

硕 士 研 究 生 读 书 报 告



题目 **CariGANs:不成对的真实图像向漫画转化模型\_\_\_\_**

作者姓名 吴迪奥

作者学号 21851055

指导教师 陈观林

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

CariGANs: Unpaired Photo-to-Caricature Translation

A Dissertation Submitted to

Zhejiang University

in partial fulfillment of the requirements for

the degree of

Master of Engineering

Major Subject: Software Engineering

Advisor: ChenGuanLin

Student ID: 21850155

By

WuDiAo

Zhejiang University, P.R. China

2018.12

摘要

CariGANs 旨在创造一个生成式对抗网络，将人脸图片转化成夸张的手绘漫画风格的图片。它由两个GAN组成，CariGeoGAN和CariStyGAN . CariGeoGAN 只负责转化人脸的几何形状达到夸张的效果，CariStyGAN负责将手绘漫画的风格转化成人脸照片(没有几何信息).将一个复杂的跨域转化问题转化成两个稍简单的任务。这项感性研究显示，通过该模型生成的手绘漫画图片与真实手绘图片非常近似，且在与其他图像转化模型的比较中有更好的效果，用户还可以调整参数改变生成的漫画的夸张程度。

#### 关键词：漫画，图像转化，GAN

Abstract

In this paper, we propose the first Generative Adversarial Network (GAN) for unpaired photo-to-caricature translation, which we call “CariGANs". It explicitly models geometric exaggeration and appearance stylization using two components: *CariGeoGAN*, which only models the geometry-to-geometry transformation from face photos to caricatures, and *CariStyGAN*, which transfers the style appearance from caricatures to face photos without any geometry deformation. In this way, a difficult cross-domain translation problem is decoupled into two easier tasks. The perceptual study shows that caricatures generated by our *CariGANs* are closer to the hand-drawn ones, and at the same time better persevere the identity, compared to state-of-the-art methods. Moreover, our *CariGANs* allow users to control the shape exaggeration degree and change the color/texture style by tuning the parameters or giving an example caricature.

**Keywords：**Caricature; Image translation; GAN

1. **介绍**

尽管业界已经有一些对交互式合成漫画的尝试，但是都需要专业的技能，且大多数只对应一种艺术家风格或者特定的卡通风格，预定义了夸张化的程度。

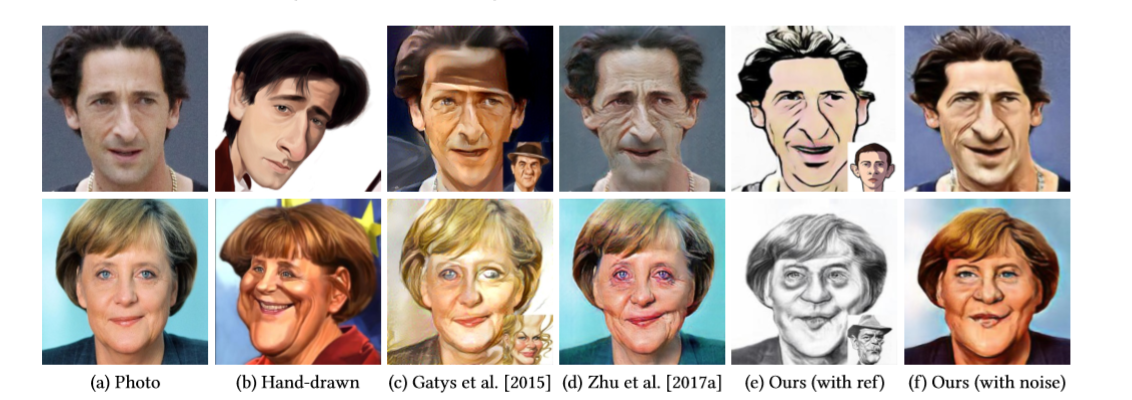
近年来随着深度学习技术的发展，在图片到图片转化的工作中被成功地应用。但是大多数的图片和漫画都不是一一对应的，所以用有监督的学习算法代价会很大。

本篇论文创造了第一个不成对的真实图片转化成漫画图片的生成式对抗网络，称为“CariGAN”.它将图片转化任务分解成几何信息的转化和图像风格的转化，分别由CariGeoGAN和CariStyGAN完成。为了建立不成对图片之间的联系，CariGeoGAN和CariStyGAN都使用了在跨域和无监督图片转化中广泛使用的循环一致性网络结构[12]。最后，通过“wraping” 将CariGeoGAN获得的夸张形状应用到由CariStyGAN获得的图片上。

