中国科学院自动化研究所
INSTITUTE OF AUTOMATION
CHINESE ACADEMY OF SCIENCES

情感计算
–表情生成

中国科学院自动化研究所

刘斌
liubin@nlpr.ia.ac.cn

目录

- 背景及意义
- 研究主要机构与数据库
- 传统的表情生成方法
- 基于深度学习的表情生成方法
- 多模态情感生成
- 音视频深度伪造
- 展望

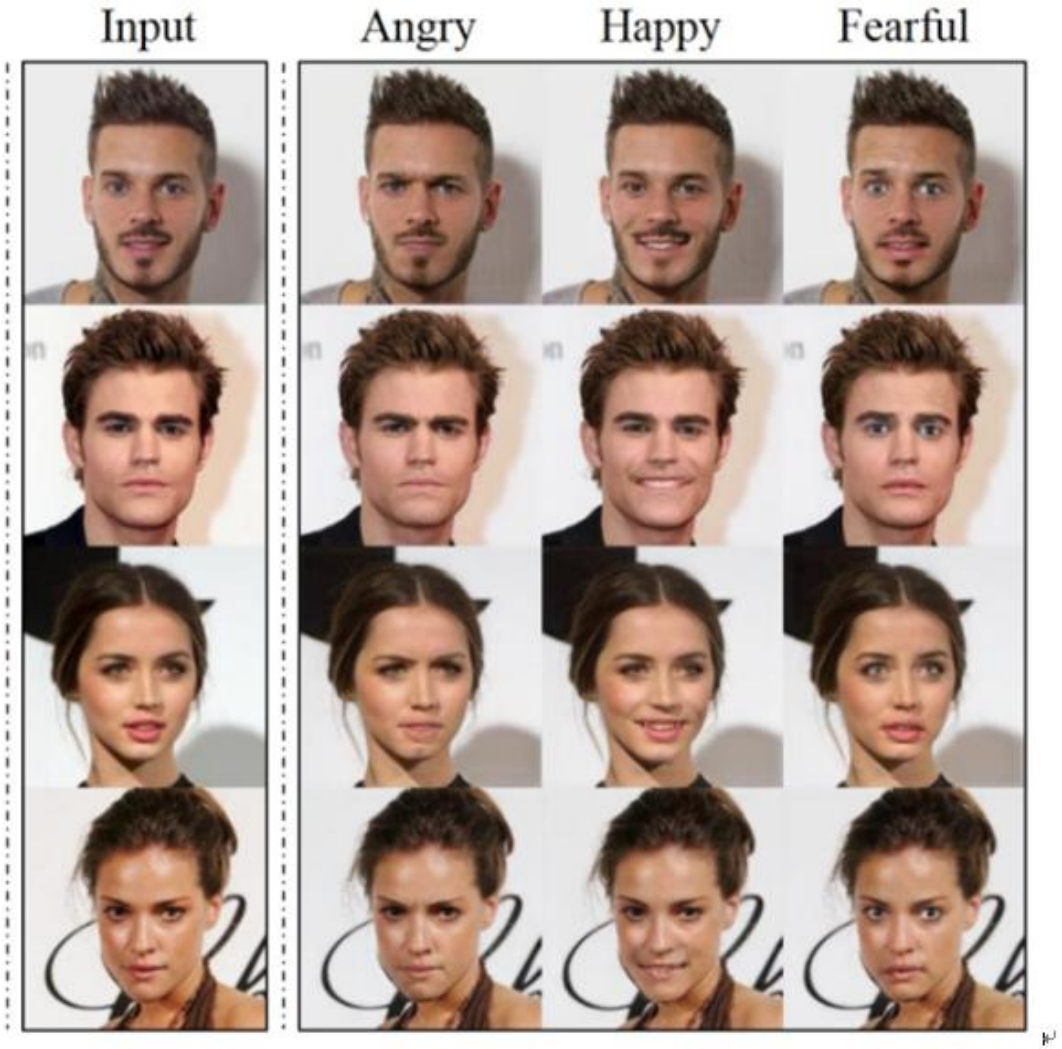
目录

背景及意义

- 表情生成的目的是通过某种表情计算方法产生出有表情的人脸图像
- 表情生成得到了计算机图形学、计算机视觉和模式识别领域的广泛关注
- 表情生成在人脸编辑、影视制作、社交网络和数据扩增方面应用广泛
- 合成高逼真度的人脸图像仍然是一个挑战性难题

背景及意义

- 表情生成示例



目录

主要研究机构

表情生成研究机构

清华大学

Xu F. A data-driven approach for facial expression synthesis in video[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2012:57-64.

韩国大学

Choi Y, Choi M, Kim M, et al. StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation[J]. 2018.

马里兰大学

Ding H, Sricharan K, Chellappa R. ExprGAN: Facial Expression Editing with Controllable Expression Intensity[J]. 2017.

表情生成数据库

- 表情生成常用数据库

CK+
http://www.consortium.ri.cmu.edu/ckagree/

RaFD Dataset
http://www.socsci.ru.nl:8180/RaFD2/RaFD?p=main

CAS(ME)3
http://casme.psych.ac.cn/casme/c4/

Oulu-CASIA NIR&VIS facial expression database
http://www.cse.oulu.fi/wsgi/MVG/Downloads/Oulu-CASIA

CAS-PEAL face database
http://www.jdl.ac.cn/peal/

表情生成数据库

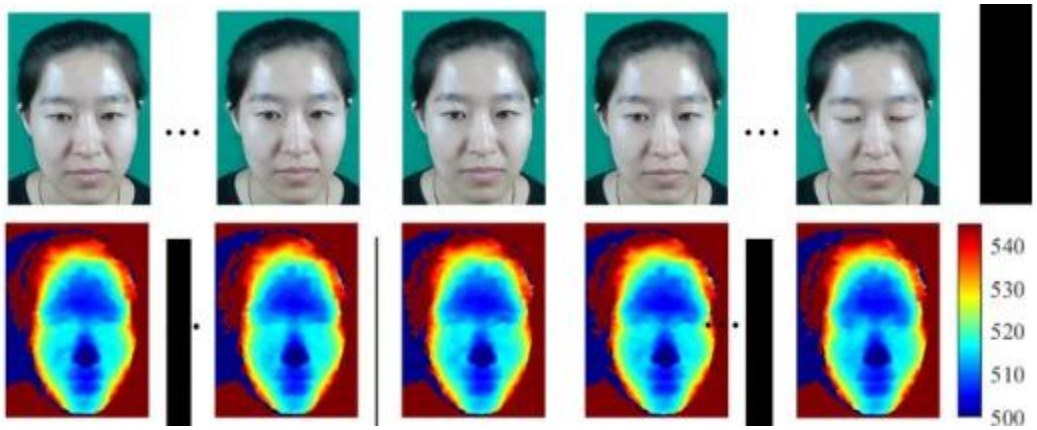
- CK+数据库:
  - 发布机构：Carnegie Mellon University
  - 表情类别：Anger, disgust, fear, happiness, sadness, surprise, contempt, and neutral
  - 数据规模：327个有情感标签的图片

表情生成数据库

- RAFD数据库:
  - 发布机构：Radboud University Nijmegen
  - 表情类别：Anger, disgust, fear, happiness, sadness, surprise, contempt, and neutral

表情生成数据库

- CAS(ME)3数据库:
  - 发布机构：中科院心理研究所
  - 表情类别：Anger, Disgust, Fear, Happiness, Sadness, Surprise, and Others



