

1、前言

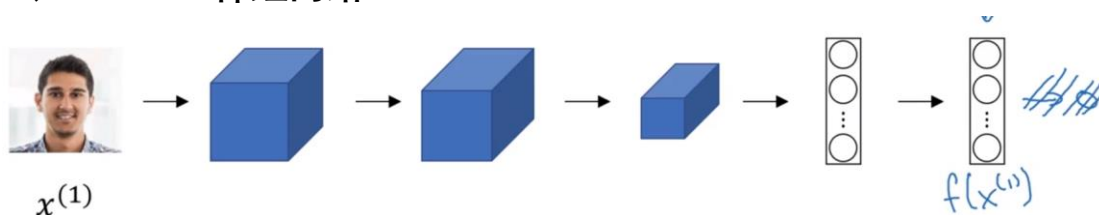
之前学习了图像检测以及图像定位，在具备了一定的基础知识之后，现在来看看人工智能在计算机视觉领域中比较火的一个应用——面部识别。同通常情况，面部识别首先要做到最基础的部分是人脸验证，即给定输入图片，看能否分析出里面的特征进而判断输出结果是否与给定对象一样，这在人脸识别问题中称之为1对1验证问题，而更高级的是解决输出结果与对象集中的对应关系，这称之为1对多识别问题。

2、一次学习

在人脸识别中，有一个非常严峻的问题，即训练对象有可能只有一个样本。譬如公司的门禁系统，每个员工的照片可能都只有一张，意味着你只能通过“一次”训练去获得对象与结果的映射关系。另外一个挑战是，输出结果随时要改变（员工的更新）难度要不断训练吗？

因此，这里引入 similarity 函数 d 。该函数的作用是给定两张图片（一张训练图片，一张测试图片）输出它们的差异值，如果输出结果小于某个阈值，则认为是同一个人。

3、Siamese 神经网络



在原来卷积神经网络中，最后一层 FC 的输出，将用于判别输出为哪一类。而在新的神经网络中，去掉了最后的 softmax 输出结果，转而使用 d 函数，该输出结果极为 $d1$ 。此时，又有另外一张图片的 d 函数输出结果为 $d2$ ，则比较两者的差值，进行差异化分析，小于约定的某个阈值，则认为是同一个对象。

接下来的问题是，如何形成这个 d 函数的输出过程，同样利用反向传播的思想计算出 d 函数的超参数，让我们利用三元损失函数计算它吧。

4、三元损失函数

首先定义几个术语，Anchor 图片为标准图片，Positive 为测试集中同一个对象，Negative 为测试集中不同对象集合，简称为 APN，三元组的名称就是由此而来的。

那么 $d(A, P, N)$ 的值即是差异值输出结果，它的定义如下图所示：

$$\|f(A) - f(P)\|^2 - \|f(A) - f(N)\|^2 + \alpha$$

训练它的目的是使整个式子的输出 ≤ 0 ，现在来看其中的参数设置， α 为间隔值，即同一对象差异值和不同对象差异值的间隔，当然 α 设置的越大，识别越精确。对于一次识别问题，同一人的样本可能只有一个，但是这里在训练 Siamese 神经网络的过程中，需要 A 对应的 P 有若干个，这样才能达到训练的目的。

最后，对于 A, P, N 的选择，通常遵循的是随机选择，但这样的效果不大好，因

为这样很容易就会达成上方的式子 ≤ 0 ，因此，得选择难以训练的样本，即 A、N 和 A、P 尽可能接近，这样 d 函数会使它们尽可能拉开距离。

在了解了这些原理之后，接下来要做的就是训练样本，很多 FaceNet 软件都是通过百万级的训练样本才得出这样的 Siamese 神经网络，因此，如果自己要说的话，最好是下载开源项目上的模型进行学习。