在CariGeoGAN 中，使用了PCA 表达面部特征点作为输入。使在信息进入神经网络之前限制了脸部的结构。另外，在CariGeoGAN中使用了一个新的特殊的损失函数，旨在鼓励生成截然不同的面部效果，而避免面部任意畸变。

在风格化方面，为了在训练CariStyGAN的过程中排除几何信息，创造一个通过CariGeoGAN的反向几何映射来扭曲原图像的形状得到的中间数据集。这样两个GAN就被解耦了。用户还可以通过输入特定的数据或者噪声来指导模型输出的图片的风格。

下图是本篇论文的生成图片与其他一些方法的效果比较。



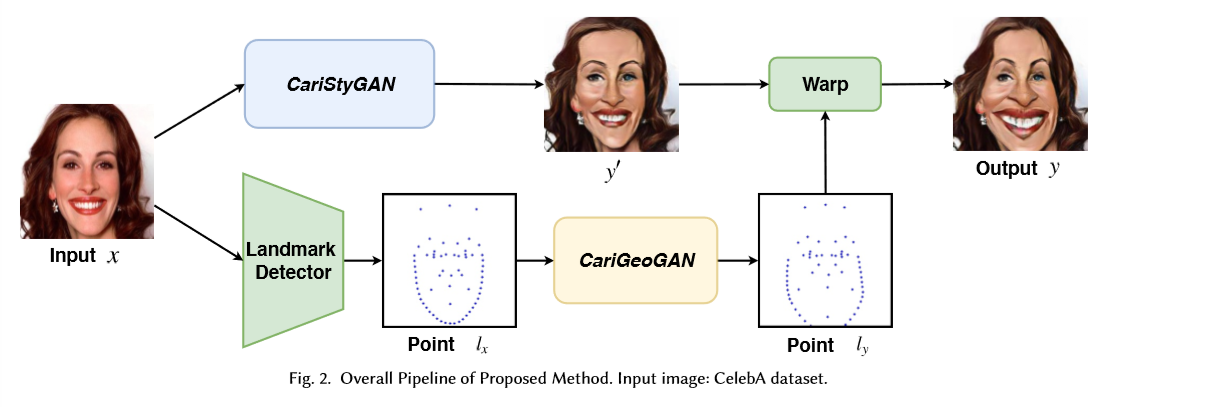
**图1.模型效果对比**

1. **方法**

因为成对的图片数据需要消耗艺术家大量的时间和金钱，因此论文的目标是将不成对的训练数据，使模型能够将普通的人脸照片转化成漫画同时保存其面部特征。

令X，Y分别表示人脸图片集和漫画图片集，两个图片集之间没有成对的数据。X是从CeleA数据集中随机选取的10000张人脸图片[24]，涵盖了各性别，年龄，姿势，表情。Y是从网上搜集的8451张手绘漫画图片，有各种形式如铅笔画，卡通画等，涵盖各种夸张手法。

论文的目的是训练一个映射 ，这是一个典型的跨域问题, 因为图片在几何形状和风格上都截然不同。不能直接训练一个图片到图片的映射。因此作者将解耦成两个映射 分别处理几何形容和风格。



**图2 CariGAN整体管道结构**

图二描述了CariGAN的结构和流程。接下来详细介绍两个CariStyGAN和CariGeoGAN实现。

* 1. 几何夸张

**训练数据：**不管是真实图片还是漫画图片，脸的形状都可以由2维的特征点表示[39]。

因此对X和Y中的图片标注了63个面部特征点。为了中心化面部轮廓，所有X,Y中的图片通过三个特征点(两个眼睛特征点和嘴的中心)，使用仿射变换向平均脸部对齐。并且所有图片都裁剪成256x256 ，标准化缩放比例。对于CariGeoGAN，使用了主要成分分析方法对

