华东师范大学软件学院

2015 年软件工程学士学位论文

基于深度卷积网络的面部特征点定位技术

Facial Landmark Localization based on Deep Convolutional Neural Networks

姓　　名：贺珂珂

学　　号：10112510119

班　　级：11 级 1 班

指导教师姓名：王晓玲

指导教师职称：教授

2015 年 5 月

目　录

摘 要

为解决多姿态下的面部特征点的定位问题，本文利用深度卷积网络抽取有区分性的层次化特征，获得了面部特征点的级联深度卷积网络模型。

首先，训练了预测 5 个面部特征点的单个卷积网络模型，该模型能初步进行特征点定位，验证了网络的确有进行位置回归的能力。

接着，为进一步提高特征点定位的准确率，本文应用了 3 阶段级联的深度卷积网络实现了由粗到细的特征点定位。第一阶段的网络将人脸的整体图像作为作为输入，直接预测所有的特征点。因为这一阶段的网络利用了全局的纹理信息且保持了所有特征点的全局约束，能够得到可靠的预测结果。后两个阶段的网络截取人脸的各个部件作为输入，对前阶段的网络预测结果进行优化，得到了更准确的定位。

同时，本文也对深度卷积网络的训练进行了探索，通过应用不同的训练技术，使网络模型有更强的学习能力。

之后，进行测试，3 阶段级联的深度卷积网络模型将验证集上平均每个点的误差降低至 1.67%。同时在现实数据上测试结果也表明此模型有高检测准确率且鲁棒。

最后，本文对 5 个面部特征点的深度卷积网络模型进行扩展，获得了高准确率的 68 个面部特征点的深度卷积网络模型。

关键词：深度卷积网络，特征点定位，级联，网络训练

Abstract

To address the challenge of facial landmark localization with complex face pose, this paper leverage the deep convolutional neural network extract the high-level discriminating features and train the cascade deep convolutional network model.

First, this paper train one convolutional network model to predict 5 facial landmarks. This model can locate the landmarks preliminarily which indicate the network can regress the position.

Secondly, to further improve the landmark localization accuracy, this paper carefully apply a three-level convolutional network cascade to achieve facial landmark localization with a course-to fine network cascade. The first level networks take the whole face region as input, directly predict all the landmarks simultaneously. This level networks utilize the global texture context information and keep the geometric constraints, can get the reliable predictions. The networks at the following two levels take the local part as input to refine the first level network predictions and get the more accurate result.

Meanwhile, this paper investigate in convolutional network training and apply a diversity of training tricks to get more learnable networks.

Next, test the model. The three-level cascade deep convolutional neural network model achieve 1.67% per point error rate on validate set. And the experiments on real test data show this three-level cascade deep convolutional neural network model is accurate and robust.

In the end, this paper extend the 5 facial landmarks deep convolutional network model and obtain the high accuracy deep convolutional network model which can predict 68 facial landmarks.

Keywords: Deep Convolutional Network, Landmark Localization, Cascade, Network Training

一、 绪论

(一)研究意义

随着生物认证技术的发展，人脸，作为最为自然和普遍的身份特征，吸引了大量的研究。而其中面部特征点的定位在人脸识别，人脸追踪，人脸属性分析，3D 人脸建模方面都起着关键性的作用[1][2]。图 2-1 为一种特征点标注的示意图，此种格式共需要标注 29 个特征点。在无约束情况下，人脸的有较大的变化，如人脸的不同的朝向，夸张的表情，多变的光照，部分的遮挡。这些情况使特征点定位的问题变得更加复杂，如图 2-2 为无约束情况下的人脸。左上图由于光照半边人脸都处于阴暗中，右上图人脸有明显的右侧，左下图的人处于哭泣中，表情较为夸张，右下图的人的眼睛被卷发遮盖，这些变化多样的情况大大增加了特征点定位的难度。因此鲁棒的特征点定位依旧是一个很大的挑战。

同时，特征点定位本身也是一项很关键的技术。目前有论文[3]通过精确地定位手关节的 14 个特征点，建立了人手姿势模型，有论文[4]通过精确地定位人体的关节信息，判断人的动作。鲁棒的特征点定位有助于计算机视觉模型取得更精确的结果。

图 1-1 一种特征点标注示意图

Figure-1-1 One Annotated Facial Landmarks Diagram

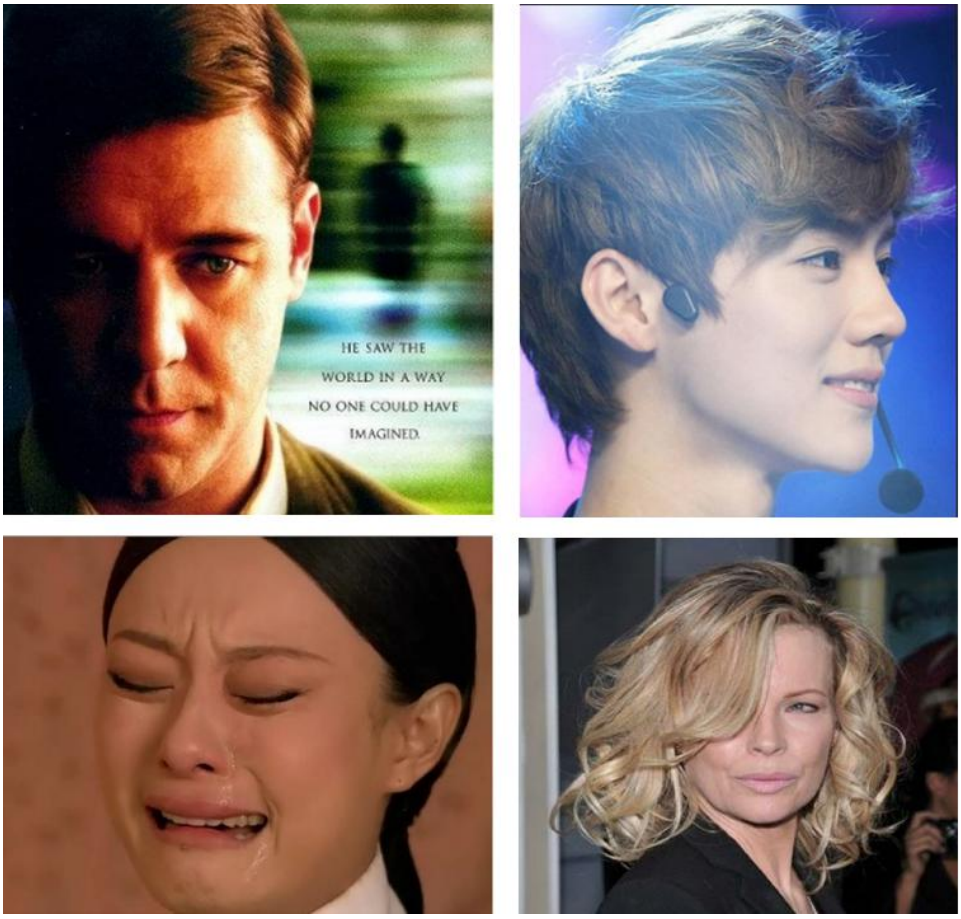

图 1-2 变化多样的人脸

Figure 1-2 Face Under Different Situations

(二)相关工作

在过去几十年，有大量的定位方法被提出，这些方法基本可以被分为 2 类，

第一类方法是基于局部特征的方法，通过对特征点附近的图像块训练部件级的分类器，在测试的时候，通过滑动窗口，如图 1-2 来定位特征点[5]。由于这种方法只利用了局部特征，所以当图像有遮挡或者模糊时，要么会定位到多个特征点，要么定位不到特征点，定位是不准确的。

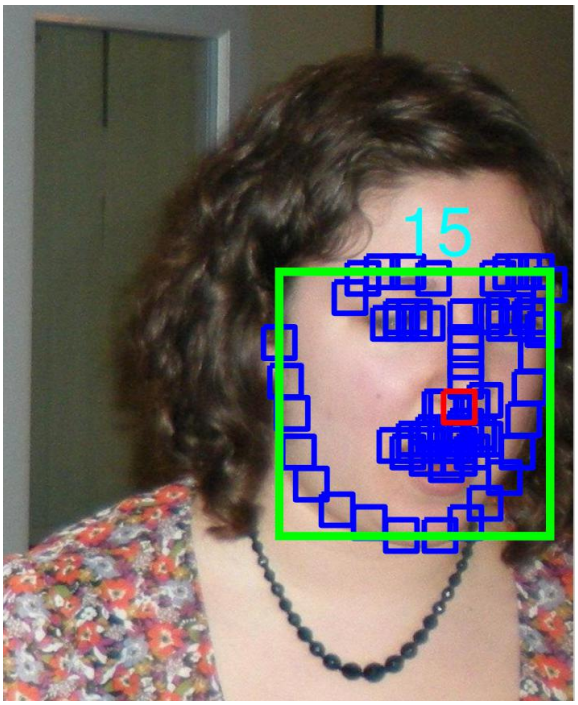

图 1-3 利用部件分类器进行特征点检测

Figure 1-3 Landmark Detection Based on Part Classifiers

第二种方法是基于整体特征的方法，利用图像的全局纹理信息，对特征点的形状进行一定约束。一种经典的定位方法是基于主动形状模型（Active Shape Model , ASM ）[6]和主动外观模型（Active Appearance Model , AAM ）[7]的，依赖于可变的模型，拟合面部特征点。之后也有大量的策略[8][9][10]来改善 ASM 和 AAM 的效果。由于要迭代地估计可变模型的参数，有较高的计算代价。

在近年来，Cao 提出了直接进行回归的定位框架[11]，将特征点定位的问题转变为回归问题，这种方法利用了全局信息，显然更加的可靠，而且不需要迭代，计算代价较低。但因为增加了特征点的候选域，定位的难度也有所增加。目前有利用随机蕨(random fern, 跟随机森林类似)进行特征点回归定位[11]，也有利用卷积网络做级联回归，实现粗到细的面部特征点定位[12][13]。

