计算机视觉课程报告——StarGAN论文分析及思考

曾京乐
21824058
航空航天学院
zjl_utopia@163.com

摘要

本次课程报告选择CVPR2018 ORAL文章StarGAN进行详细研读并做分析及思考。由于实验室最近主要研究GAN方面的课题，得知近期image-to-image translation方向主要是针对两个domain之间转换，主要算法主要有pix2pix、CycleGAN等，但需要训练多个模型，而对于多domain之间图像转换更好的思路是StarGAN，仅需一个模型就可以实现多domain任意对转换，且效果较好。通过对论文的进一步思考，结合最近实验室的人脸交换项目，我认为可以借鉴StarGAN的思路，实现人脸图片的多角度人脸生成算法，并编写代码实现。

1. 介绍

图像image-to-image translation任务是给定一张图像后，转换成另一风格的图像。比如人脸任务，就是做表情变换，肤色、发型等风格变换。

近期的算法对于多domain图像转换任务需要训练多个模型，如图 1 所示，训练耗时长且繁琐，对于变换一种状态效果可能较好，但当多种状态一起变换时效果不好。而StarGAN对于多domain图像转换任务仅需生成一个模型，就可以在多种状态同时变换，且当叠加多种状态时生成的效果也较好。

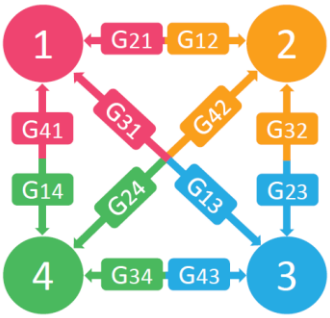

图 1 多domain交叉生成模型

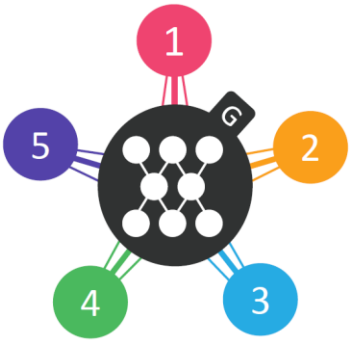

图 2 StarGAN

2. StarGAN

文章最开始提出了StarGAN，解决了多domain image-to-image translation任务，但只处理单个数据集来详细说明算法原理；之后由于要在多数据集上实施，提出了mask vector来解决不同数据集标签不一致的问题。

2.1. Multi-Domain Image-to-Image Translation

本节为了解决是如何仅用一个生成器，来实现multi-domain的图像转换任务的问题，在模型输入端除了图像，还引入了目标domain label *c*, *G(x, c)=y*。我们随机产生*c*，使得*G*能够经过学习后生成指定标签的图像。为了判断学习是否成功，我们引入了auxiliary classifier，他的作用是让鉴别器除了判断生成图像real/fake，还能判断生成图像所属的类别标签，*D:x={Dsrc(x)., Dcls(x)}*，图 3 阐述了整个训练过程。

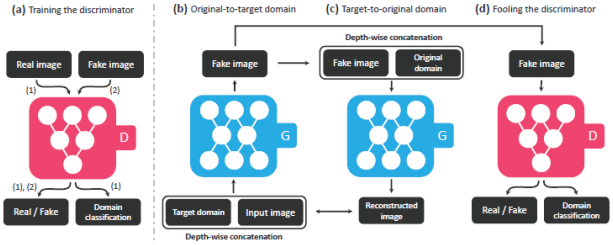

图 3 StarGAN训练单一数据集过程

整个StarGAN有两个模块，鉴别器D和生成器G。(a)D要学习如何区分生成图像real/fake，并且将real图像进行标签分类。(b)G将图像和目标domain label同时作为输入来生成fake图像。(c)G还要根据原图label对fake图像进行重构回原始图像，使重构的原始图像和真实原始图像差异尽可能小。(d)G要使生成图像尽可能真实，并且可以被D成功分类到对应标签。

