软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2021,32(9):2755−2768 [doi: 10.13328/j.cnki.jos.005986]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

面向图像场景转换的改进型生成对抗网络*

肖进胜[1], 周景龙[1], 雷俊锋[1], 李 亮[1], 丁 玲[2], 杜治一[1]

[1](武汉大学 电子信息学院,湖北 武汉 430072)

[2](湖北第二师范学院 计算机学院,湖北 武汉 430205)

通讯作者: 肖进胜, E-mail: xiaojs@whu.edu.cn

摘 要: 设计了新的生成器网络、判决器网络以及新的损失函数,用于图像场景转换.首先,生成器网络采用了带跨层连接结构的深度卷积神经网络,其中,多个跨层连接以实现图像结构信息的共享;而判决器网络采用了多尺度全域卷积网络,多尺度判决器可以区分不同尺寸下的真实和生成图像.同时,对于损失函数,该算法借鉴其他算法提出了 4 种损失函数的组合,并通过实验对比证明了新损失函数的有效性,包括 GAN 损失、L_1 损失、VGG 损失、FM 损失.从实验结果显示,该算法能够实现多种转换,且转换后图像的细节保留较为完整,生成图像较为真实,明显消除了块效应.

关键词: 图像生成;深度学习;生成对抗网络;跨层连接;场景转换

中图法分类号: TP183

中文引用格式: 肖进胜,周景龙,雷俊锋,李亮,丁玲,杜治一.面向图像场景转换的改进型生成对抗网络.软件学报,2021,32(9): 2755−2768. http://www.jos.org.cn/1000-9825/5986.htm

英文引用格式: Xiao JS, Zhou JL, Lei JF, Li L, Ding L, Du ZY. Improved generative adversarial network for image scene transformation. Ruan Jian Xue Bao/Journal of Software, 2021,32(9):2755−2768 (in Chinese). http://www.jos.org.cn/1000-9825/5986.htm

Improved Generative Adversarial Network for Image Scene Transformation

XIAO Jin-Sheng[1], ZHOU Jing-Long[1], LEI Jun-Feng[1], LI Liang[1], DING Ling[2], DU Zhi-Yi[1]

[1](Electronic Information School, Wuhan University, Wuhan 430072, China)

[2](College of Computer, Hubei University of Education, Wuhan 430205, China)

Abstract: This study designs a new generator network, a new discriminator network, and a new loss function for image scene conversion. First, the generator network uses a deep convolutional neural network with a skip connection structure, in which multi-skip connection is used to share the structure information of the image. For the discriminator network, it uses a multi-scale global convolutional network which can distinguish between real and generated images of different sizes. At the same time, the new loss function is a combination of four loss functions referring to other algorithms, including GAN loss, L_1 loss, VGG loss, and feature matching loss. Moreover, the validity of the new loss function is demonstrated through experimental comparisons. The experimental results show that the proposed algorithm can achieve multi-image transformations, and the details of generated images are preserved completely, the generated image is more realistic, and the block effect is obviously eliminated.

Key words: image generation; deep learning; generative adversarial networks; skip connection; scene conversion

许多计算机视觉问题可以被看作是一个图像到图像的翻译问题,是映射一个域中的映像到另一个域中的对应映像,实际上都是像素到像素之间的映射.例如:超分辨率可以认为是将低分辨率图像映射到相应的高分辨

* 基金项目: 国家重点研发计划(2017YFB1302401); 国家自然科学基金(61471272)

Foundation item: National Key Research and Development Program of China (2017YFB1302401); National Natural Science Foundation of China (61471272)

收稿时间: 2019-07-19; 修改时间: 2019-09-03, 2019-10-21; 采用时间: 2019-11-19

率图像的问题,而图像着色可以看作是将灰度图像映射到相应的彩色图像.这个问题可以在有监督和无监督的学习环境中进行研究.在无监督学习中,只有两组独立的图像,其中一组图像组成一个域,另一个域包含另一组图像,但训练图像不匹配,即不是成对的训练集.由于缺乏相应的图像,无监督的图像到图像转换问题更难考虑也更难实现.在有监督学习中,可在不同的域中训练配对相应的图像[1,2],有监督学习能够使生成图像与输入图像像素之间的映射关系更加准确,能够避免类似无监督学习中出现的生成图像不可控的现象.

