DeepID 系列Paper总结

PRML上的神经网络总结：

前向网络：

误差反向传播

**其一是通过非常难的大规模人脸分类任务训练神经网络，在网络的隐藏层学到了富含人脸身份属性的特征**

**是在人脸分类的同时加入另一个人脸比对的训练信号。**这一额外的比对信号的加入进一步将LFW上人脸识别的错误率减小了67%，人脸识别首次在LFW上突破99%的准确率。我们称这种通过联合人脸分类-比对学习到的人脸特征表示为DeepID2

1. 第一篇：Deep Learning Face Representation from Predicting 10,000 Classes

文章中首先介绍了

DeepID2+文章为脸部验证设计了一个高性能的深度卷积神经网络。仍然是通过识别和验证两种监督信号

DeepID2+摘要

本文设计了一种高性能深层网络（DeepID2 +）进行人脸识别。使用识别验证监控信号来学习。 DeepID2 +通过增加隐藏表现的维度并加强对早期卷积层的监督，在LFW和YouTube Faces基准上实现了新的最先进的技术。

通过实证研究，我们发现其深层神经激活的三个性质对于高性能至关重要：稀疏性，选择性和鲁棒性。 （1）观察到神经激活中度稀疏。适度的稀疏度可以最大限度地提高深层网络的辨别力以及图像之间的距离。令人惊讶的是，即使在神经反应二值化之后，DeepID2 +仍然可以实现高识别精度。 （2）其较高层的神经元对身份和身份相关属性具有高度的选择性。当存在不同的身份或属性时，我们可以识别出不断激发或抑制的神经元的不同子集。尽管DeepID2 +没有被教导在训练中区分属性，但它隐含地了解了这样的高级概念。 （3）尽管闭塞模式不包括在训练集中，但是对于闭塞来说更为鲁棒。

DeepID2+的未整理内容：

面孔识别取得了巨大进展，这得益于该领域的广泛研究工作[36,30，18，14，1，6，15，26，7，5，27，31，29，32，28]。追求更高绩效是一个中心议题，了解其背后的机制同样重要。当深层神经网络开始接触人类面临的挑战性面貌基准[31,29,32,28]，如LFW [16]时，人们渴望知道这些神经元学到了什么，以及如何实现高性能。在认知科学中，有很多研究[34]分析视觉皮层神经元的面部处理机制。在这些作品的背景下，我们分析了人造神经网络中神经元的行为，试图解释人脸识别过程深层网络，什么信息在神经元中被编码，以及它们对破坏程度的强大容忍度（鲁棒性）。

我们的研究是基于本文提出的高性能深层神经网络（ConvNet深入[19,20]），被称为DeepID2 +。通过增加隐藏表示的维度并对早期的卷积层增加监督，改进了最先进的DeepID2网[28]。最好的单一DeepID2 +网络（采用原始和水平闪烁的脸部图像作为输入）在LFW上获得98.70％的验证精度（与DeepID2相比为96.72％）。结合25个DeepID2 +网络在多个基准上设置了新的最先进的技术：面向验证的LFW为99.47％（与DeepID2相比为99.15％[28]），封闭和开放式LFW的95.0％和80.7％ （通过网络规模培训（WST）[32]分别为82.5％和61.9％），面孔验证（35％）为93.2％，而DeepFace [31]为91.4％）。

随着最先进的深层ConvNets和广泛的经验评估，我们调查神经激活的三个属性对于高性能至关重要：稀疏性，选择性和鲁棒性。大面积数据培训后，他们自然拥有deepID2 +，我们没有对模型和培训过程进行任何额外的规范化来实现。因此，这些结果对于理解深层网络的内在属性很有价值。

观察到DeepID2 +的神经激活中度稀疏。如图所示。如图1所示，对于输入面部图像，顶部隐藏层中的大约一半的神经元被激活。另一方面，每个神经元在大约一半的脸部图像上被激活。这种稀疏分布可以最大化深网的辨别力以及图像之间的距离。不同的身份有不同的神经元子集激活。相同身份的两个图像具有相似的激活模式。这促使我们将顶层隐藏层中的神经反应二值化，并使用二进制代码进行识别。其结果令人惊讶。其在LFW上的验证精度仅略微下降1％以下。它对大面积搜索具有重大影响，因为节省了巨大的存储和计算时间。这也意味着二进制激活模式比深层神经网络中的激活幅度更重要。

