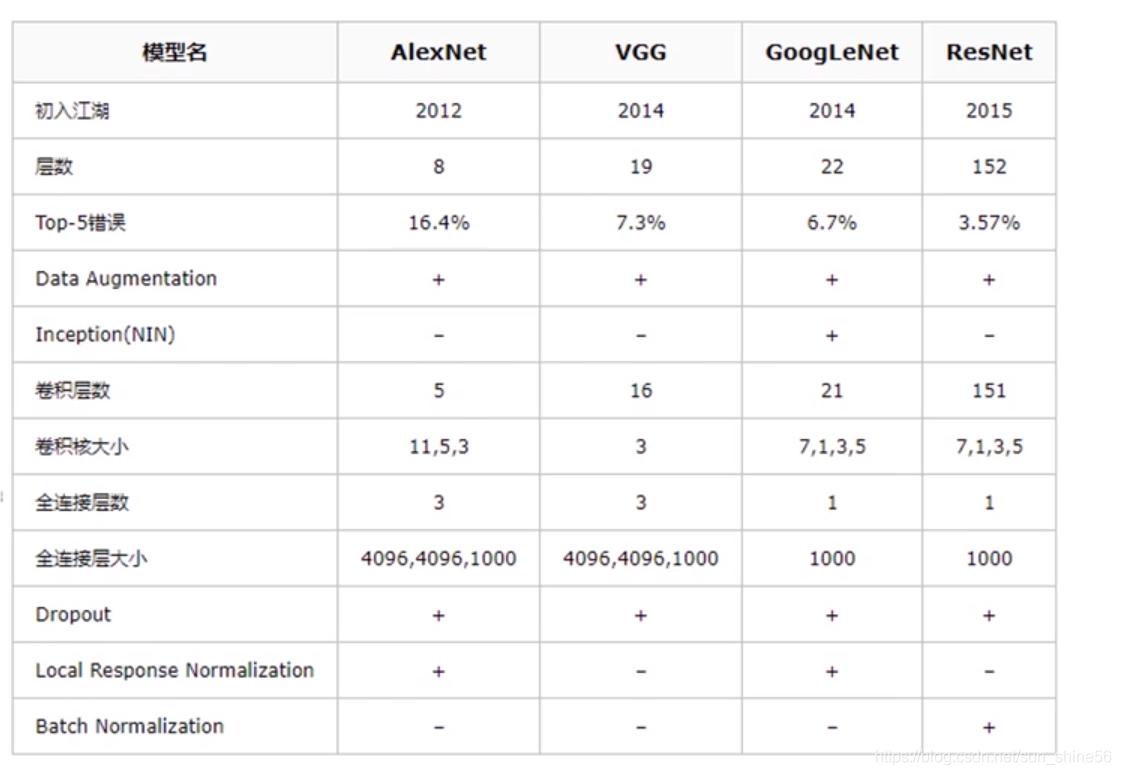


1. Inception Net。2014 年，Google 提出了 Inception 网络结构，并以此构建了 GoogLeNet。与 VGG 不同，Inception 结构虽然也倾向于加“深”网络结构，从而提高模型的表达能力，但是作者充分考虑到现实中计算资源的限制，所以对模型中局部结构进行了精心设计，并且抛弃全连接层。后来Google在此基础上提出了Inception Net V3，Inception Net V4等网络。对此类型的网络进行了有意义的探索。



1. ResNet。2015 年，Kaiming He 提出了 ResNet（拿到了 2016 年 CVPR Best Paper Award），不仅解决了神经网络中的退化问题（Degrade Problem，即相较于浅层神经网络，深层神经网络的深度到达一定深度后，拟合能力反而更差，训练/测试误差更高），还在同年的 ILSVRC 和 COCO 竞赛横扫竞争对手，分别拿下分类、定位、检测、分割任务的第一名。





三、模型方法

**详细说明**本工程所使用的模型方法和理论等

本工程所使用的的模型是VGG19和ResNet。两个模型都在这个项目中表现出来，本来我们只是使用了VGG19,但是为了追求更高的准确率，我们尝试使用了ResNet。

两个模型的说明如下：

四、系统设计

系统的**详细**设计，系统流程，系统的每一步的具体流程，例如，如何处理语料，如何训练模型，如何测试模型，最后对模型进行评估等。

我们使用SGD算法来训练模型。

五．实验结果分析、对比和讨论

对实验结果进行分析说明，测试一些课程中演示的样例，根据结果说明为什么对或者为什么错等具体分析；对不同模型和参数的对比进行分析；最后对系统提出改进方案等。

改进方案就是在之后的

六．对本门课的感想、意见和建议

希望听到你的心声。

1. 华为实验指导书相关实验的复现报告书

（如果选择第一大类的题目）