指导教师：杨涛　　提交时间：2016/03/18

CVPR2015 Paper Translation

No: 01

姓名：杜婧

学号：2013302540

班号：10011304

利用多任务深度神经网络旋转人脸

Junho Yim Heechul Jung ByungIn Yoo

Changkyu Choi Dusik Park Junmo Kim

电气工程学院，KAIST,韩国

三星尖端技术研究所

{junho.yim, heechul, junmo.kim}@kaist.ac.kr

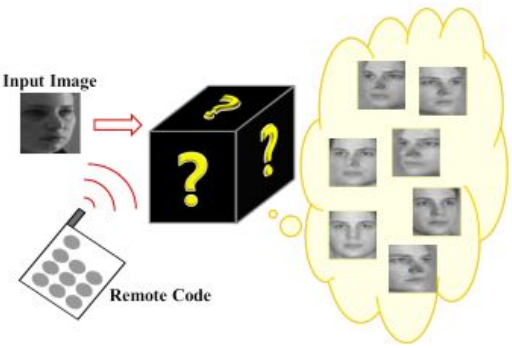
{byungin.yoo, changkyu choi, dusikpark}@samsung.com

摘要

在变化的视点和亮度条件下的人脸识别是一个十分困难的问题，许多学者已经尝试过利用产生位姿不变和光照条件不变特性的方法来解决这个问题。朱艾斓将所有任意位姿和光照条件下的图像改变为正面视角的图片以用于不变特征。在这个方案中，保留图片标识的同时旋转位姿图片是一个极其重要的问题。这篇论文提出了一个新式的基于一种新型的多任务深度学习结构。这种结构在将一个任意的位姿和亮度条件下的图片旋转称为目标姿态图片的时候表现出了卓越的效果，并且在变换的同时保留了图片的标识。而所需要的目标姿态可以由使用者的意图来进行控制。相比起单任务模型而言，这一种新兴的多任务模型显著的提升了特征保护。利用所有为位姿光照不变特征而人工合成的且受控制的位姿图片，即受控姿态图像(CPI)，并且从多人脸识别的结果中进行投票，我们发现在多数据集方面该种算法已经能够显著地超越了当前国家最先进的算法超过 4%~6%。

1. 简介

最近，我们在面部识别技术方面取得了十分重要的发展，而这份发展要尤其归功于深度学习。朱艾斓提出了一种深度模型，这个模型能够将一个具有任何位姿和亮度条件的面部图像转变为一种所谓的标准面部图像。这种图象就好像是从人脸的正面并在标准亮度条件下去观察面部。深度人脸识别是第一个在如此具有挑战性的 LFW 数据集中的面部识别方面表现出了人类水平，并且这一记录在最近被一个更加有利的深度学习理论以难以



图表 1.所提出模型的概念性示意图。输入的一个任意位姿和亮度条件下的图像被转变成为了另一个位姿下的图像。远程代码代替目标位姿代码后的输出图像与原输出图像是一致的。通过输入图像与远程代码的相互作用，我们的模型就可以产生出所需要的位姿图像。

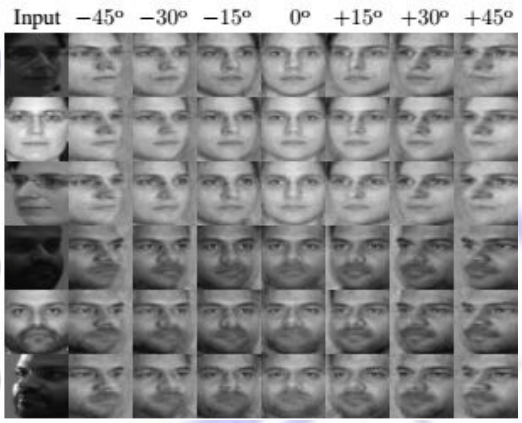
置信的 99%的 LFW 数据集识别率而刷新。一个与面部识别有关的重要挑战是是在保持面部特征的同时改变面部图像的视点或者合成一个新的视角下的面部图像。举例而言，深度人脸识别也依赖于一种预处理阶段。这个预处理阶段将输入的脸部图像旋转到一个标准视角。最近的一项研究，即多视角感知机，一个学习人脸特征和视角表示的深度模型，在进行拓展之后，不仅可以生成一个标准视角的面部图像，而且可以产生出许多具有任意位姿的面部图像，并且保持图像的特征。

这篇论文在最近研究成果的基础上改善了人脸识别，通过提出一种简单但有效的方法来将一个二维空间的脸部图像旋转到一个用户所需要的不同位姿下的脸部图像。更具体的来说，一个任意的位姿和任意的亮度被作为输入，而一个在正面光照条件下的受控位置被当作输出而生成。这一概念在图表 1 中已经进行了展示和说明。我们训练了一个深度神经网络(DNN)，该系统可以在输入一个面部图像以及一个被称为远程代码的二进制代码的条件下，将目标位姿编码并产生一个具有相同特征的从目标位姿观察的面部图像，而这些特征均由远程代码来进行标示。这就像用户具有一个能远程操作的黑匣子旋转器，这个旋转器可以根据用户的远程代码将一个给定的面部图像进行旋转。这个旋转器的质量可以用输出面部图像与所需图像的一致程度和面部特征背包吃的程度来衡量。图表 2 给出了模型的最终结果。我们的模型从输入的具有相同特征点的不同亮度和不同位姿的图片的条件下的除了几乎和在正面光照下的检测图片完全一致的结果。