考虑到定位的准确性和性能，本文选择利用回归直接进行特征点的定位。

在特征的选择上，考虑到手工地选取特征是一件非常费力、需要先验知识的方法，很大程度上需要依靠经验和运气才能选取到好的特征，而且为了便于手工调参数，特征的设计中往往限制少量的参数，特征的表示能力是有限的。而深度学习可以从大量的数据中自动学习到特征的表示，可以包含上万个参数。因此选择用深度网络提取特征。

(三)论文的主要工作

1. 分析现有定位技术，选择用深度卷积网络进行特征点定位

通过对已有技术的分析，本文选择利用深度学习的相关技术提取特征。而深度学习中的卷积网络能直接将原始图像做为网络的输入，能抽取有较强区分性、不同层次级的特征，所以利用深度卷积网络抽取的特征来进行回归。

2. 利用级联的卷积网络实现了更精确的定位

考虑到为实现更高的定位准确率，可以选择更深的网络或更多的网络。更深的网络意味着更多的训练数据和更长的检测时间。而更多的网络，会增加计算时间。但网络进行测试计算时，花费的时间是相当小的，联合不同网络学习到的内容，能够得到更可靠的结果。为实现定位的精确率和效率的平衡，采用级联网络的思想，训练了三阶段级联的卷积网络，能够精确地定位面部特征点。

在第 1 阶段的网络，利用了整体的特征，同时预测了所有的特征点。保留了点与点之间的整体信息，即使人脸有不同的朝向，夸张的表情，多变的光照，部分的遮挡，依旧能够基本保证定位的准确性。后两阶段的网络，根据前阶段的网络的输出，获得局部图像，进行特征点定位修正，获得了更为精确的位置。

3. 深度卷积网络的训练

如果将神经网络看成黑盒，认为只要给定一系列训练数据和监督信息，进行神经网络的训练，就能够直接取得结果,那么网络可能会学习到信息，也可能学不到信息[14]，而且往往无法取得很好的精确率。设计和训练网络会面临大量的选择，如学习率，激活函数，卷积层数等等。

在进行面部特征点训练时，本文应用了不同的训练技巧以取得更好的准确率。在这里同时对网络的部分训练技巧进行归纳。

4. 级联卷积网络的定位效果检测

在测试集上进行测试。观察在人脸有不同的朝向，夸张的表情，多变的光照，部分的遮挡等情况下的定位效果，同时对定位的准确率进行具体的分析，便于更进一步地提升定位效果。

(四)论文的组织结构

本文主要由七个部分组成。

第一部分，介绍面部特征点定位的研究意义，相关工作和论文主要的工作。

第二部分，对深度学习及其中的卷积神经网络的相关知识进行介绍。

第三部分，介绍了级联的深度卷积神经网络的设计思想。

第四部分，描述网络的实现细节和运用的训练技术。

第五部分，进行级联网络模型的准确性和效率的测试，并对结果进行分析。

第六部分，扩展训练好的 5 个特征点的模型，得到 68 个特征点的模型。

第七部分，总结本文的工作和未来的扩展。

二、 深度学习及卷积神经网络的介绍

在这一章节，将会对深度学习进行介绍。内容包括深度学习的优势，特征表示，通用的神经网络模型和特别适合处理图像的卷积网络模型。

(一)人工神经网络的发展

人工神经网络在上世界四十年代兴起，因其类似人脑信号处理的结构。引起了研究的热潮。1986 年 Rumelhart 提出了著名的反向传播算法来训练神经网络[15]。由于数据和计算能力的有限，神经网络在测试集上效果不好。而且，局部最优和梯度扩散会出现在模型的训练过程中，训练难度较大。神经网络的研究有一段时间的停滞。

(二)深度学习的兴起

神经网络的再一次兴起，神经网络研究学者 Hinton 在其中起着至关重要的作用。2006 年，Hinton 提出了深度学习的概念。在网络的训练上，经过逐层的预训练后，在利用反向传播优化网络时，网络的参数已经达到了一个较好的初始值，这种训练方法避免了网络容易陷入局部最优值[16]。

而且，在当前互联网时代，获得大量的数据更加容易，神经网络不容易过拟合，能够在测试集上也取得较好的效果。

同时，计算机硬件的发展使得训练大规模的神经网络成为可能。GPU 计算的出现为网络的训练提供了强大的计算能力。神经网络中一个经典的模型是 LeCun 于 1998 年提出的数字手写字符识别网络[17]，这个网络在当时需要 2 个星期才能训练好。而现在在 GPU 上只须 2 分钟，就能获得同样的结果。

新的网络训练技术，大数据，硬件的发展使得深度网络的训练成为可能。那深度

学习到底有何优势？

深度学习跟传统的浅层学习相比有以下 2 个优势：

1.强调了模型结构的深度。理论研究表明,对于特定的任务，如果模型的深度不够,模型的参数会呈指数增加。

2.明确突出了特征学习的重要性。即通过逐层特征变换，将样本从原空间的特征表示变换到一个新特征空间，从而使分类或回归问题更加容易。

(三)特征表示

一般的学习算法都会提到特征，那什么是特征呢？什么样的特征是好特征呢？在这里，本文将会对特征进行介绍且对深度学习中的特征进行重点分析。

1. 特征表示的概要

在对数据进行分析的时候，一般原始数据都比较复杂，冗余，数量级上又很大，非常不便于计算机直接进行处理。因此需要有特征提取算法，对原始数据进行变换。比如在文本处理方面，我们可能会用单词的分布来近似表示文本，在图像处理方面，我们会用像素的直方图表示图像。我们期望的特征提取算法能使得相近的原始数据，提取特征后的结果是相近的，不同的原始数据，提取特征后是有差异的。特征可以看成是对原始数据的另一种表示。提取包含关键信息，有判别能力的特征是分类、回归等问题中关键的步骤。

2. 特征表示的粒度

那么对于特定学习问题，什么粒度上的特征表示是有效的呢？

比如，我们的学习算法需要区分摩托车和非摩托车，一些底层的视觉特征（如颜色，纹理）等不足以区分，我们可能需要一些高层的、带有语言的信息，如是否有轮子，是否有车把手等特征对图像进行表示，我们的算法才能有效工作。

3. 人脑视觉特征

为了让学习算法从原始的数据学习到高层的特征，先来看看人类的大脑是如何工作的。

1981 年的诺贝尔医学奖，颁发给了 David Hubel 和 Torsten Wiesel，他们的主要贡献是，发现了人的视觉系统的信息处理是分级的。首先从视网膜接受原始信号（像素），经过低级的 V1 区提取边缘和方向，到 V2 区进行抽象，获得基本形状或局部的目标（如椭圆形），再到高层的整个目标（如人脸），之后到更高层的前额叶皮层进行分类判断等。

这个过程说明了人的认知是深度的，从低层特征不断抽象，获得高层语义信息。

4. 层次化特征表示：

我们从人脑的视觉系统学习到人脑的特征表示是不断抽象，有层次化的。图 2-1 显示了层次化的特征。

先从原始的像素中学习到不同的边，这些边就是基(bias)，我们对这些边进行线性组合，得到了不同的部件，这些部件又可以视为高层的基，之后对这些部件进行线性组合，我们得到了不同的对象。

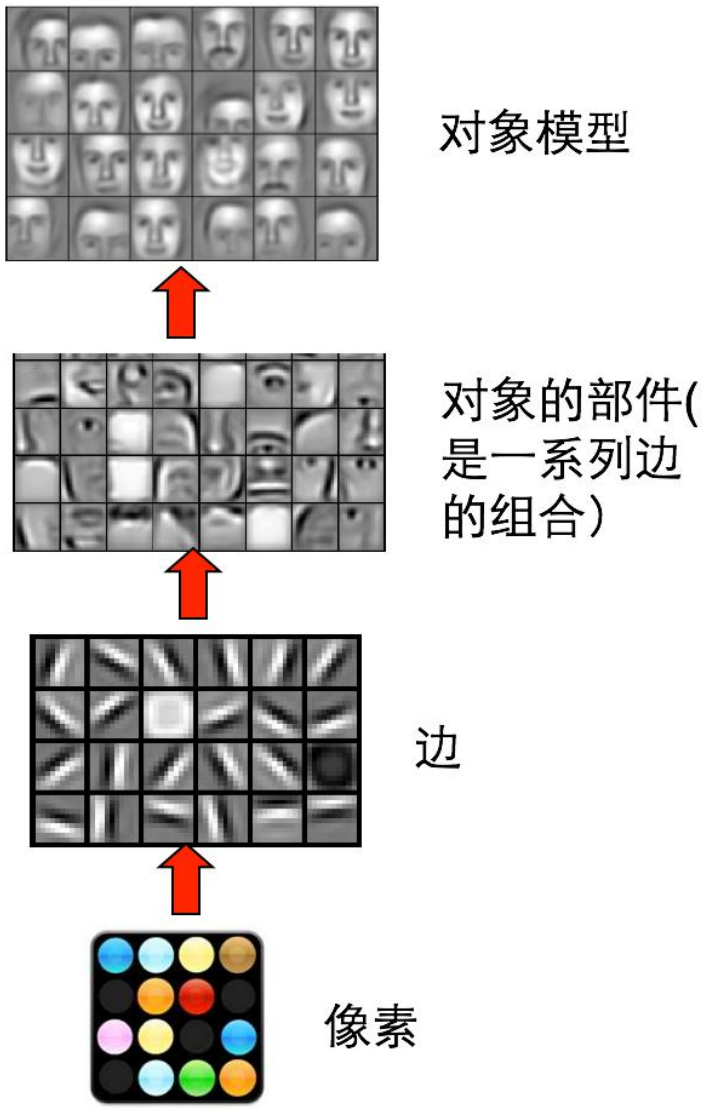

图 2-1 层次化特征

Figure 2-1 Hierarchical Features

5. 深度学习的特征

深度学习的目的是像人脑一样建立从低到高的层次化特征。那么这些特征如何学习得到呢？

在之前的学习算法中，特征提取和分类器是分开的，首先要跟根据应用选择合适的特征提取算法，然后选择某种分类器（如 SVM ）进行类别判断。而现在，深度学习框架将特征和分类器结合到一个框架中，形成一个从原始的像素到标签的端到端的非线性系统，利用数据，自动地进行特征的学习。这个非线性系统有很多层，每层都是简单的计算模块，因此会有很多中间的特征，通过对前一层特征的线性组合学习到当前层的特征，构成了层次化的特征。

