

中 国 地 质 大 学

研究生课程报告

课 程 名 称_____模式识别_____

教 师 姓 名_____杨林权_____

研究生姓名_____徐鸿飞_____

研究生学号_____1202110887_____

研究生专业_____软件工程_____

所 在 院 系_____地理与信息工程学院_____

类 别_____学术硕士_____

日 期_____2021 年 11 月 22 日_____

评 语

课程论文的评语:

平时成绩:	课程论文成绩:
总 成 绩:	评阅人签名:

注: 1、无评阅人签名成绩无效;

2、必须用钢笔或圆珠笔批阅, 用铅笔阅卷无效;

3、如有平时成绩, 必须在上面对评分表中标出, 并计算入总成绩。

人脸识别方法最新研究综述

摘要:

本文主要研究了 2021 年计算机视觉与模式识别会议 (CVPR2021) 中几篇关于人脸识别的论文, 并对其进行了综述。首先是人脸图片的生成, 使用深度卷积神经网络进行人脸识别依赖于大量人脸图像数据集的可用性。需要许多身份的示例, 对于每个实体, 需要大量不同的图像, 以便网络学习对类内变化的鲁棒性。在实践中, 这样的数据集很难获得, 特别是那些包含足够的姿势变化的数据集, 所以将 3D 可变形模型合并到 GAN 的生成器中, 以便从自然场景下的图像中学习非线性纹理模型。然后是 MagFace: 人脸识别和质量评估的通用表示法, 一类学习通用特征的损失, 其大小可以衡量给定面的质量。然后是 WebFace260M, 提供了一个新的百万规模的人脸基准测试, 其中包含噪音为 4M 的身份/260M 的人脸 (WebFace260M) 和清洁的 2M 身份/42M 人脸 (WebFace42M) 训练数据, 以及一个精心设计的时间约束评估协议。最后是解决年龄不变人脸识别满足人脸年龄合成的一个多任务学习框架同时处理年龄不变人脸识别和人脸年龄合成这两个任务。

Abstract:

This paper mainly studies several papers on face recognition in the 2021 Computer Vision and Pattern Recognition Conference (CVPR2021), and reviews them. The first is the generation of face images. The use of deep convolutional neural networks for face recognition relies on the availability of a large number of face image data sets. Many examples of identities are needed, and for each entity, a large number of different images are needed in order for the network to learn robustness to changes within the class. In practice, such data sets are difficult to obtain, especially those that contain enough pose changes, so the 3D deformable model is incorporated into the GAN generator to learn non-linear texture models from wild images. Then there is MagFace: a general representation for face recognition and quality evaluation, a type of loss of learning general features, the size of which can measure the quality of a given face. Then there is WebFace260M, which provides a new million-scale face benchmark test, which contains 4M identities/260M faces (WebFace260M) and clean 2M identities/42M faces (WebFace42M) training data, and a Well-designed time constraint evaluation protocol. Finally, it is a multi-task learning framework that solves age-invariant face recognition and meets face age synthesis, and

simultaneously processes age-invariant face recognition and face age synthesis tasks.

关键词：模式识别；机器学习；人脸识别；

引言

随着信息化的发展和网络服务的普及，生物特征识别技术在现代社会的应用越来越重要。在这些技术中，人脸识别技术是最方便和实用的技术之一，因为它可以实现远距离认证，而且不需要手动进行任何认证操作。众所周知，人脸识别容易受到由于年龄、周围光线和姿势造成的人脸外观变化的影响。还有一些技术挑战需要解决^[1]。由于深度学习方法的出现，人脸识别技术也取得了显著的进步。本文根据 CVPR2021 中最新的几篇论文对人脸识别技术进行了综述，介绍了人脸识别技术的最新发展，包括人脸生成、人脸识别和质量评估的通用表示法、百万级深度人脸识别能力的基准和同时处理年龄不变人脸识别和人脸年龄合成这两个任务的多任务学习框架。最后还总结了关于人脸识别技术仍然需要处理和解决的研究挑战。

1 将3D可变形模型合并到GAN的生成器的人脸图像数据集的生成

1.1 简介

最新的人脸识别算法使用数百万张图像进行训练。有了互联网作为资源，人脸图像相对容易获得。然而，这些图像的语义分布通常是高度不平衡的。例如，大多数可用的照片都是正面肖像，其中包含较大信息的图像相对清晰。目前，照片角度的稳健性被认为是人脸识别面临的最大挑战之一。一些研究人员试图通过首先正面化问题图像来避免这个问题，而其他则试图通过综合增加训练数据集来学习角度的额外稳健性。