对抗损失：为了使生成图像尽可能真实，我们采用对抗损失如下，和传统GAN对抗损失相同

$$\mathcal{L}_{adv} = \mathbb{E}_x [\log D_{src}(x)] + \mathbb{E}_{x,c}[\log (1 - D_{src}(G(x,c))] \quad (1)$$

Domain分类损失：对于给定图像x和目标domain label c，我们的目标是使x变换到y，并且能够被分类到标签c。我们提出了domain分类损失来优化D

$$\mathcal{L}^r_{cls} = \mathbb{E}_{x,c'}[-\log D_{cls}(c'|x)] \quad (2)$$

为了优化G，我们提出了针对fake图片的domain分类损失

$$\mathcal{L}^f_{cls} = \mathbb{E}_{x,c}[-\log D_{cls}(c|G(x,c))] \quad (3)$$

重构损失：通过最小化对抗损失和分类损失，G生成的图像将尽可能真实并且能够被准确分类到指定标签，但是不一定能保证和原图是同一个事物，因此我们引入了重构损失，和CycleGAN中一样，将fake图片和原图标签输入到G，使得生成的图像G(G(x))和x尽可能相等，这就保证了fake图像中还是和原图同样的事物

$$\mathcal{L}_{rec} = \mathbb{E}_{x,c,c'}[||x - G(G(x,c),c')||_1] \quad (4)$$

总体损失：最终，将上述所有损失综合考虑，组合称为生成器损失和鉴别器损失，来分别优化G和D

$$\mathcal{L}_D = -\mathcal{L}_{adv} + \lambda_{cls}\,\mathcal{L}^r_{cls},$$

$$\mathcal{L}_G = \mathcal{L}_{adv} + \lambda_{cls}\,\mathcal{L}^f_{cls} + \lambda_{rec}\,\mathcal{L}_{rec}, \quad (5)$$

在实际执行时，对抗损失没有采用传统GAN的损失，而是采用WGAN中的对抗损失来代替

$$\mathcal{L}_{adv} = \mathbb{E}_x[D_{src}(x)] - \mathbb{E}_{x,c}[D_{src}(G(x,c))] - \lambda_{gp}\,\mathbb{E}_{\hat{x}}[(||\nabla_{\hat{x}} D_{src}(\hat{x})||_2 - 1)^2], \quad (6)$$

2.2. 对多数据集训练

在单数据集中domain标签对全部图像数据统一，但多数据集中可能会存在图像标签与其他数据集图像标签不一致的情况，如CelebA数据集中只有人物的肤色、发型等标签，但RaFD数据集中包含人脸表情变化的标签，如高兴、生气等。对于这样的数据，domain label无法确定，对单纯包含所有标签计算损失会出现问题。如CelebA中图片虽然没有表情标签，但也会有高兴、生气等变化，这样就导致CelebA的domain label不能准确表示。

Mask Vector：为了解决这个问题，文章提出了mask vector m，使StarGAN能够不受不同数据集标签不同的影响，计算损失时仅针对该图片所在的数据集来计算，结果会更加准确。Mask vector m是一个n维的one-hot向量，代表了所有数据集的所有标签。其中Ci代表是否选中i数据集，是一个二进制序列，而计算损失时仅对Ci=1 的数据集进行计算，选择m中指定数据集的one-hot向量。

$$\tilde{c} = [c_1, ..., c_n, m] \quad (7)$$

训练策略：当在多数据集上训练StarGAN时，我们使用domain label c作为生成器的输入，此时生成器仅关注Ci=1 的标签。一方面生成器生成目标标签的fake图像，交给鉴别器判断是否为真，当鉴别器判断为真时再通过auiliary classifier对生成图像进行分类，与真实标签对比计算损失反向传播；另一方面生成器再对fake图像和原始标签作为输入，生成图像与原始图像对比满足差异最小，使生成的图像尽可能与原始图像相一致，且不改变图像背景区域。