利用卷积神经网络(convolutional neural networks,简称 CNN)进行有监督学习,在生成图像时也需要最小化损失函数,并作为网络调优的标准.然而,在采取了这种方法时,要求 CNN 尽量减少预测图像与真实图像之间的欧氏距离,它可能会产生模糊的结果[3,4],其原因是欧式距离通过平均所有像素的输出而导致模糊.因此,要让 CNN 网络针对特定的转换任务就需要制定特定的损失函数,但这是一个棘手的难题.如果可以指定网络只有一个高层次的目标,比如“使生成图像难辨真伪”,然后自动学习一个损失函数以实现此目标,这种方式也就是生成对抗网络的思路(generative adversarial nets,简称 GAN)[5−8].GAN 尝试分类输出图像是真实或者伪造的,同时训练生成模型,其损失函数可以应用于传统上需要种类差别很大的任务.在这样的背景下,如何利用优化 GAN 网络进行有监督学习、进行图像的各种转换,都已经渐渐成为研究热点.

图像转换包含多种类型,比如图像的风格转换,将水墨画转换成山水画、将真实图像卡通化;图像的色彩转换,比如彩色与黑白图像之间;图像的内容转换,比如卫星图像与地图的转换、斑马与马的转换;图像的场景转换,比如白天到黑夜等等.这些对图像的变换、纹理调整、风格化编辑,在艺术、科研、工程领域均有所应用.然而,由于时间、地点和相机参数等限制,通过人工方法采集同一景物不同场景的图像有很大的困难;而通过图像处理的方法,比如进行超分辨率、锐化、去噪[9]等方式对图像进行优化,提升图像质量,是一条可行性较高的途径.

作为图像转换领域的代表,图像风格转换相关领域研究趋于成熟.现有的图像风格转换有两类:一类是基于全局[10],通过匹配像素颜色的均值和方差或其柱状图来实现样式化;另一类是基于局部[11,12],通过利用低层次或高层次特征内容和风格照片之间的密集对应关系对图像进行风格化.这些方法在实践中很耗时,并且通常是为特定场景来设定的.Gatys 等人[13]提出了艺术风格的神经风格转换算法,其主要步骤是解决从内容图像和风格图像中提取深层特征与 Gram 矩阵匹配.目前已有了许多方法,在此算法上[14−16]进一步提高其性能和速度.然而,这些方法有时生成的图像不够真实,所以还需要在此基础上进行后处理[17],来匹配输入图像与输出图像的梯度.

高保真的图像风格化与图像到图像的翻译问题[18−22]有关,目标是学习将图像从一个域翻译到另一个域.然而,真实照片图像的风格化并不需要学习翻译功能的内容和风格图像的训练数据集.照片写实图像的风格化,可以看作是一种特殊的图像到图像的转换,用来把照片翻译成不同的领域(例如从白天到晚上).Luan 等人[23]通过在优化目标中加入一个新的损失函数,提高了风格转换算法计算出的风格化输出的真实感,从而更好地保留图像内容中的局部结构.然而,它通常会产生不一致的风格化;此外,该方法的计算成本也很高.Pix2pix[1]将条件 GAN[24]用于不同的图像转换,例如将谷歌地图转换为卫星视图等.在没有训练对的情况下,实现图像到图像翻译的各种方法[19,21,25]也陆续被提出.而 Chen 等人[26]指出:由于训练的不稳定性和优化问题,条件 GAN 训练难以生成高分辨率图像.为了避免这种困难,提出了感知损失[27].生成的图像是高分辨率的,但往往缺乏细节和现实的纹理.

基于以上的研究,本文提出了一种新型的基于生成对抗网络的图像场景转换算法,主要有如下 3 点创新.

首先,设计了新的生成器网络结构.主要采用带跨层连接结构的深度卷积,通过跨层连接能够实现底层卷积与顶层卷积的信息共享,更好地保留了图像的内容结构,最终使输出图像与输入图像的结构和边缘保持一致;

其次,设计了多尺度判决器网络结构,分别对图像的不同尺度进行判决.当判决器的输入图像为大尺度时更关注图像的细节,小尺度时更关注图像的结构.这样将大小尺度相结合的方式,能够在判决时兼顾图像的细节和结构;

最后,提出了新的损失函数.基于常用的损失函数 GAN 损失和 L_1 损失,加入了 VGG 损失和 FM(特征匹配)损失,以利用 VGG 网络和判决器网络来增加对生成对抗网络的控制,最终使生成图像与目标图像更加接近.

1 相关工作

图像转换是一个经典的计算机视觉任务,而近些年,以卷积神经网络为代表的深度学习算法的流行,让这一任务有了显著的突破.2014 年,Goodfellow 提出了生成对抗网络[5],基于 GAN 的算法在图像转换上表现良好,Pix2pix,CycleGAN,MUNIT 等模型陆续被提出.参考这些算法,本文提出了一种新型的基于生成对抗网络的图像场景转换算法.