为了提高深度神经网络的特征保持能力，我们介绍了一种辅助的深度神经网络和一个辅助的任务，他们需要主深度神经网络和辅助深度神经网络串联连接，主深度神经网络用于生产所需的位姿图像，而辅助深度神经网络用来重建原始的输入图像，即辅助深度神经网络将主深度神经网络生成的输出图像重建为原始的输入图像。而问题在于如果主深度神经网络和辅助深度神经网络的串联连接可以重建原始输入图像，那所用的主深度神经网络的输出图像必须是所有的特征点都被保持的并且要包含输入图像特征的

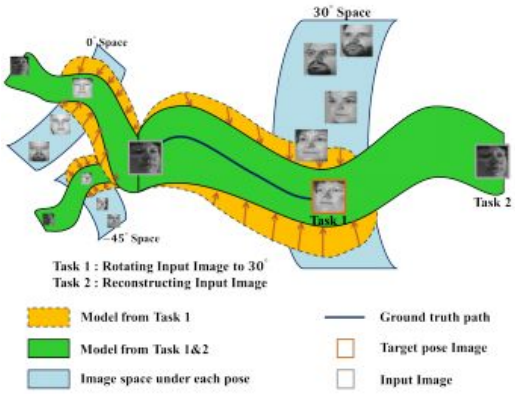
有效信息。如果主深度神经网络没有将其特征保持下来，那么主深度神经网络的输出图像就已经包含了和原始的输入图像不同的特征，那么接下来的辅助深度神经网络的结果将会偏离甚至完全脱离所设定的标准面部图像最终的特征。

这个多任务学习方法的另一概念性图表在图表 3 中展示。假设你想要将一个给定的面部图像旋转到 30 度，一个用典型单任务方法培训的深度神经网络将会以某种程度上的一种脱离参考标准的路径的一种方法来对人脸进行变形。这一方法的路径在图中以黄色的形式进行了标出。这时，输出图像将在某种程度上位于符合 30 度参数的位姿的区域与黄色区域的交界或



图表 2. 第一列展示了从多重 PIE 数据集中提取的两个人的输入测试图像。剩下的列是在不同的远程代码下的输入图像产生的输出。例如，第三列展示了从第一列图像和-30 度的远程代码中得到的-30 度位姿的图片。前三行具有相同的特征，最底部的三行是在相同特征条件的情况下的不同亮度和不同位姿。

其子空间。如果有了将输入图像从输出图像中恢复出来的额外任务则变形路径将更接近于描画的参考标准路径的绿色区域。而这种变化的原因则归功于已改进的特征保持能力。相似的，所需的目标位姿可以为 0 度或 45 度等等，如图表 3 所示。先前的多任务学习模型已经展示了用于决定普通特征的依据，在已展示出的层次以后，余下的层次将会费城多个多任务处理。然而我们将多任务处理模型设计成为一种与众不同的方式。如图表 4 所述，多任务处理模型在主深度神经网络中分享所有的层次并且辅助深度神经网络正确连接在主深度神经网络之后，用于提高特征保持能力。为了评价我们所设计的模型的表现，我们准备了一个面部识别挑战。我们在大的多处理 PIE 脸部数据库进行训练和测试，该数据库涵盖了在不同高度下拍下的不同位姿的脸部图像。我们在每一个被测试的图像样本中使用了一些位子不同的图像来测试位姿不变特性和亮度不变特性。我们得出的有效结论如下：1. 我们提出了新的体系结构和远程代码，他们可以高效的将图像变换到所需要的位姿。与多视角感知机不同，我们所提出的理论可以在单个实验中产生出具有所需位姿的新的面部图像，在该实验中，许多志愿者提供的面部图像的面部旋转结果将会产生，而具有最佳对照位姿的图像的志愿者将会从这许多的志愿者中被选出。2. 我们介绍了



图表 3.多任务学习的概念性示意图。通过与第二任务做结合，以输入图像到目标位姿的路径比单任务学习更接近于参考标准路径。

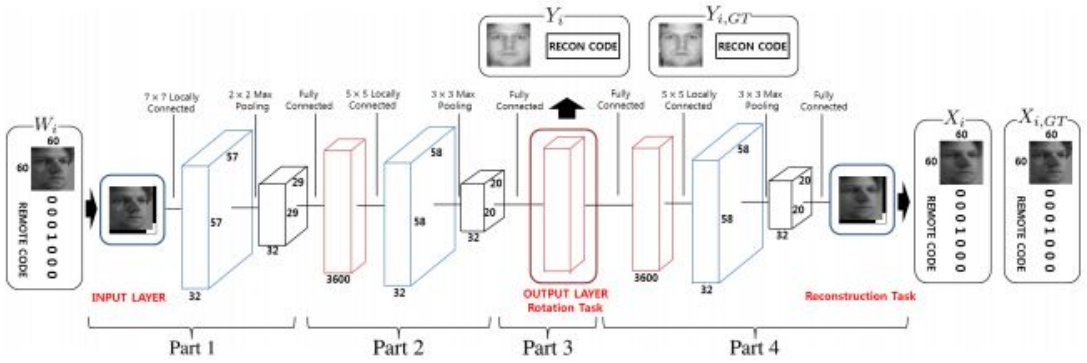
一种新类型的多任务学习策略，这个策略可以更加深刻地提升深度神经网络的特征保持能力。3.我们得到一个比深度学习特征保持脸空间更好的面部识别率。多视角感知机在多视点组成中运用了所有的合成图像并且在多种面部识别结果中进行选择。

这篇论文之后的部分是按如下的结构组织的。在第二部分将会对之前关于面部识别和多任务学习的研究进行解释。关于我们模型的描述和我们在设计时的关注点也将在第三部分进行解释。模型的参数将在第四部分进行描述。第五部分描述了各种各样用于证明模型有力性的实验，紧随其后的第六部分概括了结论。