(四)神经网络模型

下面介绍通用的神经网络模型。

神经网络可以看做将数据 x 通过一系列的嵌套函数 f 的映射，得到结果 y 的过程。$y = f_n(...(f_2(f_1(x, w_1); w_2)...; w_n)$（$w$ 为函数的参数）可以视为深度为 n 的神经网络。

图 2-2，是一个 3 层的神经网络，一般将图中的节点称为神经元。Layer L_1 是输入层，Layer L_2 为输出层。称 Layer L_2 为隐层。Layer L_1 经过函数 f_1 的映射得到 Layer L_2，Layer L_2 经过函数 f_2 的映射得到 Layer L_3。图中 +1 的结点表示偏置，线表示神经元之间的连接，每个神经元都会跟前一层的神经元有连接关系。每层的网络结果都会经过一个非线性激活函数，最后，需要给网络模型设定损失函数，梯度下降算法可以优化损失函数的参数，神经网络模型利用反向传导算法进行参数的优化。

下面将会对非线性激活函数，损失函数，梯度下降算法，反向传导算法进行具体的展开。

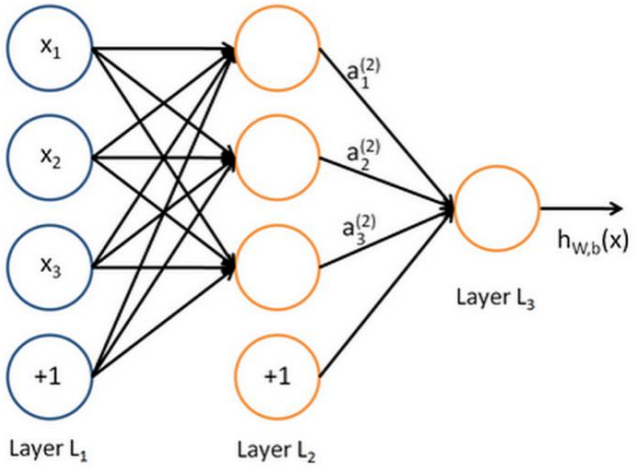

图 2-2 单隐层的神经网络

Figure 2-2 One Hidden Layer Neural Network

1. 非线性激活函数

因为线性模型的表达能力不够，利用非线性的激活函数来加入非线性因素。网络的每一层都会有激活函数。如图 2-3-2 为一个线性不可分模型，一条直线不能把点和叉分开。

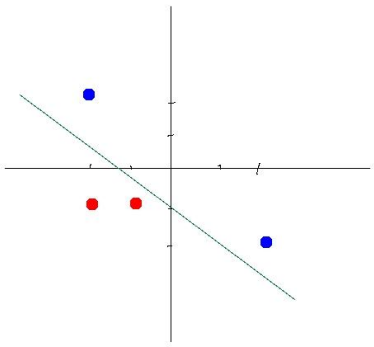

图 2-3-1 线性可分模型

Figure 2-3-1 Linear Separable Model

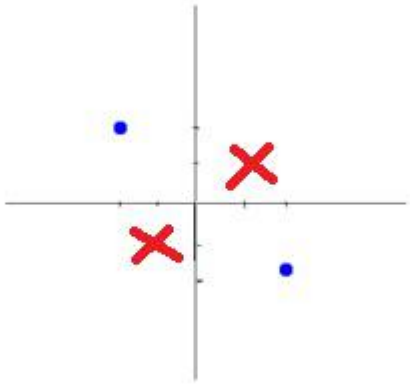

图 2-3-2 线性不可分模型

Figure 2-3-2 Linear Inseparable Model

2. 损失函数

损失函数可以用来衡量模型跟训练数据的拟合程度，按照模型具体类型应用特定的损失函数 l。有监督的神经网络需要给出标注的真实值（ground truth），记为 y_{target}。

如对于回归问题，可以选择平方损失函数$l=1/2*(y-y_{target})$^2。

对于分类问题，可以选择逻辑损失（logistic loss）函数。

$l(y,y_{target})=\mathrm{y}\ln(1+e^{-y_{target}})+(1-y)\ln(1+e^{y_{target}})$

求得　使损失$l(y;y_{target})$最小的模型参数 $(w_1;...;w_n)$。

3. 梯度下降算法

我们的目标是求得一系列的参数，这些参数能使损失函数最小。因此神经网络的学习的问题基本上可以转化为优化问题。一般的优化问题算法都需要计算梯度，也就是对参数的一阶偏导。将参数往梯度的反方向变化，能使损失函数逐步减小。

将损失函数比作山，我们的目的是从山顶走向山谷，梯度下降算法即我们站在山上某一点，向四周观察，选择梯度最大的方向下降，这是一种能够最快下降的方法，学习率（learning rate）即人脚跨的步长。

算法: 梯度下降算法

输入: 损失函数$l(\theta)$,步长 n，（θ可以是一个向量）

输出: 优化后的θ

流程:

1:随机初始化θ

2:检测是否满足终止条件（终止条件可以是 2 次的参数变化小于某个阈值，或者迭代到了一定的次数）

3:　若满足，进入 6

4:　若不满足，计算 L 相对于θ的一阶导数，　$\mathrm{gr}adient=\frac{\partial}{\partial\theta}\mathrm{J}(\theta)$

5:　更新　$\theta=\theta-\mathrm{n}*gradient$，进入 2

6: return　θ

图 2-4 是某参数为$\theta_0\theta_1$的函数J的损失平面，纵轴代表函数J的值。从一个初始点开始，不断计算梯度，向函数的最小值靠近的过程。梯度下降算法容易受初始值影响，不同的初始值，可能会下降到不同的极小点。

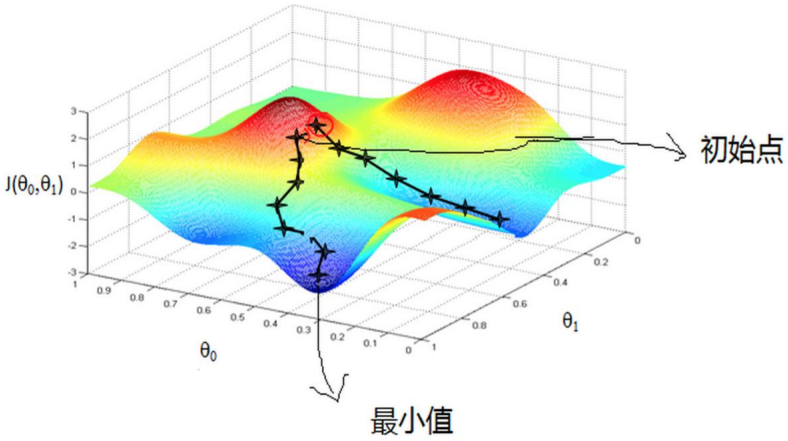

图 2-4 梯度下降示意图

Figure2-4 Gradient Descent

4. 反向传导算法

多层神经网络的后一层依赖于前一层计算得到，每层都有大量的参数，因此多层神经网络无法得到显式的梯度表达式。反向算法解决了这个困难，它是一种近似的最优解决方案。它将多层转变为一层接一层的优化，就可以利用梯度下降算法优化每一层的参数了。但我们需要反向进行优化——由输出层开始优化前一层的参数，然后再优化前一层。这样反向一轮下来，所有的参数都优化过一次了[19]。

反向传播神经网络的学习过程可以分为两步：数据的前向计算和误差的反向传播。通过数据的前向计算，将输出层结果与训练样本标签数据对比，计算误差以衡量现有网络对数据的拟合能力。通过误差的反向传播，利用梯度下降算法对网络权值进行调整，使得网络的输出结果与样本相匹配。在进行测试的时候。利用已训练好的网络模型，直接进行前向计算就能够获得结果。

(五)卷积神经网络模型

对于全连接的神经网络而言，需要优化大量的参数。如一张 500*500 的图片，输入就有 250000 维，如果经过第一个隐层降到 10000 维度，那么第一层共需 500*500*10000=2.5*10^9 个参数，大量的参数往往导致网络的过拟合，即在训练数据上效果好，在测试数据上效果会很差。深度卷积神经网络是一种特殊的深层的神经网络，它的特殊性体现在四个方面，通过这四个方面来降低网络中的参数。

1. 层与层间的神经元是局部连接的

在生物神经网络，认为视觉皮层的神经元只接受部分信息，通过局部的信息组合得到全局的视觉信息。因此，每个神经元也不必要对全局的图像进行感知。即后层的神经元不必跟前层的所有神经元关联，只关联前层部分几个神经元。

如每个神经元只跟 10*10 的像素相连，这个参数就相当于卷积，这样参数变为 10000*10*10 = 10^6。这种网络结构使得映射函数的参数大大减少，降低了网络模型的复杂度。