1.1 生成对抗网络

生成对抗网络(GAN)[5]是一种无监督的机器学习方法,有两个网络模型:生成器(generative model)和判决器(discriminative model),两个网络相互对抗相互牵制.判决器是判定一个样例是来自数据集还是生成器合成的图像,生成器目的是尽可能使生成图像以假乱真以迷惑判决器难辨真伪.两个网络模型相互对抗来提升各自的算法能力,直到判决器无法分辨出合成图像与真实图像.在数据集中真实图像中,生成器想要从 y 中学习其分布,定义输入噪声变量 $p_z(z)$,则损失函数定义为

$$\min_G \max_D V(D,G) = E_{y\sim p_{data}(y)}[\log D(y)] + E_{z\sim p_z(z)}[\log(1-D(G(z)))] \tag{1}$$

z 表示输入生成器 G 的噪声,而 $G(\cdot)$表示 G 网络生成的图片.$D(\cdot)$表示判决器 D 网络判断真实图片是否真实的概率,E 为数学期望.由于其为无监督学习,该方法应用范围十分小,无法实现像素与像素之间的转换.

1.2 CycleGAN

CycleGAN[21]利用非成对图像进行训练,主要贡献在于提出了循环一致性损失.该损失要同时学习正向和反向两个映射,设正向映射也即 $G:X\to Y$,反向映射 $F:Y\to X$.并要求图像能够从一个方向转换后,还可以反向转换,实现一个循环.即 $F(G(x))\approx x$ 和 $G(F(y))\approx y$.循环一致性损失可以定义为

$$L_{cyc}(F,G,X,Y)=E_{x\sim pdata(x)}[\|G(F(x))-x\|_1]+E_{y\sim pdata(y)}[\|G(F(y))-y\|_1] \tag{2}$$

同样还引用 GAN 损失.正向映射 $G:X\to Y$,定义其判决器为 D_Y,则其 GAN 损失为

$$L_{GAN}(G,D_Y,X,Y)=E_{y\sim pdata(y)}[\log D_Y(y)]+E_{x\sim pdata(x)}[\log(1-D_Y(G(x)))] \tag{3}$$

由于循环一致性,则定义反向映射的判决器为 D_X,由此可以同样定义 $L_{GAN}(G,D_X,X,Y)$.最终的损失就由 3 部分组成:

$$L=L_{GAN}(G,D_Y,X,Y)+L_{GAN}(F,D_X,X,Y)+\lambda L_{cyc}(F,G,X,Y) \tag{4}$$

该方法在 Pix2pix 基础之上修改损失函数.由于图像的训练集不需要成对的数据进行训练,该方法的应用范围更加广泛.但是由于训练集不匹配,只能通过训练去猜测真实的映射关系,因此其学习到的映射关系可能会出现偏差.

1.3 MUNIT模型

MUNIT(multimodal unsupervised image-to-image translation)即多模态非监督图像翻译算法[22],算法中通过图像编码分别获得图像集 X_i 的风格空间 S_i 和共同的内容空间 C.实现从图像 x_1 转换到 x_2,将输入图像的内容 c 与转换目标的风格 s_2 相结合.不同的风格得到不同的转换结果.该网络的损失函数包括两部分:一是双向重建损失,二是 GAN 损失.

双向重建损失有两部分——图像重建和潜在重建.生成网络用 G 表示,E 表示 G 的反向操作.图像重建损失表示为

$$L_{recon}^{x_1} = E_{x_1\sim p(x_1)}[\| G_1(E_1^c(x_1),E_1^s(x_1)) - x_1 \|_1] \tag{5}$$

潜在重建损失表示为

$$L_{recon}^{c_1} = E_{c_1\sim p(c_1),s_2\sim q(s_2)}[\| E_2^c(G_2(c_1,s_2)) - c_1 \|_1] \tag{6}$$

$$L_{recon}^{s_2} = E_{c_1\sim p(c_1),s_2\sim q(s_2)}[\| E_2^s(G_2(c_1,s_2)) - s_2 \|_1] \tag{7}$$

GAN 损失表示为

$$L_{GAN}^{x_2}=E_{c_1\sim p(c_1),s_2\sim q(s_2)}[\log(1-D_2(G_2(c_1,s_2)))]+E_{x_2\sim p(x_2)}[\log D_2(x_2)] \tag{8}$$