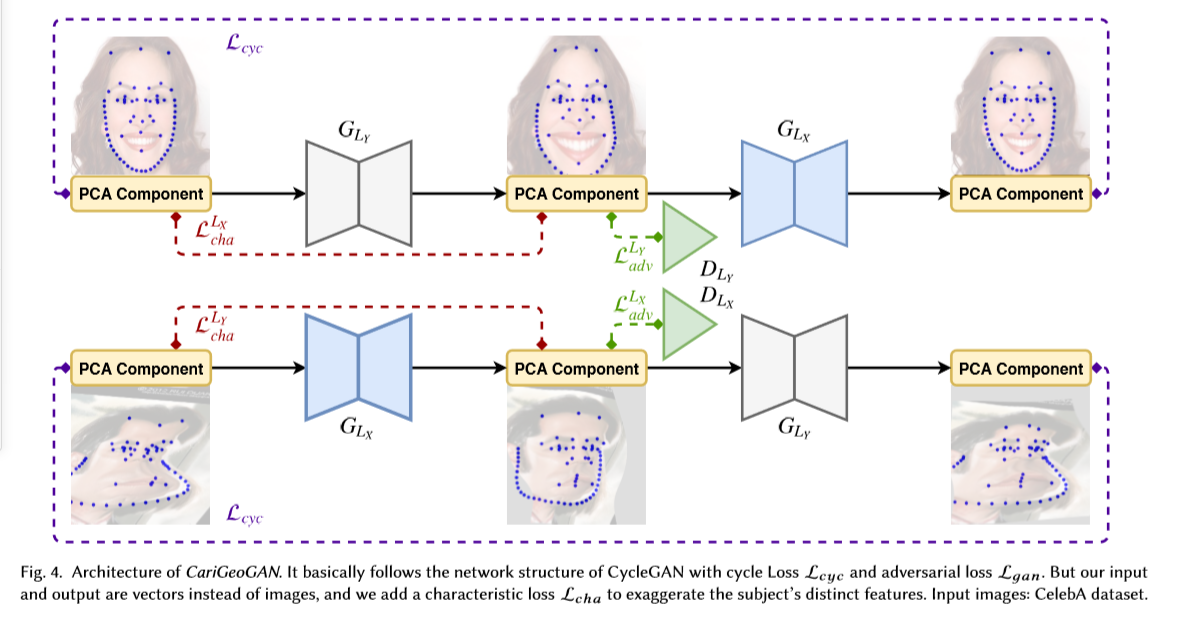
特征点进行降维，选取了前32个特征，能够覆盖99.03%的信息。因此63个特征点就变成了长度为32的向量。令Lx和Ly分别表示X和Y的 PCA 特征点域。CariGeoGAN转而学习从Lx到Ly的转化。

**CariGeoGAN:** 因为Lx和Ly 是不成对的，CycleGan通过一个反向的映射

: Ly -> Lx 将他们联系起来。受其启发，CariGeoGAN包含两个生成器和判别器，前一个生成器 学习映射 合成漫画轮廓  ,后一个生成器学习反向映射合成脸部轮廓. 判别器 学习判别图片是来自(或)的真实图片还是来自于或()合成的图片。

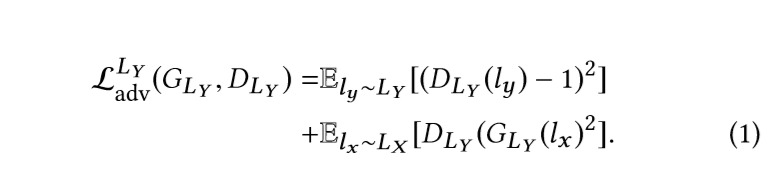
CariGeoGAN 的结构如下图所示包含两条路径。第一条路径是图4中的第一行。给一个脸部轮廓，可以通过生成一个漫画脸部轮廓 。一方面，该轮廓被喂给判别器。另一方面还可以通过与输入脸部轮廓靠近。另一条路径也是类似，如图4底部一行所示，对反向映射建模。值得注意的是， 和 在各个路径共享权重。

与CycleGAN[37]不同的是，CariGeoGAN采用了PCA向量替代了图片作为输入，且在生成器和判别器中使用FC-Relu 替代CONV-Relu。



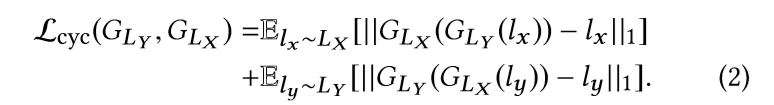
**图4 CariGeoGAN整体架构**

**Loss：**CariGeoGAN定义了三种不同的损失，均在图4中有所展示。第一个是在GANs中被广泛使用的对抗损失函数，采用了LSGAN[4]的对抗损失。

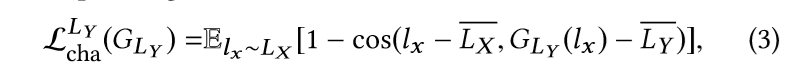


第二个是双向循环一致性损失，在CycleGAN中被使用，限制前向映射和反向映射的一致性。

如果使用对进行夸张化操作，该结果用反向映射返回应当与原输入 接近，反之亦然。



第三个是一个新的特殊的损失函数，用于惩罚输入的特征向量和预测向量减去其平均向量后的cos



因此CariGeoGAN总的损失函数如下图所示：



**训练细节:** 采用了CycleGAN[37]相同的训练策略。设置 , .使用Adam解决方案，batch\_size 设为1，所有训练都从头开始，开始的学习率设为 0.0002

* 1. 外观风格化

**训练数据：**

为了训练一个纯粹的不带几何信息的外观风格，需要创建一个中间域，它与X 贡献相同的几何信息，与Y有贡献相同的外观风格。对Y中的所有漫画图片的特征点用CariGeoGAN 转化得到Y’。CariStyGAN 转而学习从X到Y‘ 的转化。