与稀疏相关，还观察到，较高层中的神经元对身份和身份相关属性具有高度的选择性。当身份（谁可以在训练数据之外）或属性被呈现时，我们可以识别不断激发的神经元的子集，并且还可以发现不断被抑制的另一个神经元子集。来自这两个子集中的任一个的神经元具有关于该身份或属性的存在/不存在的强烈指示，并且我们的实验表明单个神经元对于特定身份或属性具有高识别精度。换句话说，神经激活对身份和属性具有稀疏性，如图1所示。尽管DeepID2 +没有被教导在训练中区分属性，但它隐含地了解了这样一个高层次的概念。直接使用DeepID2 +学习的面部表情，导致身份相关属性的分类精度高于广泛使用的手工功能，如高维LBP [7,5]。

我们的实证研究表明，较高层中的神经元对于人脸识别中的图像损伤要比手工制作的特征（如高维LBP或下层神经元）更加鲁棒。作为图1所示的例子。如图1所示，当面部图像被部分遮挡时，其二进制激活模式保持稳定，尽管幅度可能改变。我们推测原因可能是更高层的神经元捕获全局特征，对局部变化较不敏感。 DeepID2 +由自然网络脸部图像训练，并且没有将人工遮挡模式添加到训练集中。

**相关工作的内容，以上都是介绍**

属性学习是一个活跃的话题[10,2,25,4,24,40]。 已经有关于第一个学习属性分类器和使用属性预测进行面部识别的作品[18,8]。 我们在本文中尝试的是首先预测身份的逆，然后使用学习的身份相关特征来预测属性。

基于稀疏表示的分类[36,37,22,39,9,17]被广泛研究用于具有闭塞的面部识别。 唐等 [33]提出了Robust Boltzmann Machine来区分损坏的像素并学习潜在的表示。 这些方法设计的组件明确地处理遮挡物，而我们显示DeepID2 +学习的特征具有隐式编码的不变性。 自然而然地，在不增加对模型或人工遮挡模式的调节以训练数据的情况下。

**DeepID2+net**

我们的DeepID2 +网络继承自DeepID2网络[28]，它们具有四个卷积层，具有20,40,60和80个特征图，随后是与第三个和第四个卷积完全相连的160维特征层层。 160维特征层（DeepID2特征层）由面部识别和验证任务进行监督。给定一对训练图像，通过DeepID2网络前向传播两个图像，获得两个DeepID2特征向量（f i和f j）。然后，每个DeepID2特征向量被归类为训练集中的8192个身份之一，并产生分类（识别）错误。如果f i和f j来自相同的身份，则验证错误2为1 2 kf i - f j k 2 2或1 2 max0，m - kf i - f j k 2否则。结合识别和验证监控信号有助于更有效地学习功能[28]。

与DeepID2相比，DeepID2 +做了三个改进。首先，在四个卷积层中的每一个中，128个特征图更大。最终的特征表现从160到512个维度增加。第二，我们的训练数据通过合并CelebFaces +数据集[29]，WDRef数据集[6]和一些从LFW排除的新收集的身份来扩大。大型的DeepID2 +网络训练有来自12,000个身份的大约290,000张脸部图像，而来自用于训练DeepID2网络的8,000个身份的160,000张图像。第三，在DeepID2网络中，监控信号仅被添加到连接到第三和第四卷积层的一个完全连接的层，而较低的卷积层只能通过从较高层反向传播的梯度得到监视。我们通过将512维全连接层连接到四个卷积层（最大池）之后来加强监督，在n = 1,2,3,4时表示为FC-n，并对这四个完全连接层进行监控，连接层与识别验证监控信号[28]同时进行，如图1所示。以这种方式，监控信号变得越来越接近早期卷积层，并且更有效。