因此,网络的优化目标可以表示为

$$\min_{E_1,E_2,G_1,G_2}\max_{D_1,D_2}L(E_1,E_2,G_1,G_2,D_1,D_2)=L_{GAN}^{x_1}+L_{GAN}^{x_2}+\lambda_x(L_{recon}^{x_1}+L_{recon}^{x_2})+\lambda_c(L_{recon}^{c_1}+L_{recon}^{c_2})+\lambda_s(L_{recon}^{s_1}+L_{recon}^{s_2}) \tag{9}$$

1.4 Pix2pix

Pix2pix 算法[1]是一个条件 GAN 框架,用于图像到图像的转换,由生成器 G 和判决器 D 组成.生成器 G 网络目的是学习输入图像 x 到目标图像 y 的映射 $G:x\rightarrow y$,使生成图像与目标图像十分接近,难辨真假;判决器 D 网络的目的是尽可能判断出图像是生成图像还是真实图像.对以下公式进行网络优化:

$$\arg\min_G\max_D(loss_{GAN}+\lambda loss_{L_1}) \tag{10}$$

其中,$loss_{GAN}$ 为生成对抗网络损失函数,$loss_{L_1}$ 为 L_1 损失,λ为可调参数.$loss_{GAN}$ 定义如下:

$$loss_{GAN}=E_{y\sim pdata(y)}[\log D(y)]+E_{x\sim pdata(x)}[\log(1-D(G(x)))] \tag{11}$$

其目的是使 GAN 网络生成器与判决器相互制约,共同优化. $loss_{L_1}$ 定义如下:

$$loss_{L_1}=E_{x,y\sim pdata(x,y)}[\| y-G(x)\|_1] \tag{12}$$

因为图像生成本质上是回归问题,所以使用 L_1 损失对生成图像进行限制.Pix2pix 方法采用 U-net 作为生成器以及 patchGAN 的卷积网络[28]作为判决器.

2 基于生成对抗网络的图像转换算法

本文提出的基于生成对抗网络的图像场景转换算法主要分训练和测试两个阶段.在训练阶段将 GAN 网络模型进行优化,使得在测试阶段输入图像通过 GAN 网络模型得到输出图像.通过生成网络和判决网络不断迭代,优化网络参数.算法流程图以图像加雾实验作为示例,如图 1 所示.本节将对生成器、判决器及损失函数进行阐述.

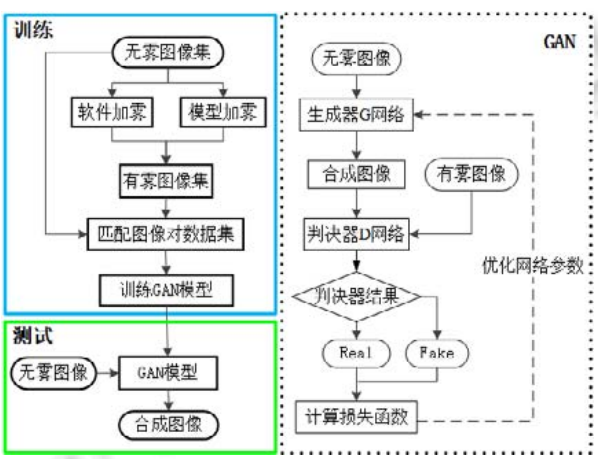


Fig.1 Algorithm flowchart

图 1 算法流程图

2.1 生成器结构

2.1.1 跨层连接生成网络

以图像加雾为例.本文在生成器网络 G 设计上采用跨层连接,是由于在图像转换中有大量的信息在输入和输出之间共享,并需要直接在网络上传输这些信息.例如进行场景转换时,输入和输出共享突出边缘的位置.网络结构如下.

如图 2 所示:网络整体呈现左右对称的结构,左侧为卷积操作,右侧为反卷积操作.