然而，随着生成性对抗网络发展的最新进展，出现了一种可行的替代方案。已经证明，GANS能够生成新的人的真实图像，因此不需要将数据扩展限制到现有的人类照片。然而，为了生成完全合成的训练数据，有必要将身份与其他特征（如姿势）分离。最近的研究表明，2D GAN方法难以实现这种解分离。在这项工作中，Richard T. Marriott^[2] 等人将一个3D变形模型（3DMM）合并到一个GAN中，这样就可以生成新的合成身份的图像，并且在不使用标识符的情况下修改姿势。

他们主要介绍一种从自然场景图像中学习非线性纹理模型的方法，该方法可用于生成具有完全分离姿势的合成身份的图像，无需对面部结构进行特殊捕获

扫描；并演示通过使用合成的3D GAN图像和CPLFW的最先进精度扩充数据集来改进大姿势面部识别。

1.2 三维GAN

生成性对抗网络（GAN）通常由卷积生成器和鉴别器组成，它们在mini-max游戏中交替训练：鉴别器训练用于区分生成的图像和一组真实图像的图像，并训练生成器以最小化鉴别器的成功。虽然生成的图像似乎代表真实世界的3D对象，但它们当然不过是由生成器的2D卷积滤波器学习的2D特征的组合。由于这个原因，当线性穿越GAN发生器的最新空间时，人们往往会看到“惰性”的2D形式之间的变换，而不是3D空间中物理上真实的变换。例如，即使识别出潜在空间中影响生成图像中人脸姿势的方向，人脸的外观也不可能保持一致。实际上，生成器甚至可能无法以不同姿势生成同一张脸。为了确保在操纵姿势时在合成图像中保持外观（和身份），通过集成3DMM来增强生成器。典型地，GAN的输入是随机向量。3D GAN的输入是随机纹理和背景向量，但也有随机的3DMM形状、表达式和位置参数。然后使用可微分渲染器将随机头部形状渲染为生成的“背景图像”，面部纹理由纹理生成器提供，纹理生成器从训练图像中从头学习其模型。无论随机模型实例的形状、表情或姿势如何，对于鉴别器，渲染图像必须看起来逼真。为了实现这一点，纹理生成器开始生成具有与模型形状正确对应的特征的真实纹理。

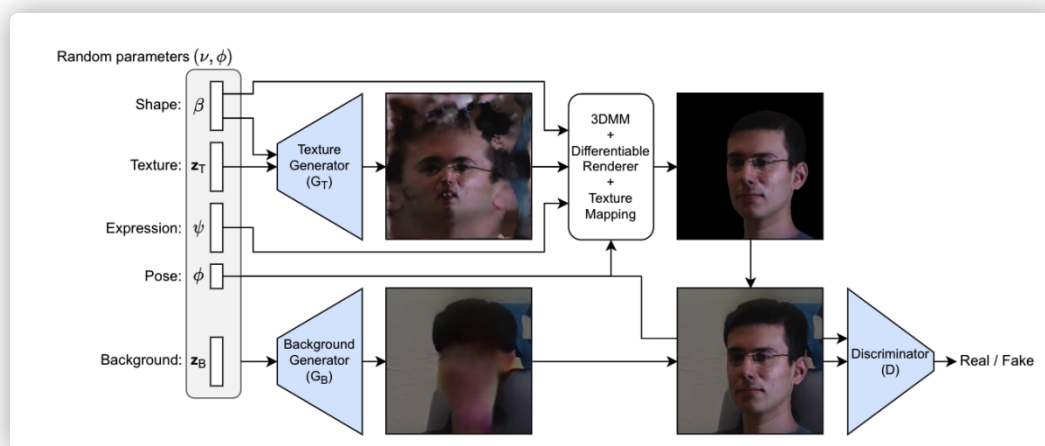


图1 GAN的生成器包括两个CNN，用于生成面部纹理和背景

图1描述了提出的3D GAN的架构。图的下半部分描绘了一个标准的条件图，其中一些图像是由随机参数生成的，然后与这些参数一起馈送到鉴别器，以与同样带有姿势的训练图像相对应。在实现中，姿势信息在空间上重复，并连接为图像的附加通道。该图的上半部分描述了3DMM的集成，其中学习的纹理通过可微分渲染器渲染到该图像中。通过渲染纹理提供图像的主要对象，背景生成器将学习仅生成背景和未由3DMM建模的特征，例如眼镜、衣服和头发的边缘。由于纹理生

成器不受姿势或表情参数的限制,因此可以在不影响3D模型纹理的情况下对图像的这些方面进行操作。因此,不需要额外的身份约束。

经过实验, 3D GAN能够更有效地利用塞莱巴的图像,其效率是前者的三到六倍。禁用SH光照,并将姿势的变化限制在CelebA本身的范围内,这两种方法都会降低CFP FP和CPLFW的精度,其中姿势的限制影响最大。