表情生成评价方法

- 表情生成评价指标
  - 客观评价:身份信息是否保持、合成表情是否正确、其他客观评价指标
  - 主观评价：MOS（Mean Opinion Score）：邀请被试者，人工对合成图像打分，评判身份信息是否保持、合成表情是否正确、合成图像是否符合审美

表情生成评价方法

- 客观评价 -- 身份信息是否保持
  - 选取Rank-1准确率指标：就是第一次命中；Rank-k，就是在第k次以内命中。人脸识别中，Rank-k就代表与目标人脸最相似的k个人脸中，成功命中的概率
- 客观评价 -- 合成表情是否正确
  - 对于表情生成后的效果，设计或利用现有模型对生成的表情进行识别验证，指标为分类准确率（ACC）
- 客观评价 --直接比较生成图像和标准答案图像之间的差异
  - 峰值信噪比（PSNR）、结构相似性（SSIM）、均方误差（MSE）、Inception Score（IS）

目录

传统的表情生成方法

- 表情生成方法
  - 渐变法
  - 表情映射
  - 几何驱动
  - 表情系数
  - 五官移植
  - 统计学方法

传统的表情生成方法

- 渐变法
  - 通过同一时域变形函数，完成相关联的两个表情状态图像的帧间插值坐标转换构造出渐变图像
  - 可以是二维/三维/纹理空间的坐标值
  - 渐变技术是产生人脸表情的直观方法，按其特点可分为：基本渐变(morphing)、基于视点的渐变(view morphing) 、三维渐变(3D morphing)

传统的表情生成方法

- 基本渐变
  - 定义一个在单位时间区间上的形变函数，通过对同一对象两种不同的表情图像进行帧间插值，计算生成中间状态特征点的位置坐标，从而产生两个指定的脸部表情图像之间的光滑过渡，即在两个已有的表情之间生成新的表情

传统的表情生成方法

- 基本渐变
  - 给定源对象S光滑的变化到目标对象T
  - 中间的对象既有S的特征，也有T的特征
  - S和T可以具有不同的拓扑

S　　T

  - 要求确定两个图像特征点之间点对点的对应关系
  - 当试点和姿势发生变化时，会产生不真实的脸部表情图像

传统的表情生成方法

- 基于视点的渐变
  - 克服了图像变化对视点和头部姿势的敏感性，图像中目标对象可视度的变化对变形结果有一定的影响

传统的表情生成方法

- 三维渐变
  - 三维渐变是两维图像渐变和三维几何模型变形相结合的产物，需要计算三维位置坐标和纹理空间坐标值，可以附加物体的物理特性描述
  - 实质是用三维插值实现脸部表情之间的形状变化，用两维渐变实现对应纹理图像的变化
  - 三维渐变获得了独立于视点的真实性，但是动画还是受预先定义的关键表情之间插补的限制

传统的表情生成方法

- 表情映射
  - 将某个人脸对象的表情重新定位到其他特定人脸上的方法，广泛应用于表演驱动的脸部动画中。可分为两类：一般表情映射和表情比率图
- 一般映射法
  - 给定某人的中性脸和表情脸图像，确定两幅图像中的特征点，然后计算这两组特征点的差向量，并将它作用到另一个人中性脸的特征点上，使该中性脸依此进行图像变形，从而得到新的表情
  - 其实质是利用已经存在的顶点运动向量等数据，将其他人脸对象的表情映射到或者说定位到新的特定人脸上

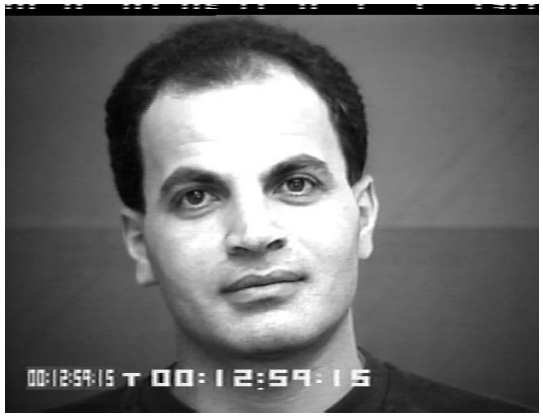
传统的表情生成方法

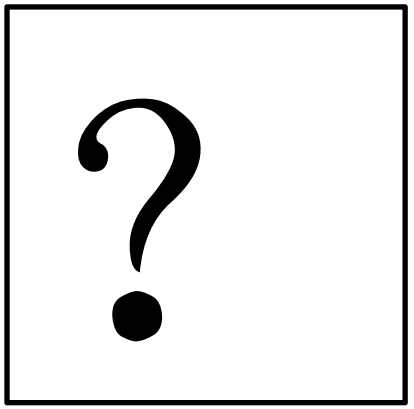
- 一般表情映射：示例图

中性

确定两幅图像中的特征点，然后计算这两组特征点的差向量

开心





传统的表情生成方法

- 一般表情映射
  - 优势：借助两幅参考图像的帮助实现了任意对象新表情的生成，弥补了单纯的表情渐变方法的缺陷
  - 劣势：整个过程仅针对人脸表情进行，没有考虑皮肤变形挤压产生的皱纹等变化丰富的表情细节，因而影响了表情的真实感程度

传统的表情生成方法

- 表情比率图（ERI）
  - 用于捕获由于扰动而引起的光照变化的且与脸部皮肤颜色无关的数据结构。一个人脸的表情比率图能被应用到任何其他人脸来得到正确的光照改变，从而将一个人的表情细节更好地整体转移到其他人的脸部
  - 给定某人的中性脸和表情脸图像，计算两幅图像各对应像素光亮度之比或RGB 3个成分的比值，然后结合表情变化前后特征点的移动，将这组比例作用到另一个人的中性脸图像上，进行变形操作，从而获得另一个人的表情图像

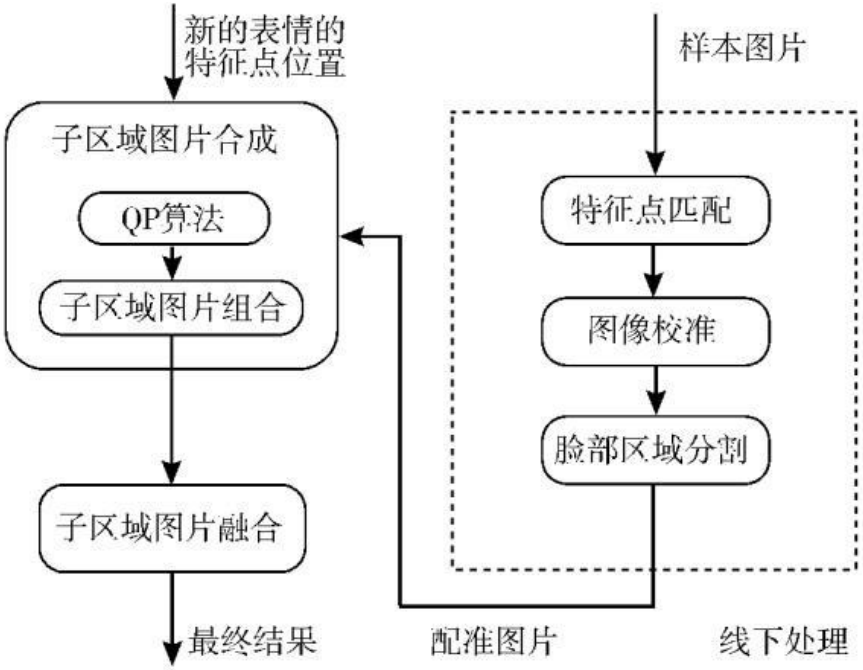
传统的表情生成方法

- 表情比率图（ERI）
  - 优势：解决了一般表情映射无法合成表情细节的缺陷
  - 劣势：要获得某个人某种特定表情脸，必须有一幅已知的表情脸图像作为一个样本与之对应，因此需要大量的样本