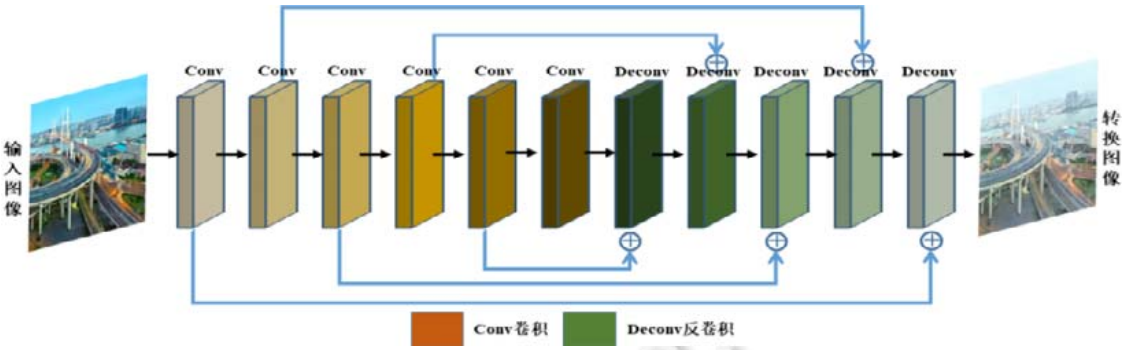


Fig.2 Generator network structure

图 2 生成器网络结构

将卷积层-batch Normalization(BN 层)-prelu 看作一个模块,记为一层.输入图像经过多层卷积操作,得到中间层,这时如图右侧表示,相应的卷积层再进行反卷积,此时,将反卷积层-batch Normalization(BN 层)-prelu 看作一个模块,记为一层.同时,再与左侧对应的卷积层信息直接相加,最后得到图像的输出.每个卷积层和反卷积层的参数设置为卷积核大小为 4×4,padding 为 0,stride 为 1.

2.1.2 生成器网络模型对比

本文设计的生成器网络与 pix2pix 网络同样利用的跨层连接,但网络整体结构不同:pix2pix 网络采用编码器-解码器结构,先进行多层下采样再进行多层上采样,每经过一层下采样,图像的长宽各减小一半;本文的生成器网络没有进行上采样和下采样,而是单纯地进行多层卷积和反卷积操作.为了证明本文提出的算法生成器结构的优越性,比较了两种算法的实验结果.实验时,网络的判决器个数为 1,损失函数为 GAN 损失和 L_1 损失.本文以图像加雾的训练集进行训练,加雾结果如图 3 所示.

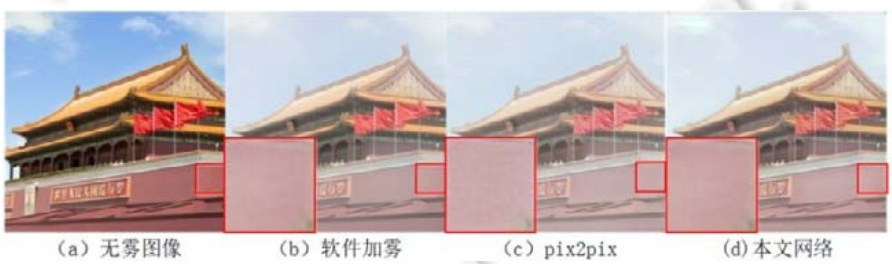


Fig.3 Comparison of different generator structures

图 3 不同生成器结构结果对比

由图 3 可见,两种算法的加雾效果在细节上有所不同.Pix2pix 在平坦的区域有块效应出现,而本文效果较为平滑,如图中的右下部分实线框区域.

2.2 判决器结构

2.2.1 多尺度判决网络

GAN 生成图像的难点在于让生成图像的过程可控,即生成更加真实和清晰的图像,而这对 GAN 判决器设计提出了重大挑战.为了区分真实图像和生成图像,判决器需要具有大的感受野.这就需要设计更深的网络结构或者采用更大的卷积核,但两者都会增强网络能力并可能导致过拟合.此外,这两种选择由于增加了网络复杂程度,都需要更大的内存占用.为了解决这个问题,本文采用多尺度判决器,也即多个判决器在不同尺寸的输入图像下进行.本文最多使用了 3 个判决器,分别记作 D_1~D_3.当使用 3 个判决器时,是分别对图像下采样一倍和两倍

再进行判决.

D_1~D_3 的网络结构如图 4 所示,由多个下采样层和一个输出判决层组成,只是输入图像的尺寸不同.输入图像为大尺度时更关注图像的细节,小尺度时更关注图像的结构.

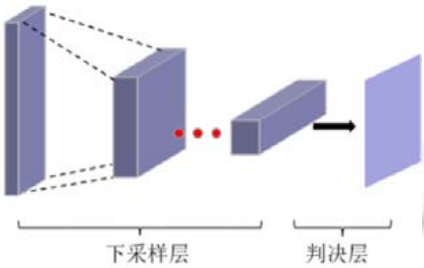


Fig.4 Discriminator network structure

图 4 判别器网络结构

2.2.2 多尺度判决器数目

本文采用多尺度判决器,即多个判决器在不同尺寸的输入图像下进行.因此,判决器个数的选择至关重要.因为在不同尺度下进行判决,小尺度图像作为输入时能够更多关注图像的整体结构和边缘,大尺度图像作为输入则更多关注图像的细节保留.判决器个数少,则影响生成器效果.理论上是判决器越多越好,但也并非如此.判决器越多,一是增加了网络的复杂度和计算量,影响训练时间;二是判决器个数与输入图像本身尺寸有关,如果输入本身尺寸适当,非大尺寸或超大尺寸,判决器没必要过多.于是,针对判决器的个数选择多少较为合适,在本文研究的特定情况下,进行了以下实验.本文在加雾训练集上进行训练,图像输入大小为 256×256,分别测试判决器个数为 1~3 的情况下,迭代 60 个 epoch 的效果.