2. 相关工作

面部识别

在过去的二十年中，一些具有典型手工特征的方法，例如LBP，SIFT和CABOR已经被用于面部识别的任务中。最近，通过位姿来进行的面部识别及证明已经成为了主要问题。这些研究被分为3D 理论研究和 2D 理论研究。对于 3D 理论，通过 3D 位姿标准化进行的全自动位姿不变的面部识别方法通过 3D 模型和标志性特征点来将非正面的图像旋转到正面。基于形变位移场的姿态人脸识别图像匹配通过运用筛选器和3D 模型产生了位姿鲁棒性的特征。另一方面，2D 理论在没有 3D 信息的情况下获得了位姿不变特性。通过用一个图像加权和来代替被测试图像，交叉姿态人脸识别的耦合偏差-方差权衡运用这些加权和来作为位姿不变特性。深度神经网络已经用于在没有手工制作特征的情况下发现位姿鲁棒性特征。深度学习特征保持脸空间通过运用 CNN 和直接运用他们的输出图像来作为位姿不变特性将各种各样的位姿图片变成了正面拍摄的图片。这种理论同样在通过一步步调整位姿来最小化各种各样的因素的影响后被运用在将位姿叠进自动编码器（空间）的人脸识别中。多视图感知：学习人脸识别与视图表示的深层模型提出了一个多视角感知器(MVP)，这个感知器可以通过利用随即隐藏的神经元来识别身份和位姿。



图表 4. 我们的 DNN 模型包含四个主要部分：特征提取，特征旋转，成像和重建任务。其中重建任务是辅助任务。每个部分由局部连接层、最大汇聚层和全连接层组成。在第三部分中，红盒子表示生成目标图像的输出层。

多任务下的深度学习

最近，许多的深度神经网络结构已经通过多任务学习模型改进了部分计算机视觉任务的表现。因为全局权重可以从多样化的任务中获得特征，所以为了获得全局的权重，R. Collobert and J. Weston 迭代地在各个不同任务的每个训练集上训练了单任务模型。一个深度神经网络在前面的部分中有共担层次，之后的分开的层次用于执行不同的任务，而不是将所有的权重平均分配。然而我们的模型和第二任务共同分担了主任务的所有层次。

3. 模型描述

我们所构建的模型的两个重要目的分别是根据远程代码川建一个新的位姿图像，和在位姿改变的情况下保持输入图像的特征。我们的模型是为了在这些目的中得到超常的表现而小心设计的。图表 4 描述了网络的最终设计。我们的模型使用了一张 $M \in R^{N\times N}$ 的图像和一段 $C \in \{0,1\}^{2N+1}$ 的代码作为输入 $W \in R^{(N+1)\times(N+1)}$，并定义为：

$$W_{(x,y)} = \begin{cases} M_{(x,y)}, if\, 1 \le x, y \le N \\ C^{N+1-x+y}, otherwise \end{cases}$$

其中，(x,y) 和 C^i 代表了像素坐标 (x,y) 和远程代码 C 的第 j 位。正如图表 4 中所示，远程代码包围着输入图片，目的是为了做出一个正方形的输入集。实验显示，连接远程代码和图像的方法并不能对提高性能表现起到作用。

之前的许多工作已经高效地利用了 CNN 来训练 DNN 模型对于图片的作用。CNN 在所有图片之上公用筛选器然而对于连接了图片的远程代码而言，他们均分了权重，所以提供这样的筛选器已经是不够合适的了。因为这个原因，我们为第一部分运用局部连通无权重均摊的层。对于第二部分，我们

使用完全连通层来改变特征以包含远程代码所代表的目标位姿信息。局部连通层和汇聚层被应用于完全连接层，以使功能能更有效地包含位姿信息和保留身份特征。第二部分之后，输出层，由完全连接层构成，其功能是构建新姿态图像。此外，附加任务部分的新颖的元素附加在第三部分之后。第四部分的详细的解释将会包含在 3.2 部分。

这一系列的参数被作为输入(61*61)-L(7,32)-P(2,2)-FC(3600)-L(5,32)-P(3,3)-FC(3729)-FC(3600)-L(5,32)-P(3,3)-FC(3721)，其中 L,P,FC 分别作为局部连通层，汇聚层和全连通层。L(7,32)，P(2,2)和 FC(3600)分别代表着这一层提供了 32 个尺寸为 7 且无权重分享的筛选器，最大汇聚层的大小是 2，步长为 2，和 3600 个神经元的完全连接层。FC(3729)是输出层，它产生出了目标位姿和包含着输入图像信息的代码。除此之外，FC(3721)表示一个第二任务层，他就像输入层一样重建了一个输入图片和远程代码。局部连通层和全连通层使用了一种热鲁激活功能，整个局部连通层的步长为 1，而所有的汇聚层的步长被设置为和筛选器大小一致的大小。然而，输出层和最底层包含了无需校正的线性激活函数参数设置可以基于输入图像大小和目标位姿的数量来进行自由地变换。上述参数设置是为了 60*60 的输入图像和 7 个位姿所设计的，这些都在第 5.2.1 部分中进行了描述。

3.1 远程代码

我们在输入层和输出层使用了两种特殊的代码来控制改变输入图片的位姿。在输入层的代码被称为远程代码 $C_i, i=1,2,...,\mathrm{n}$，用于训练让输入图片能够从所给的 n 个位姿改变成第 i 个位姿，并保持相同的身份特征。远程代码是一种简单的重现代码，$C_i \in \{1,0\}^l$ 中总长度 1 被定义为

$$C_i^j = \begin{cases} 1, if\,(i-1)\times k < j \le i\times k \\ 0, otherwise \end{cases},$$