传统的表情生成方法

几何驱动

- 针对表情比率图方法中存在的不能很好反应皱毛发光照等细节纹理的缺陷。计算一系列样本表情的凸组合来生成照片真实感的脸部表情，然后从几何信息反推出纹理信息

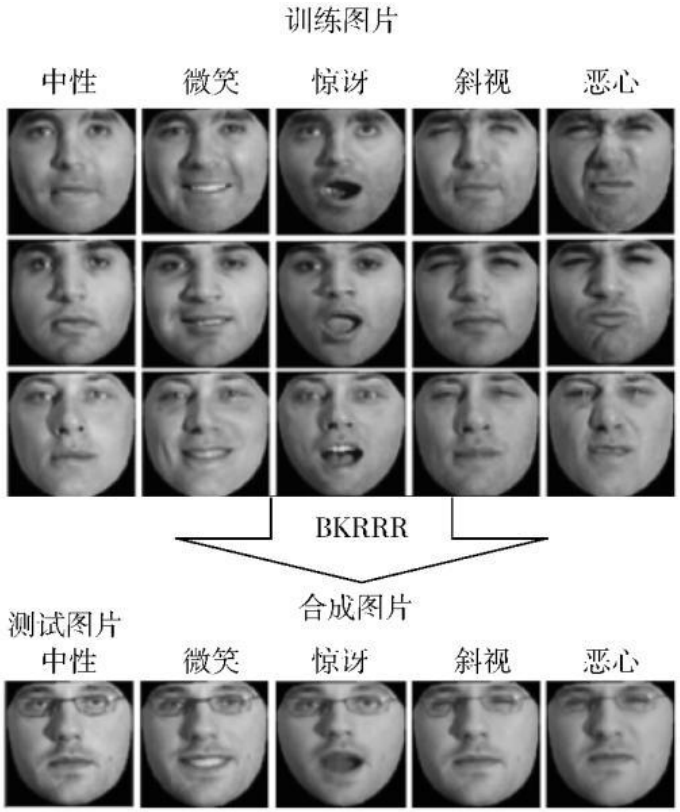


传统的表情生成方法

- 几何驱动
  - 优势：生成结果纹理细节特征丰富，且光照准确真实
  - 劣势：需要对特征标记点进行逐幅图像的追踪，工作量大；生成数据库时需要准备目标人脸一整套样本表情

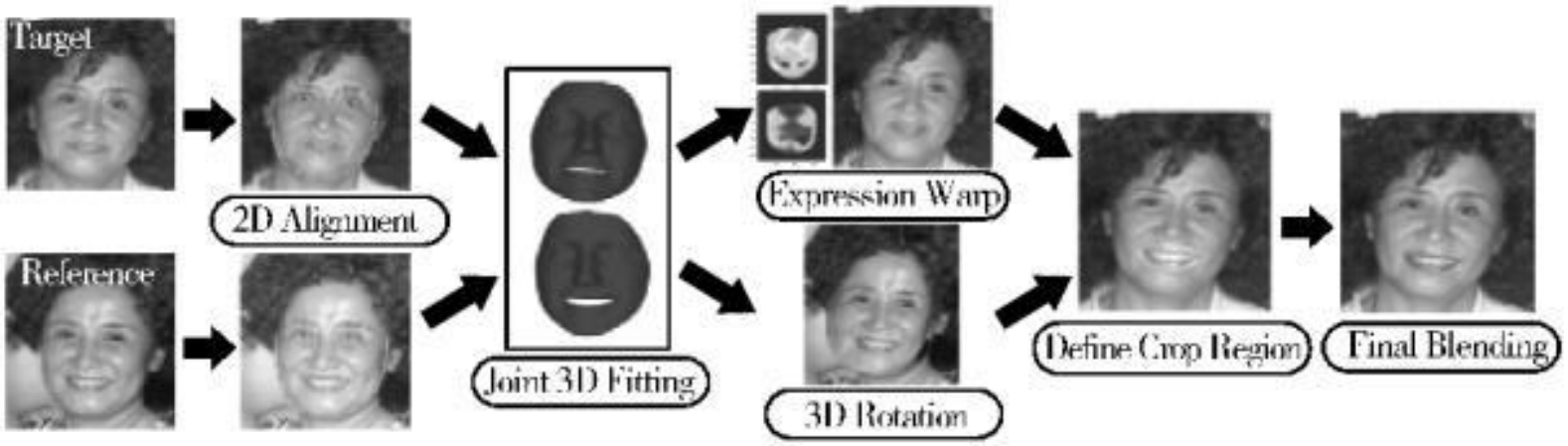
传统的表情生成方法

- 表情系数
  - 采用双线性核降秩回归（BKRRR）方法来学习中性表情和其他表情之间的变形系数，从而生成目标人脸的表情



传统的表情生成方法

- 五官移植
  - 使用五官移植生成算法，可以把输入人脸图片中的某部分五官组件（比如鼻子）移植到另一张照片上，并得到整体自然的效果



传统的表情生成方法

- 统计学
  - 利用样本库中的人脸图像，以线性组合或其他组合方式表示新的人脸
  - 通过总结人脸对象的一般规律，对特定人脸图像进行模型匹配与表达，可以结合不同熟悉特征的人脸图像数据库实现不同的脸部图像处理效果

目录

- 背景及意义
- 研究主要机构与数据库
- 传统的表情生成方法
- 基于深度学习的表情生成方法
- 多模态情感生成
- 音视频深度伪造
- 展望

基于深度学习的表情生成方法

基于深度学习的表情生成方法

- 基于深度学习的表情生成方法，随着生成模型的发展而发展。大量用于图像翻译任务的模型，都可以用于表情生成
- 图像翻译是指图像内容从一个域迁移到另一个域，可以看成是图像移除一个域的属性，并赋予另一个域的属性

Training data ~ $p_{data}(x)$

Generated samples ~ $p_{model}(x)$

Want to learn $p_{model}(x)$ similar to $p_{data}(x)$

基于深度学习的表情生成方法

- 基于深度学习的表情生成方法
  - PixelRNN
  - GAN
  - GAN部分变体

基于深度学习的表情生成方法

- PixelRNN
  - PixelRNN是使用概率链式法则来计算一张图片出现的概率
  - 每一项为给定前i-1个像素点后第i个像素点的条件概率分布
  - 分布通过神经网络RNN来建模，再通过最大化训练数据x的似然来学习出RNN的参数

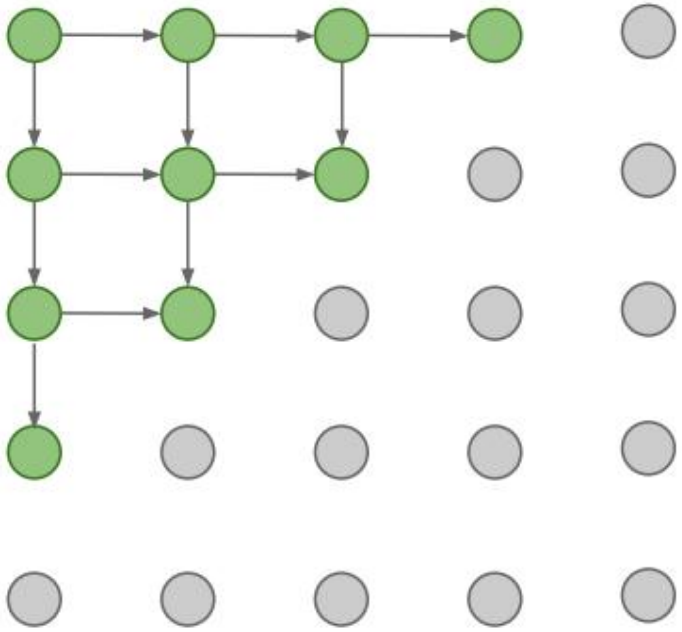
$$p(x) = \prod_{i=1}^{n} p(x_i | x_1, \ldots, x_{i-1})$$