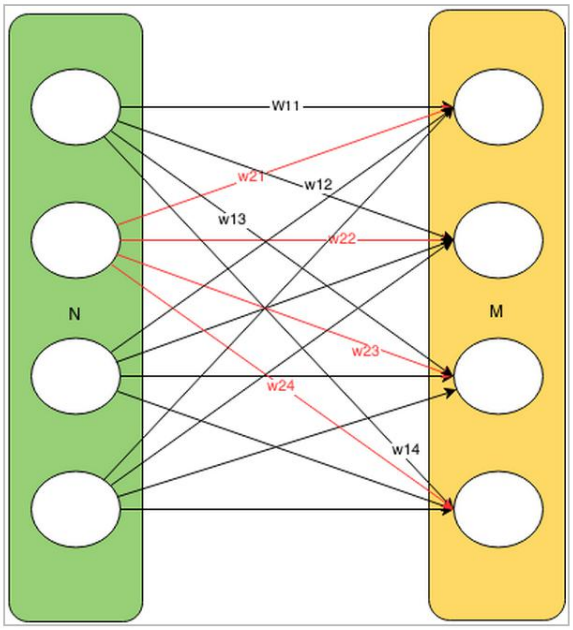

图 2-5-1 全连接的网络

Figure 2-5-1 Fully Connected Network

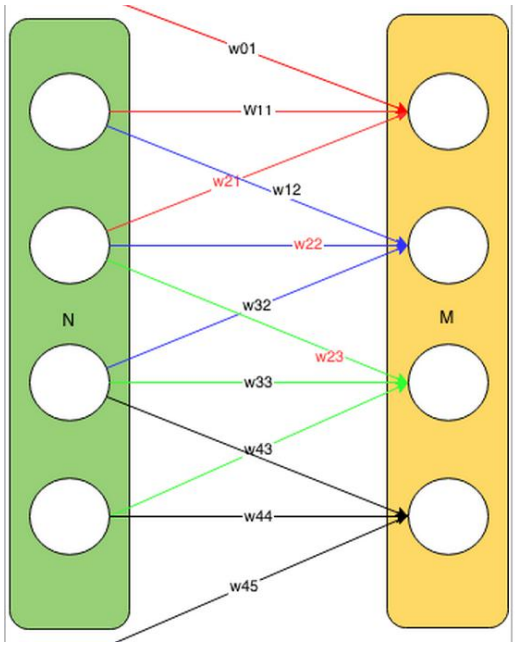

图 2-5-2 局部连接的网络

Figure 2-5-2 Locally Connected Network

2. 参数共享

在局部连接思想中，每个神经元都跟 10*10 个像素相连，每个神经元都有 100 个参数。这里大胆地使每个神经元的 100 个参数为相同的。因此网络从 10^6 个参数将到 100 个参数。

我们可以将 100 个参数看成卷积，卷积相当于一种特征检测器，可以检测是否有边，纹理，颜色等特征。利用同一个卷积在图像的各个区域进行特征检测的原理是:图像各个部分的统计特征是一样的，因此我们能够在图像的各个位置使用相同的特征检测器。

如图 2-6，展示了一个 3×3 的卷积核在 5×5 的图像上做卷积后得到的值，我们将得到的值称为一个特征图。3×3 的卷积核在图像上从左到右，从上到下滑动，与每个 3×3 的图像进行卷积运算,即进行点积运算，再相加。如最后一个值通过 1×1＋1×0＋1×1＋1×0＋ 1×1 ＋ 0×0＋ 1×1 ＋ 0×0 ＋ 0×1 计算得 4。

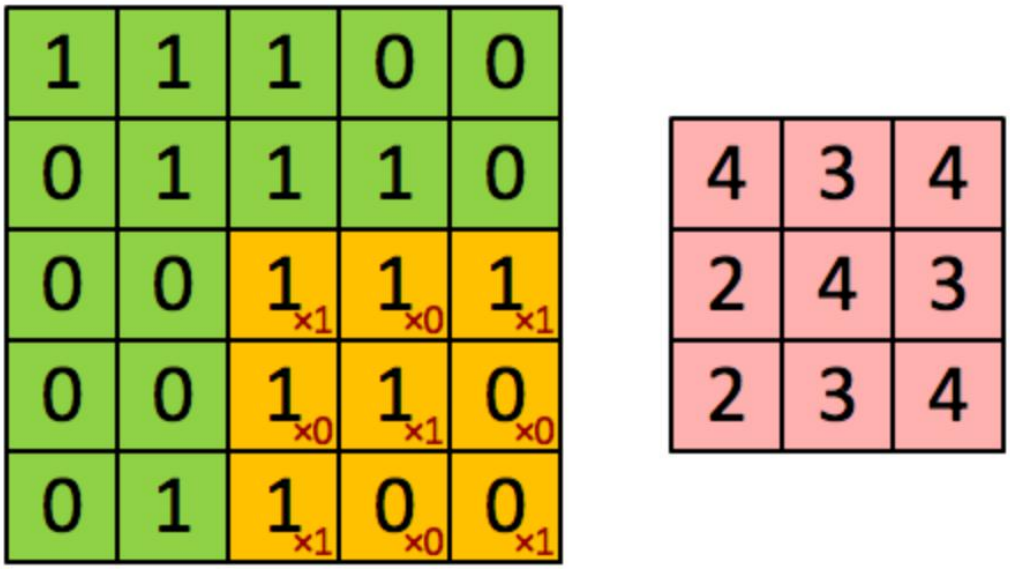

图 2-6 卷积的计算

Figure 2-6 Convolution

3. 多卷积核

在上面的例子中，我们假定只有一个卷积核，因此共有 10*10 = 100 个参数，但只有一个特征检测器，对图像的特征提取能力是比较小的，我们可以增加卷积核的个数，如 20 个卷积核，那么我们就能提取到 20 种不同的特征，使网络有更强的学习能力。图 2-7 显示了 4 个不同的卷积核提取得到 4 个特征图的过程。

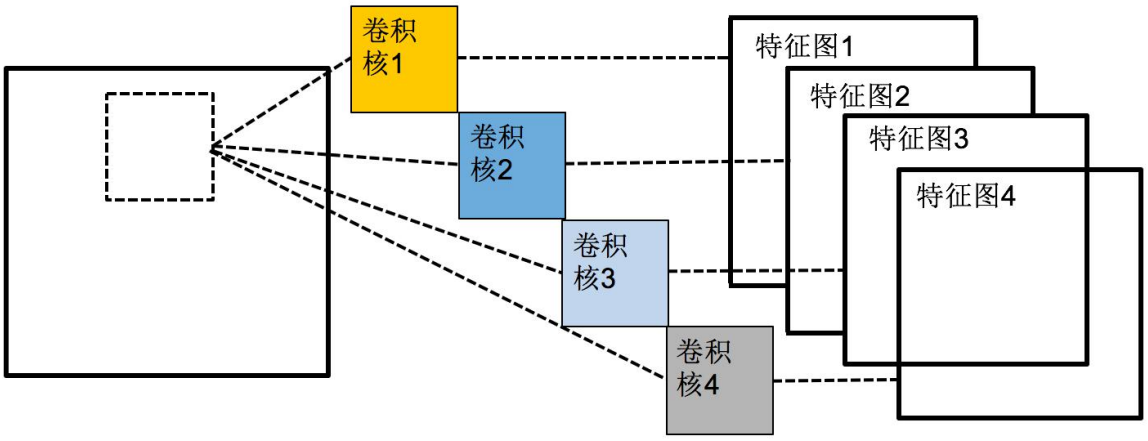

图 2-7 多卷积核的计算

Figure 2-7 Multi-Convolution Kernel

4. 降采样层

一个输入为 500*500 的图片，若经过 5*5 的卷积后，得到的特征图的大小为(500-5+1)*(500-5+1) = 496*496，神经元的个数还是较多。因为图像中相邻像素是类似的，我们考虑聚合图像来降低神经元的个数。具体就是不单独考虑一个像素点，而是将相邻的部分像素点进行计算，得到合并后的结果。这种方法降低了特征图的分辨率，得到的结果有较好的平移不变性。根据区域的点合并的原则有最大降采样(max-pooling)和平均降采样(mean-pooling) 2 种。

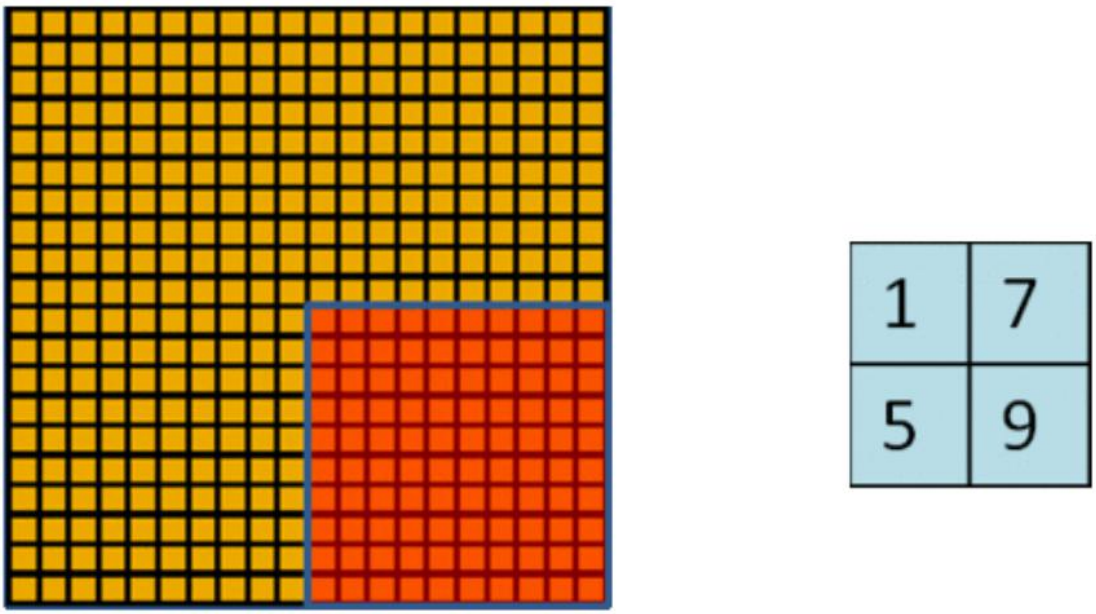

图 2-8 降采样的计算

Figure 2-8 Pooling

三、 网络的设计思想介绍

在这一章节，主要进行网络模型结构的试验，首先设计单隐层神经网络和 2 层卷积网络进行定位，之后应用级联的思想，设计级联的网络结构，以进一步提高定位准确率。

(一)利用神经网络定位的试验

在本文中，目的检测 5 个面部特征点，这些点分别是，左眼中心点，右眼中心点，鼻尖点，左嘴角点，右嘴角点。为了确定定位问题的复杂程度，本文运用不同结构的网络模型，对网络的结构进行探索。

1. 一个单隐层的神经网络。

图 3-1 是一个简单的单隐层神经网络，输入是人脸的原始像素值，即 39 * 39 = 1521 个像素值。中间的隐藏层大小为 100，需要一个 100*1521 大小的权重，来实现映射，再加上大小为 100 的偏置，经过非线性的激活函数，得到隐藏层。同样，经过 10 * 100 权重的映射，得到输出的 10 维数据。我们应用 Alex 在[18] 中提出的 ReLU 作为每层的激活函数。网络的参数共有：100*1521 + 100 + 10 * 100 + 10 = 153210 个。

观察训练误差和验证误差，当训练误差一直减少，而验证误差基本不再减少时，认为网络已经接近于过拟合。当网络过拟合，在验证集上的测试效果会较差，选择停止迭代。在 2.5hz CPU 上，迭代 400 个 epoch。epoch 代表训练的次数,400 个 epoch 表示所有训练数据会被学习 400 次，共花费 160 秒.