（2）

其中 C_i^j 为代码 C_i 的第 j 位，$k=\lfloor l/n \rfloor$。(l,n) 在第 5.2.1 和第 5.2.2 部分中的实验里被分别置为(121,7)和(65,9)。由于输出层用来从无数亮度条件下的图像中产生正面光照下目标位姿图像，所以我们在输入层代码中不需要光照信息。辅助深度神经网络以主深度神经网络输出层作为开始，然而他所需要的不仅仅是位姿的信息，而且有输入图像的亮度条件的信息来重建输入图像，我们设置了输出层的代码，被称为重构代码，$\{Q_i, S_i\}, i=1,2,...,n, t=1,2,...,m.$，它代表着输入图像在 n 个位姿中的第 i 个位姿在 m 个光照变化中的第 t 种光照

条件的代码。与远程代码相似，我们设置了姿态码 $\mathrm{Q}_i \in \{0,1\}^l$，他们被定义为

$$Q_i^j = \begin{cases} 1, if\,(i-1)\times k < j \le i \times k \\ 0, otherwise \end{cases},$$

（3）

其中 Q_i^j 是代码 Q_i 的第 j 位并且 $k = \lfloor l/n \rfloor$。(l,n) 在第 5.2.1 和第 5.2.2 部分中的实验里被分别置为 $(49,7)$ 和 $(72,9)$。此外，包含了总长度 1 的光照代码 $S_t \in \{0,1\}^l$ 被定义为：

$$Q_t^j = \begin{cases} 1, if\,(\mathrm{t}-1)\times k < j \le t \times k \\ 0, otherwise \end{cases},$$

（4）

其中 Q_t^j 是代码 Q_t 的第 j 位并且 $k = \lfloor l/\mathrm{m} \rfloor$。$(l,\mathrm{m})$ 在第 5.2.1 和第 5.2.2 部分中的实验里被分别置为 $(80,20)$ 和 $(60,20)$。

最后，我们可以定义训练数据集。由于我们可以为一张图象配置 n 个远程代码，所以数据集将是原始数据集的 n 倍大。我们可以设置训练数据集，将每个图像的输入输出作为一对 $M, L = \{\{M, C_i\}, \{\mathrm{M}_i, Q_j, \mathrm{S}_t\}\}$，其中 $i = 1,2,...,n$，M_i 待变第 i 个正面光照下与 M 具有相同身份特征的位姿图像。Q_j 和 S_t 分别代表图像 M 的位姿和光照代码。

3.2 多任务学习

我们使用了图表 4 中所描述的多任务学习模型。虽然我们的模型的主要任务是重建新的位姿图像，但是我们还附加了一个第二任务，即重建输入图像，也就是在第一个任务模型中保持输入的身份的同时旋转输入图像。我们将 L2 的平方规范作为两个任务的成本函数。对于第一个任务输出层，即新的位姿图片和重构代码，的成本函数被定义为：

$$E_c = \sum_{i=1}^{N} \| Y_{i,GT} - Y_i \|_2^2, \qquad (5)$$

其中 $Y_{i,GT}$ 和 Y_i 是地面实况和所产生的分别包含改变的姿态图像和输入图像的姿态和照明信息的输出。此外，i 和 N 分别代表培训投入指数和总批次大小。

第二任务的成本函数，即重建输入图像和远程代码，被定义为

$$E_r = \sum_{i=1}^{N} \| \mathrm{X}_{i,GT} - X_i \|_2^2, \qquad (6)$$

其中 $\mathrm{X}_{i,GT}$ 和 X_i 分别是地面实况和所构造的包含输入图像和远程代码的输出。我们最后的成本函数是第一任务和第二任务的加权和。

$$E = \lambda_c E_c + \lambda_r E_r, \qquad (7)$$

其中 λ_c 和 λ_r 分别是第一任务和第二任务的权重。我们假设两个任务具有相同的重要性，那么这两个是将在

所有实验中被同时置为 1.

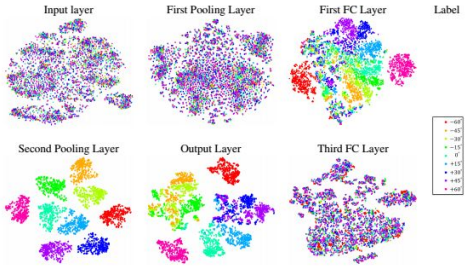
4. 训练

我们所有的实验都用到了 CUDA ConvNet，即最流行的深度神经网络工具箱之一。我们可以控制一些参数，其中包括初始权重(iniW)，权重学习率和偏差学习率(epsW,epsB)，权重和偏差动量(monW, momB),和 L2 权重衰减(wc)。对于所有的实验设置，我们使用相同的参数。对于本地连接层和完全连接层，我们将 iniW 分别设置为 0.001 和 0.01.此外，对于除了第一局部连通层，我们将 epsW, epsB, momW, momB, 和 wc 分别设置为 0.0001，0.0002，0.9，0.9 和 0.04.为了第一局部连通层，我们将 epsW 设置为 0.001,将 epsB 设置为 0.02.我们使用小批量梯度下降和反向传播来训练我们的模型，其中批量大小为 100.

