

文章编号:1007-2780(2022)06-0746-12

基于双专用注意力机制引导的循环生成对抗网络

劳俊明¹, 叶武剑^{2*}, 刘怡俊², 袁凯奕¹

(1. 广东工业大学 信息工程学院, 广东 广州 510006;

2. 广东工业大学 集成电路学院, 广东 广州 510006)

摘要: 现有基于循环生成对抗网络的图像生成方法通过引入独立通用的注意力模块, 在无匹配图像转换任务中取得了较好的效果, 但同时也增加了模型复杂度与训练时间, 而且难以关注到图中关键区域的所有细节, 图像生成效果仍有提升的空间。针对上述问题, 提出一种基于双专用注意力机制引导的循环生成对抗网络(Dual-SAG-CycleGAN), 分别对生成器和判别器采用不同的注意力机制进行引导。首先, 提出一种名为 SAG(Special Attention-mechanism Guided)的专用注意力模块来引导生成器工作, 在提升生成图像质量的同时降低网络的复杂度; 然后, 对判别器采用基于 CAM(Class Activation Mapping)的专用注意力机制引导模块, 抑制生成器生成无关的噪声; 最后, 提出背景掩码的循环一致性损失函数, 引导生成器生成更加精准的掩码图, 更好地辅助图像转换。实验证明, 本文方法与现有同类模型相比, 网络模型参数量降低近 32.8%, 训练速度快 34.5%, KID 与 FID 最低分别可达 1.13 和 57.54, 拥有更高的成像质量。

关键词: 生成对抗网络; 无匹配图像转换; 专用注意力机制; 循环一致性损失; 图像生成

中图分类号: TP391.4 **文献标识码:** A **doi:** 10.37188/CJLCD.2021-0293

Cycle generative adversarial network guided by dual special attention mechanism

LAO Jun-ming¹, YE Wu-jian^{2*}, LIU Yi-jun², YUAN Kai-yi¹

(1. College of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China;

2. College of Integrated Circuit, Guangdong University of Technology, Guangzhou 510006, China)

Abstract: The existing image generation methods based on cycle generative adversarial network have achieved excellent results in unpaired image to image translation tasks by introducing a separate generic attention module, but they also increase the model complexity and training time, and it is difficult to focus on all the details of key regions in the image, and there is still room for improvement in image generation results. To solve these problems, this paper proposes a dual special attention-mechanism guided cycle generative adversarial network architecture for unpaired image transformation translation (Dual-SAG-CycleGAN). Firstly, to improve the quality of the generated images and reduce the complexity of the

收稿日期: 2021-11-24; **修订日期:** 2022-01-06.

基金项目: 广东省教育厅创新人才广东工业大学青年百人项目(No. 220413548); 广东省重点区域研究开发计划(No. 2018B010115002, No. 2018B010107003, No. 2018B030338001)

Supported by Guangdong Provincial Education Department Innovative Talents Guangdong University of Technology Youth 100 Project (No. 220413548); Key Area Research and Development Program of Guangdong Province (No. 2018B010115002, No. 2018B010107003, No. 2018B030338001)

*通信联系人, E-mail: yewjian@126.com

network, a special attention module called SAG (Special Attention-mechanism Guided) is proposed to guide the generator. Then, to suppress the generation of extraneous noise by the generator, a discriminator based on special attention mechanism of CAM (Class Activation Mapping) is introduced. Finally, a cyclic consistency loss function for the background mask is introduced to guide the network to generate a more accurate mask map, which can better aid for image translation. Experiments demonstrate that compared with similar existing models, our proposed method can reduce the maximum number of parameters by 32.8%, speed of train 34.5% faster, and generate higher quality images with a minimum KID and FID of 1.13 and 57.54, respectively.

Key words: generative adversarial networks; unpaired image-to-image translation; attention mechanism; cycle consistency loss; image generation

1 引言

图像到图像的转换任务一直以来都是人们的研究热点,其目标是建立图像源领域到目标领域的映射,被广泛应用于图像超分辨率重建^[1]、风格迁移^[2]、图像着色^[3]、图像去雾^[4]等多个领域。

2014年,基于博弈思维的生成对抗网络^[5]开辟了一个新的研究领域,大为促进了图像转换领域的发展。此后,诞生了一系列优秀的图像转换算法,例如 Pix2pix^[6]、DualGAN^[7]、DiscoGAN^[8]、UNIT^[9]、MUNIT^[10]、DRIT^[11]、CycleGAN^[12]等算法。其中, Pix2pix 算法属于有监督的图像转换算法,该算法需要匹配的图像组才能完成图像转换的任务,然而,匹配的数据集制作难度较大、成本较高。为了实现无配对图像之间的转换, DualGAN、DiscoGAN 以及 CycleGAN 等网络模型引入了循环一致性约束。实验结果表明,上述模型在无匹配的图像转换任务取得了较好的效果。尽管 CycleGAN 等算法在无匹配的图像转换任务表现优越,但是,由于生成器既需要维持图像背景区域不变,又要对目标前景进行转换,从而导致生成器对目标图像转换的效率低下。

近年来,注意力机制在深度学习中的应用引起了研究人员的高度重视,为解决 CycleGAN 网络等算法存在的问题,研究人员相继提出了基于注意力机制的无匹配图像转换算法 UAIT^[13]、Attention-GAN^[14]、SAT^[15]等,但上述算法将独立的通用注意力结构加到生成器网络中,使得模型的整体结构变得非常庞大,导致其训练时间变长,而且也难以关注到图中关键区域的所有细节。AttentionGAN^[16]提出了专用注意力机制与生成

器共享编码器的结构,虽然能够缩小整体的网络结构大小,但是,由于鉴别器没有注意力机制引导,从而引入背景元素的干扰,影响了生成图像的质量。U-GAT-IT^[17]提出了基于类激活 CAM^[18]注意力机制引导的生成器和鉴别器,虽然提升了成像质量,但整体网络的判别器需要从2个增加至4个,不仅大幅增加了模型的复杂度,并且其提出的基于 CAM 注意力机制引导的生成器不够优越,仍然会对整体图像进行修改,图像转换效率不高,图像生成质量仍然有提升的空间。

2 基本原理

为解决上述存在的问题,本文提出了一种新的用于无匹配图像转换任务的循环生成对抗网络(Dual-SAG-CycleGAN),采用不同的专用注意力模块分别引导生成器和判别器,达到提升生成图像质量同时降低模型复杂度的目的。

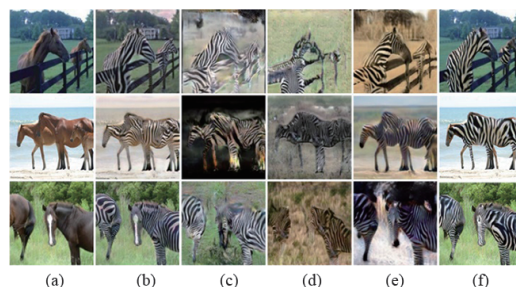


图1 不同模型在马转斑马任务上的表现。(a)原图;(b) CycleGAN;(c) UNIT;(d) MUNIT;(e) DRIT;(f) 本文模型。

Fig. 1 Performance of different models on the horse to zebra task. (a) Origin image; (b) CycleGAN; (c) UNIT; (d) MUNIT; (e) DRIT; (f) Ours.