最后，在验证集上进行测试的效果见图 3-2：可以发现，网络输出的坐标尽管距离真实的值有一定偏移，但基本上还是准确的。可见，简单的反向传播神经网络的确有直接将输入图像回归出坐标的能力。

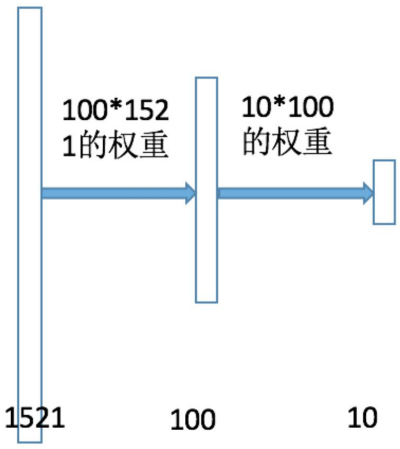

图 3-1 单隐层神经网络结构图 图 3-2 单隐层神经网络检测结果

Figure 3-1 One Hidden Layer Neural Network Diagram Figure 3-2 One Hidden Layer Detection Result

2. 尝试更复杂的卷积网络模型

图 3-3 是一个 2 层卷积，1 层全连通，1 层输出的网络。网络的输入为 39 乘以 39 大小的人脸图像的像素值。卷积块的大小为 4 乘以 4，从输入中扫过，每应用一个卷积块可以得到 36 乘以 36 的特征图，设定第一个卷积层要学习 20 个卷积块的权值，因此能得到 20 个 36 乘以 36 的特征图。之后进行 2 乘以 2 的最大降采样，降采样层不改变特征图的个数，得到 20 个 18 乘以 18 的特征图。接着进行第二个卷积层，这里设定卷积块的权值大小为 3 乘以 3，得到 16 乘以 16 的特征块，这一层学习 40 个卷积块。之后又跟着最大降采样层。接下来是 2 个全连接层，分别为 500 和 10，最后的 10 即为网络的坐标输出。此卷积网络需要的估计的参数为 1293090 个。

在这个网络中，每层卷积层后都会跟降采样层，因为本身利用网络做跟位置相关的定位问题，我们很容易想到，如果进行降采样，会不会损失了定位的精度。在论文[13]中，作者说明了网络做位置定位问题时加降采样层的合理性。因为对整个网络而言，降采样层所带来的平移不变性能够补偿这种损失，而且特征点的形状和相对位置比像素级信息更重要的。所以在本文的网络结构中，依旧保留了降采样层。相较于简单的单隐层神经网络，卷积网络的计算复杂度更大，训练难度更大。在 CPU 上训练 500 轮花费了 8.3 个小时。

图 3-4 为利用 2 层卷积网络进行定位的结果。对比单隐层的神经网络，发现利用卷积网络，定位更加准确。但依旧有部分的偏移。可见对于二维的图像，卷积能够更好地进行特征抽取。

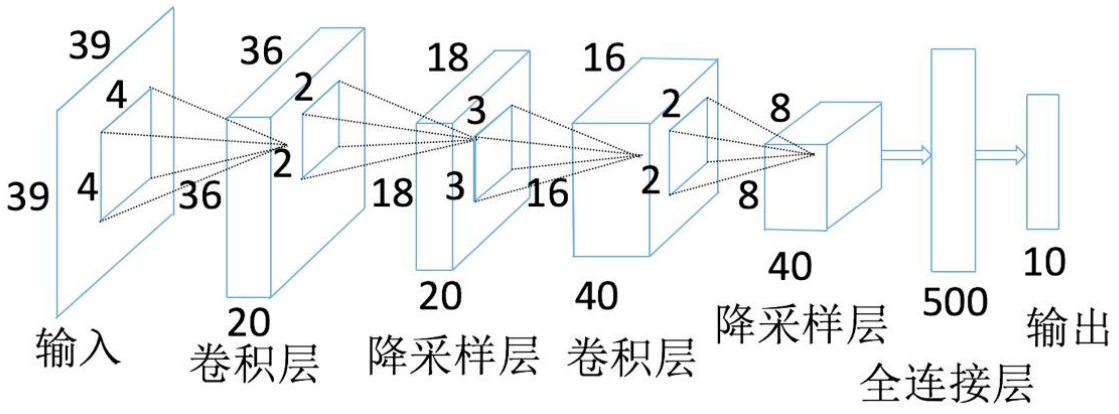

图 3-3 卷积网络示意图

Figure 3-3 Convolutional Network Diagram

图 3-4 卷积网络检测结果

Figure 3-4 Convolutional Network Detection Result

(二)级联的思想介绍

经过上面实验，学习到神经网络除了做图像分类任务外，的确能够处理回归问题。同时，对于图像数据，卷积网络相比普通的全连接的神经网络有更好的特征抽取能力，能够降低训练误差和验证误差。接下来，考虑如何进一步提升定位的准确率。

若使用单一的卷积网络，为提高准确率，需要更深的网络，更深的网络意味着更多的训练数据和更长的检测时间。经典的人脸检测算法 Viola–Jones[20]，采用基于 Haar 特征的级联分类器策略，可有快速且有效的找到多种姿态和尺寸的人脸图像。闫鹏采用级联的卷积网络进行车牌检测[21]，级联的检测方法能够应对自然场景下的车牌检测对检测准确率和检测效率的挑战。因此，本文考虑利用级联回归策略进行特征点定位，在增加一小部分网络计算代价下，提升定位的准确率。

本文先利用前面的网络进行粗粒度的特征点定位，前面的网络采用整体的图像信息进行训练，能够利用整体的纹理信息和保留特征点之间的全局位置约束，所以网络得到的位置是基本准确，可以信赖的。后面的网络对前面的网络的定位结果进行修正，

得到细粒度的特征点定位。级联多个卷积神经网络最后得到精确的位置，是一个有效的方法。

初始的网络的先验信息较少，只能凭借人脸框来限制 5 个特征点的位置。后续的网络预知到特征点就在第一阶段的预测点的附近，预测的范围更小，因此能得到更精确的预测位置。级联网络的思想能够平衡特征点定位的准确率和效率。后面的实验也将显示利用级联有效地提高了准确率。

(三)级联的网络结构

图 3-5 显示了级联卷积网络用于面部特征点定位的过程。先利用人脸检测工具得到图片中的人脸框，在这里截取 3 个部分，整个人脸（包含 5 个特征点，上半部分人脸（包含眼睛和鼻尖 3 个特征点），下半部分人脸（包含鼻尖点和嘴角点 3 个特征点），分别将这 3 个图片输入到 3 个不同的网络中去，将得到后的点取平均，获得第一阶段的级联网络的预测效果，经过第一阶段的定位，取得了粗粒度的特征点位置。在第一阶段，眼睛的定位效果相对鼻尖点和嘴角点而言较为准确，预计眼睛周围的纹理信息，如眼眶之类的较为丰富。

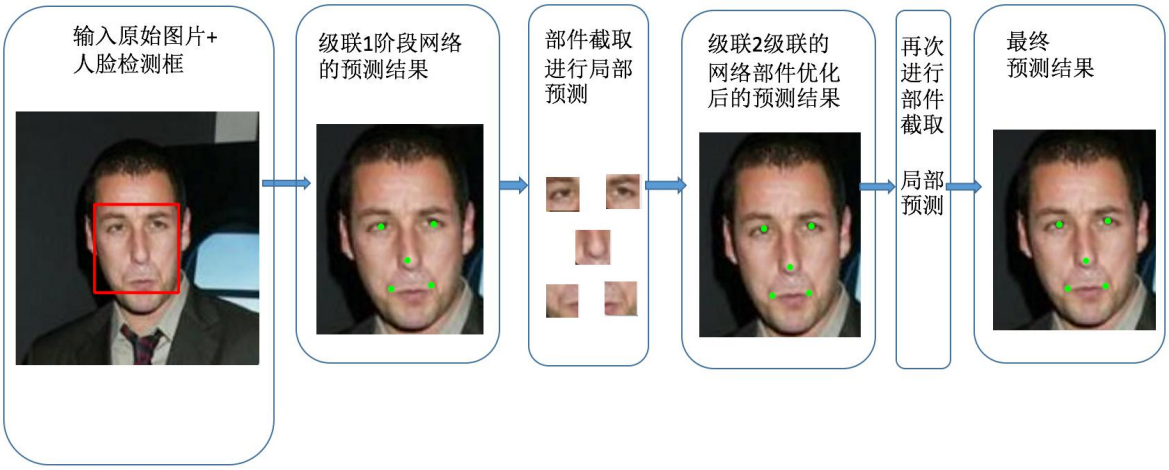

图 3-5 3 阶段级联卷积网络进行面部特征点检测的过程

Figure 3-5 Process of Three–level Cascade Convolutional Network for Facial Landmark Localization