Likelihood of image x

Probability of i'th pixel value given all previous pixels

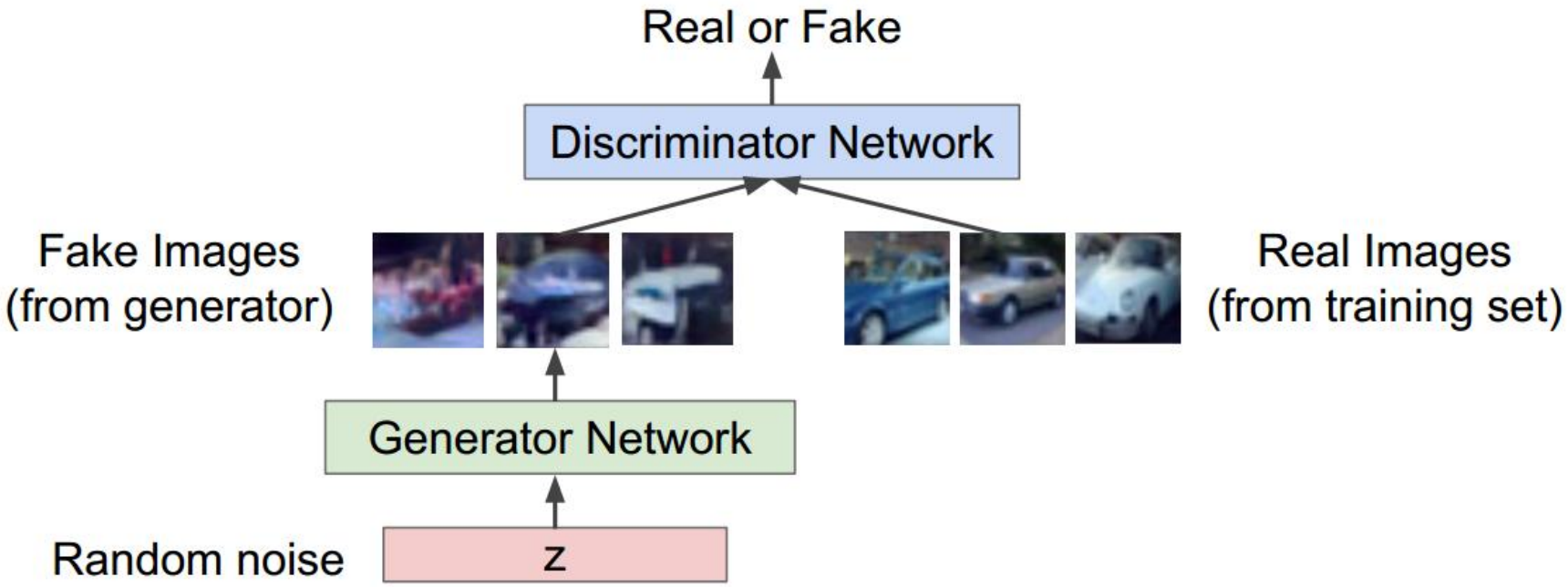
基于深度学习的表情生成方法

- PixelRNN
  - 从左上角开始生成图像。由于RNN每个时间步的输出概率都依赖于之前所有输入，因此能够用来表示上面的条件概率分布
  - 计算量大，耗时。训练这个RNN时，一次前向传播需要从左上到右下串行走一遍，然后根据上面的公式求出似然，并最大化似然以对参数做一轮更新



基于深度学习的表情生成方法

- GAN示例
  - 生成器（Generator network）：试着生成和真实图像很相似的数据
  - 判别器（Discriminator network）：试着区分真实图像和生成图像



基于深度学习的表情生成方法

- GAN损失函数：

1. **Gradient ascent** on discriminator

$$\max_{\theta_d} \left[\mathbb{E}_{x \sim p_{data}} \log D_{\theta_d}(x) + \mathbb{E}_{z \sim p(z)} \log(1 - D_{\theta_d}(G_{\theta_g}(z))) \right]$$

2. **Gradient descent** on generator

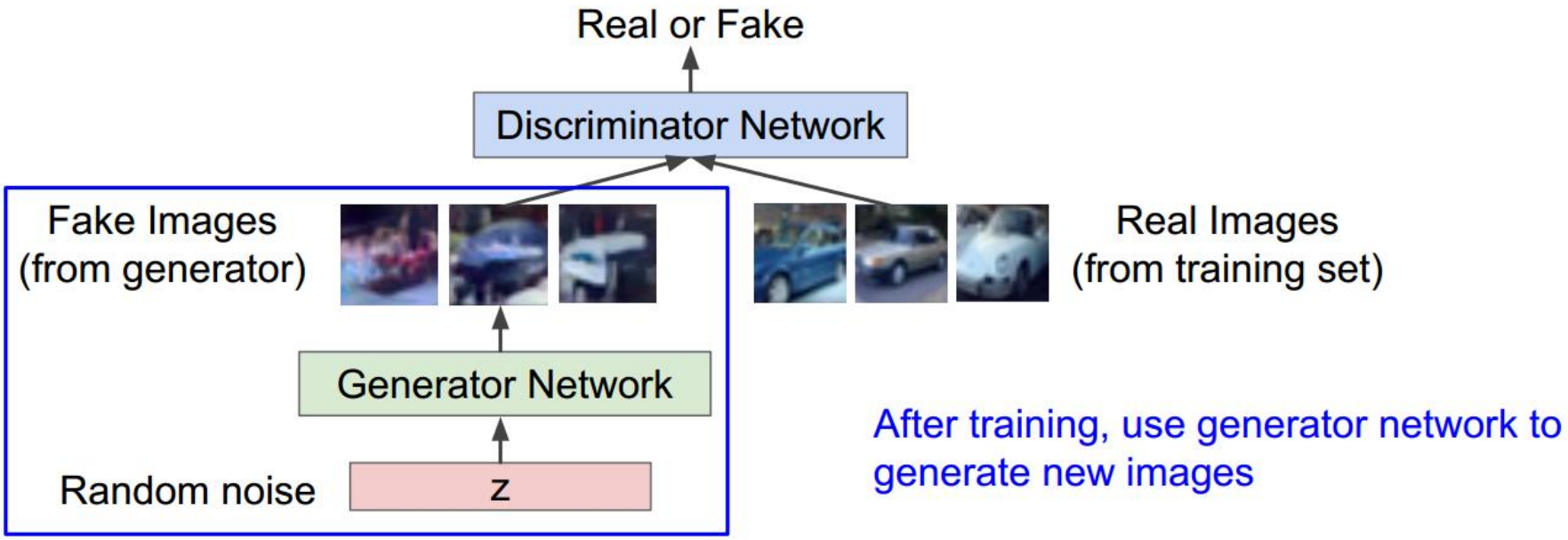
$$\min_{\theta_g} \mathbb{E}_{z \sim p(z)} \log(1 - D_{\theta_d}(G_{\theta_g}(z)))$$