**CariStyGAN:**

该模型继承了CycleGAN[37] 中图片级别的循环一致性限制保持了面部结构，并且受到MNUIT[12]的启发，将图片的内容编码和风格编码分成两个不同域的属性，通过组合内容编码和风格编码得到不同的转化效果。

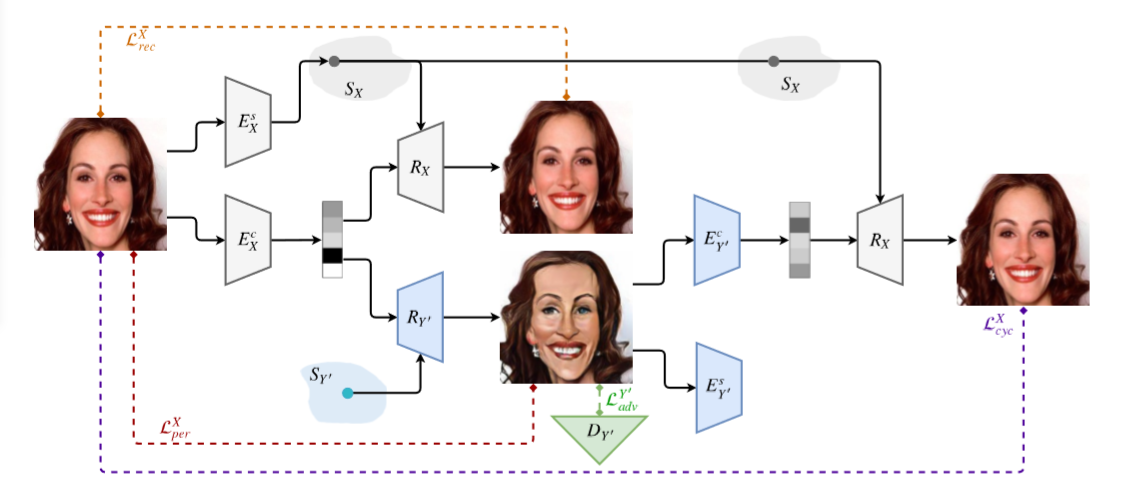
论文作者设计了自动编码结构，对于每个域I（I = X,Y’）包含两个编码器和一个解码器。

内容编码器和风格编码器 将输入图片 分解成内容编码 和风格编码 . 解码器 将内容编码 和风格编码 重新构建。内容编码 假设服从标准正太分布。

图8显示了网络的前向循环结构。比如给一张肖像图片x，首先分离出它的内容编码,再从 中随机取一个风格编码.接着使用解码器而不是它的原始解码器 生成一个图片, 该图片同样受到判别器的限制。

MUNIT在两个编码域要求循环一致，而本论文要求在图片域循环一致。即返回回来的图片  应与原输入图片x接近

通过该结构，前向映射 通过 实现，而反向映射 通过 实现



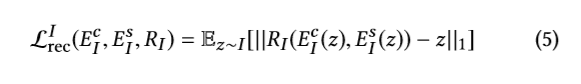
**图8 CariSTYGAN结构**

**损失函数：**

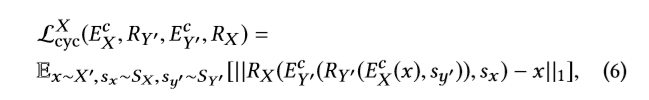
在图8中显示了CariStyGAN需要的四种损失。

第一个对抗损失函数 ,目的是使转化都得结果 与Y中真实的样本相同，让判别器难以判断。另一个对抗损失函数，在反向映射中， 的定义与之类似。

第二个重建损失函数，对输入图片和由内容编码和风格编码构建得到的结果进行L1惩罚

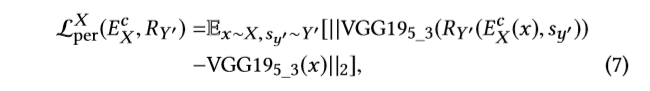


第三个是循环一致性损失，强制图片经过前向和反向映射后返回。具体地说，给一张图片x，拿到结果. 该结果被喂给编码器,得到它的内容编码，通过联合中一个随机的风格编码，使用解码器得到最终的结果。该结果应当与源输入x接近。