文章提出了一种新的3D GAN公式,用于从原始图像中学习非线性纹理模型,然后生成具有完全分离姿势的新身份的合成图像。3D GAN合成的图像可以成功地用于提高大姿态人脸识别的准确性。3D GAN的优点是不需要特殊的封盖纹理扫描。最后,由于3D GAN可以生成新身份的图像,因此它提供了一种从嘈杂的数据集(如FFHQ)提取有用信息的途径。

2 人脸识别与质量评价的通用表示法

2.1 简介

在自然场景下识别人脸很困难,主要是因为非压缩环境下采集的人脸图像显示出很大的可变性。这种可变性与图像采集条件(如照明、背景、模糊度和低分辨率)、面部因素(如姿势、遮挡和表情)或彩色人脸识别系统的偏差有关^[6]。

为了应对这些挑战,无约束环境下最相关的人脸分析系统(如监控视频)包括三个阶段:1)人脸采集从一组原始图像中选择或从视频流中捕获最适合识别的人脸图像;2)特征提取,从每个人脸图像中提取有区别的表示;3)面部应用程序将参考图像与一个给定的图片集匹配,或将面部分组。

2.1 MagFace^[3]

人脸图像质量是实现高性能人脸识别系统的一个重要因素^[7],传统方法,如ISO/IEC 19794-5标准、ICA09303标准、Brisque、Niqe和Piqe,从基于图像的方面(例如,距离、照明和遮挡)或基于对象的测量(例如,附件)描述质量。基于学习的方法,如FaceQNet和Best Rowden通过在人类评估和基于相似性的标签上训练的网络回归质量。然而,这些质量标签容易出错,因为胡曼可能不知道恢复系统的最佳特性,因此不能考虑所有合适的因素。近年来,人们提出了几种基于不确定性的方法,通过特征的不确定性来表达人脸的质量。SER-FIQ多次将图像转发到网络,并通过提取特征的变化来测量人脸质量。置信度感知人脸识别方法建议将每个人脸图像表示为潜在空间中的高斯分布,并学习特征值中的不确定性。MagFace是一种将质量度量编码到人脸表示中的新框架。通过优化幅度

$AI = ||f_i||$ ，而无需对每个特征进行归一化，从而实现极简设计。这种设计主要有两个优点：

- (1) 可以继续使用大多数现有推理系统广泛采用的基于COSINE的度量；
- (2) 通过同时增强其方向和幅度，学习的人脸表征对野外人脸的可变性的鲁棒性更好。

这是首次将特征幅值统一为人脸识别中的质量指标。MagFace的设计不仅遵循直观的动机，而且产生具有理论保证的结果，MagFace通过基于图片大小的自适应边界和正则化来优化特征。

实验对比结果如图2，IJB-B数据集包含1845个主题，其中包含21.8K静态图像和7011个视频中的55K帧。作为IJB-B的扩展，IJB-C数据集包含约3500个身份，总共31334个图像和117542个无约束视频帧。在比较苦难的基准数据集上，可以看出MagFace的精准度是最高的。

Method	IJB-B (TAR@FAR)			IJB-C (TAR@FAR)		
	1e-6	1e-5	1e-4	1e-6	1e-5	1e-4
VGGFace2* [6]	-	67.10	80.00	-	74.70	84.00
CenterFace* [44]	-	-	-	-	78.10	85.30
CircleLoss* [32]	-	-	-	-	89.60	93.95
ArcFace* [9]	-	-	94.20	-	-	95.60
Softmax	46.73	75.17	90.06	64.07	83.68	92.40
SV-AM-Softmax [42]	29.81	69.25	84.79	63.45	80.30	88.34
SphereFace [20]	39.40	73.58	89.19	68.86	83.33	91.77
CosFace [40]	40.41	89.25	94.01	87.96	92.68	95.56
ArcFace [9]	38.68	88.50	94.09	85.65	92.69	95.74
MagFace	40.91	89.88	94.33	89.26	93.67	95.81
MagFace+	42.32	90.36	94.51	90.24	94.08	95.97

图2 基准数据集上的验证精度

通过将模棱两可的样本推离类中心，MagFace改进了以前基于边缘的人脸识别工作中的类内特征分布。充分的理论和实验结果证明，MagFace可以同时获得输入人脸图像的质量。作为一个通用框架，MagFace可以潜在地扩展到其他分类任务，如细粒度对象识别、人员重新识别。此外，提出的探索特征量的原则为估计其他对象的质量铺平了道路，例如，reid中的人体或活动分类中的动作片段。