（1）判别器：最大化目标函数，从而使得对于真实数据，D(x)接近1；对于生成数据，D(G(z))接近0

（2）生成器：最小化目标函数，使得D(G(z))接近1。使得判别器能够误认为生成图像为真实图像

基于深度学习的表情生成方法

- GAN生成图片
  - 在训练完成之后，利用生成器，生成接近训练集数据分布的图片
  - 先训练k轮判别器，再训练一轮生成器，但是k取多少比较好，并没有定论



基于深度学习的表情生成方法

- GAN延生模型

“The GAN Zoo”

- GAN - Generative Adversarial Networks
- 3D-GAN - Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling
- acGAN - Face Aging With Conditional Generative Adversarial Networks
- AC-GAN - Conditional Image Synthesis With Auxiliary Classifier GANs
- AdaGAN - AdaGAN: Boosting Generative Models
- AEGAN - Learning Inverse Mapping by Autoencoder based Generative Adversarial Nets
- AffGAN - Amortised MAP Inference for Image Super-resolution
- AL-CGAN - Learning to Generate Images of Outdoor Scenes from Attributes and Semantic Layouts
- ALI - Adversarially Learned Inference
- AM-GAN - Generative Adversarial Nets with Labeled Data by Activation Maximization
- AnoGAN - Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery
- ArtGAN - ArtGAN: Artwork Synthesis with Conditional Categorial GANs
- b-GAN - b-GAN: Unified Framework of Generative Adversarial Networks
- Bayesian GAN - Deep and Hierarchical Implicit Models
- BEGAN - BEGAN: Boundary Equilibrium Generative Adversarial Networks
- BiGAN - Adversarial Feature Learning
- BS-GAN - Boundary-Seeking Generative Adversarial Networks
- CGAN - Conditional Generative Adversarial Nets
- CaloGAN - CaloGAN: Simulating 3D High Energy Particle Showers in Multi-Layer Electromagnetic Calorimeters with Generative Adversarial Networks
- CCGAN - Semi-Supervised Learning with Context-Conditional Generative Adversarial Networks
- CatGAN - Unsupervised and Semi-supervised Learning with Categorical Generative Adversarial Networks
- CoGAN - Coupled Generative Adversarial Networks
- Context-RNN-GAN - Contextual RNN-GANs for Abstract Reasoning Diagram Generation
- C-RNN-GAN - C-RNN-GAN: Continuous recurrent neural networks with adversarial training
- CS-GAN - Improving Neural Machine Translation with Conditional Sequence Generative Adversarial Nets
- CVAE-GAN - CVAE-GAN: Fine-Grained Image Generation through Asymmetric Training
- CycleGAN - Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks
- DTN - Unsupervised Cross-Domain Image Generation
- DCGAN - Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks
- DiscoGAN - Learning to Discover Cross-Domain Relations with Generative Adversarial Networks
- DR-GAN - Disentangled Representation Learning GAN for Pose-Invariant Face Recognition
- DualGAN - DualGAN: Unsupervised Dual Learning for Image-to-Image Translation
- EBGAN - Energy-based Generative Adversarial Network
- f-GAN - f-GAN: Training Generative Neural Samplers using Variational Divergence Minimization
- FF-GAN - Towards Large-Pose Face Frontalization in the Wild
- GAWWN - Learning What and Where to Draw
- GeneGAN - GeneGAN: Learning Object Transfiguration and Attribute Subspace from Unpaired Data
- Geometric GAN - Geometric GAN
- GoGAN - Gang of GANs: Generative Adversarial Networks with Maximum Margin Ranking
- GP-GAN - GP-GAN: Towards Realistic High-Resolution Image Blending
- IAN - Neural Photo Editing with Introspective Adversarial Networks
- iGAN - Generative Visual Manipulation on the Natural Image Manifold
- IcGAN - Invertible Conditional GANs for image editing
- ID-CGAN - Image De-raining Using a Conditional Generative Adversarial Network
- Improved GAN - Improved Techniques for Training GANs
- InfoGAN - InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets
- LAGAN - Learning Particle Physics by Example: Location-Aware Generative Adversarial Networks for Physics Synthesis
- LAPGAN - Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks

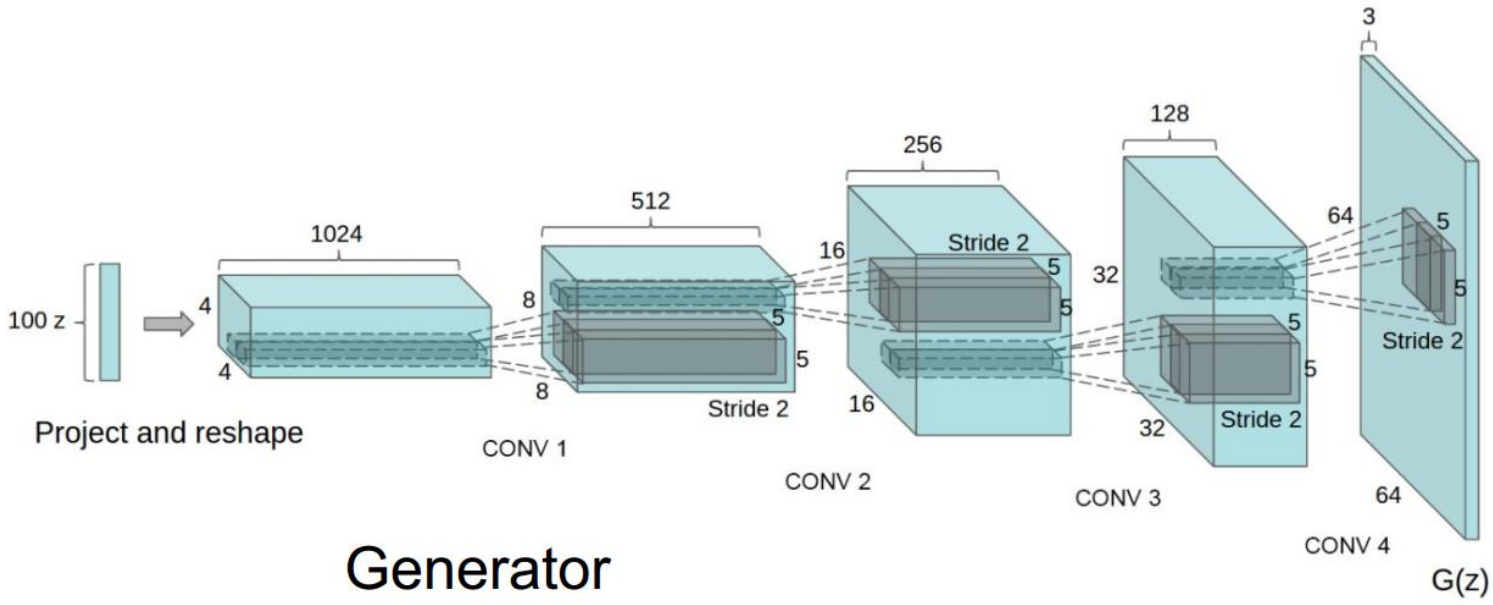
https://github.com/hindupuravinash/the-gan-zoo

基于深度学习的表情生成方法

- GAN延生模型
  - DCGAN
  - Pix2Pix
  - CycleGAN
  - StarGAN
  - GANimation

基于深度学习的表情生成方法

- DCGAN
  - 相较原始的GAN，DCGAN几乎完全使用了卷积层代替全链接层
  - 判别器几乎是和生成器对称的
  - 整个网络没有池化层和上采样层的存在，实际上是用带步长的卷积代替了上采样，以增加训练的稳定性
  - 在生成器和判别器中都添加了批量归一化操作