为了要获得的输入和输出的训练集，我们进行了图像集而不是远程代码的 2 个预处理步骤。首先，为了能对光照变化有鲁棒性，每个图像被减去，并除以每个图像的平均值和方差。其次，我们也在计算过训练图像减去了每个像素的平均值的数据后再根据像素的方差分开。

5. 实验

实验部分由四部分组成，以证明我们的模型的强度。第 5.1 部分显示了每



图表 5.从多 PIE 数据集的每一层提取的测试图像的 6000 个特征构成的特征空间。具有同样颜色的每一个点待变具有相同远程代码的输入集中的特征。例如，红色圆点是包含着代表 60 度的远程代码的特征。

个层的特征空间，分析了如何在这个深度体系结构中旋转的输入面图像。我们用 T-SNE 方法，即将高维空间转换到低位空间的著名方法之一，5.2 部分包含了与国家的最先进的程序，以证明的能力来保护身份的人脸识别实验的结果。我们精心设计了一个模型来构建一个目标图像，并在保持他的特征同时在需要执行的两项任务中认真表现。因此，我们设计了一个实验来证明我们的模型的有效性。在 5.3 部分我们比较了我们的多任务模型和单任务模型。最后，在 5.4 部分，我们设计了一个实验来说明在最开始放置一个完全连通层的优点。

5.1 特征空间

正如在图表 5 中所示，在第一汇聚层的特点和在输入层的特点以类似的方式混合在一起。这表明第一局部连通层和汇聚层将输入图像的有用信息进行了提取，而不仅仅是改变位姿。

具有相同远程代码的特征插入到了输入层之中并开始和第一完全连通层的特征合并。然而，一些不同颜色的圆点相互混杂表明仅仅依靠一层完全连通层并不能够完美的改变位姿。局部连通层和汇聚层在完全连通层之后清楚地联系在了一起来对改变位姿的目标作出努力。正如图表 5 所示，在第二汇聚层，特征已经被完美的与其他颜色区分开来。由于输出层是作为一个完全连通层，他操作的目标是为了将特征对应到目标位姿图片中，所以特征与和那些相似位姿的特征混在了一起。例如，-45 度，-30 度和-15 度的图像很相似。从第二汇聚层提取出的特征比那些从其他层提取出的特称在面部识别任务中具有更好的表现，载入表格 3 所示。

5.2 面部识别

为了说明我们的模型是如何保持了输入图像的特征，我们在多 PIE 数据集上进行了试验。我们为实验准备了两个实验设置：设置 1，我们仅仅在多 PIE 数据集中使用第一部分的图像，这一部分包含了 249 项。有 100 项（ID 从 001 到 100）在 7 个位姿和 20 中光照，他们本用来训练这个模型去分析人脸。在训练之后，我们选择了剩下的 149 项（ID 从 101 到 250，除去 213）在 6 种位姿和 19 种光照（ID01~20 除去正面光照，ID07）下作为探针测试。对于画廊的图像，每一个有照明的正面图像将被使用。因此，14000 的图像被用于训练，16986 个图像被用于测试。设置 2，我们准备了更大数据规模的多 PIE 数据集。我们使用 200 项（ID 为 001~200）在 9 种位姿 20 种光照下的情况作为训练。作为测试，我们使用剩下的 139 项在 9 种位姿 20 种光照条件下的项目。一共 137*9*20 个图像。对画廊图像的选择过程与设置 1 相同。对于测试步骤，我们从输出层中提取被称为被控位姿图像（CPI）的特点，在图表 4 中他是一个红色盒子。此外，我们从在输出层之前的第二汇聚层中提取的特征被称为控制位姿特征（CPF）。为了评估所有的誓言都是用耳机距离标准比较测试图像和画廊图

	−45°	−30°	−15°	+15°	+30°	+45°	*Avg*
Li[11]	63.5	69.3	79.7	75.6	71.6	54.6	69.3
Z.Zhu[26]	67.1	74.6	86.1	83.3	75.3	61.8	74.7
CPI	66.6	78.0	87.3	85.5	75.8	62.3	75.9
CPF	**73.0**	**81.7**	**89.4**	**89.5**	**80.4**	**70.3**	**80.7**

表格 1. 在设置 1 的条件下对于不同位姿的识别率。最好的结果在表格中用黑色粗体字标出。

	00	01	02	03	04	05	06
Li[11]	51.5	49.2	55.7	62.7	79.5	88.3	**97.5**
Z.Zhu[26]	**72.8**	**75.8**	75.8	75.7	75.7	75.7	75.7
CPI	66.0	62.6	69.6	73.0	79.1	84.5	86.6
CPF	59.7	70.6	**76.3**	**79.1**	**85.1**	**89.4**	91.3
	08	09	10	11	12	13	14
Li[11]	**97.7**	**91.0**	79.0	64.8	54.3	47.7	67.3
Z.Zhu[26]	75.7	75.7	75.7	75.7	75.7	**75.7**	73.4
CPI	86.5	84.2	80.2	76.0	70.8	65.7	76.1
CPF	92.3	90.6	**86.5**	**81.2**	**77.5**	72.8	**82.3**
	15	16	17	18	19	*Avg*	
Li[11]	67.7	75.5	69.5	67.3	50.8	69.3	
Z.Zhu[26]	73.4	73.4	73.4	72.9	**72.9**	74.7	
CPI	78.2	80.7	79.4	77.3	65.4	75.9	
CPF	**84.2**	**86.5**	**85.9**	**82.9**	59.2	**80.7**	