本文的贡献具体如下:(1)提出了一种名为SAG的改进专用注意力模块来引导生成器工作,其中,生成器结构由内容图生成子模块与注意力掩码图生成子模块组成,两者高度共享参数权重。(2)为减少生成器对无关内容的生成以及削弱判别器对背景元素的判断,本文引入了基于CAM注意力模块引导的鉴别器。(3)为了生成更加精准的掩码图来辅助图像转换,本文提出了背景掩码循环一致性损失函数约束掩码生成器的工作。图1所示为本文所提出的模型以及现有对比模型在马转斑马任务上的生成效果图。

2.1 整体框架

图2是本文所提出的基于双专用注意力机制引导的循环生成对抗网络的示意图。本文在CycleGAN整体网络框架的基础上分别对生成器和判别器采用不同的注意力机制进行引导,其中,生成器采用本文提出的SAG专用注意力机制进行引导,判别器使用了基于CAM技术的专用注意力机制进行引导,两者构成紧密的整体。SAG专用注意力机制引导生成器修改图像的目标前

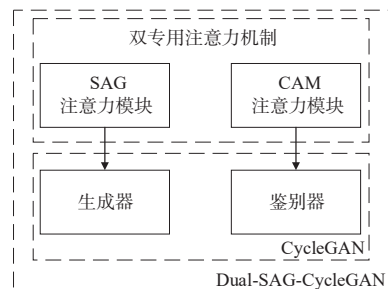


图2 网络示意图

Fig. 2 Network diagram

景,而CAM专用注意力机制则引导鉴别器关注两组图像之间差异最大的部分。

图3为本文的整体框架图。在图像与图像转换的任务中,通常有两组不同的图像领域,其目标是得到域与域之间的映射。为了达成该目标,网络中需要两组对称的生成器与鉴别器,如图3所示,一组生成器 G 的功能是将 X 域(马)图像生成至 Y 域(斑马)图像,该组的鉴别器 D_Y 则将生成器生成的 $G(Y)$ 域(斑马)与真实的 Y 域(斑马)区分开来。同样,第二组生成器 F 的功能是

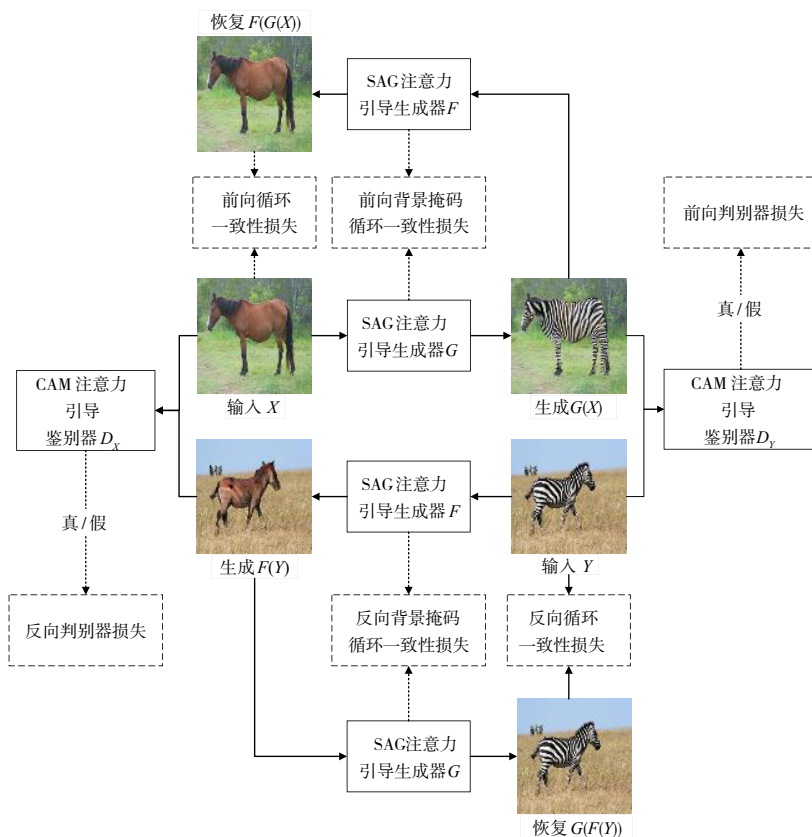


图3 整体框架示意图

Fig. 3 Overall framework

将Y域(斑马)图像生成至X域(马)图像,该组的鉴别器 D_X 则将生成器生成的 $F(X)$ 域(马)与真实的X域(马)区分开来。

为了防止网络在转换过程中,出现将X域(Y域)所有图像都映射到Y域(X域)中同一张图片的情况,本文与CycleGAN一样,引入了两个循环一致性损失。所谓循环一致性损失就是X域的图像经过生成器 G 转换至Y域后,仍然可以通过生成器 F 恢复至原来的X域中,并且经过恢复的图像与原图之间差异不能过大,即前向映射 $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$ 。同样,反向映射 $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$ 亦是如此,本文遵循了CycleGAN的整体基本原理。

由于在前向或反向循环映射过程中,需要修改各自的目标前景相同,即各自的背景掩码一致,故本文引入循环一致性损失函数约束掩码生成器的工作,使其生成更加精准的掩码图来辅助图像转换。通过两组对称的生成器与鉴别器动态博弈,最终得到相对理想的X域与Y域之间的映射。