3 百万级人脸识别数据集和基准与多任务学习框架

3.1 百万级人脸识别数据集和基准：WebFace260M^[4]

人脸基准测试使研究人员能够培训和评估高性能的人脸识别系统。尽管越来越多的人致力于研究复杂的网络和损失，但学术界受到有限的训练集和近乎饱和的测试协议的限制。如表1所示，就身份和面孔而言，公众最大的训练集分别是MegaFace2和MS1M。MegaFace2包含672K名受试者的470万张面孔，这些人是从Flickr收集的。MS1M由10M面组成，面宽100K，但噪音率约为50%。相比之下，来自行业的公司可以访问更大的私有数据来训练人脸识别模型：谷歌利用800万个身份的200万张图像来训练FaceNet，face book通过500万张1000万个身份的人脸进行训练。这一数据缺口阻碍了研究人员推进深度人脸识别的前沿。大量显示数据的主要障碍在于大规模身份收集、有效且可扩展的清理和有效的培训。

Dataset	# Identities	# Images	Images/ID	Cleaning	# Attributes	Availability	Publications
CASIA-WebFace [84]	10 K	0.5 M	47	Auto	-	Public	Arxiv 2014
CelebFaces [57]	10 K	0.2 M	20	Manual	40	Public	ICCV 2015
UMDFaces [7]	8 K	0.3 M	45	Semi-auto	4	Public	IJCB 2017
VGGFace [41]	2 K	2.6 M	1,000	Semi-auto	-	Public	BMVC 2015
VGGFace2 [8]	9 K	3.3 M	363	Semi-auto	11	Public	FG 2018
MS1M [21]	0.1 M	10 M	100	No	-	Public	ECCV 2016
MS1M-IBUG [14]	85 K	3.8 M	45	Semi-auto	-	Public	CVPRW 2017
MS1MV2 [12]	85 K	5.8 M	68	Semi-auto	-	Public	CVPR 2019
MS1M-Glint [1]	87 K	3.9 M	44	Semi-auto	-	Public	-
MegaFace2 [38]	0.6 M	4.7 M	7	Auto	-	Public	CVPR 2017
IMDB-Face [64]	59 K	1.7 M	29	Manual	-	Public	ECCV 2018
Facebook [59]	4 K	4.4 M	1,100	-	-	Private	CVPR 2014
Facebook [60]	10 M	500 M	50	-	-	Private	CVPR 2015
Google [47]	8 M	200 M	25	-	-	Private	CVPR 2015
MillionCelebs [87]	0.6 M	18.8 M	30	Auto	-	Private	CVPR 2020
WebFace260M	4 M	260M	65	No	-	Public	-
WebFace42M	2 M	42M	21	Auto	7	Public	-

图3 深度人脸识别的训练数据

另一方面，评估协议和测试集在分析人脸识别性能方面起着至关重要的作用。常见的人脸识别评估包括LFW families、CFP、AgeDB、RFW，MegaFace，IJB家族主要致力于追求精确性，这一点几乎得到了认可。在现实世界的应用场景中，人脸识别总是受到推理时间的限制，例如解锁体验流畅的手机。轻量级人脸识别挑战朝着这一目标迈出了一步，但它忽略了检测和对齐的时间成本。NIST-FRVT是唯一的时间约束人脸识别协议。然而，严格的提交政策（每四个日历月不超过一次提交）阻碍了研究人员自由评估他们的算法。

为了解决上述问题，WebFace260M构建了一个新的大规模人脸基准测试系统，该系统由4M identities/260M人脸和一个时间受限的评估协议组成。首先，通过搜索引擎收集了400万名人的名单，下载了2.6亿张图片。然后，通过自训练（CAST）管道自动执行清洗，这是可伸缩的，不需要任何人工干预。提出的铸造程序产生了高质量的2M标识和42M面（WebFace42M）。有了这样的数据量，开发

了一个分布式训练框架来进行高效的优化。参考各种实际应用，设计了推断时间约束下的人脸识别（FRETS）协议，该协议使SACADEMIA能够全面评估深度人脸匹配器。FRETS协议由3个轨道组成：100、500和1000毫秒。由于公共评估的satv评分最高，并且可能包含噪音，因此人工构建了一个新的测试集，该测试集具有丰富的属性，以支持各种结果，包括不同的年龄、性别、种族和场景评估。此测试集将被积极维护和更新。

3.2 多任务学习框架^[6]