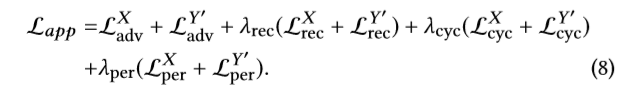


反向映射的循环一致性损失定义类似。

如果只使用上述三个损失函数，风格编码和内容编码并不能完全解耦，在转化阶段会造成失败的效果。为了解决这个问题，论文中添加了一个新的感性损失函数[13]，如公式7所示。



其中 [31]是在图像识别任务中预训练的relu5\_3 特征映射。最终的损失函数如公式8

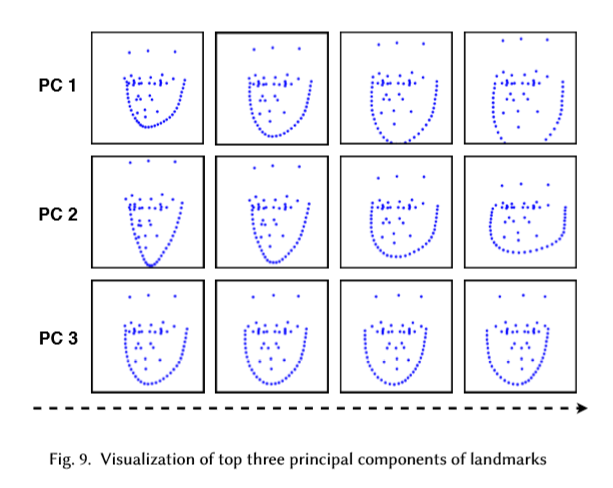


**训练细节：**

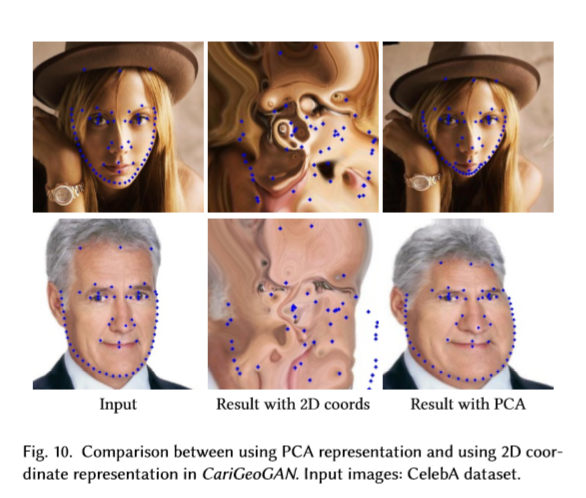
设置,使用Adam [16]算法，batch\_size =1 .所有训练从头开始，学习率为 0.0001 .

1. **讨论**
   1. PCA对CariGenGAN的决定性作用

通常情况下几何形状的变换比图片的转化要难。首先，特征向量被传入全连接层而不是卷积层，所以它们在学习过程中是去了局部空间限制。第二，特征向量结果对细微的错误比图片像素更加敏感，它们可能会造成严重的几何变形，比如折叠。如果使用原生的2维特征点，脸部结构很难保持。PCA可以做到这一点，每一个主成分代表了变量的一个方向，比如姿势，形状，大小。图九显示了前三个主要成分的特征点。图十比较了使用原生二维特征点和使用PCA向量的结果。



**图9 前三个主要成分脸部特征点**



**图10 使用PCA和使用传统二维特征效果对比**

* 1. 中间数据集对CariStyGAN的关键作用

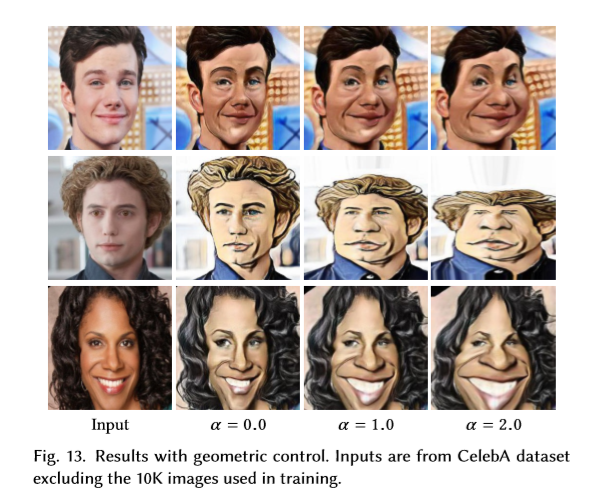
建立中间数据集对CariStyGAN的成功起到了重要作用，它联系了图片集X和漫画集之间的几何差异，让GAN能够集中与风格的转变。为了理解的作用，作者直接用Y训练了CariStyGAN，发现一些因素会混乱脸部结构如图11所示。一个可能的原因是网络企图同时学习两个混合的映射(几何结构和风格)。