**High-performance DeepID2+**

为了验证改进，我们基于DeepID2 +网络中n = 1,2,3,4的四个512维特征向量（神经激活）FC-n中的每一个，了解了用于面部验证的联合贝叶斯模型[6]。联合贝叶斯在我们的培训班（不包括LFW人员）中接受了2000人培训，不用于培训DeepID2 +网络，并在LFW的6 000名给定面部对中进行面部验证。当训练DeepID2 +网络以确定学习率和训练迭代时，这2000个身份也作为验证集。将DeepID2 + net（建议）与三个网络进行比较，分别删除了三个改进之一，如图1所示。对于没有分层监督的网络，只有FC-4的梯度反向传播到卷积层。对于具有较少训练数据的网络，仅使用来自CelebFaces +的训练数据。对于较小的网络，卷积层中的特征图的数量与DeepID2中的特征映射的数量相同，并且在n = 1,2,3,4中，将160维特征用于FC-n。所有被比较的网络在固定区域覆盖整个脸部。我们可以清楚地看到图3的三个方面的改进。

为了与DeepID2网络进行比较[28]，我们训练了25个DeepID2 +网络，采用了DeepID2选择的相同的25个脸部区域，如图1所示。 [28]中的2。提取了DeepID2 +的FC-4层的特征，以训练贝叶斯联合面进行验证（如[28]中的图2所示，在原始水平面或水平面上提取特征）。 LFW表面验证任务的25个深度对比之间的比较如图1所示。 DeepID2 +网络在DeepID2平均提高了约2％的精度。当分别从25个脸部区域提取的FC-4层特征及其与25个DeepID2 +网络的水平对照相结合时，我们分别在LFW和YouTube Faces数据集上实现了99.47％和93.2％的面部验证精度。标签。 1和Tab。 2是与两个数据集之前的最佳结果的精度比较。图。 5， 6是中华民国的比较。我们的DeepID2 +网络在两个数据集上都胜过以前的所有结果。 LFW和YouTubeFaces中有几个错误标签的测试脸对。修正后，我们的脸部验证精度在LFW上提高到99.52％，在YouTubeField上增加了93.8％。

面部识别是评估高性能面部识别系统的一项更具挑战性的任务[32]。因此，按照[3]中的协议，我们进一步评估了LFW上关闭和开放面部识别任务的25个DeepID2 +网络。封闭式标识报告了1级识别精度，而开放式识别以1％的误报率（FAR）报告了Rank-1检测和识别率（DIR）。比较结果如表。我们的结果显着优于以前最好的[32]，分别为95.0％和80.7％的闭合和开放识别精度。

**Moderate sparsity of neural activations**

对于每个图像，大约有一半被激活（具有正激活值）的神经元，并且对于每个神经元，大约一半的图像被激活，这意味着神经激活是适度稀疏的。图像的适度稀疏使不同身份的脸部最大限度地区分开来，而神经元的适度稀疏使他们具有最大的辨别能力。我们通过在我们的验证数据集（图7左边）中计算46 594个图像中的每个图像上激活的神经元的直方图，以及激活每个神经元的图像数量的直方图来验证这一点（图7右侧）。评估基于单个DeepID2 +网络中的FC-4层神经元，以整个面部区域作为输入。与FC-4层的所有512个神经元相比，图像上激活的神经元数量的平均值和标准偏差为292±34，而与所有46,594个验证图像相比，图像数量的平均值和标准偏差每个神经元被激活的是26,565±5754，两者都以所有神经元/图像的一半近似为中心。

我们进一步验证激活模式，即神经元是否被激活，比精确的激活值更重要。我们将神经激活转换为二进制代码，通过阈值和比较LFW的面部验证能力与原始表示的面部验证能力。如表。 4，当与联合贝叶斯联合时，二进制表示法分别为1％或更低的精度（分别为97.67％和99.12％，单个网络或组合25个网络）。更有趣的是，即使通过直接计算汉明距离，二进制代码仍可以通过单个网络或组合25个网络分别达到96.45％和97.47％的精度。这表明神经元的激发或抑制状态已经包含了大多数歧视性信息。二进制代码是经济的存储和快速的图像搜索。我们相信这将是未来工作的有趣方向。