之后进行更细粒度的特征点定位。以第一阶段预测的人脸特征点为中心，按照一定的比例截取人脸部件，经过这一部件的截取，显著降低了特征点的可行域，如人眼特征点的范围已经被限制在截取出来的小图片上。在级联的第二阶段，分别将部件块输入到不同的部件网络中，得到预测的位置，预测位置经过一定的计算可以恢复到最原始的位置。如果第一阶段的网络预测点稍微偏离目标点，第二阶段的网络能够对第

一阶段的网络进行部分修正，获得了更细粒度的特征点定位。

最后，进行第三阶段的级联定位，第三阶段的网络跟第二阶段的网络类似，取前一阶段的网络，即第二层网络的预测点为中心，区别在于截取的比例小于第二阶段。因为定位是越来越精细的，特征点的可能范围也越来越小。同时，因为截取的人脸部件信息是局部的，局部的信息可能会有遮挡，模糊等，因此局部的信息是不完全可靠的。所以要限制特征点位置的变化在一个小的范围内，防止损害定位的精度。最后，得到了最终的定位。通过由粗到细的级联卷积网络定位，在增加了一小部分部件级的修正网络后，获得了显著的定位精度。

四、 网络的实现细节

在这一章节，将详细描述实现的细节。首先探索了如何训练卷积网络，能在验证集中的误差较小，使网络有更强的学习能力。然后，实现了级联的深度卷积网络，进一步提升了检测效果。

训练的数据：

用论文[12]中提供的 10000 张图片，作为训练集，3466 张图片作为验证集。训练集与验证集中的照片没有重叠。

数据的预处理：

先用人脸检测器检测到人脸框，之后截取人脸的图像，然后将所有的人脸图像缩放到指定的大小，并且对人脸图像进行灰度化。在这里，定义人脸的大小为 39 * 39。根据经验，将人脸的像素值都除以 255，使输入的像素值在(0,1)范围。将标定的特征点的坐标值减去 20，再除以 20，使作为监督信息的特征点坐标在(-1,1)。使输入和输出在相同的数量级，能够提高网络的学习能力。

(一)网络训练的探索

卷积神经网络利用反向传播算法来优化网络中的参数，一般网络可以学习到特征，使误差到一个较小点，但往往无法取得很好的准确率。设计和训练网络会面临大量的选择，如学习率，激活函数，卷积层数等等。

在进行面部特征点训练时，为了使网络更快地收敛和取得更好的检测准确率，使用了不同的网络训练技巧，在这里对部分网络的训练技巧进行总结。因为级联第一阶段的网络需要同时预测 5 个点的位置，后面阶段的网络依靠第一阶段的网络进行调整，所以第一阶段的网络定位的准确性对整个级联系统定位的准确性起着至关重要的

作用。同时，因为此网络需要预测的位置多，所以训练也是最为困难的。在这里，将第一阶段同时预测面部 5 个特征点的网络作为例子，进行网络训练。

在之前为了验证网络是否有回归位置的能力，本文分别训练过 2 个网络，记第一个网络为 net1,它是单隐层的全连接神经网络，将得到的模型在 3466 张图片的验证集上进行测试，平均每个点的误差为：0.0481。

其中每个点的误差计算公式为：

$$err = \sqrt{(x - x')^2 + (y - y')^2} \ / \ l$$

其中 (x, y) 为标注值，(x', y') 为网络监测到的位置。

同样的，将 2 层卷积网络的模型进行测试，记为 net2, 得到平均每个点的误差为：0.0313。在下面，进行了各种实验，降低验证集上每个点的误差。

1. dropout

网络的输入和输出间有隐层神经元, 通过调整输入和隐层神经元，隐层神经元与隐层神经元之间的权重来学习特征检测器。当隐层神经元过多，训练数据又很少时，很容易过拟合。2012 年 Hinton 提出了 dropout 思想[22]，dropout 是一种的正则项技术，防止网络的过拟合，使网络模型在非训练数据上有更好的泛化效果。

dropout 的思想是在训练的时候，随机地漏掉网络中一些隐层神经元（包括他们的连接），即随机地将网络的部分权重置为 0。漏掉的比例需要设置，如设置为 0.5，即每个隐层神经元都会有 0.5 的概率被漏掉，所以每个隐层神经元不能依赖其他的隐层神经元（防止了训练数据的联合变化）。隐层神经元会随机漏掉，所以网络的结构也在随机变换。如图 4-1-2 为一个 dropout 后的网络，带叉的神经元被扔掉了。

所以可以通过结合大量不同的网络的预测结果来减少测试误差。训练不同的网络会增加计算开销，但漏掉了部分神经元又节约了时间，所以综合起来 dropout 是计算合理的。在 net2 的网络的基础上增加了 dropout 后，验证集上平均每个点的误差为 0.0302。

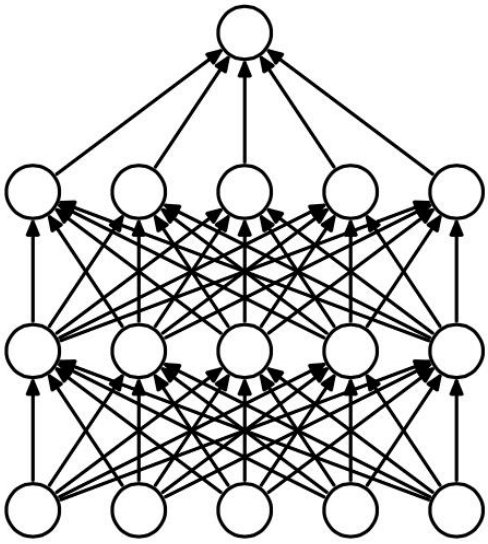

图 4-1-1 标准的神经网络

Figure4-1-1Standard Neural Net

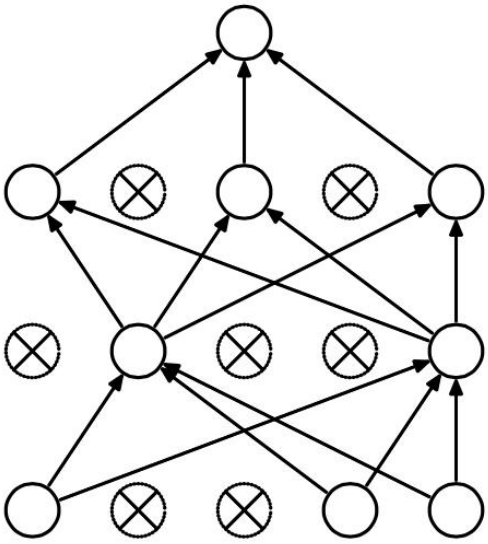

图 4-1-2 应用 dropout 后的神经网络

Figure4-1-2 After Applying Dropout.

看看加 dropout 前后的误差变化。左图是 net2，未加 dropout 的，训练误差在不断下降，验证误差一开始也处于不断下降中，但之后变得平稳开始有上升的趋势。右图是 net3,加了 dropout 的，右图中训练误差和验证误差基本处于一致下降的状态，且误差变化更加平滑，可见 dropout 的确有防止网络的过拟合的作用。

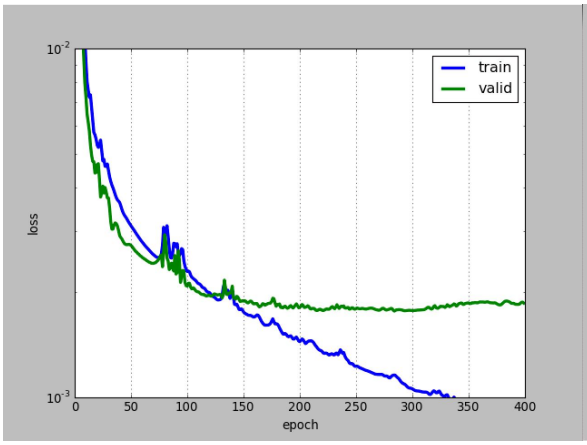

图 4-2-1 Net2 的误差变化

Figure4-2-1 Loss Curve of Net2

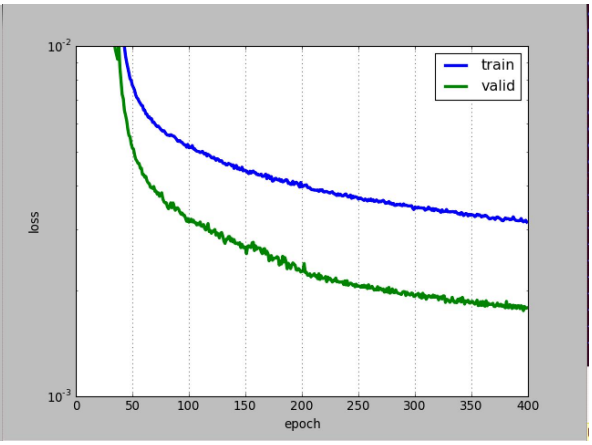

图 4-2-2 Net3 的误差变化

Figure4-2-2 Loss Curve of Net3

2. 动态调整网络的学习率和动量因子

在利用随机梯度下降算法进行优化时，学习率表示参数更新的比率。

动量因子表示后一时刻的梯度的更新依赖于前一阶段的梯度更新的程度。

此参数可以保证权重一直向同一个方向进行变化，有助于收敛。随机梯度下降算法可以理解为人在山顶走向山谷的过程中，随机向四周观察，选择梯度最大的方向下降，这是一种快速的方法，学习率即人脚跨的步长。

在利用随机梯度下降方法优化网络的参数的过程中，一开始参数是随机化初始化的，所以离最优化点是很远的，这时候，可以选择较大的学习率，然后随着不断的训练，将会离最优化点越来越近，这时候要选择较小的学习率。例如，人一开始一开始离山谷很远，可以用较大的步长下山。越靠近山谷，越需要小心，用较小的步长靠近山谷点。而动量因子使人当前步的下山的方向会依赖于前一步下山的方向，这样可以防止我们下山方向有较大的偏移，有助于更好地到达山谷点。因此，本文设计了动态变化的学习率和动量因子，对于学习率，一开始初始值为 0.03, 最小的学习率为 0.001，网络的动量因子的初始值为 0.9,最大值为 0.999。