表格 2. 在设置 1 的条件下对于不同亮度的识别率。最好的结果在表格中用黑色粗体字标出。

像。由于我们的模型可以创造不同的姿势图像，我们创造 n，即训练位姿的个数（设置 1 为 7，设置 2 为 9），个图象，每个探针图像 PI（i，2，。。。，n）个。除此之外，我们制成了 n 个图像，每个画廊图像 G_i^j 个。对于每一个 i，以下的等式：

$$\min \| \mathrm{P}_i - G_i^j \|_2^2, \qquad (8)$$

的计算结果将会被相加。每一个 i 的结果将会作为最终结果的一部分。

5.2.1 设置 1 的结果：包含 7 种位姿

在这个设置中，我们使用了 60*60 大小的图片作为输入，正如在图表 4 中所描述的一样。我们比较了我们的结果和当前发展水平的结果，以及交叉姿态人脸识别的耦合偏差-方差权衡的结果。对于不同位姿的识别率的结果已经展示在了表格 1 中.有人类判断作为对比，我们的模型发现在 45 度到-45 度之间，要想从大量的旋转图片中想象出面部特征是非常困难的。然而，表格 1 显示 CPI 和 CPF 在许多位姿上都比当前发展水平表现得更好。表格 2 展示了 20 种不同光照条件下的识别率。由于我们在除了正面光照（ID 为 07）外的 19 种光照条件下进行测试，所以只有 19 个可显示的结果。CPF 在 19 个部分的表现比其他 12 种方法都要更好。

5.2.2 设置 2 的结果：包含 9 个位姿

由于当前的发展水平采用了 32*32 大小的图片作为训练和测试，我们准备了同样的设置条件。将图片的大小改变后，我们就需要设置不同的参数。

	−60°	−45°	−30°	−15°	0°	+15°	+30°	+45°	+60°	*Avg*
Landmark LBP[3]	35.5	52.8	71.4	83.9	94.9	82.9	68.2	48.3	32.1	63.2
FIP+LDA[26]	49.3	66.1	78.9	91.4	94.3	90.0	82.5	62.0	42.5	72.9
RL+LDA[26]	44.6	63.6	77.5	90.5	94.3	89.8	80.0	59.5	38.9	70.8
MTL+RL+LDA[27]	51.5	70.4	80.1	91.7	93.8	89.6	83.3	63.8	50.2	74.8
Z.Zhu+LDA[27]	60.2	75.2	83.4	93.3	95.7	92.2	83.9	70.6	60.0	79.3
CPI	55.8	71.8	80.0	90.1	98.4	90.2	82.7	71.0	52.9	77.0
CPF	**63.2**	**80.4**	**88.1**	**94.5**	**99.5**	**95.4**	**88.9**	**79.4**	**60.6**	**83.3**
CPF-FC1600	45.4	72.7	80.8	885	96.8	90.3	79.6	70.22	42.5	74.1
CPF-Pool1	9.7	39.1	51.6	69.9	92.5	70.8	51.1	39.4	9.3	48.1

表格 3. 在设置 2 调价下不同位姿的识别率。CPF-FC1600 和 CPF-Poo11 分别表示从第一个 FC(1600）层和第一个汇聚层中提取的特征。最好的结果已经用黑色粗体字标出。

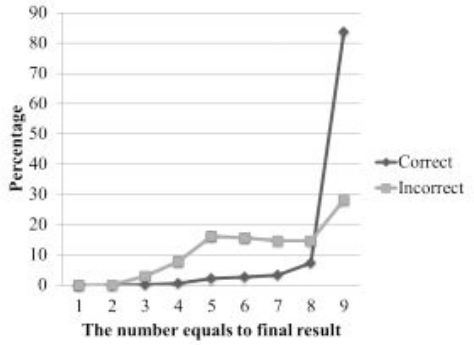
输入改变为这样的公式：（33*33）-L(5,16)-P(2,2)-FC(1600)-L(5,16)-P(2,2)-FC(1156)-FC(1600)-L(7,16)-P(2,2)-FC(1089)。我们将我们得到的结果与某些特征进行了对比，结果放在了【27】。所有之前的设置都使用了 LDA 来减小特征的损失。如表格 3 所示，CPF 在所有不同的位姿条件下都比其他的所有方法要表现得更好。从不同的层次中提取的特征将得到不同的结果。因为输出层将高级特征转变进了输出图像中，一些对于辨别面部特征有用的判别特征将会在输出图像中被丢掉。这就是为什么在输出层之前的高级特征，CPF，表现的最好。我们的模型在 0 度的条件下获得了标志性的表现。因为我们的模型表现出了

99.5%的书识别率，所以它清楚地表现出了比当前发展水平下的算法更好的表现，而所报道的当前发展水平的算法在 0 度条件下的识别率为 99%。事实上，我们的理论下的识别率是在 2740 个图片之中有 14 个错误分类。

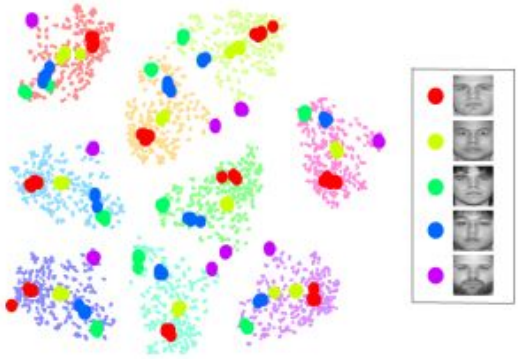
虽然上述的最终的结果都是由每一个 CPF 产生的结果中投票产生的，但是其中最正确的结果是由大量具有高置信度的选择选中的，如表格 6 所示。这一结果表明，在目标姿态的人脸图像的基础上合成了相当一致的结果，并且提出了深度神经网络可以通过位姿来生成高质量的多视角人脸图像。除此之外，如图表 7 所示，所提出的模型可以在改变位姿的情况下同样保存身份特征。