**图11 中间数据集的重要性**

1. **结果与对比**
   1. 可控的结果

CariGAN 提供两个方面的控制。第一，系统允许用户通过设置变量 来控制脸部形状的变化程度、取值范围[0.0,2.0] ，不同参数的效果比较如图13所示。



**图13 不同a值效果对比**

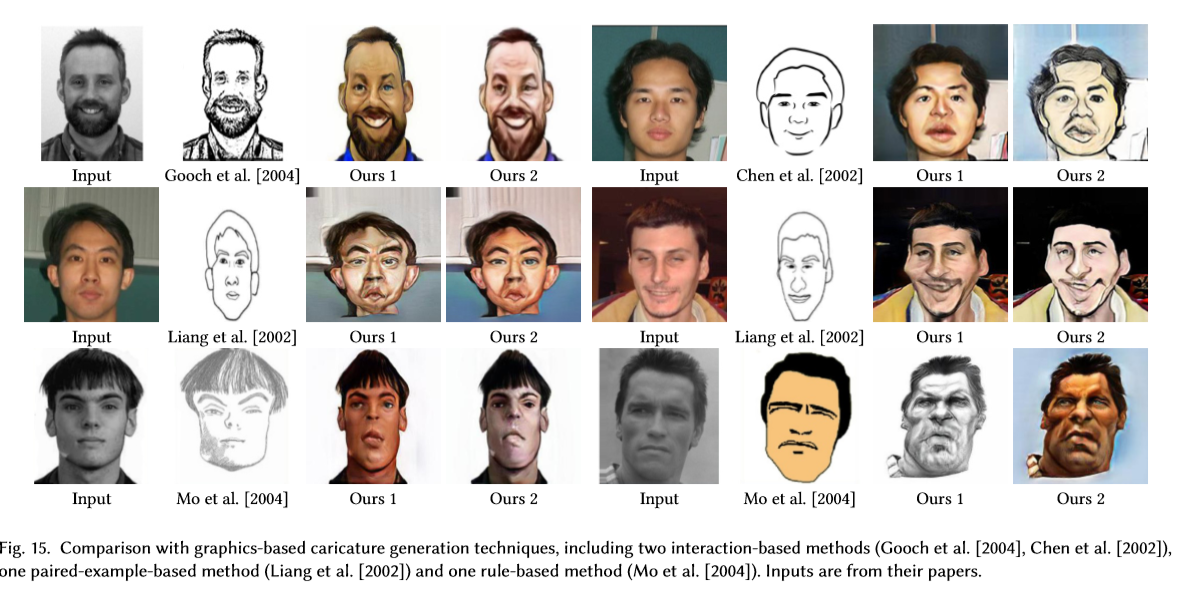
除此以外，用户也可以控制风格，一方面CariStyGAN是一个多模型图片转化网络，可以将一个图片转化成不同类型的漫画。另一方面，在CariStyGAN中，风格近似的漫画可以被编码，生成的效果也会类似，所以用户可以通过提供指定类型或一类相似风格的漫画来达到控制的目的。图14 显示了随机选取四个风格和两个相似风格得到的结果对比。