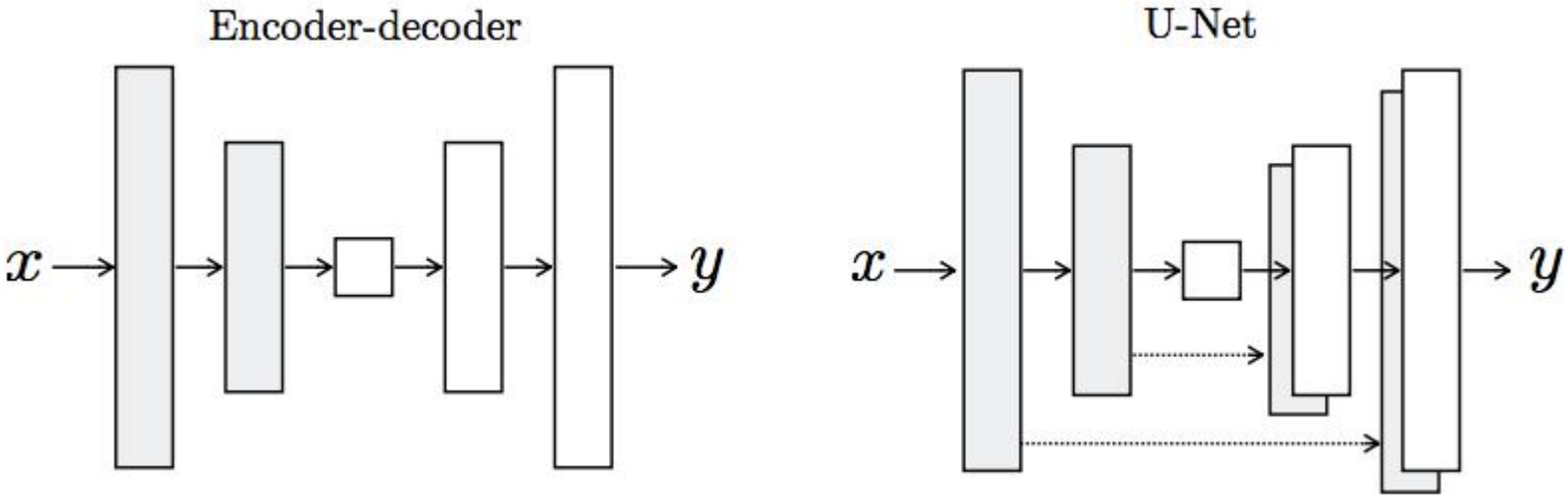


基于深度学习的表情生成方法

- Pix2Pix
  - Pix2Pix使用的是Conditional GAN（cGAN）
  - 它的G输入显然应该是一张图x，输出当然也是一张图y
  - D的输入却应该发生一些变化，因为除了要生成真实图像之外，还要保证生成的图像和输入图像是匹配的（即两者具有一定相似性）

基于深度学习的表情生成方法

- Pix2Pix
  - 生成器：输入和输出之间会共享很多的信息。因而，使用U-Net结构，使得信息得以更好的传播
  - 所谓的U-Net是将第i层拼接到第n-i层，这样做是因为第i层和第n-i层的图像大小是一致的，可以认为他们承载着类似的信息



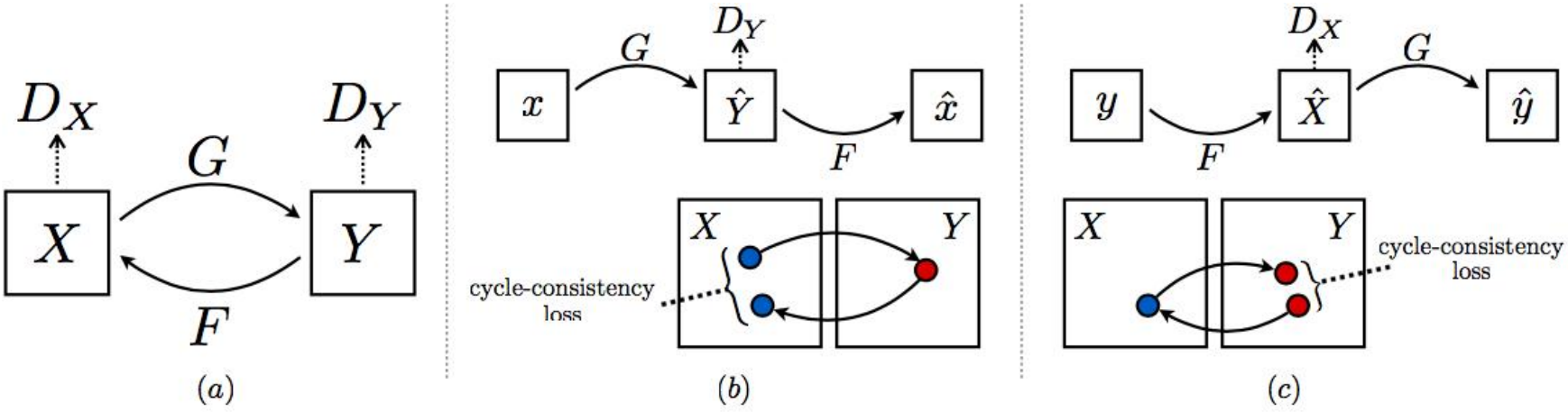
基于深度学习的表情生成方法

- Pix2Pix
  - 判别器：图像的变形分为两种，局部的和全局的。在损失函数中既要防止全局的变形，也要保证局部能够精准即可
  - Pix2Pix中的D被实现为Patch-D，所谓Patch，是指无论生成的图像有多大，将其切分为多个固定大小的Patch输入进D去判断
  - 因为G本身是全卷积的，对图像尺度没有限制。而D如果是按照Patch去处理图像，也对图像大小没有限制。就会让整个Pix2Pix框架对图像大小没有限制。增大了框架的扩展性

基于深度学习的表情生成方法

CycleGAN

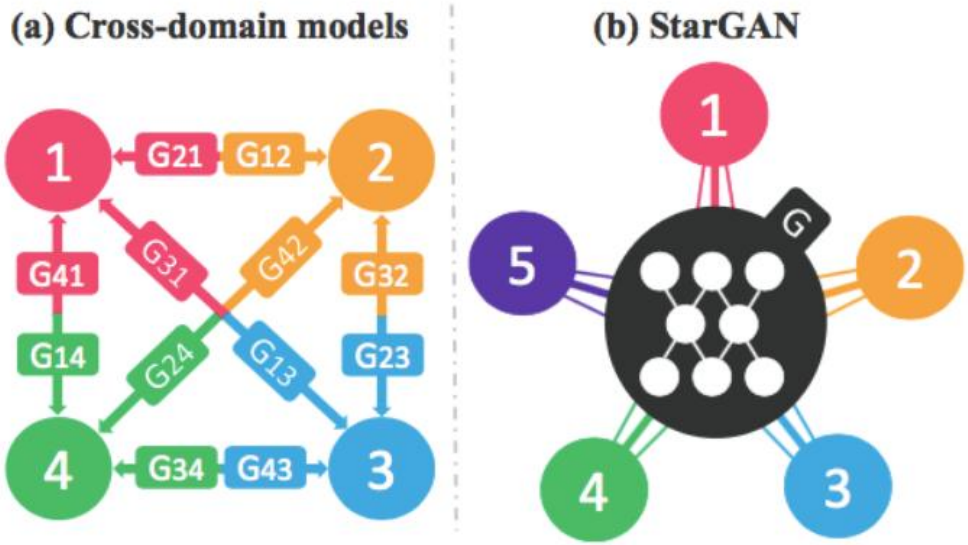
- CycleGAN算法就是将这种Cycle一致性思维引入到图像翻译任务上来，用于处理unpaired图像翻译问题。CycleGAN本质上是两个镜像对称的GAN，构成了一个环形网络
- 如果我们同时训练两个GAN，其中一个是生成器GA2B的鉴别器DB，另一个是GA2B的鉴别器DA ，那么一张A类型的图片x，通过两次变换，应该能变回自己