5.3 与单任务模型的对比

我们构造了一个新的实验来说明在输出层之后附加重建任务层的有效性。我们准备了两个模型，多任务模型和图表 4 中所示相同，单任务模型就像是只有第一个任务的第一个模型。因为 CPF 超越了 CPI，所以我们给每个模型都采用了 CPF 并将条件设置为设置 1。两个模型对于不同位姿和不同亮度条件下的识别率分别在表格 4 和表格 5 中展示。对于所有的位姿和亮度设置，



图表 6. CPF 对最终结果所作出的贡献的百分比。大多数的错误结果是由于低置信度产生的。例如 9 个 CPF 中的 5 个进行了投票。另一方面，最正确的结果是从高置信度的情况下产生的。因此我们可以推断，每个 CPF 都有能力来保持身份。



图表 7. 设置 2 第二汇聚层的 6000 个特征的特征空间。每个淡点的颜色代表了一种不同的远程代码。54 个具有相同深色的点代表了同一个身份的特征。在特征空间里，不仅是特征靠着远程代码聚集统一，深色的点也在相同身份的不同位姿下聚集统一。

多任务模型要比单任务模型更好。多任务模型的第一个任务是创建一个所用的远程代码所代表的目标位姿图像，这个目标和单任务模型是相同的。然而，多任务模型的第二个任务是要根据输出图像的特征重建输入图像和

对应的远程代码。由于多任务模型的输出层必须要包含身份保持的特征以便于在第二任务中重建输入层，所以多任务模型能够比单任务模型更加有效的保留身份特征。

5.4 早期 FC 层的有效性

大多数的深度神经网络由两大部分组成，即特征提取和图特征组合。因为大比例的输入大小很难处理，所以卷积层或局部连接层通常用于特征提取的网络的开始。除此之外，完全连通层用于在后面进行特征结合。然而，正如在图表 4 中所述，我们的模型在最开始就用到了一个完全连通层。为了检测一个早期完全连通层模型(EFC)的有效性，我们构造了一个在设置 2 下的实验。该模型在 5.2.2 部分已经进行了描述。我们设计了另一个模型(LFC)，在这个模型中，完全连通层恰巧位于输出层之前，而不是在一开始。整个实验的参数设置被定义为如下输入：

(33*33)-L(5,16)-P(2,2)-FC(1600)-FC(1156)-FC(1600)-L(7,16)-P(2,2)-FC(1089).除了完全连通层，即第一个 FC(1600)层次，的位置，其他的所有参数都相同。两种模型的识别率在表格 6 中被标出。我们也包含了每个模型在使用 CPI 和 CPF 时的结果。CPF-LFC 和 CPI-LFC 是分别从 FC(1600)和 FC(1156)中提取出来的。

因为局部连通层筛选器在某些部分起作用，所以只有完全连通层在全局起作用。于是，在完全连通层，远程代码就开始将改变特征，将远程代码代表的目标位姿包含进特征之中。因此在早期完全连通层模型中包含有目标位姿的特征比晚期完全连通层模型。正如结果所显示，虽然 CPI-LFC 的结果比其他 CPI-EFC 要好，但是早期完全连通层在具有最佳性能特点 CPF 的情况下要比晚期完全连通层更好。

	-45°	-30°	-15°	$+15^{\circ}$	$+30^{\circ}$	$+45^{\circ}$	*Avg*
Single	65.4	76.5	85.9	85.8	76.3	63.2	75.5
Multi	**73.0**	**81.7**	**89.4**	**89.5**	**80.4**	**70.3**	**80.7**

表格 4.在设置 1 条件下，单任务模型与多任务模型对不同位姿的识别率。最好的结果在表格中用黑色粗体字标出。

	00	01	02	03	04	05	06
Single	45.4	64.3	72.9	74.9	82.0	86.9	89.8
Multi	**59.7**	**70.6**	**76.3**	**79.1**	**85.1**	**89.4**	**91.3**
	08	09	10	11	12	13	14
Single	89.7	87.9	81.7	76.5	72.2	66.7	76.9
Multi	**92.3**	**90.6**	**86.5**	**81.2**	**77.5**	**72.8**	**82.3**
	15	16	17	18	19	*Avg*	
Single	80.9	82.7	79.9	76.5	47.1	75.5	
Multi	**84.2**	**86.5**	**85.9**	**82.9**	**59.2**	**80.7**	

表格 5.在设置 1 条件下，单任务模型与多任务模型对不同光照的识别率。最好的结果在表格中用黑色粗体字标出。

	CPI-EFC	CPF-EFC	CPI-LFC	CPF-LFC
Result	77.0	**83.3**	78.3	79.7

表格 6.我们的模型，即一开始具有 FC 层，与晚期 FC 模型，即在一开始没有 FC 层，的识别率。最好的结果在表格中用黑色粗体字标出。

6. 结论

在这篇论我莪们，我们提出了一种新的多任务网络，这种网络可以通过分析用户的远程代码代表的信息来合成想要的位姿和正面光照下的面部图像。通过将一个重建输入图像的第二任务接在第一任务，即将一张输入图像旋转到一个确定的位姿，之后，我们所以出的多任务网络得到一个能够比单任务模型更好保持身份特征的结果。第二汇聚层的变量在第一人物模型中可以被用做位姿不变和光照不变特性。在任意位置和光照条件下的面部识别任务中，我们的模型完全可以超过当前发展水平的模型至少 4%~6%。

特此鸣谢

该项工作由三星尖端科技研究所（SAIT）以及技术创新计划，10045252，机器人任务智能技术的发展所支持，由贸易工业能源部（MOTIE，韩国）赞助。此外，这项研究也同样受到了韩国国家基金研究会（NRF）的支持和韩国政府(MEST) (2014-003140)以及(MSIP)(2010-0028680)的赞助。