**图14 采用不同编码结果对比**

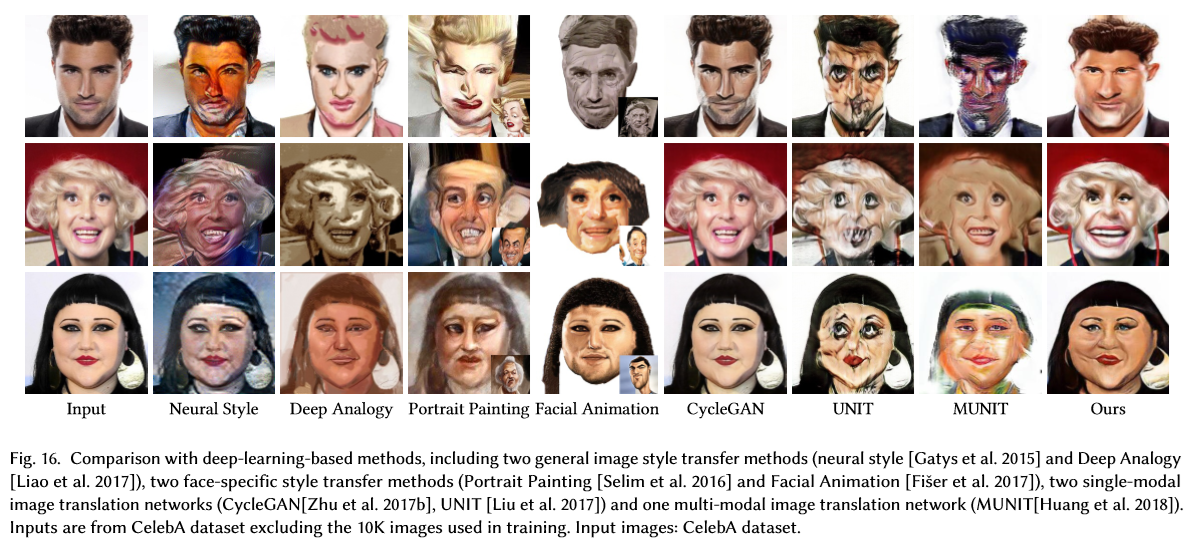
* 1. 与基于图的方法比较

作者将其方法与四种典型的基于图的方法进行了效果上的对比，如图15所示。显示这些方法在外表风格上效果一般，只能生成简单的风格比如卡通画，而本CariGAN可以学习生成上千种手绘风格。在几何变形方面，一些方法需要的人工部分较多，难以实际操作，有一些虽然是自动生成，但是在手绘领域难以完全表述清楚所有的几何变换规则，对比来说，通过自学习的方法更加可行。



**图15 与基于图的方法对比效果**

* 1. 与基于深度学习的方法比较



**图16 与其他基于深度学习的方法效果对比**

图16 显示了CariGAN和现存的基于深度学习的方法效果比较。所有数据均从本论文的手绘图片数据集中随机取。可以看到的是，它们一定程度上可以转化成漫画，但是不能完成几何变换。

接着与三个图片与图片直接转化的网络，包括典型的CycleGAN和UNIT和MUNIT，用和本论文相同的数据集进行训练，虽然他们都能完成几何变换和图片风格的转化，UNIT和MUNIT无法保证面部结构，CycleGAN保证了面部结构但是漫画风格种类的夸张变换程度低。对比来看，CariGAN的表现更加好。