学习率随着迭代的次数，逐步减少，动量因子逐步增加，较快地使网络收敛。但发现测试误差为 0.0307，加了这个后，误差反而上升了，又结合之前看到的损失曲线，猜想网络可能没有学习完全，400 个 epoch 太少了。

3. 更多的 epoch

经过前面阶段的网络训练，发现在有 dropout 的情况下，网络是不容易过度拟合的，猜想如果增加网络训练的 epoch 的数量,这个模型可能会变得更好。于是将 epoch 数量增加至 10000，进行训练来验证我们的猜想。果然，经过 10000 个 epoch 后，网络的训练误差和验证误差都降低到了一个新的水平，之后，将点恢复到原始的空间，发现每个点的误差，都有降低了，最后平均误差降低至 0.0297。

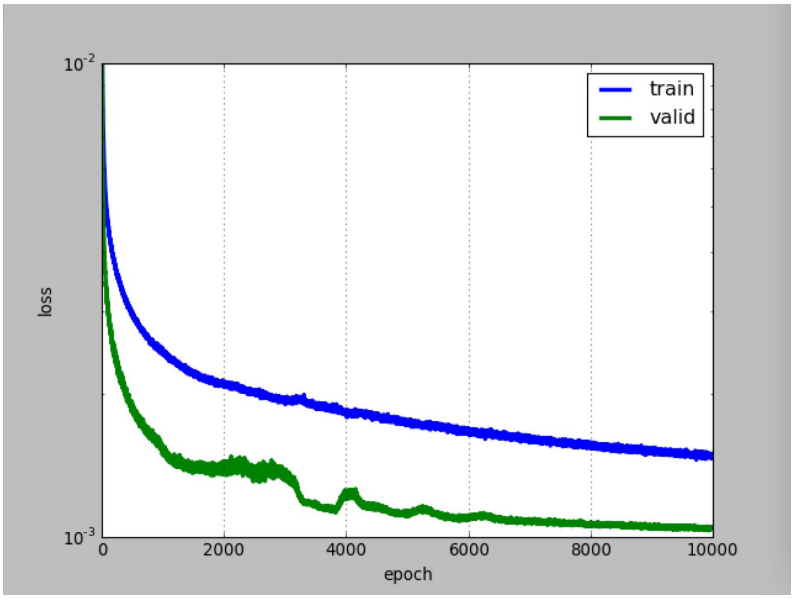

图 4-3 10000 个 epoch 的误差曲线

Figure4-3 Loss Curve of 10000 Epoch Network

4. 数据增强

对于深度卷积网络的训练来说，数据越多，网络的学习素材就越多，就能学习到更好的特征。所以，考虑在已有的 10000 张数据的基础上增加训练数据。对训练图像进行较小范围的上下左右的平移和小角度的旋转，使网络有更强的学习能力，能够更好的应对复杂多变的姿态下的面部特征点定位挑战。

最后对级联第 1 层的网络的训练过程进行总结，训练时间是在 GPU 下的。

表 4-1 不同网络的训练结果

Table 4-1 The Result of Different Net

网络名	描述	Epoch	训练时间	平均误差
Net1	单隐层神经网络	400	80s	0.0481
Net2	卷积网络	400	25.3min	0.0313
Net3	Net2 + dropout	400	26.2min	0.0302
Net4	Net3 + 动态调整网络的学习率	400	26.2min	0.0307
Net5	Net4 + 更多的 epoch	10000	10.8h	0.0297
Net6	Net5 + 数据增强	10000	20.5h	0.0263

(二)级联网络的训练过程

图 4-4 为级联网络各个阶段的示意图，此图参考于 Sun 在[12]中的网络结构。第一阶段的网络利用整体信息获得 5 个点的位置。后两阶段的网络利用局部信息对每个点的位置进行修正。

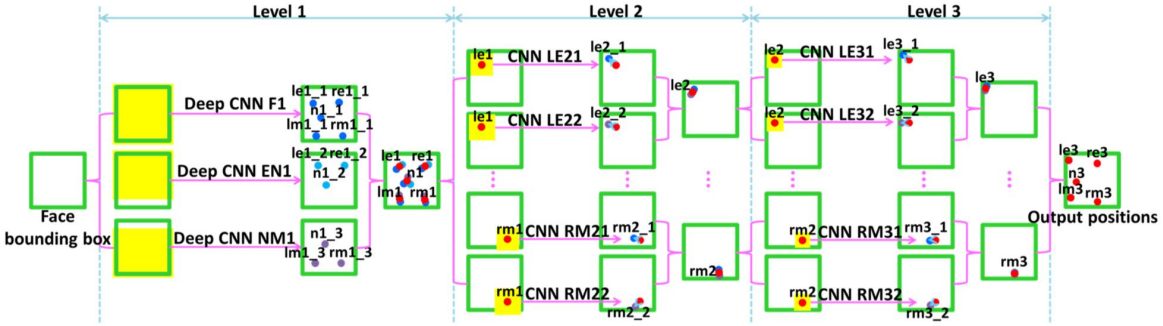

图 4-4 级联网络的结构

Figure 4-4 The Structure of Cascade Network

1. 第一阶段的网络训练

根据人脸框的位置，按照比例截取整个人脸，上半部分人脸，下半部分人脸，分别进行 3 个网络的训练，分别记为 F1，F2, F3，网络的模型如图 3-3，这三个网络的结构类似，最后一层的大小有区别，因为后 2 个网络，都只要输出 3 个点。

2. 第二阶段的网络训练

然后以标注的特征点位置为中心，截取周围的人脸部件块，并对标注的特征点位置进行上下左右的随机移动。目的就是让网络在一个局部的范围内学习到特征点的位置。因为部件的大小较小，用较浅的网络就可以学到相关特征。为了使网络有更好的学习能力，进行多尺度网络学习，即每个点都截取 2 个不同大小的人脸部件块，此这阶段共需训练 10 个网络。

3. 第三阶段的网络训练

第三阶段的网络与第二阶段类似，因为此阶段的网络目的是对前一阶段的网络做更细的调整，所以随机移动的程度也要小于第二阶段的网络。且截取的人脸部件块要小于第二阶段。同样，这阶段共需训练 10 个网络。

所有的 23 个网络，都采用随机梯度下降方法进行训练，每次迭代的大小为 128。需要学习的参数包括卷积的权重 w 和偏置 b 。

五、 级联网络模型的测试

进行级联的卷积网络模型的测试，比较级联对最终准确率的影响。

先进行网络准确率的检测。

(一)网络的准确率的分析

按照 4 中的网络的实现细节，本文训练了第二阶段和第三阶段的网络对第一阶段的网络预测结果进行修正。发现通过级联的方法，显著地降低了检测误差。图 5-1 显示了经过第二阶段的网络，误差大大降低。第三阶段的网络降低了一小部分的检测误差。

最终，经过 3 阶段的级联卷积网络，将每个点的平均误差降低至 1.67%。

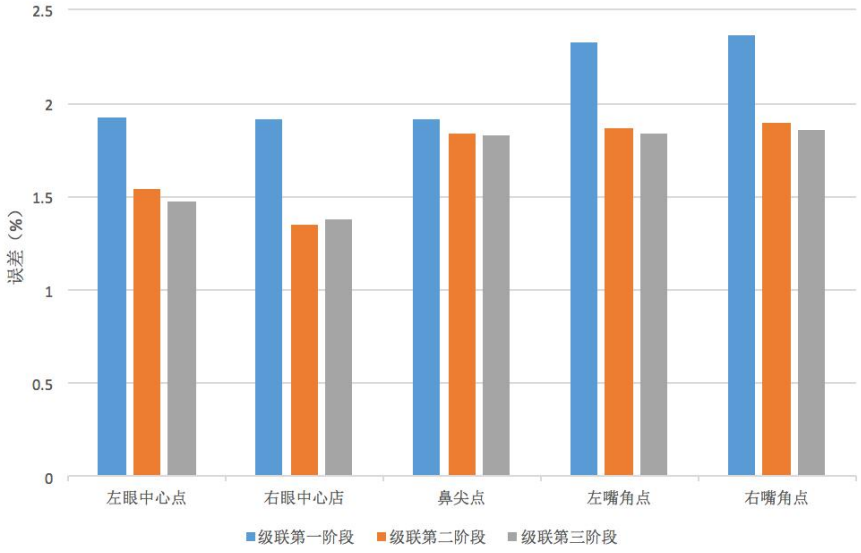

图 5-1 不同级联阶段的 5 个特征点的误差

Figure 5-1 5 Point Detection Error with Different Cascade Network

在本文初始阶段提到因为人脸有不同的朝向，夸张的表情，多变的光照，部分的遮挡等各种情况，面部特征点定位有很大的挑战，现对之前的 4 张图片应用获得的模型进行测试，测试结果见图 5-2。左上角图由于光线问题，左半部分人脸几乎是不可见的。右上角图为侧脸，右眼几乎是不可见的。左下角图的人在哭泣，表情是比较夸张的。右下角图的人的头发遮住了研究。定位的结果显示了本文的面部特征点定位模型能够适应变化多样的人脸条件下高准确率的定位的需要。

图 5-2 变化多样人脸的特征点定位的结果

Figure5-2 Localization Result Under Natural Scenes

从验证集中选了几张图来显示模型定位的效果。

前 4 张为遮挡的人脸，有戴了眼镜，看不到眼睛的，有因为抽烟，挡住了嘴角点的，也有因为物体挡住了人脸的。

中间 4 张为不同姿势的人脸，有向左侧、向右侧的人脸，有向下的人脸，也有朝上的人脸。

最后 4 张为有夸张表情的人脸。因为悲伤，愤怒，跟正常表情下的人脸有差别。

定位的结果表明级联的深度卷积网络模型能够进行准确的定位。

图 5-3 遮挡的人脸的特征点定位的结果

Figure5-3 Localization Result with Occluded Faces

图 5-4 不同朝向的人脸的特征点定位的结果

Figure5-4 Localization Result with Different Face Poses

图 5-5 表情夸张的人脸的特征点定位的结果

Figure5-5 Localization Result with Exaggerated Expressions

(二)网络的效率的分析

之后来测试每张图片的处理时间，在 CPU 上进行测试 3466 张图片。

包括图片的的处理和网络预测，3466 张图片共花费 84.5 秒，约 0.0244 秒每张图片，有较好的检测性能。

六、 特征点定位的扩展

在本文中，实现了 5 个面部特征点的精确定位，但对于有些应用，特征点的数量是不够的，如利用特征点进行 3D 人脸建模。因此考虑定位更多的特征点。现在，目标是定位 68 个特征点。这些特征点覆盖到了眉毛，眼睛，鼻子，嘴巴，脸的轮廓。