2.2 基于SAG专用注意力引导的生成器

图4为本文所设计生成器 G 详细结构,其中,卷积函数后的数字分别为特征输出通道数、卷积核大小、步长、四边填充行数,受X. Chen^[14]以及T. Hao^[16]所启发,本文将X. Chen^[14]所设计的独立内容生成器和注意力掩码生成器与T. Hao^[16]共享编码器的思路结合在一起,提出了一个新的生成器结构。其中,内容生成子模块与注意力掩码生成子模块形成高度共享参数的一个整体结构,使得网络不仅共享编码器,并且也共享解码器的一部分。这样设计的优点以及依据在于,由于生成器最终需要修改目标前景并且维持背景图像不变,所以内容生成器与注意力掩码生成器之间具有非常相似的任务,区别在于内容生成器生成三维的目标前景图像矩阵,而注意力掩码生成器生成一维的目标前景掩码图像矩阵。采用内容生成器与注意力掩码生成器高度共享权重参数的方式不仅能够大幅降低模型的参数量以及复杂度,而且使得图像映射的效率更高,网络学习的速度也更快。

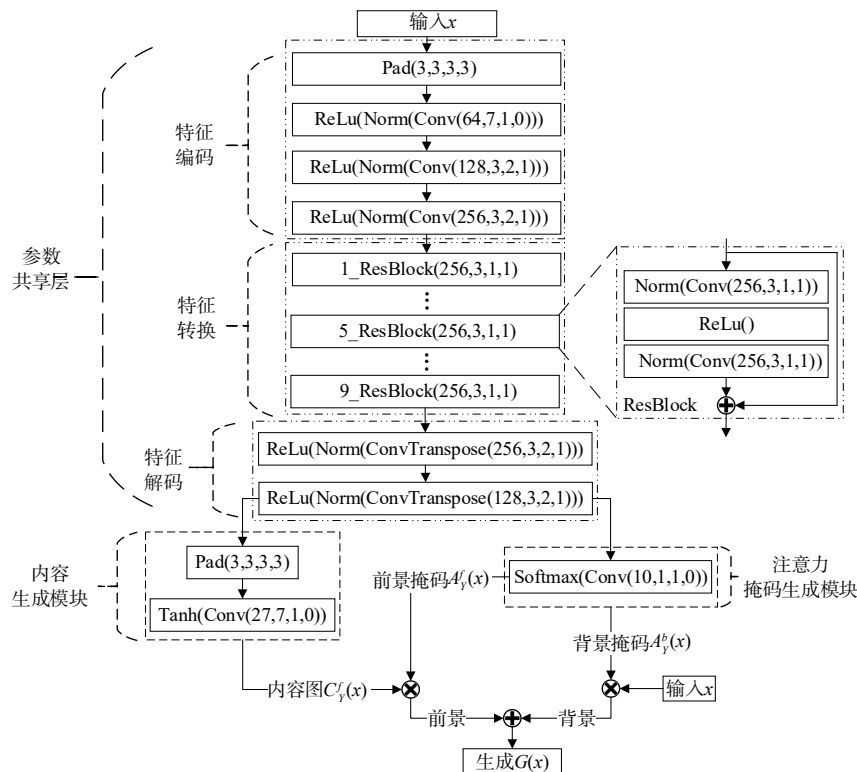


图4 基于专用注意力机制引导的生成器 G 结构

Fig. 4 Structure of generator G based on special attention mechanism-guide

与其他基于注意力机制引导的模型相比,本文设计的基于专用注意力引导的生成器高度共享模型参数的结构,模型计算复杂度更低,而且由于我们的内容生成子模块卷积输出 27 维图像矩阵,相较于其他模型直接输出 3 维图像矩阵来说,可计算的映射路径更多。对于复杂场景下的图像,由于注意力机制的引入,使得网络只需要专注于生成前景目标,所以映射更加简单,网络更容易收敛,其生成图像细节更好。

由图 4 可知,一个生成器能够生成内容图($C_Y^f(x)$ 或 $C_X^f(y)$)和注意力掩码图(前景掩码图 $A_Y^f(x)$ 或 $A_X^f(y)$ 以及背景掩码图 A_Y^b 或 A_X^b),内容图与前景注意力掩膜相乘产生前景图,背景注意力掩膜与输入图像相乘产生背景图,而前景图与背景图相加可以得到最终的输出图像,其中, x 和 y 代表两组不同的图像矩阵。因为生成器 G 与 F 具有相同的结构,故所生成的图像可由式(1)、式(2)表示,其中,我们将 N 设置为 10,则 $N-1=9$,表示输出 9 维的内容图像矩阵与相对应的前景掩码图像矩阵进行相乘运算后进行累加, b 为常数 1:

$$G(x) = \sum_{j=1}^{N-1} (C_Y^f(x) * A_Y^f(x)) + x * A_Y^b, \quad (1)$$

$$F(y) = \sum_{j=1}^{N-1} (C_X^f(y) * A_X^f(y)) + y * A_X^b, \quad (2)$$

同样,在 $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$ 以及 $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$ 图像的循环一致性映射中,生成器 F 与 G 生成的图像可由式(3)、式(4)表示:

$$G(F(y)) = \sum_{j=1}^{N-1} (C_Y^f(F(y)) * A_Y^f(F(y))) + F(y) * A_Y^b, \quad (3)$$

$$F(G(x)) = \sum_{j=1}^{N-1} (C_X^f(G(x)) * A_X^f(G(x))) + G(x) * A_X^b. \quad (4)$$

2.3 基于 CAM 专用注意力引导的鉴别器

为了防止判别器对生成图像的背景元素进行判断,减少生成器对无关或者虚假内容的生成,本文采用了 U-GAT-IT 中的鉴别器结构,如图 5 所示,其利用了辅助鉴别器基于类激活图 CAM 的原理对前景目标进行权重加权,影响最终鉴别器的输出,其中,卷积函数 Conv 后的参数分别为特征输出通道数、卷积核大小、步长、四边填充行数。