3. 实验

由于本文后续会对自己的思考进行编写代码测试，因此对于StarGAN的实际效果本文就引用论文中的测试图片进行展示，效果如下图所示。

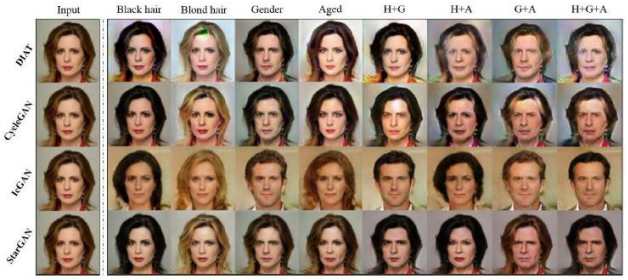

图 4 StarGAN实验数据

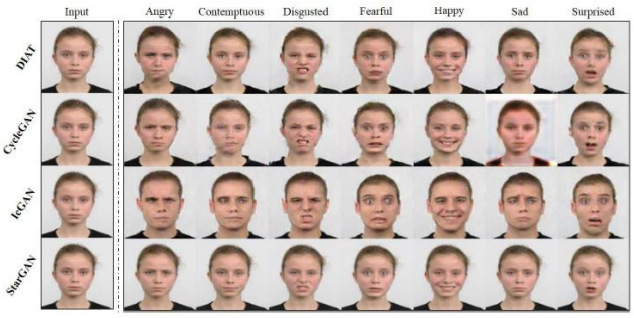

图 5 在RaFD数据集上实现人脸表情变换

第一列是原图，未经图像转换，第二列到第五列是CelebA中的生成的对应标签结果图片，后四列是RaFD的标签和CelebA标签混合后所得到的结果图片。经过DIAT、CycleGAN、IcGAN、StarGAN这四种算法对比后，可以看出DIAT效果相对较差；IcGAN生成图的质量较好，但和原图差异较大；CycleGAN算法对于但类别标签生成图像较好，但多标签综合生成图像最好的是StarGAN。

4. 分析与思考

在multi-domain image-to-image translation任务上，StarGAN巧妙的结合了监督学习分类任务中输入标签的思想，且在鉴别器的设计上也加入了分类问题常用的classifier思想，对图像做分类，与图像转换任务完美结合，使生成器能够自行学习对于输入图像将要转换的风格标签，仅用一个生成器完成了传统图像转换算法需要multi-cross生成器的问题。且文章对于multi-datasets情况下不同数据集分类标签不一致的问题仅仅引入了一个mask vector向量就完美解决，不可否认越简洁的数学思想越能使人们信服。

通过使用StarGAN算法，不仅可以解决人脸风格转换问题，还可以继续衍生到穿着服装、头戴手势等图像风格变换，甚至可以在多种说话人语音之间随意变换，可实践的领域可谓数不胜数。

由于我们实验室最近在研究人脸交换问题，二维图像人脸交换任务现有算法已经能够很好解决，但若扩展至三维，问题将变得复杂许多。除了算法的复杂度指数上升，最大的问题是三维人脸数据的缺乏，和二维人脸数据相比简直少之又少，因此，通过二维人脸生成三维人脸的算法就至关重要。

通过研读StarGAN论文受到的启发，虽然直接由二维人脸生成三维人脸问题不能马上解决，但我想通过借鉴该算法，首先实现利用单张二维人脸图像生成多角度人脸图像，而生成另一个不同的角度人脸图像可以通过pix2pix或CycleGAN来实现，但衍生出来进而生成任意多角度人脸图像不就可以通过StarGAN来实现吗。