(一)网络设计结构

同样的，先开始考虑网络结构的设计。在 68 个特征点的定位中，网络需要学习到更多的信息，考虑增加网络的输入，从 39*39 大小增加至 60*60，同时，增加卷积的层数，以便于网络更好地进行学习到层次化的特征。同时，在实验的时候发现，脸的轮廓特征点的误差远大于内部特征点（即轮廓内的特征点）的误差。于是，考虑将轮廓点（共 17 个）与内部点（共 51 个）分开进行训练。本文设计的网络模型见表 6-1。5*5*20 表示卷积核的大小为 5*5，产生 20 个特征图。

表 6-1 网络结构

Table 6-1 Network Structure

网络	轮廓点	内部点
输入	60*60	60*60
卷积 1	5*5*20	5*5*20
卷积 2	5*5*40	5*5*40
卷积 3	3*3*60	3*3*60
全连接	200	200
输出	34	102

(二)数据处理

数据集和 68 个标注点来自 AFW, LFPW, HELEN 和 IBUG [23]，共有 3837 张图片。选择 3200 张图片进行训练。然后，我们进行数据增强，对图片进行上下左右的平移和旋转，也对点进行相应变换。最后获得 22400 张图片。数据的预处理技术同第四章。

(三)利用 5 个特征点模型预训练

本文先处理好数据进行 51 个内部点的网络的训练，希望此网络能够直接回归出 102 个数值，即 51 个坐标。但在训练的时候发现了一个很奇怪的现象，网络的验证误差下降的非常缓慢，如图 6-1。在之前训练网络的过程中，前几个 epoch 的误差下

降都是很快的，一般可以从 0.02 直接降到 0.001 这样的级别。然后进行学习率的调整，增大学习率，网络的验证误差下降的会快一点，但经过 20 个左右 epoch 后就不再下降了，而减少学习率，网络的误差又下降的非常慢。本文考虑到直接回归出 102 个数值本身是个比较困难的事情，而在网络的初始化的时候，权重和偏置等信息又是随机初始化的，因此离的要优化的目标有很大的距离。

Epoch	Train loss	Valid loss	Train / Val
1	0.066014	0.022744	2.902514
2	0.025313	0.022741	1.113136
3	0.024691	0.022738	1.085859
4	0.024259	0.022736	1.067004
5	0.024024	0.022735	1.056671
6	0.023814	0.022731	1.047653
7	0.023673	0.022726	1.041673
8	0.023545	0.022659	1.039093

图 6-1 训练 68 特征点时的网络损失

Figure6-1 The Loss of Training 68 Facial Point

所以开始考虑，怎样能使网络的权重有一个比较好的初始值呢？这样网络就可以较快收敛，大大减少训练时间。一个很自然的想法就是，本文之前已经训练过 5 个面部特征点的网络，这个网络已经能够提取人脸从低层次到高层次的特征。

因为都是人脸数据，68 个特征点的网络和 5 个特征点的网络应该是可以共享一些低层次的特征。所以本文将 68 个特征点的网络权重以 5 个面部特征点的网络学习到的权值来初始化。第三层的卷积和 2 个全连接的权重依旧随机初始化（因为 5 个面部特征点的网络只有 2 层卷积，且因为输出的不同，全连接的权重的维数也有差异）。利用随机梯度下降算法训练，直至收敛。通过利用之前 5 个特征点网络的权值，本文的网络参数有一个较好的初始值，避免网络陷入局部最小值，也提升了训练速度。

(四)测试结果

本文应用卷积网络模型进行 68 个特征点的检测，发现在人脸有不同朝向，不同的表情和遮挡情况下，定位都是准确且鲁棒的。

图 6-2 HELEN 测试集上的结果

Figure 6-2 Detection Result of HELEN Test Set

图 6-3 LFPW 测试集上的结果

Figure 6-3 Detection Result of LFPW Test Set

七、 总结和展望

(一)总结

本文为解决人脸因为不同的朝向，夸张的表情，多变的光照，部分的遮挡，导致面部特征点的定位困难这个问题，设计和实现了基于深度卷积网络的定位模型。此模型有较高准确率和鲁棒性。

主要有以下工作：

1.利用深度卷积网络抽取了有区分性的层次化特征，训练了卷积网络模型，进行特征点定位。

2.为进一步提高特征点定位的准确率，训练了 3 阶段级联的深度卷积网络实现了由粗到细的特征点定位。第一阶段的网络将人脸的整体图像作为作为输入，直接预测所有的特征点，这一阶段的网络利用了全局的纹理信息且保持了所有特征点的全局约束，能够得到可靠的预测结果。后两个阶段的网络截取人脸的部件作为输入，对前阶段的网络预测进行优化，得到更准确的定位。

3.对深度卷积神经网络的训练进行了探索，通过应用不同的训练技术，使网络模型有更强的学习能力。

4.通过利用已训练好的 5 个特征点的网络权重，本文训练得到了 68 个特征点的卷积网络模型。

(二)未来的工作

在本文中，利用 3 阶段级联的深度卷积网络实现了 5 个面部特征点的精确定位，但对于有些应用，特征的数量还是太少，如利用特征点进行 3D 人脸建模，因此训练了基于 68 个特征点的网络。但目前，网络对于眉毛，鼻子等部件的定位不够准确，考虑对这些部件的定位进一步的优化，提高定位准确率。

参考文献

[1]P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, and N. Kumar. Localizing parts of faces using a consensus of exemplars. In Proc. CVPR, 2011.

[2] X. Cao, Y. Wei, F. Wen, and J. Sun. Face alignment by explicit shape regression. In Proc. CVPR, 2012.

[3]Tompson J, Stein M, Lecun Y, et al. Real-time continuous pose recovery of human hands using convolutional networks[J]. ACM Transactions on Graphics (TOG), 2014, 33(5): 169.

[4] Toshev A, Szegedy C. Deeppose: Human pose estimation via deep neural networks[C]//Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on. IEEE, 2014: 1653-1660.

[5] X. Zhu and D. Ramanan. Face detection, pose estimation, and landmark localization in the wild. In Proc. CVPR, 2012.

[6]T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models- their training and application[J]. Computer Vision and Image Understanding.1995,61(1):38–59.

[7]T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models[J].IEEE Transactions on Pattern Analysis & Machine Intelligence. 2001,23(6) : 681–685.

[8]Stephen Milborrow, Fred Nicolls. Locating Facial Features with an Extended Active Shape Model[J].Computer Vision – ECCV.2008, 5305:504-513.

[9]D. Cristinacce and T. F. Cootes. Boosted regression active shape models[C]. In BMVC, UK, 2007.

[10]吴证. 人脸特征点定位研究及应用[D].上海交通大学,2007.

[11]Cao X, Wei Y, Wen F, et al. Face alignment by explicit shape regression [J]. International Journal of Computer Vision, 2014, 107(2): 177-190.

[12]Sun Y, Wang X, Tang X. Deep convolutional network cascade for facial point detection[C]. Computer Vision and Pattern Recognition (CVPR), US, 2013.

[13] Zhou E, Fan H, Cao Z, et al. Extensive facial landmark localization with coarse-to-fine convolutional network cascade[C]//Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on. IEEE, 2013: 386-391.

[14]LeCun Y A, Bottou L, Orr G B, et al. Efficient backprop[M]//Neural networks: Tricks of the trade.

Springer Berlin Heidelberg, 2012: 9-48.

[15] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representations by error propagation. Nature, 323(99):533– 536, 1986.

[16]Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507.

[17]LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[18]Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. 2012: 1097-1105.

[19]Bouvrie J. Notes on convolutional neural networks[J]. 2006.

[20]Viola P, Jones M. Rapid object detection using a boosted cascade of simple features[C]//Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. IEEE, 2001, 1: I-511-I-518 vol. 1.

[21]闫鹏,牛常勇,范明. 基于级联卷积网络的自然场景下的车牌检测[J]. 计算机工程与设计,2014,12:4296-4301.

[22]Hinton G E, Srivastava N, Krizhevsky A, et al. Improving neural networks by preventing co-adaptation of feature detectors[J]. arXiv preprint arXiv:1207.0580, 2012.

[23]Sagonas C, Tzimiropoulos G, Zafeiriou S, et al. 300 faces in-the-wild challenge: The first facial landmark localization challenge[C]//Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on. IEEE, 2013: 397-403.

致谢

在华东师范大学软件学院学习的四年时间，给我留下了深刻的印象。我得到了许多老师和同学们的帮助和支持，值此论文完成之际，特向他们表示由衷的感谢。

我要感谢我的论文指导老师王晓玲老师，她在学业以及毕业论文上给与了我细心指点，王老师严谨的治学精神以及精益求精的工作态度深深地感染着我，同时对于我的论文初稿在内容以及格式方面提出了建议，使我不断完善论文直至完稿。我也要感谢薛向阳老师在论文的选题和论文实现遇到困难时给了我明确的指点，使得问题一一化解。同时也要感谢大学四年所有的老师，正是他们传授的知识给了我完成这篇论文的知识基础。感谢师兄师姐，室友以及各位同学在论文信息方面提供了及时有效的信息。感谢母校和学院为我提供的优美自在的学习环境，使得我能够安心并且专心地完成所有的学业。感谢以上所有，帮助我顺利完成这篇论文。