从整体上来看,判决器的个数对生成图像的内容影响不大,但在细节上会有所差别.由图 5 可见,当 *num_D*=1 时,景物的细节会出现缺失.比如图 5 中建筑的横线,而当 *num_D* 为 2 或 3 时,两张图片相差不多,细节保持都较好.*num_D*=3 时,图像颜色更亮一点,只是略有提升,但是效果并不明显.同时,图 5 中的天空部分,当 *num_D*=1 时会出现失真.而由于原始生成图像天空区域较亮,放大后依然很难观察到差别.于是,本文对天空区域进行了处理,变换公式如下:

$$im\_new_{r,g,b}=(im_{r,g,b}-240)\times 15 \tag{13}$$

其中,$im\_new_{r,g,b}$ 为输出的图像,$im_{r,g,b}$ 为变换前图像.变换后图像如图放大区域所示,当 *num_D*=2 或 3 时,天空颜色则较为均匀.考虑到网络复杂度和计算量,在本文所有实验中,判决器的个数 *num_D*=2.

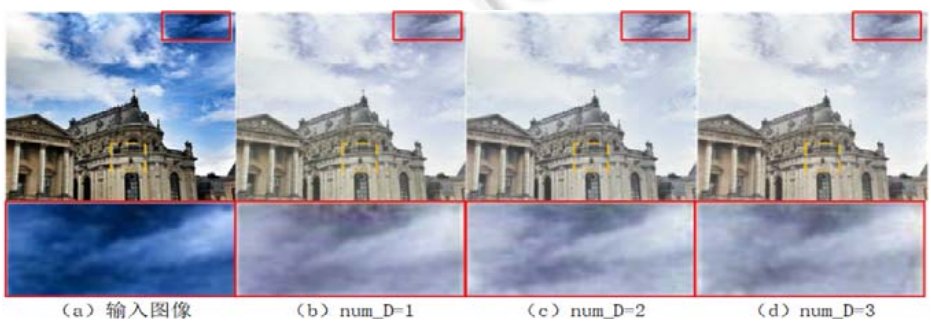


Fig.5 Comparison of different number of discriminators

图 5 不同个数判决器结果对比

2.3 损失函数

2.3.1 损失函数组成

本文损失函数共使用了 4 种,GAN 损失、L_1 损失、VGG 损失和 FM(feature matching,特征匹配)损失.首先,对于生成对抗网络,由于本文使用多尺度判决器,因此生成对抗网络的优化问题表示为

$$\min_{G}\max_{D_1,D_2,D_3}\sum_{k=1,2,3}L_{GAN}(G,D_k) \tag{14}$$

x 为输入图像,y 为目标图像.其中,GAN 损失表达为

$$L_{GAN}(G,D_k)=E_{(x,y)}[\log D_k(y)]+E_x[\log(1-D_k(G(x)))] \tag{15}$$

该算法对于生成器的生成结果还要加入限制,对生成图像进行评价,引入 L_1 损失:

$$L_{L_1}(G)=E_{(x,y)}[\| y-G(x)\|_1] \tag{16}$$

为了使输出图像更加逼近真实图像,引入特征匹配(FM)损失.具体来说,从判决器多个层来提取特征并学习来匹配真实图像和合成图像的中间特征.表示第 i 层特征提取器 $D_k^{(i)}$(从输入到 D_k 判决器的第 i 层).然后计算 FM 损失 $L_{FM}(G,D_k)$如下:

$$L_{FM}(G,D_k)=E_{(x,y)}\sum_{i=1}^{T}\frac{1}{N_i}[\| D_k^{(i)}(x,y)-D_k^{(i)}(x,G(x))\|_1] \tag{17}$$

其中,T 是判决器总层数,N_i 表示层数每层中的元素数量.为了两个图像特征之间的差距,引入 VGG 损失,通过预先训练的 VGG 网络,提取图像的特征,定义 FM 损失为

$$L_{VGG}(G)=\sum_{i=1}^{N}\frac{1}{M_i}[\| F^{(i)}(y)-F^{(i)}(G(x))\|_1] \tag{18}$$

其中,$F^{(i)}$表示 VGG 网络的第 i 层,M_i 表示该层的元素个数.因此,本文算法最终总的损失函数优化目标表示为

$$\min_{G}\left(\left(\max_{D_1,D_2,D_3}\sum_{k=1,2,3}L_{GAN}(G,D_k)\right)+\lambda_1L_1(G)+\lambda_2\sum_{k=1,2,3}L_{FM}(G,D_k)+\lambda_3L_{VGG}(G)\right) \tag{19}$$

而对于损失函数各个部分的作用,本文将在第 2.3.2 节中进行实验分析,验证本文算法改进的损失函数的有效性和必要性.