**Selectiveness on identities and attributes**

6.1。神经元的辨别力

我们测试了两个二进制分类任务的DeepID2 +功能。首先是将一个人的脸部图像与所有其他人或背景的脸部图像进行分类。第二个是将脸部图像分类为具有属性或不属性。 DeepID2 +功能分别来自整个脸部区域的单个DeepID2 +网络的FC-4层及其水平闪烁的对应物。实验在LFW下进行，人们在训练期间被DeepID2 + net看不到。 LFW被随机分为两个子集，交叉验证精度被报告。准确度归一化w.r.t.正面和负面课程中的图像编号。我们还比较了具有各种特征尺寸的高维LBP特征[7，5]。如图所示。 8，DeepID2 +特性在属性分类中显着优于LBP（DeepID2 +具有良好的身份分类结果并不奇怪）。图。 9， 10显示了仅选择一个最佳特征的身份和属性分类精度。为不同的身份（属性）选择不同的最佳特征。使用单个功能（神经元），DeepID2 +对于某些身份和属性达到约97％。这就是DeepID2 +特征是身份和属性选择的证据。显然LBP没有。

6.2。兴奋性和抑制性神经元

我们发现，对身份和面部属性的歧视是由于某些身份或属性上的神经元的激发和抑制模式。例如，当看到乔治·布什（George Bush）看到科林·鲍威尔（Colin Powell）时，看到乔治·布什（George Bush）变得抑制，或者一个神经元可能会被西方人激动，而对亚裔人士的抑制，一个神经元可能会激动起来。图。 11a比较DeepID2 +神经激活对属于特定单个身份（左列）和剩余图像（中间列）的图像的平均值和标准偏差，以及显示每个给定身份与每个特定身份的每个神经元分类的准确性剩余图像（右列）。在LFW中具有最多脸部图像的顶级身份被评估（其他身份具有相似的结果）。所有三列中图像的评估身份的平均神经激活对神经顺序进行排序。对于每个给定的身份，存在强烈兴奋的神经元（例如，具有小于200的神经元的神经元）或被抑制（例如具有大于600的神经元的神经元）。对于激发的神经元，其激活分布在更高的值，而其他图像在这些神经元上具有显着较低的平均值。因此，兴奋性神经元可以很容易地将身份与其他身份区分开来，这是通过右栏列出的图形中具有小神经元的红点所示的高分类精度来验证的。

对于位于中间的神经元（例如，具有约400个神经元的神经元），它们在给定身份上的激活分布与其他身份上的激活分布大部分重叠。对于给定的身份，他们具有弱的识别能力，由两种颜色的连接点附近的红色和蓝色点的低精度证实。这些神经元的激发或抑制状态有很大的不确定性。

当平均激活进一步降低（例如，高于600的神经元）时，神经元显示出抑制性质，并且与其他特征相比较，对于给定身份很少被激活。这些抑制性神经元也具有相对较高分类精度的鉴别能力。

然而，如图1所示，在LBP特征上不能发现类似的现象。 11B。给定身份的LBP功能的范围和剩余的图像对于所有功能重叠。与DeepID2 +神经激活相比，LBP特征具有低得多的分类精度，其中大部分在50％随机猜测线上累积。 12a比较包含特定属性（左列）和剩余图像（中间列）的面部神经激活的范围，以及显示每个属性与其余图像（右列）区分的每个神经元分类的准确性。与身份相似，下列和更高级别的神经元对于属性的选择表现出如图所示，包括男性，白人，黑人，亚裔，儿童，高级，秃头和灰色头发。这些属性与身份有区别。其他属性的选择性相对较弱，如印度，青年，中年，黑发，白发和褐发（未显示）。这些属性在视觉上是不明确的或对身份的歧视较少。例如，印度人有时看起来像亚洲人，而且经常看到在青年和中年时拍摄的身份相同的身份，或以不同的头发颜色拍摄。图。 12b比较了与图1相同的属性集合的LBP特征的范围和每特征分类精度。 12A。给定属性和剩余图像的LBP特征范围与所有特征重叠，分类精度累积在50％随机猜测线上。

6.3。神经激活分布

图。 13，图14示出了在给定身份或属性上的神经激活的直方图的示例。图。 13第一行还显示了五个随机选择的神经元的所有图像上的直方图。对于每个神经元，其激活的大约一半是零（或接近零），另一半具有更大的值。相比之下，给定身份的直方图表现出很强的选择性。一些神经元对于给定的身份不断被激活，激活直方图分布在正值中，如图1中每个身份的直方图的第一行所示。 13，而其他一些被持续抑制，其中激活直方图累积在零或小的值，如每个身份的第二行直方图所示。