鉴别器的整体由辅助鉴别器以及最终鉴别器组成,两者同样高度共享权重参数。其中,辅

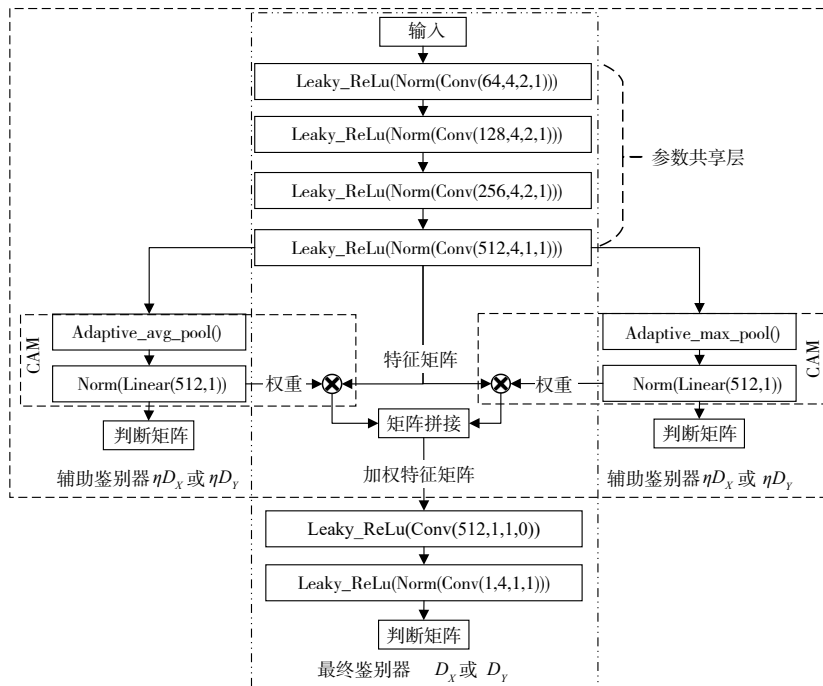


图 5 基于专用注意力机制引导的鉴别器结构

Fig. 5 Structure of discriminator based on special attention mechanism-guide

助鉴别器的构成与传统的判别器一致,经过多层卷积提取特征后,通过池化层计算后,再通过全连接层后输出大小为 $1 \times 1 \times 1$ 的判别结果矩阵,将两个 $1 \times 1 \times 1$ 的判别结果矩阵拼接在一起,形成 $2 \times 1 \times 1$ 大小的辅助鉴别判断矩阵输出,从全局的方式去判断图像的真伪,维持生成目标的整体形状。而最终鉴别器的构成则与 CycleGAN 的判别器相似,从局部的方式去判断图像中每一小块的真伪,其最终输出大小为 30×30 的判断结果矩阵,能够保持生成目标纹理的高细节、高分辨率。

最终鉴别器通过辅助鉴别器的卷积层、池化层以及全连接层计算后,从全局的方式利用 CAM 的原理来得出对目标前景加权的矩阵,再与相应的特征矩阵相乘后得到注意力特征矩阵,进而输入到最终判别器进行进一步的特征提取,使得最终判别器进一步收敛至判断前景目标而不是背景元素。

由图5可知,本文采用的基于注意力机制引导的鉴别器结构具有两组输出,其中, ηD_X 或 ηD_Y 为辅助鉴别器输出,它能够从全局的形式来判断图像的真伪,由辅助鉴别器构成的生成对抗损失函数的数学表达式由式(5)以及式(6)表示。其中, x 和 y 是真实的图像, $P_{\text{data}}(x)$ 和 $P_{\text{data}}(y)$ 表示真实图像的样本分布, $x \sim P_{\text{data}}(x)$ 和 $y \sim P_{\text{data}}(y)$ 表示样本 x 、 y 随机地从 P_{data} 分布中取出, E 则为求解数学期望。

$$L_{\text{CAM}}(G, \eta D_Y, X, Y) = E_{x \sim P_{\text{data}}(x)} \left[\left(1 - \eta D_Y(G(x)) \right)^2 \right] + E_{y \sim P_{\text{data}}(y)} \left[\eta D_Y^2(y) \right], \quad (5)$$

$$L_{\text{CAM}}(F, \eta D_X, X, Y) = E_{y \sim P_{\text{data}}(y)} \left[\left(1 - \eta D_X(F(y)) \right)^2 \right] + E_{x \sim P_{\text{data}}(x)} \left[\eta D_X^2(x) \right]. \quad (6)$$

要获得相似的样本分布,鉴别器 D 应当分辨出真实的图像与生成器 G 、 F 生成的虚假图像,也就是 $\eta D_Y(G(x))$ 及 $\eta D_X(F(y))$ 的值要接近 0, $\eta D_Y(y)$ 及 $\eta D_X(x)$ 接近 1, 即 $L_{\text{CAM}}(G, \eta D_Y, X, Y)$ 越大越好;而生成器 G 则应当使生成的图像尽量可以骗过鉴别器,即 $\eta D_Y(G(x))$ 及 $\eta D_X(F(y))$

的值要接近 1, 也即 $L_{\text{CAM}}(G, \eta D_Y, X, Y)$ 越小越好,该最优值为 0.5。

最终鉴别器输出所构成的生成对抗损失函数与 CycleGAN 一致,其数学表达式如式(7)和式(8)所示,类似的原理已在式(5)和式(6)处进行了详细的解释,故在此不再重复描述。

$$L_{\text{GAN}}(G, D_Y, X, Y) = E_{x \sim P_{\text{data}}(x)} \left[\left(1 - D_Y(G(x)) \right)^2 \right] + E_{y \sim P_{\text{data}}(y)} \left[D_Y^2(y) \right], \quad (7)$$

$$L_{\text{GAN}}(F, D_X, X, Y) = E_{y \sim P_{\text{data}}(y)} \left[\left(1 - D_X(F(y)) \right)^2 \right] + E_{x \sim P_{\text{data}}(x)} \left[D_X^2(x) \right]. \quad (8)$$

2.4 注意力引导下的循环一致性损失函数

为了防止网络在转换图像过程中,出现将 X 域(Y 域)所有图像都转换到 Y 域(X 域)中同一张图片的情况,我们引入了 CycleGAN 的循环一致性损失函数,其数学表达式由式(9)表示,其中, $\|F(G(x)) - x\|_1$ 及 $\|G(F(y)) - y\|_1$ 表示的是对模型循环复原后的图像与原图像进行 L1 范数的求解, $L_{\text{cycle}}(G, F)$ 的值越小越好。

$$L_{\text{cycle}}(G, F) = E_{x \sim P_{\text{data}}(x)} \left[\|F(G(x)) - x\|_1 \right] + E_{y \sim P_{\text{data}}(y)} \left[\|G(F(y)) - y\|_1 \right]. \quad (9)$$