注释

[1] T. Ahonen, A. Hadid, and M. Pietikainen. 人脸描述与局部二值模式：应用人脸识别. 模式识别与机器智能, IEEE Transactions on, 28(12):2037－2041, 2006. 3

[2] A. Asthana, T. K. Marks, M. J. Jones, K. H. Tieu, and M. Rohith. 三维姿态归一化全自动姿态不变人脸识别. 计算机视觉领域 (ICCV), 2011 IEEE International Conference on, pages 937－944. IEEE, 2011. 3

[3] D. Chen, X. Cao, F. Wen, and J. Sun. 维数灾难：高维特征及其有效压缩. 计算机视觉与模式识别领域 (CVPR), 2013 IEEE Conference on, pages 3025－3032. IEEE, 2013. 7

[4] R. Collobert and J. Weston. 自然语言处理的统一架构：多任务学习的深度神经网络。第二十五届机器学习国际会议的会议, pages 160－167. ACM, 2008. 3

[5] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. 多饼图像与视觉计算, 28(5):807－813, 2010. 2, 6

[6] G. B. Huang, M. Mattar, T. Berg, E. Learned-Miller, et al. 在野外贴标签：一个在无约束环境中学习人脸识别的数据库，在“现实生活”图像中的人脸检测、定位和识别方面的研究, 2008. 1

[7] M. Kan, S. Shan, H. Chang, and X. Chen. 叠进自动编码器（空间）的人脸识别的姿势. In Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, pages 1883－1890. IEEE, 2014. 3

[8] A. Krizhevsky. cuda-convnet. http://code.google.com/p/cuda-co

nvnet/. July 2012. 5

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton. 深卷积神经网络的图像网分类. In F. Pereira, C. Burges, L. Bottou, and K. Weinberger, editors, Advances in Neural InformationProcessingSystems25,pages1097–1105.Curran Associates, Inc., 2012. 3, 4

[10] Y.LeCun,B.Boser,J.S.Denker,D.Henderson,R.E. Howard, W. Hubbard, and L. D. Jackel. 反向传播算法在手写压缩编码识别中的应用. Neural computation, 1(4):541–551, 1989. 5

[11] A. Li, S. Shan, and W. Gao. 交叉姿态人脸识别的耦合偏差-方差权衡, IEEE Transactions on, 21(1):305–315, 2012. 3, 6

[12] S. Li, X. Liu, X. Chai, H. Zhang, S. Lao, and S. Shan. 基于形变位移场的姿态人脸识别的图像匹配. In Computer Vision– ECCV 2012, pages 102–115. Springer, 2012. 3

[13] S. Li, Z.-Q. Liu, and A. B. Chan. 深卷积神经网络的异构多任务学习. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on, pages 488–495. IEEE, 2014. 2, 3

[14] C. Liu and H. Wechsler. 使用增强的Fisher线性判别模型的人脸识别基于Gabor特征的分类. Image processing, IEEE Transactions on, 11(4):467–476, 2002. 3

[15] D. G. Lowe. 独特的形象特征的尺度不变特征点. Internationaljournalofcomputer vision, 60(2):91–110, 2004. 3

[16] K. Simonyan and A. Zisserman. 用于大规模图像识别的非常深卷积网络. arXiv preprint arXiv:1409.1556, 2014. 3

[17] Y.Sun,X.Wang,andX.Tang. 通过联合识别验证的深度学习人脸表示. CoRR, abs/1406.4773, 2014. 1

[18] Y.Taigman,M.Yang,M.Ranzato,andL. Wolf. Deepface: Closing the gap to human-level performance in face verification.【deepface:在人脸检测中缩小与人类水平的差距。】. In Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, pages 1701–1708. IEEE, 2014. 1

[19] L. Van der Maaten and G. Hinton. 使用T-SNE可视化数据. Journal of Machine Learning Research, 9(2579-2605):85, 2008. 5

[20] D. Yi, Z. Lei, and S. Z. Li. 姿态鲁棒的人脸识别. InComputerVisionandPatternRecognition (CVPR), 2013 IEEE Conference on, pages 3539– 3545. IEEE, 2013. 3

[21] X.-T. Yuan, X. Liu, and S. Yan. 多任务联合稀疏表示的视觉分类. Image Processing, IEEE Transactions on, 21(10):4349-4360, 2012. 3

[22] C.ZhangandZ.Zhang. 通过多任务的深度卷积神经网络改进的多视角人脸检测 In Applications of Computer Vision (WACV), 2014 IEEE Winter Conference on, pages 1036-1041. IEEE, 2014. 3

[23] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja. 结构化多任务稀疏学习的鲁棒视觉跟踪 International journal of computer vision, 101(2):367-383, 2013. 3

[24] X. Zhang, Y. Gao, and M. K. Leung. 从正面和侧面视图识别旋转面：对档案数据库的有效利用方法. Information Forensics and Security, IEEE Transactions on, 3(4):684-697, 2008. 3

[25] Z.Zhang,P.Luo,C.C.Loy,andX.Tang. 深度多任务学习的人脸标志检测. In Computer Vision-ECCV 2014, pages 94-108. Springer, 2014. 3

[26] Z. Zhu, P. Luo, X. Wang, and X. Tang. 深度学习身份保留的脸空间. In Computer Vision(ICCV),2013IEEEInternationalConferenceon, pages 113-120. IEEE, 2013. 1, 2, 3, 6, 7

[27] Z. Zhu, P. Luo, X. Wang, and X. Tang. 多视图感知：学习人脸识别与视图表示的深层模型. In Advances in Neural Information Processing Systems, pages 217-225, 2014.