**参考文献**

1. Ergun Akleman. 1997. Making caricatures with morphing. In *Proc. ACM SIGGRAPH*. ACM, 145.
2. Ergun Akleman, James Palmer, and Ryan Logan. 2000. Making extreme caricatures with a new interactive 2D deformation technique with simplicial complexes. In *Proc. Visual*. 165–170.
3. Susan E Brennan. 2007. Caricature generator: The dynamic exaggeration of faces by computer. *Leonardo* 40, 4 (2007), 392–400.
4. Dongdong Chen, Jing Liao, Lu Yuan, Nenghai Yu, and Gang Hua. LSGAN. Coherent online video style transfer. In *Proc. ICCV*.
5. Dongdong Chen, Lu Yuan, Jing Liao, Nenghai Yu, and Gang Hua. 2017b. Stylebank: An explicit representation for neural image style transfer. In *Proc. CVPR*.
6. Hong Chen, Nan-Ning Zheng, Lin Liang, Yan Li, Ying-Qing Xu, and Heung-Yeung Shum. 2002. PicToon: a personalized image-based cartoon system. In *Proc. ACM international conference on Multimedia*. ACM, 171–178.
7. Forrester Cole, David Belanger, Dilip Krishnan, Aaron Sarna, Inbar Mosseri, and William T Freeman. 2017. Synthesizing normalized faces from facial identity features. In *Proc. CVPR*. 3386–3395.
8. Jakub Fišer, Ondřej Jamriška, David Simons, Eli Shechtman, Jingwan Lu, Paul Asente, Michal Lukáč, and Daniel Sykora. 2017. Example-based synthesis of stylized facial` animations. *ACM Trans. Graph. (Proc. of SIGGRAPH)* 36, 4 (2017), 155.
9. Leon A Gatys, Alexander S Ecker, and Matthias Bethge. 2015. A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576* (2015).
10. Bruce Gooch, Erik Reinhard, and Amy Gooch. 2004. Human facial illustrations: Creation and psychophysical evaluation. *ACM Trans. Graph. (Proc. of SIGGRAPH)* 23, 1 (2004), 27–44.
11. Geoffrey E Hinton and Ruslan R Salakhutdinov. 2006. Reducing the dimensionality of data with neural networks. *science* 313, 5786 (2006), 504–507.
12. Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. 2018. Multimodal Unsupervised Image-to-image Translation. *arXiv preprint arXiv:1804.04732* (2018). Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-to-Image Translation with Conditional Adversarial Networks. In *Proc. CVPR*.
13. Justin Johnson, Alexandre Alahi, and Li Fei-Fei. 2016. Perceptual losses for real-time style transfer and super-resolution. In *Proc. ECCV*. Springer, 694–711.
14. Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. 2017. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196* (2017).
15. Taeksoo Kim, Moonsu Cha, Hyunsoo Kim, Jungkwon Lee, and Jiwon Kim. 2017. Learning to discover cross-domain relations with generative adversarial networks. *arXiv preprint arXiv:1703.05192* (2017).
16. Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
17. Hiroyasu Koshimizu, Masafumi Tominaga, Takayuki Fujiwara, and Kazuhito Murakami. 1999. On KANSEI facial image processing for computerized facial caricaturing system PICASSO. In *Proc. IEEE International Conference on Systems, Man, and Cybernetics*, Vol. 6. IEEE, 294–299.
18. Nguyen Kim Hai Le, Yong Peng Why, and Golam Ashraf. 2011. Shape stylized face caricatures. In *Proc. International Conference on Multimedia Modeling*. Springer, 536–547.
19. Lin Liang, Hong Chen, Ying-Qing Xu, and Heung-Yeung Shum. 2002. Example-based caricature generation with exaggeration. In *Proc. Pacific Conference on Computer Graphics and Applications*. IEEE, 386–393.
20. Jing Liao, Yuan Yao, Lu Yuan, Gang Hua, and Sing Bing Kang. 2017. Visual attribute transfer through deep image analogy. *arXiv preprint arXiv:1705.01088* (2017).
21. Pei-Ying Chiang Wen-Hung Liao and Tsai-Yen Li. 2004. Automatic caricature generation by analyzing facial features. In *Proc. ACCV*, Vol. 2.
22. Junfa Liu, Yiqiang Chen, and Wen Gao. 2006. Mapping learning in eigenspace for harmonious caricature generation. In *Proc. ACM international conference on Multimedia*. ACM, 683–686.
23. Ming-Yu Liu, Thomas Breuel, and Jan Kautz. 2017. Unsupervised image-to-image translation networks. In *Advances in Neural Information Processing Systems*. 700– 708.
24. Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. 2015. Deep Learning Face Attributes in the Wild. In *Proc. ICCV*.
25. Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley. 2017. Least squares generative adversarial networks. In *Proc. ICCV*. IEEE, 2813–2821.
26. Zhenyao Mo, John P Lewis, and Ulrich Neumann. 2004. Improved automatic caricature by feature normalization and exaggeration. In *ACM SIGGRAPH Sketches*. ACM, 57.
27. Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. 2017. Automatic differentiation in PyTorch. In *Proc. of NIPS*.
28. Lenn Redman. 1984. *How to draw caricatures*. Vol. 1. Contemporary Books Chicago, IL.
29. Ahmed Selim, Mohamed Elgharib, and Linda Doyle. 2016. Painting style transfer for head portraits using convolutional neural networks. *ACM Trans. Graph. (Proc. of SIGGRAPH)* 35, 4 (2016), 129.
30. Rupesh N Shet, Ka H Lai, Eran A Edirisinghe, and Paul WH Chung. 2005. Use of neural networks in automatic caricature generation: an approach based on drawing style capture. (2005).
31. Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).
32. Tamás Szirányi and Josiane Zerubia. 1997. Markov random field image segmentation using cellular neural network. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications* 44, 1 (1997), 86–89.
33. Yaniv Taigman, Adam Polyak, and Lior Wolf. 2016. Unsupervised cross-domain image generation. *arXiv preprint arXiv:1611.02200* (2016).
34. Chien-Chung Tseng and Jenn-Jier James Lien. 2007. Synthesis of exaggerative caricature with inter and intra correlations. In *Proc. ACCV*. Springer, 314–323.
35. Fei Yang, Lubomir Bourdev, Eli Shechtman, Jue Wang, and Dimitris Metaxas. 2012. Facial expression editing in video using a temporally-smooth factorization. In *Proc. CVPR*. IEEE, 861–868.
36. Zili Yi, Hao Zhang, Ping Tan, and Minglun Gong. 2017. Dualgan: Unsupervised dual learning for image-to-image translation. *arXiv preprint* (2017).
37. Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. 2017a. Unpaired Imageto-Image Translation using Cycle-Consistent Adversarial Networkss. In *Proc. ICCV*.
38. Jun-Yan Zhu, Richard Zhang, Deepak Pathak, Trevor Darrell, Alexei A Efros, Oliver Wang, and Eli Shechtman. 2017b. Toward multimodal image-to-image translation. In *Advances in Neural Information Processing Systems*. 465–476.
39. Shizhan Zhu, Cheng Li, Chen-Change Loy, and Xiaoou Tang. 2016. Unconstrained face alignment via cascaded compositional learning. In *Proc. CVPR*. 3409–3417.