由于在 $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$ 循环一致映射中,本文希望图像能够从 X 域转换到 Y 域的背景掩码图与从 Y 域恢复至 X 域的背景掩码图保持一致,这样注意力掩码生成子模块能引导生成器更加精确地修改需要转换的目标。受循环一致性损失函数所启发,本文提出了针对于背景掩膜的循环一致性损失函数,其数学表达式如式(10)所示,该损失值越小越好。

$$L_{\text{Att}_{\text{cycle}}}(A_Y^b, A_X^b) = E_{x \sim P_{\text{data}}(x)} \left[\|A_Y^b(x) - A_X^b(G(x))\|_1 \right] + E_{y \sim P_{\text{data}}(y)} \left[\|A_X^b(y) - A_Y^b(F(y))\|_1 \right]. \quad (10)$$

2.5 优化目标

上文介绍了各部分的损失函数,而模型完整的损失函数方程由 6 部分组成,如式(11)所示,分

别为 X 域、 Y 域最终鉴别器的生成对抗损失; X 域、 Y 域辅助鉴别器的生成对抗损失; 整体图像的循环一致性损失以及背景掩码的循环一致性损失。

$$\begin{aligned} L(G, F, D_X, D_Y, \eta D_X, \eta D_Y, A_Y^b, A_X^b) = & L_{\text{GAN}}(G, D_X, X, Y) + L_{\text{GAN}}(G, D_Y, X, Y) + \\ & L_{\text{CAM}}(G, \eta D_Y, X, Y) + L_{\text{CAM}}(F, \eta D_X, X, Y) + \\ & \alpha \cdot L_{\text{cycle}}(G, F) + \lambda \cdot L_{\text{Att}_{\text{cycle}}}(A_Y^b, A_X^b). \end{aligned} \tag{11}$$

在图像转换任务中, 要获得相似的样本分布, 生成器 G 、 F 应当使生成的图像可以骗过鉴别器, 而最终鉴别器 D_X 、 D_Y 以及辅助鉴别器 ηD_X 、 ηD_Y 则应当准确分辨出真实的图像与由生成器 G 、 F 生成的虚假图像, 因此, 生成器与鉴别器形成了对抗的关系。对于鉴别器 D_X 、 D_Y 、 ηD_X 以及 ηD_Y , 本文希望其辨别真实图像与生成器生成虚假图像的能力得到最大的提升, 即为最优的鉴别器。在最优鉴别器的条件下, 生成器 G 、 F 的优化目标是 minimized 图像的真实分布与模型生成图像分布之间的 JS 散度, 以获得最优的图像转换映射路径, 同时本文希望循环一致性损失函数值最小。因此, 本文网络的整体优化目标方程如式 (12) 所示:

$$\begin{aligned} \arg \min_{(G, F, A_X^b, A_Y^b)} \max_{(D_X, D_Y, \eta D_X, \eta D_Y)} L(G, \\ F, D_X, D_Y, \eta D_X, \eta D_Y, A_Y^b, A_X^b). \end{aligned} \tag{12}$$

3 实验分析

3.1 基准模型

本文对两大类无匹配图像转换算法进行对比, 一类为无注意力机制引导的图像转换模型, 包括 DualGAN^[7]、DiscoGAN^[8]、UNIT^[9]、MUNIT^[10]、DRIT^[11]、CycleGAN^[12]、RA^[19]; 另一类为有注意力机制引导的图像转换模型, 主要有 UAIT^[13]、Attention-GAN^[14]、SAT^[15]、Attention-GAN^[16]、U-GAT-IT^[17]、DA-GAN^[20]、TransGa-Ga^[21]等模型。

3.2 数据集

本文采用 3 组无匹配的图像数据集对模型进行训练与测试, 各个数据集的详细情况如表 1 所示。

表 1 各数据集详细信息

Tab. 1 Details of each dataset

数据集	训练集/张	测试集/张	总计/张
Horse2Zerba ^[12]	2 041	260	2 661
Apple2Orange ^[12]	2 016	514	2 530
Selfie2Anime ^[17]	6 800	200	7 000

3.3 参数设置

在训练阶段, 本文对训练集图像进行随机水平翻转和裁剪为 256×256 大小的操作。在测试阶段, 本文将测试集图像缩放至 256×256 大小。与 CycleGAN 一样, 本文使用 Adam optimizer^[12] 进行优化, 动量参数组设置为 $\beta_1 = 0.5$, $\beta_2 = 0.999$, Batchsize 设置为 2, 学习率为 0.000 2, 采用线性衰减策略。本文的超参数设置为 $\alpha = 12.0$, $\lambda = 1.0$ 。

3.4 评价指标

3.4.1 生成图像质量评价

Kernel Inception Distance (KID)^[22] 指标通过计算原始表征之间最大均值差异的平方来度量真实样本与假样本之间的差异, 越低的 KID 参数表示两组样本之间越相似。Frchet Inception Distance (FID)^[23] 指标使用 Inception V3 模型来提取图像的高级语义信息, 通过计算真实图像与生成图像提取特征向量后的均值以及协方差距离来衡量生成对抗网络生成图像的质量以及真假图像的相似性。当生成的图像与真实的图像特征越相近时, FID 数值越小。

3.4.2 模型复杂度评价

浮点运算 Floating Point Operations (FLOPs) 以及乘加运算 Multiply Accumulate Operations (MACs) 是常用的模型复杂度统计指标, 它们能够统计数据通过网络模型所需要计算量的大小, 即启用该模型时所需要的计算力。Parameters 模型参数量也是描述模型复杂度的指标之一, Times 为模型运行时实际消耗时间而 Memory 为模型训练时占用的实际显存空间, 三者的数值越小, 代表模型越优越。

3.5 消融实验

图 6 为有无本文专用注意力机制引导的生成器、鉴别器以及背景掩码的循环一致性损失函数对生成图像的影响。在上述选项中, 无本文专用

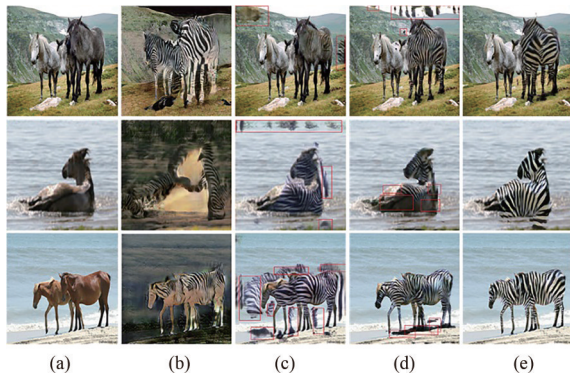


图6 有无本文生成、鉴别器以及背景掩码循环一致性损失函数对生成图像质量的影响。(a)原图;(b)无专用注意力引导的生成器;(c)无专用注意力引导的鉴别器;(d)无背景掩码循环一致性损失函数;(e)三者都有。

Fig. 6 Effect of the generated image's quality that with or without our generator, discriminator and cycle consistency loss function background mask. (a) Original image; (b) Generator without special attention-mechanism guided; (c) Discriminator without special attention-mechanism guided; (d) Without cycle consistency loss function of background mask; (e) All of three factors.