基于深度学习的表情生成方法

StarGAN

- 在StarGAN中，生成网络G被实现成星形。左侧为普通的Pix2Pix模型要训练多对多模型时的做法（多个G）。右侧可以看到，StarGAN仅仅需要一个G来学习所有领域对之间的转换

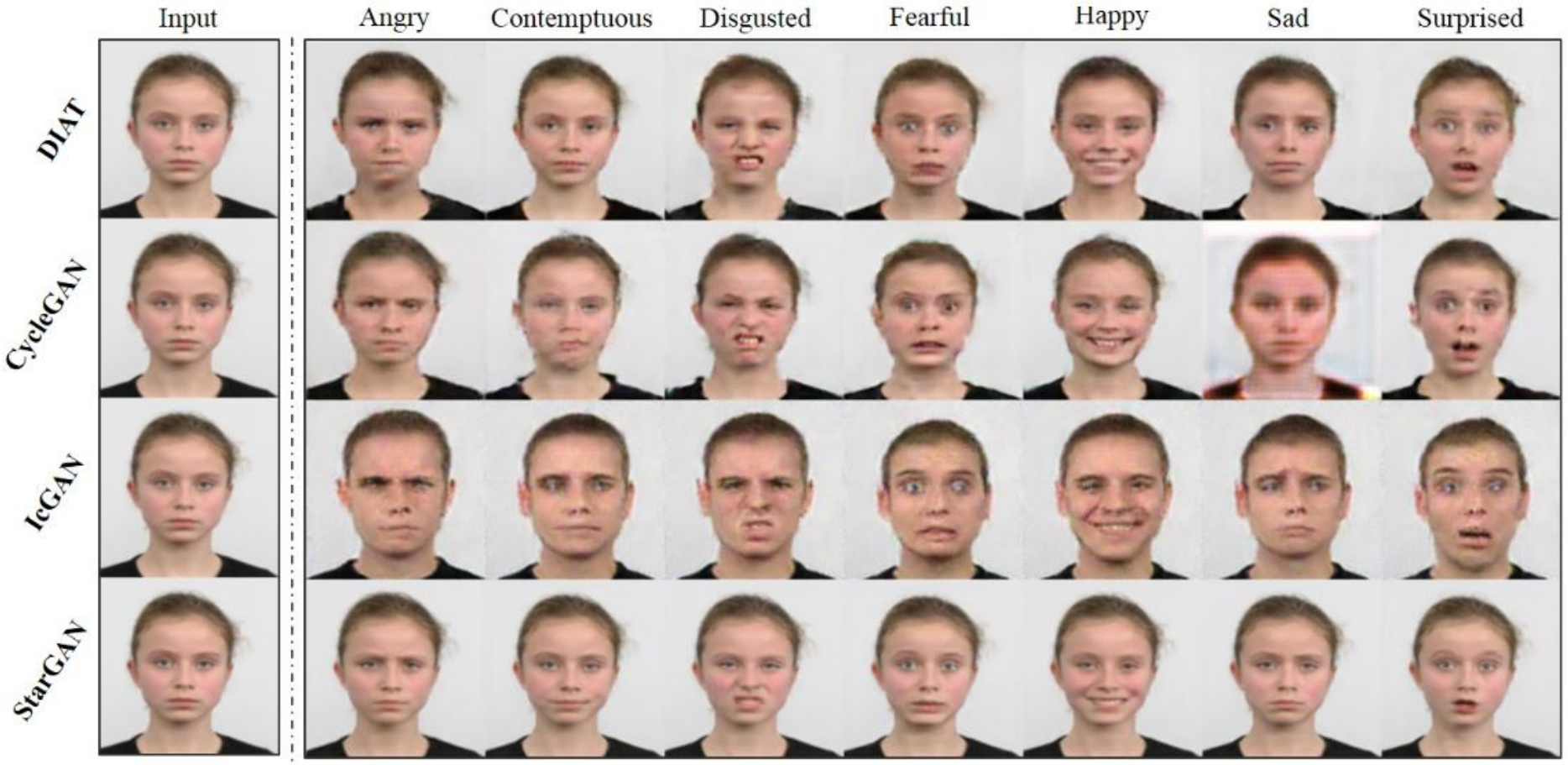


基于深度学习的表情生成方法

- StarGAN
  - G拥有学习多个领域转换的能力
  - 在G的输入中添加目标领域信息，即把图片翻译到哪个领域这个信息告诉生成模型
  - D除了具有判断图片是否真实外，还要有判断图片属于哪个类别的能力。保证G中同样的输入图像，随着目标领域的不同生成不同的效果
  - 保证图像翻译过程中图像内容要保持，只改变领域差异的那部分。图像重建可以完成这一部分，图像重建即将图像翻译从领域A翻译到领域B，再翻译回来，不会发生变化

基于深度学习的表情生成方法

- StarGAN效果图



基于深度学习的表情生成方法

- InterFaceGAN
  - InterFaceGAN可以学习到一种或多种语义特征在GAN的隐式空间的编码方式。
  - 不同属性的语义特征在子空间的线性变换后解耦，能做到分别修改性别、年龄、表情和戴眼镜与否，还可以显著改变姿态。