为了最小化年龄变化对人脸识别的影响，以前的工作要么通过最小化实体和年龄相关特征之间的相关性来提取身份相关的辨别特征，称为年龄不变人脸识别（AIFR），或者通过将不同年龄组的人脸转换为同一年龄组来消除年龄差异，称为人脸年龄合成（FAS）；然而，前者在模型解释时会丢失视觉结果，而后者则会受到影响下游识别的伪影的影响。因此，本文提出了一个统一的多任务框架来联合处理这两个任务，称为MTL人脸，它可以学习年龄不变的身份相关表示，同时实现令人满意的人脸合成。具体地说，首先通过注意机制将混合人脸特征分解为两个不相关的成分——身份和年龄相关特征，然后通过多任务训练和连续域自适应对这两个成分进行去相关。与传统的实现组级FAS的热编码相比，提出了一种新身份条件模块来实现身份级FAS，并采用权重共享策略来提高合成人脸的年龄平滑度。此外，收集并重新租赁了一个带有年龄和性别标记的大型跨年龄人脸数据集，以推进AIFR和FAS的发展。在五个基准跨年龄数据集上进行的大量实验表明，提出的MTLFace比现有的最先进的IFR和FAS方法具有更高的性能。在两个流行的通用人脸识别数据集上进一步验证了MTLFace，显示了在自然场景下人脸识别的竞争性能。

AIFR：为了鼓励年龄AFD稳健地分解特征，使用年龄估计任务和人脸识别任务来监督特征分解。具体来说，XAGE通过年龄估计任务绘制年龄变化，同时提取身份相关信息。

FAS：身份级别年龄条件是通过将身份条件模块（ICM）与一系列ICB一起应用，从区别性特征表示sXIB中推导出来的。然后，解码器根据从编码器中提取的多级高分辨率特征，在学习到的身份级别年龄条件的控制下，重建前进/倒退的人脸。

提出了一个多任务学习框架，称为MTLFace，以同时实现AIFR和FAS。提出了两个新的模块：AFD将姿态特征分解为年龄和身份相关特征，CM实现身份级别的人脸年龄合成。在跨年龄和通用人脸识别基准数据集上的扩展实验证明了该方法的优越性。

4 结束语

在本文中，概述了 2021CVPR 中人脸识别相关的技术，人脸识别已经取得了相当不错的成果，仍然需要大量的研究者进行更加深入的研究。（1）人脸识别算法适用于从婴儿到老年的整个生命过程中面部外观的变化。生命时间不变性是人脸识别的一个至关重要的因素。特别是，从技术挑战和局限性的角度来看，新生儿或儿童的面部图像注册时间将有多长的问题。（2）针对人脸识别的防护措施。虽然已经设想了在正常情况下极高的匹配精度，但对于需要认证的人戴着面罩和/或太阳镜，或者他们的脸完全被围巾或胡须覆盖的情况，还需要改进。（3）在鉴定双胞胎、兄弟姐妹或亲属时获得更高的匹配精度。这仍然是一个技术挑战^[8]。在不同年龄段，非配偶双胞胎的面部相似性高于同一人。（4）人脸欺骗难以得到保护，需要一种更加安全的技术通过人工智能对抗网络攻击，以及将它们与现有的人脸识别技术相结合以达到更好的识别效果。

5 参考文献：

- [1] Imaoka Hitoshi et al. The future of biometrics technology: from face recognition to related applications[J]. APSIPA Transactions on Signal and Information Processing, 2021, 10(9) : e9-e9.
- [2] Marriott R T, Romdhani S, Chen L. A 3D GAN for Improved Large-pose Facial Recognition[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 13445-13455.
- [3] Meng Q, Zhao S, Huang Z, et al. Magface: A universal representation for face recognition and quality assessment[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 14225-14234.
- [4] Zhu Z, Huang G, Deng J, et al. WebFace260M: A Benchmark Unveiling the Power of Million-Scale Deep Face Recognition[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 10492-10502.
- [5] Huang Z, Zhang J, Shan H. When age-invariant face recognition meets face age synthesis: A multi-task learning framework[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 7282-7291.
- [6] Philipp Terhorst, Jan Niklas Kolf, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. SER-FIQ: Unsupervised estimation of face image quality based on stochastic embedding robustness. In IEEE Conference on Computer

Vision and Pattern Recognition, 2020.

- [7] Lacey Best-Rowden and Anil K Jain. Learning face image quality from human assessments. *IEEE Trans. Information Forensics and Security*, 13(12):3064–3077, 2018.
- [8] Grother, P.; Grother, P.; Ngan, M.; Hanaoka, K.: Face recognition vendor test (FRVT) part 2: identification, US Department of Commerce, National Institute of Standards and Technology, NISTIR 8271, 2019.