注意力机制引导的生成器、鉴别器分别表示使用CycleGAN的生成器、鉴别器,而无背景掩码的循环一致性损失函数则表示不引入本文所设计的背景掩码的循环一致性损失函数。

由图6(b)可以观察到,当不使用本文的专用注意力引导的生成器时,图像的背景会被大幅度地修改,且前景目标的生成质量非常低。而当不使用本文的专用注意力引导的鉴别器时,在训练过程中,当鉴别器无法分辨生成器所生成的前景目标时,就会通过背景元素进行判断,这时候鉴别器也会引导生成器生成背景元素,如图6(c)所示,由于判别器对背景中的天空和海水进行判断,所以引导了生成器生成虚假的内容。由图6(d)观察到,当不引入背景掩码的循环一致性损失函数时,注意力掩码器所生成的背景掩码图会不完整地覆盖前景目标又或者覆盖无关的背景区域,从而使得生成器生成虚假的背景元素又或生成不完整的前景目标。

由图6(e)可以观察到,使用本文的专用注意力机制引导生成器、鉴别器以及有背景掩码循环一致性损失函数时模型生成的图像质量最好。

表2给出了有无专用注意力引导的生成器、鉴别器以及背景掩码循环一致性损失函数以及对生成图像质量的量化参数FID的影响。在本项测试中,FID越小代表图像质量越高,并且生成的图像与真实的图像越相似,可以观察到,当模型三者都拥有时能够取得最小的FID指数。

表2 在马转斑马任务上模型的消融研究

Tab. 2 Ablation studies of models on the horse-to-zebra task

设置	无注意力引导生成器	无注意力引导鉴别器	无背景掩码循环一致性损失	三者都有
FID	107.23	92.24	86.05	57.54

图7给出了不同背景掩码循环一致性损失函数的 λ 系数在马转斑马任务上对生成图像质量的影响。由图7可以观察到,当 $\lambda=0.0$ 时,模型生成的前景目标以及背景掩码不够完整,又或者将背景元素引入到了生成图像中,降低了图像的质量;当 $\lambda=1.0$ 时,模型生成的前景目标以及背景掩码最优,此时的生成图像质量最好;当 $\lambda=10.0$ 时,由于过大的权重,降低了模型生成图像的清晰度以及产生了失真的色彩,并且使得模型生成了过大的前景目标掩码图。

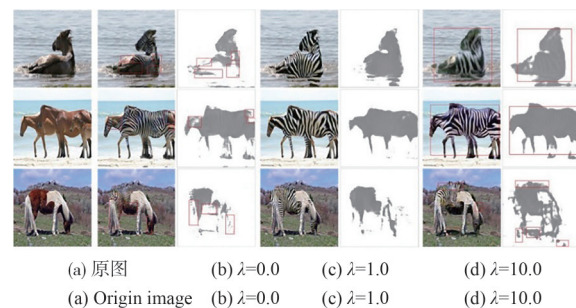


图7 不同的 λ 系数在马转斑马任务上对生成图像质量的影响

Fig. 7 Effect of different λ -factors on the quality of the generated images on the horse-to-zebra task

表3给出了不同背景掩码循环一致性损失函数的 λ 系数对模型生成图像质量量化参数FID的影响,其中,FID参数越低越好。

由表3可以观察到,当 $\lambda=1.0$ 时,模型生成图像的FID参数最低,此时的图像生成质量最优。

图8和图9展示的是本文模型在不同的转换

表 3 不同 λ 系数对图像生成质量的影响

Tab. 3 Effect of different λ -coefficients on the quality of the generated images

设置	FID
$\lambda = 0.0$	86.05
$\lambda = 1.0$ (本文)	57.54
$\lambda = 10.0$	79.58

任务上注意力掩码图生成器所生成的背景掩码图以及生成器生成的最终效果图。



图 8 本文模型在苹果转橘子和橘子转苹果任务上生成的注意力掩码以及生成效果

Fig. 8 Attention mask and images generated by ours model on the apple to orange and orange to apple tasks



图 9 本文模型在马转斑马和斑马转马任务上生成的注意力掩码以及生成效果

Fig. 9 Attention mask and images generated by ours model on the horse to zebra and zebra to horse tasks

由图 8 和图 9 可以观察到,本文模型能够准确地识别到需要转换的前景目标,做到修改前景目标的同时能够维持原背景图像不变,最大程度地提升了生成图像的质量。

3.6 量化结果

图 10 是不同模型在马与斑马互相转换任务上的效果图。由图 10 可以观察到,RA、DiscoGAN、UNIT、DualGAN、CycleGAN 等模型在生成前景目标的同时也对背景进行了转换,这大幅降低了模型所生成的图片质量。在复杂的无匹配图像转换任务上,这类模型所生成的图像质量

更低,而带有注意力机制的模型,如 UAIT、AttentionGAN 等模型则只改变或者轻微改变背景元素,提升了生成图像的质量。

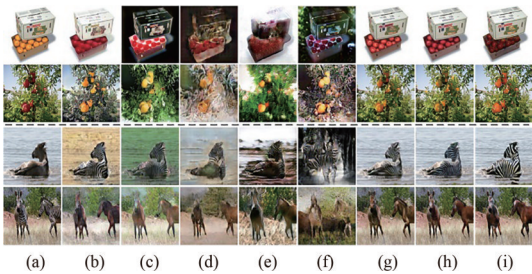


图 10 不同模型在马与斑马互换任务上的表现。(a) 原图; (b) CycleGAN; (c) RA; (d) DiscoGAN; (e) UNIT; (f) DualGAN; (g) UAIT; (h) AttentionGAN; (i) 本文模型。

Fig. 10 Performance of different models on the horse-zebra interchange task. (a) Original image; (b) CycleGAN; (c) RA; (d) DiscoGAN; (e) UNIT; (f) DualGAN; (g) UAIT; (h) AttentionGAN; (i) Ours.