基于深度学习的表情生成方法

- InterFaceGAN效果图

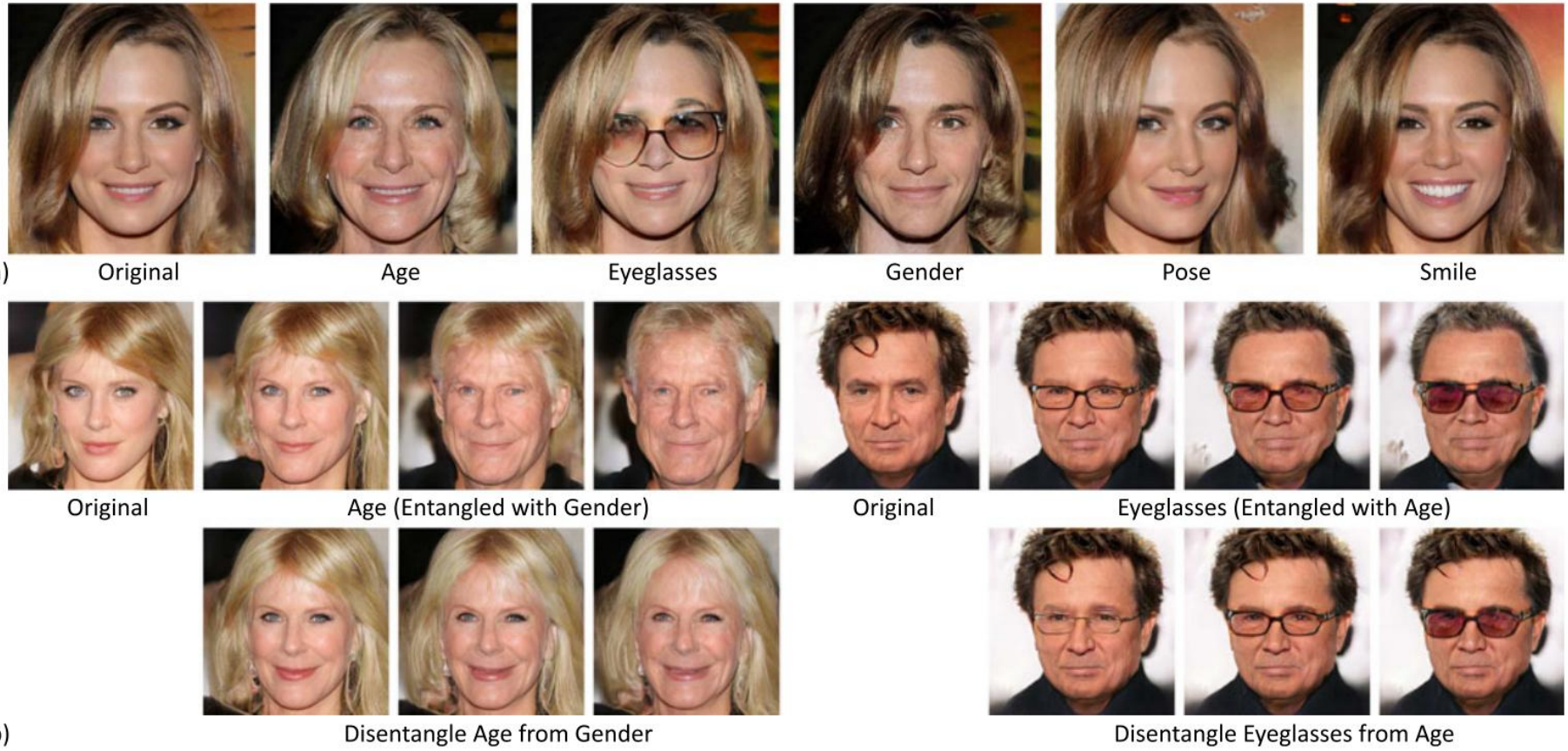


Fig. 1. (a) *Manipulating various facial attributes* through varying the latent codes of a well-trained GAN model. (b) *Conditional manipulation* results using InterFaceGAN, where we can better disentangle the correlated attributes (top row) and achieve more precise control of the facial attributes (bottom row). All results are synthesized by PGGAN [1].

目录

- 背景及意义
- 研究主要机构与数据库
- 传统的表情生成方法
- 基于深度学习的表情生成方法
- 多模态情感生成
- 音视频深度伪造
- 展望

多模态情感生成

- 个体情感的表达是可以从多个模态（语音、面部表情以及生理信号等等）感知出来的
- 语音和面部表情动作是最容易感知的模态形式
- 单模态情感生成是多模态情感生成的基础
- 多模态情感生成需要根据情感的变化使得音视频信息呈现一致
- 多模态情感生成需要考虑不同模态之间的同步问题

多模态情感生成

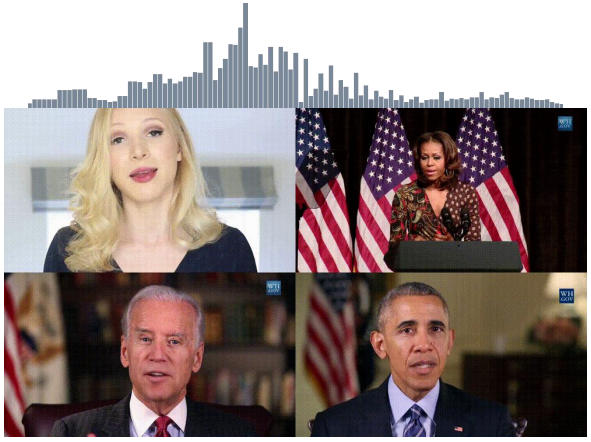
- 典型方法
  - 美国加州圣塔芭芭拉分校的Sargin等分析了头部姿势和语音韵律模式，基于隐马尔可夫模型方法完成了韵律驱动的头部姿势动画的自动生成
  - 韩国先进科技学院的Kim等致力于人类友好机器人的多模态表情生成。他们合成的机器人表情包括询问、请求、回答和解释四种类型
- 主要应用
  - 类人机器人
  - 虚拟主播
  - 虚拟现实

目录

- 背景及意义
- 研究主要机构与数据库
- 传统的表情生成方法
- 基于深度学习的表情生成方法
- 多模态情感生成
- 音视频深度伪造
- 展望

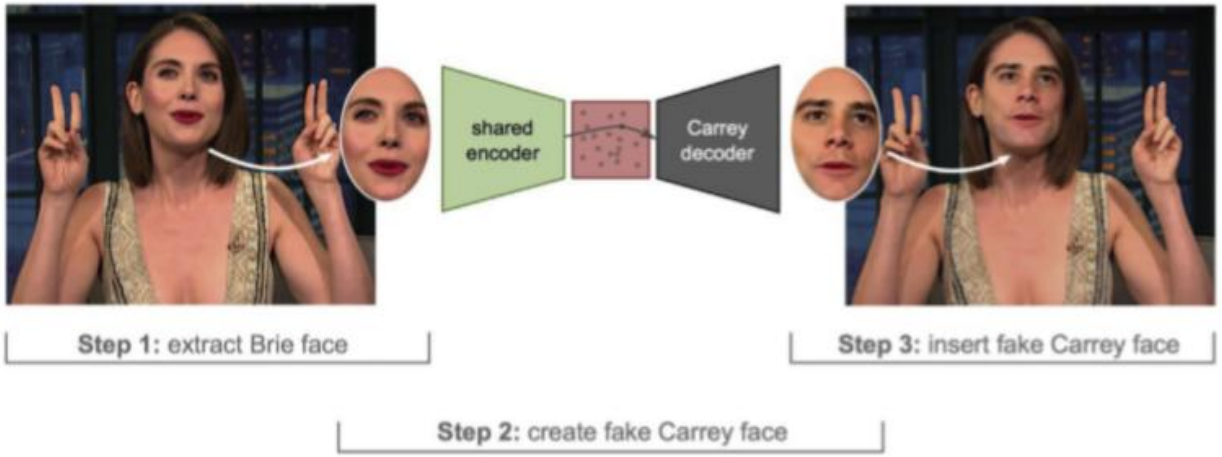
音视频深度伪造

- 使用“生成式对抗网络”深度学习模型进行大样本学习，将图片或视频合并叠加到源图片或视频上，或将声音、面部表情及身体动作拼接合成虚假内容的人工智能技术
- 截至2020年2月，伪造视频达到14,678个，伪造色情视频访问量超过1亿次。美国和中国政府均推出相应法规和政策
- 可以将音视频深度伪造的方法应用到情感生成任务中



音视频深度伪造

- 数据获取：多媒体网络数据（图像、音视频、文本），所有的伪造可以从网络获取，便利、快捷、数量庞大
- 将这些源数据作为深度神经网络学习的输入，通过ML和AI领域的技术，自动创建生成和目标内容匹配的数据
- 然后将生成的目标内容嵌入到原始的源数据内容中，创建或者篡改源数据，完成深度伪造的假视频内容



目录

总结

- 随着深度学习技术的发展，表情生成技术不断取得突破
- 细微表情的跟踪能力有待进一步提升
- 融合情境感知和用户个性化的表情生成
- 跨模态协同的表情生成

致谢

Thanks