

硕 士 研 究 生 读 书 报 告



题目 读书报告

作者姓名 孟祥宝

作者学号 21851425

指导教师 李启雷

学科专业 软件工程

所在学院 软件学院

提交日期 二○一八年十二月

Report of Hand Gesture Recognition Using Acoustic Signals

A Dissertation Submitted to

Zhejiang University

in partial fulfillment of the requirements for

the degree of

Master of Engineering

Major Subject: Software Engineering

Advisor: Li Qilei

By

Meng Xiangbao

Zhejiang University, P.R. China

2018

摘要

根据侧面照片合成正面人脸一直是个难题，现在，由中科院自动化所（CASIA）、中科院大学和南昌大学的 Rui Huang、Shu Zhang、Tianyu Li、Ran He 合作的一项研究，首次解决了这一个问题，他们受人类视觉识别过程启发，结合对抗生成网络（GAN）的强大性能，提出了一个双路径 GAN（TP-GAN），能够在关注整体结构的同时，处理人脸面部细节，在不同的角度、光照条件都取得了很好的结果。不仅如此，这种方法还能够处理大量不同姿势的照片。

**关键词**：图像识别， 图像合成，对抗神经网络

Abstract

Synthesizing front faces based on side photos has always been a problem. Now, a study by the Institute of Automation of the Chinese Academy of Sciences (CASIA), the Chinese Academy of Sciences and Rui Huang, Shu Zhang, Tianyu Li, and Ran He of Nanchang University solved this problem for the first time. Inspired by the human visual recognition process, combined with the powerful performance of the anti-generation network (GAN), they proposed a dual-path GAN (TP-GAN) that can handle facial details at different angles while focusing on the overall structure. The lighting conditions have achieved good results. Not only that, but it also handles a large number of photos in different poses.

**Keywords：**image recognition, image synthesis, GAN

1论文主要内容

论文作者第一次成功使用合成的人脸图像进行图像识别任务。作者在论文中写道，他们的这项工作主要贡献在于三个方面：

1）提出了一个像人类一样能够考虑整体和局部信息的 GAN 结构，能够根据单一的图像合成正面人脸视图，合成的图像非常逼真且很好地保留了身份特征，而且可以应对大量不同的姿势。

2）将从数据分布（对抗训练）得来的先验知识，和人脸领域知识（对称性、身份保留损失）结合起来，将从三维物体投射到二维图像空间时固有的缺失信息精确地恢复了出来。

3）展示了一个“通过生成进行识别”（recognition via generation）的框架的可能性，并且在大量不同姿势下取得了目前最好的识别结果。

**2 图像识别现状**

虽然计算机识别图像已经在多个基准数据集中超越了人类，但真实应用场景中，对于不同姿势的识别问题仍然没有得到很好地解决。

现有方法可以分为两类，一类是采用手绘的（hand-crafted）特征或学习不同姿态的特征，另一类则是致力于在大量不同姿态的人脸中获取（recover）一个正面人脸视图，然后用这个视图进行人脸识别。

但是，第一类方法由于要在不变和可识别之间做出权衡，往往无法有效处理大量不同的姿势。

第二类方法，早期的尝试是先将二维图像与通用或有确切身份的3D模型对齐，然后利用三维几何变换渲染正面人脸视图。但是，这种方法遇到大量不同姿势的图像时，纹理损失严重，性能也不好。

近来，有研究者提出了由数据驱动的深度学习方法，让系统在学习估计正面视图的同时，分辨身份和姿势表征。虽然结果喜人，但合成的图像在细节方面有所欠缺，再一次地，这种方法也无法很好应对大量不同的姿势，因此合成的图像也无法用于法医取证或属性估计。

更重要的是，从优化的角度看，从观察到的不完全侧面脸部恢复正面视图，本身就是一个不合理而且也没有很好定义的问题。恢复的图像的质量严重依赖于训练过程中的先验或约束条件。

以往的方法通常采用配对监督学习的方式（pairwise supervision），极少在训练过程中引入约束条件（constraints），因此，才导致合成的图像模糊不清。。

**3 TP-GAN**

当人类在进行视觉合成的时候，我们首先是通过观察到的侧脸，在以往的经验/知识基础上，推测出整张脸的结构（或草图）。然后，我们会将注意力转向脸部的细节，比如眼睛、鼻子、嘴唇，将这些细节在刚才那张草图上“填满”。

受此启发，作者提出了一个有两条路径的深度架构（TP-GAN），用于正面人脸图像合成。这两条路径，一条专注于推理全局结构，另一条则推理局部的纹理，分别得到两个特征地图。这两个特征图会融合在一起，用于接下来的最终合成。

不仅如此，作者还将正面人脸分布的信息并入了一个生成对抗网络（GAN），由此对恢复过程进行了很好的约束。

GAN 在二维数据分布建模方面的卓越性能（capacity）极大地改善了很多不合理的低级视觉问题，比如超分辨率和修复（inpainting）。

**4 图像合成**

作者还根据人脸是对称结构这一点，提出了一个对称性损失（symmetry loss），用于补全被遮挡住的部分。

为了忠实地保留一个人脸部最突出的特征，作者在压缩特征空间中除了像素级别的 L1 loss，还使用了一个感知损失（perceptual loss）。

最后，关键一环，将身份保留损失（identity preserving loss）整合进来，实现忠实的正面脸部合成，图像质量得到大幅提升。

在不同的光线条件下合成的结果。上面一行是合成结果，下面一行是原始照片。

作者指出，这些图像有可能用于人脸分析的任务。

**5 总结**

使用单一脸部图像合成逼真的正面脸部视图在人脸识别领域中有着广泛的应用。尽管此前有研究试图从大量面部数据中寻求解决方案，也即数据驱动的深度学习方法，但这个问题仍然具有挑战性，因为它本质上是个不合理的问题（ill-posed）。

本文提出了双通道生成对抗网络（Two-Pathway Generative Adversarial Network，TP-GAN），通过同时感知全局结构和局部细节，合成逼真的正面人脸视图。

除了常用的全局编码器-解码器网络之外，论文还提出了４个定位标记的补丁网络（landmark located patch networks）处理局部纹理。除了全新的架构，我们将这个不合理的问题进行了很好的转化，引入了对抗性损失（adversarial loss）、对称性损失（symmetry loss）和身份保留损失（identity preserving loss）的组合。这一损失的组合能够利用正面脸部的分布和预训练识别深度脸部模型（pre-trained discriminative deep face models），指导身份保留推理从正面脸部视图合成侧面照。不同于以往的深度学习模型主要依靠中间特征用于识别的方法，我们的方法直接利用合成的、保留身份的图像用于下游任务，比如人脸识别和归因估计。实验结果表明，我们的方法不仅在视觉上令人信服，也在多种人脸识别中超越了现有最佳方法。

参考文献

[1] Huang, Rui et al. “Beyond Face Rotation: Global and Local Perception GAN for Photorealistic and Identity Preserving Frontal View Synthesis.” 2017 IEEE International Conference on Computer Vision (ICCV) (2017): n. pag. Crossref. Web.