由于 UAIT 等模型引入了独立的通用注意力机制结构,大幅增加了模型的复杂度。虽然 AttentionGAN 模型提出的注意力机制与生成器共享参数层的结构,减少了模型的参数,但是,由于无基于注意力机制引导的鉴别器,使得生成器生成了许多虚假的幻纹,而本文提出的基于专用注意力机制引导的生成对抗网络不仅能够抑制和减少虚假的幻纹的生成,并且进一步缩小了模型的结构以及提升了生成图像的质量。

图 11 是不同模型在自拍照与漫画脸转换任



图 11 不同模型在自拍与动漫互换任务上的表现。(a) 原图; (b) CycleGAN; (c) UNIT; (d) MUNIT; (e) DRIT; (f) U-GAT-IT; (g) AttentionGAN; (h) 本文模型。

Fig. 11 Performance of different models on the selfie-anime interchange task. (a) Original image; (b) CycleGAN; (c) UNIT; (d) MUNIT; (e) DRIT; (f) U-GAT-IT; (g) AttentionGAN; (h) Ours.

务上的效果图,可以观察到,本文的模型不仅能够仅转换前景目标区域图像,亦可以修改整张图像。由图 11(h)的脸型、头发可以观察到,本文的模型能够最大程度地保留原有的语义信息的基础上来对图像进行转换。

表 4 是不同模型在马与斑马以及苹果与橘子转换任务的 KID 指标,表 5 是不同模型在自拍与

漫画脸转换任务的 KID 指标,表 6 是不同模型在马与斑马转换任务上的 FID 指标,其中, KID 与 FID 的值越小,代表网络生成的图像与真实的图像越相似,生成图像质量越高,即数值越小越好。由表 4~6 可知,本文的模型在不同的转换任务上都能取得最优的 KID 以及 FID 参数,即本文模型生成的图像与真实的图像最接近。

表 4 不同模型在不同任务上的 $KID \times 100 \pm \text{std.} \times 100$ 指标

Tab. 4 KID $\times 100 \pm \text{std.} \times 100$ metrics for different models on different tasks

模型	苹果转橘子	橘子转苹果	马转斑马	斑马转马
DiscoGAN ^[8]	18.34 ± 0.75	21.56 ± 0.80	13.68 ± 0.75	16.60 ± 0.50
RA ^[19]	12.75 ± 0.49	13.84 ± 0.78	10.16 ± 0.12	10.97 ± 0.26
DualGAN ^[7]	13.04 ± 0.72	12.42 ± 0.88	10.38 ± 0.31	12.86 ± 0.50
UNIT ^[9]	11.68 ± 0.43	11.76 ± 0.51	11.22 ± 0.24	13.63 ± 0.34
CycleGAN ^[12]	8.48 ± 0.53	9.82 ± 0.51	10.25 ± 0.25	11.44 ± 0.38
UAIT ^[13]	6.44 ± 0.69	5.32 ± 0.48	6.93 ± 0.27	8.87 ± 0.26
AttentionGAN ^[16]	10.03 ± 0.66	4.38 ± 0.42	2.03 ± 0.64	6.48 ± 0.51
Ours	7.28 ± 0.71	3.02 ± 0.29	1.13 ± 0.23	5.06 ± 0.60

表 5 不同模型在自拍转动漫任务上的 $KID \times 100 \pm \text{std.} \times 100$ 指标

Tab. 5 KID $\times 100 \pm \text{std.} \times 100$ metrics for different models on selfie to anime task

模型	自拍转动漫
CycleGAN ^[12]	13.08 ± 0.49
UNIT ^[9]	14.71 ± 0.59
MUNIT ^[10]	13.85 ± 0.41
DRIT ^[11]	15.08 ± 0.62
U-GAT-IT ^[17]	11.61 ± 0.57
AttentionGAN ^[16]	12.14 ± 0.43
Ours	10.01 ± 0.29

表 7 为基于注意力机制引导的各个模型的复杂度指标。由表 7 可知,与文献[14]相比,本文的网络模型参数量降低近 32.8%,训练速度快 34.5%。

表 7 基于注意机制引导的不同模型复杂度分析

Tab. 7 Complexity analysis of different models guided by attention mechanism

模型	参数量/M	浮点运算量/G	乘法累加运算量/G	训练时间(100张)/s	占用显存(1批量)/MiB
X. Chen ^[14]	43.94	102.01	204.02	55	5 069
T. Hao ^[16]	29.22	74.67 G	149.31	47	5 463
Ours	29.49	65.50	131.00	36	3 669

表 6 不同模型在马与斑马转换任务上的 FID 指标

Tab. 6 FID metrics for different models on horse to zebra task

模型	马转斑马
UNIT ^[9]	241.13
CycleGAN ^[12]	109.36
DA-GAN ^[20]	103.42
TransGaGa ^[21]	95.81
SAT ^[15]	98.90
AttentionGAN ^[16]	68.55
Ours	57.54

得益于本文所提出的注意力掩码生成子模块与内容生成子模块高度共享参数的生成器结构,本文的模型复杂度最小,占用硬件显存资源最少,训练速度最快。

4 结 论

本文提出了一种新的用于无匹配图像转换任务的基于双专用注意力机制引导的循环生成对抗网络(Dual-SAG-CycleGAN)。生成器由端到端的、基于专用注意力机制引导的前背景掩码图生成子模块和内容图生成子模块所构成,两者高度共享权重参数。其中,前景、背景掩码图生成子模块的组成仅需要在原有的内容生成器结构中添加一个额外的卷积层以及 Softmax 激活函数。为了引导前背景掩码图生成子模块生成更加精准的掩码图以抑制生成器生成降低生成图像质量的无关背景内容,我们提出了背景掩码循环一致性损失函数并引入了基于 CAM 专用注意力机制引导的判别器结构。

实验证明,本文结构不仅能够实现自动转换图像相关目标前景的功能,并且拥有更短的训练时间、更低的硬件开销、更好的生成图像质量。本文方法与现有同类模型相比,网络模型参数量降低近 32.8%,训练速度快 34.5%,KID 与 FID 最低分别可达 1.13、57.54,拥有更高的成像质量。