思路 1：最初始的想法就此确立，首先利用CelebA人脸数据集，通过三维人脸特征点检测算法进行三维landmark计算，再利用相机矫正算法得到人脸的欧拉角，得到人脸姿态的水平方向偏航角，进而根据人脸偏航角的不同，将人脸朝向分为left、lefttop、top、righttop、right这四个方向，放入对应的文件夹中。将人脸数据集制作成满足官方StarGAN算法规定的格式后，就可以开始训练。

结果 1：训练完成后，输入人脸图像和对应五个标签来生成对应标签的人脸多角度图像，如图 6 所示

图 6 StarGAN训练人脸多角度图像

可以看出结果不尽如人意，人脸角度并没有发生明显变化。通过观察结果，我认为问题可能出在对于image-to-image translation问题中，转换前后的图像整体差异不应该很明显，而仅仅应该是图像风格细节的变化，针对这样的数据集才能训练收敛到较好结果。而本文是将人脸图像直接分为五个大类标签，StarGAN在训练挑选原始图像和目标图像时是随机选择， 可能会出现原始图像是左脸，但目标图像是右脸的情况，这样训练出来的效果就会很差。另一个可能存在的问题是不同角度类别数据集是连续的，比如top和lefttop数据集中人脸角度可能会有一些完全相同的情况，这样就会导致auiliary classifier分类不准确，导致训练的生成器效果不好。

思路 2：对于这两问题我准备循序渐进，先使用人脸角度变化较少两组人脸角度图像，且对于这两组人脸角度，要保证不会存在两个角度由重合或近似重合部分，这样auiliary classifier分类器才能准确分类。

改变数据集后再一次利用StarGAN进行训练，得到的结果如图 7 所示

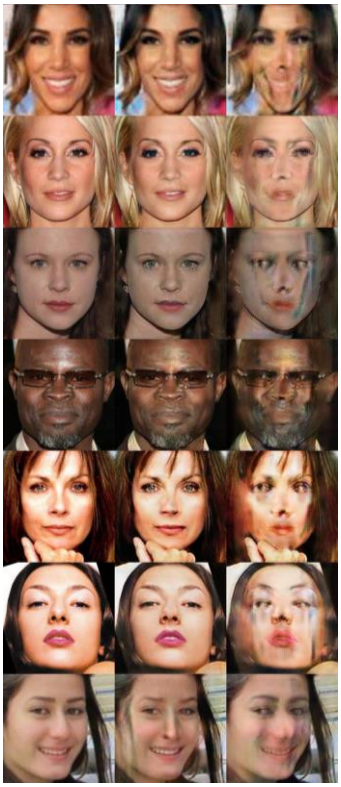

图 7 思路 2 利用StarGAN训练得到的图片

通过结果图可以看出，人脸角度变化依旧不太明显，左侧是输入图片，中间是top人脸角度图片，右侧是lefttop人脸角度图片，可以看出lefttop人脸发生扭曲变形，但仍未实现人脸角度变化。因此我开始怀疑StarGAN是否能够处理此类人脸角度变化的image-to-image translation问题，若不能原因到底是出在哪里，具体的原因还有待分析。明确原因后，再通过对网络结构、损失函数进行修改相信就能够提出另一套image-to-image translation的算法。

References

[1] Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, Jaegul Choo. StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation. arXiv PREPRINT arXiv: 1711.09020 2018.

[2] [2] A. Brock, T. Lim, J. M. Ritchie, and N. Weston. Neural photo editing with introspective adversarial networks. arXiv preprint arXiv:1609.07093, 2016. 3

[3] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D.Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In Advances in Neural Information Processing Systems (NIPS), pages 2672–2680, 2014.

[4] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR), pages 770–778, 2016.

[5] X. Huang, Y. Li, O. Poursaeed, J. Hopcroft, and S. Belongie. Stacked generative adversarial networks. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.

[6] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim. Learning to discover cross-domain relations with generative adversarial networks. In Proceedings of the 34th International Conference on Machine Learning (ICML), pages 1857–1865, 2017.

[7] D. Kingma and J. Ba. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.