2.3.2 损失函数的设置

本文的损失函数一共由 4 部分组成,分别为:GAN 损失,L_1 损失,FM 损失,VGG 损失.本文算法是基于 GAN 框架,故对比了总损失(total loss)、不使用 VGG 损失(no_VGGloss)、不使用 L_1 损失和 FM 损失(no_matchingloss)这 3 种情况下的实验结果.

由图 6 可以看出:在不使用 VGG 损失时,则会出现图像失真,比如图 6 中,线框区域,在天空、跑道等位置会出现不规则椭圆形的近似白色的“异物”.没有了 VGG 损失的限制,会出现图像的失真.当没有 L_1 和 FM 损失时,图像不会出现失真,但是图像的色彩会出现偏差.在没有 L_1 和 FM 损失时,雾气整体偏深色;而使用了 L_1 和 FM 损失后,颜色正常.

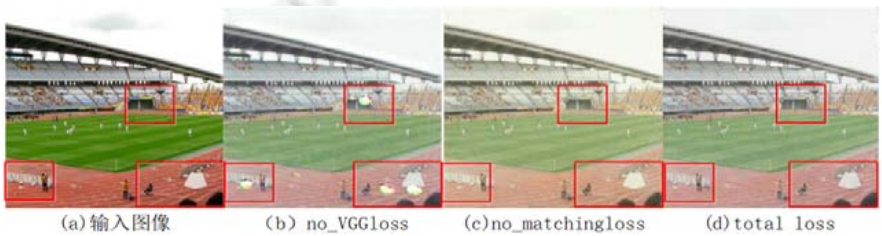


Fig.6 Comparison of different loss function

图 6 不同损失函数结果对比

2.4 训练过程参数设置

- 生成器网络:卷积核大小为 4×4,步长为 1,padding 为 0,网络为左右对称的卷积,设置网络左侧卷积层 6 层,右侧带跨层连接反卷积层 5 层,整个网络共 11 层;
- 判决器网络:下采样层卷积核大小为 4×4,步幅为 2,下采样层个数 3 层,判决器的个数为 2;
- 损失函数:λ_1=10,λ_2=10,λ_3=10,学习率(learning rate)为 0.0002.

3 实验结果

3.1 实验环境

算法的实验环境如下.

- 硬件设备:CPU:Intel Core i7-5820K @ 3.30GHz x 12;GPU:NVIDIA GeForce TITAN X;内存:16GB;
- 软件配置:操作系统为 64 位 ubuntu 14.04 LTS;CUDA Toolkit 7.0.

本文使用深度学习的框架为 Pytorch.

- 加雾训练集[29]:利用软件 Adobe lightroom CC 加雾功能,对 Middlebury Stereo Datasets 和网上收集的无雾图像进行加雾.分别对 76 张无雾图像集加浓度为 30,40,50,60,70,80,90,100 的雾,其中,室外场景 26 张,室内场景 50 张,最终形成 608 对含不同浓度雾的有雾图像与无雾图像的匹配图像对做训练集;
- SAR 图像训练集:网络上匹配图像裁剪,共 1 048 对 256×256 匹配图像;
- 白天黑夜转换训练集[11]共 17 112 张;
- 谷歌地图训练集[1]共 1 096 张.

3.2 主观结果分析

3.2.1 SAR 图像生成

之所以进行 SAR 图像合成,是由于目前通过可见光图像和 SAR 图像获得一致的匹配图像对有一定的难点.由于时间、地点、噪声干扰等问题限制,再加上图像的校准也需要消耗大量的人力物力,因此可以尝试通过图像生成的方法,从可见光图像中生成 SAR 图像,来获得特殊的地形地貌在 SAR 图像下的成像效果.