但是,由于本文的基于 CAM 专用注意力机制引导的判别器不能非常精确地关注到感兴趣的区域,在某些复杂的场景上会以背景物作为判别依据,所以会导致生成器生成虚假的幻纹,这也是限制生成图像质量进一步提升的重要因素,所以,在后续的工作中,我们会重新设计一种新的判别器结构,设计一种识别能力更加精准的注意力机制模块来引导判别器,以提升判别器识别感兴趣区域的能力,使得网络生成图像的质量有更进一步提升。

参 考 文 献:

- [1] 陈宗航,胡海龙,姚剑敏,等. 基于改进生成对抗网络的单帧图像超分辨率重建[J]. 液晶与显示,2021,36(5):705-712.
CHEN Z H, HU H L, YAO J M, *et al.* Single frame image super-resolution reconstruction based on improved generative adversarial network [J]. *Chinese Journal of Liquid Crystals and Displays*, 2021, 36(5): 705-712. (in Chinese)
- [2] KARRAS T, LAINE S, AILA T. A style-based generator architecture for generative adversarial networks [C]//*IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach: IEEE, 2019: 4396-4405.
- [3] CAO Y, ZHOU Z M, ZHANG W N, *et al.* Unsupervised diverse colorization via generative adversarial networks [C]//*European Conference on Machine Learning and Knowledge Discovery in Databases*. Skopje: Springer, 2017: 151-166.
- [4] 周文舟,范晨,胡小平,等. 多尺度奇异值分解的偏振图像融合去雾算法与实验[J]. 中国光学,2021,14(2): 298-306.
ZHOU W Z, FAN C, HU X P, *et al.* Multi-scale singular value decomposition polarization image fusion defogging algorithm and experiment [J]. *Chinese Optics*, 2021, 14(2): 298-306. (in Chinese)
- [5] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, *et al.* Generative adversarial networks [C]//*Advances in Neural Information Processing Systems (NeurIPS)*. Cambridge, MA: MIT Press, 2014: 2672-2680.
- [6] ISOLA P, ZHU J Y, ZHOU T H, *et al.* Image-to-image translation with conditional adversarial networks [C]//*IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu: IEEE, 2017: 5967-5976.
- [7] YI Z L, ZHANG H, TAN P, *et al.* DualGAN: unsupervised dual learning for image-to-image translation [C]//*IEEE International Conference on Computer Vision (ICCV)*. Venice: IEEE, 2017: 2868-2876.
- [8] KIM T, CHA M, KIM H, *et al.* Learning to discover cross-domain relations with generative adversarial networks [C]//*Proceedings of the 34th International Conference on Machine Learning (ICML)*. Sydney: ACM, 2017: 1857-1865.
- [9] LIU M Y, BREUEL T, KAUTZ J. Unsupervised image-to-image translation networks [C]//*Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach: ACM, 2017: 700-708.
- [10] HUANG X, LIU M Y, BELONGIE S, *et al.* Multimodal unsupervised image-to-image translation [C]//*15th European Conference on Computer Vision (ECCV)*. Munich: Springer, 2018: 179-196.

- [11] LEE H Y, TSENG H Y, HUANG J B, *et al.* Diverse image-to-image translation via disentangled representations [C]//15th *European Conference on Computer Vision (ECCV)*. Munich: Springer, 2018: 36-52.
- [12] ZHU J Y, PARK T, ISOLA P, *et al.* Unpaired image-to-image translation using cycle-consistent adversarial networks [C]//*IEEE International Conference on Computer Vision (ICCV)*. Venice: IEEE, 2017: 2242-2251.
- [13] MEJJATI Y A, RICHARDT C, TOMPKIN J, *et al.* Unsupervised attention-guided image-to-image translation [C]//*Advances in Neural Information Processing Systems (NeurIPS)*. Cambridge, MA: MIT Press, 2018: 3697-3707.
- [14] CHEN X Y, XU C, YANG X K, *et al.* Attention-GAN for object transfiguration in wild images [C]//15th *European Conference on Computer Vision (ECCV)*. Munich: Springer, 2018: 167-184.
- [15] YANG C, KIM T, WANG R Z, *et al.* Show, attend, and translate: unsupervised image translation with self-regularization and attention [J]. *IEEE Transactions on Image Processing*, 2019, 28(10): 4845-4856.
- [16] TANG H, XU D, SEBE N, *et al.* Attention-Guided Generative Adversarial Networks for Unsupervised Image-to-Image Translation [C]//*International Joint Conference on Neural Networks (IJCNN)*. Budapest: IEEE, 2019: 1-8.
- [17] KIM J, KIM M, KANG H, *et al.* U-GAT-IT: unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation [C]//*International Conference on Learning Representations (ICLR)*. Berlin: ICLR, 2020.
- [18] ZHOU B L, KHOSLA A, LAPEDRIZA A, *et al.* Learning deep features for discriminative localization [C]//*IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas: IEEE, 2016: 2921-2929.
- [19] WANG F, JIANG M Q, QIAN C, *et al.* Residual attention network for image classification [C]//*IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu: IEEE, 2017: 6450-6458.
- [20] MA S, FU J L, CHEN C W, *et al.* DA-GAN: Instance-level image translation by deep attention generative adversarial networks [C]//*IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Salt Lake City: IEEE, 2018: 5657-5666.
- [21] WU W, CAO K D, LI C, *et al.* TransGaGa: geometry-aware unsupervised image-to-image translation [C]//*IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach: IEEE, 2019: 8004-8013.
- [22] BIŃKOWSKI M, SUTHERLAND D J, ARBEL M, *et al.* Demystifying MMD GANs [C]//*International Conference on Learning Representations (ICLR)*. Berlin, German: ICLR, 2018: 871-878.
- [23] HEUSEL M, RAMSAUER H, UNTERTHINER T, *et al.* GANs trained by a two time-scale update rule converge to a local nash equilibrium [C]//*Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach: ACM, 2017: 6629-6640.

作者简介:



劳俊明(1996—),男,广东茂名人,硕士研究生,2015年于五邑大学获得学士学位,主要从事计算机视觉与深度学习方面的研究。E-mail: 1135804550@qq.com



叶武剑(1987—),男,广东韶关人,博士,讲师,2015年于檀国大学(Dankook University)获得博士学位,主要从事计算机视觉与深度学习、机器学习应用等方面的研究。E-mail: yewjian@126.com