对于属性，在图3的每一列中。在14a，14b，14c和14d中，我们分别显示了几个相关属性，即与性别，种族，年龄和头发相关的单个神经元的直方图。神经元选择为兴奋性（红色帧）或抑制性（绿色框架），并且可以最佳地分类每行左侧所示的属性。如这些图所示，神经元对于某些属性表现出强烈的选择性（被激活或抑制），其中对于给定属性，神经元被激活（抑制），同时对同一类别中的其他属性被抑制（激活）。

**7. Robustness of DeepID2+ features**

我们测试DeepID2 +功能对具有遮挡的脸部图像的鲁棒性。在第一个设置中，面部被局部遮挡10％到70％的区域，如图1所示。第15排。在第二设定中，如图所示，面部被10×10〜70×70像素的随机块遮挡。 15秒排。在闭塞实验中，DeepID2 +网和联合贝叶斯模型在我们的训练集中的原始脸部图像上被学习，没有任何人工添加的闭塞，而闭塞面仅用于测试。我们还测试了高维LBP特征加上联合贝叶斯模型[7]进行比较。图。 16比较了具有不同程度部分遮挡的LFW测试集上的DeepID2 +和LBP特征的面部验证精度。将DeepID2 +功能从单个DeepID2 +网络中的FC-1到FC-4层，并以整个面部区域作为输入。我们还使用25个DeepID2 +网来评估我们的整个人脸识别系统。比较的高维LBP功能是从21个面部地标中提取的99，120个维度。如图所示。 16，LBP的表现显着下降，即使轻微的10％-20％的闭塞。相比之下，对于具有两个卷积（FC-2，FC-3和FC-4）的DeepID2 +功能，性能在很大的范围内缓慢下降。当40％的面部被遮挡（除了FC-1层）之后，DeepID2 +的面部验证精度仍然高于90％，而LBP特征的性能已经降到70％以下。当关键眼睛区域被遮挡时，DeepID2 +的表现只会降低50％以上的闭塞。它还表明，较高层（其应该是更全球分布）的特征对于闭塞更为鲁棒，而LBP和FC-1都是局部特征，对闭塞敏感。结合从25个脸部区域提取的DeepID2 +网络实现了最强的鲁棒性，具有93.9％的面部验证精度，40％的遮挡和88.2％的精度，即使只显示额头和头发。

我们还对具有随机块闭塞的面部图像进行DeepID2 +和LBP特征的面部验证，n×n块大小分别为n = 10到70。这种设置是具有挑战性的，因为待验证的两个面中的封闭区域的位置通常是不同的。因此，在像素距离的意义上，同一人的图像将会看起来有很大的不同。图。图17显示了比较结果，当块大小大于20×20时，LBP特征的准确性开始下降，而DeepID2 +特征（FC-1除外）在较大范围内保持了性能。具有50×50块的遮挡，LBP功能的性能已经下降到70％左右，而单个DeepID2 +网络的FC-4层仍具有89.2％的准确性，而25个DeepID2 +网络的组合更高达92.4 ％ 准确性。再次，浅层FC-1层的特征行为更接近LBP特征。上述实验表明，深层结构使神经元对图像的破坏更为鲁棒。这种鲁棒性在深层ConvNets中是固有的，没有明确的建模。

图。图18和19分别显示FC-4层神经元对具有不同程度的部分和随机块闭塞的单个身份的图像的平均激活。神经元根据每个身份的原始图像的平均激活进行排序。对于两种类型的闭塞，激活模式保持大致不变，直到大程度的闭塞。

8.结论

本文设计了一个高性能的DeepID2 +网络，为LFW和YouTube Faces设置了面向身份识别和验证的新技术。通过实证研究发现，DeepID2 +学习的面部表征中等稀疏，对身份和属性的选择性很高，对图像的破坏也很具有鲁棒性。在过去，已经做了许多研究工作，旨在通过向模型或系统显式添加组件或规则化来实现这种有吸引力的属性。然而，他们可以通过大规模的训练自然地通过深层次的模式实现。这项工作不仅显着提高了人脸识别的表现，而且提供了宝贵的见解，帮助人们了解深入学习及其与许多现有计算机视觉研究（如稀疏表征，属性学习和闭塞处理）的联系。这样的见解可能会激发未来更激动人心的研究。作为一个例子，在这项工作中，我们已经表明二进制神经激活模式是高效率和有效的面部识别。