在 SAR 图像生成上,其他相似的 GAN 图像生成的算法并未有类似的转换测试,也无法评测用何种方法生成 SAR 图像更加真实.为了更客观地评价从可见光图像向 SAR 图像转换,本文对比了其他 GAN 图像生成算法.以下几种算法分别各有特点,都能够实现图像场景和内容的转换:Pix2pix[1]是利用匹配图像对进行图像生成;CycleGAN[21]能够利用非匹配图像训练集进行训练提取特征;MUNIT[22]同样可以提供图像的内容转换,可以从场景、内容上进行变化,在 CycleGAN[21]基础上可能实现多种映射,同时生成多幅不同的转换图像.本文通过对可见光图像和真实 SAR 图像组成的训练集对 CycleGAN[21],MUNIT[22],Pix2pix[1]和本文算法进行训练,在相同训练集,不同算法得到如下对比结果,如图 7 所示.

由图 7 可见:

- CycleGAN 能在该训练集下生成呈黑白色图像,且在图像的内容上与可见光图像保持高度一致,但并不能够学习到 SAR 图像的特定特征.比如图中树的形态在真实 SAR 图像和可见光图像中有很大差别;再如街道在可见光图像中呈现近白色,而在真实 SAR 图像中是近黑色.而 CycleGAN 则并不能够学习到这些 SAR 图像的特点,且 CycleGAN 更类似于把彩色图像转换成黑白;
- 而 MUNIT 则表现得更糟,甚至对于图像的内容都不能生成,这主要是由于 MUNIT 更多地会自己生成一些内容;
- Pix2pix 和本文算法则更接近与真实的 SAR 图像.从图像的内容和景物上都能够明显体现,对于道路、树、房屋的生成纹理和颜色都能够以假乱真.但是,Pix2pix 算法相比本文算法图像整体偏模糊且图像会有一些不必要的纹理出现,如图中右下角的草坪.

Fig.7 Comparison of SAR image synthesis results

图 7 SAR 图像合成结果对比

3.2.2 图像加雾

图 8 是分别利用 GAN 算法在本文的训练集下的效果,对比算法包括 Pix2pix[1],CycleGAN[21],DRPAN[6]和软件加雾效果.

由图 8 可见:Pix2pix 处理后图像呈现加雾效果,且图像的内容较为清晰,细节没有丢失,但图像加雾后导致图像的整体色彩有偏差(从树干部分可以看出);CycleGAN 效果则最差,内容模糊,色彩严重失真,其整幅图像色彩有偏差;软件加雾效果与本文效果十分相近,加雾均匀,雾的颜色没有偏差,且图像的细节保留较好;DRPAN 效果则略差,整幅图像虽然色彩鲜艳度有所下降,但是其图像较模糊,尤其图像上方树干、树叶部分,没有边界,十分模糊.

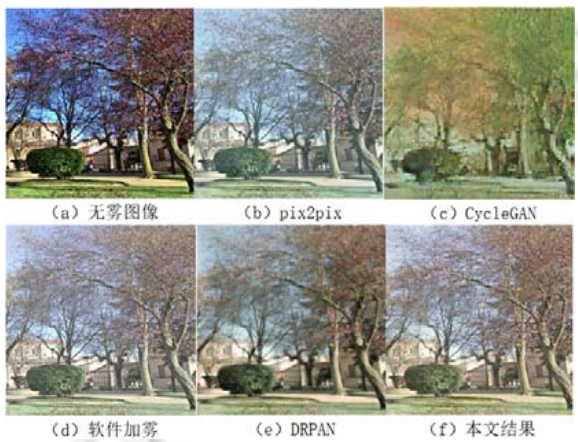


Fig.8 Comparison of imagehazing results

图 8 加雾结果对比

3.2.3 卫星图像到地图转换

同时,本文测试了卫星图像到地图的转换,训练集和测试集采用 Pix2pix[1]的公共训练集,并测试了本文算法与其他基于生成对抗网络算法的转换效果.对比算法包括 Pix2pix[1]、CycleGAN[21]和 DRPAN[6],如图 9 所示.

从整体来看,均能生成类似地图效果的图像较为逼真.DRPAN 算法对图像进行的增强处理,对比度较强,但不影响整体的比较.Pix2pix 中,对于左下草坪区域大部分能够恢复出来,且草坪与道路相连的虚线框中区域、道路恢复得比较直,同时,最下面的湖水区域边界明显,而对于草坪中的小路则出现内容缺失;而 CycleGAN 算法对于草坪区域均不能够着色,草坪与道路相接的虚线框中虽然能够恢复道路,但道路不直且没有连贯,而对于草坪

中的小路同样没有转换成功;DRPAN 算法对于湖水、草坪、草坪间的小路均明显地生成,但草坪与道路相接的区域,道路的内容模糊缺失;本文算法能够生成草坪与道路之间的路,且道路较连贯,对于草坪区域、湖的区域则着色不均,也有所缺失.

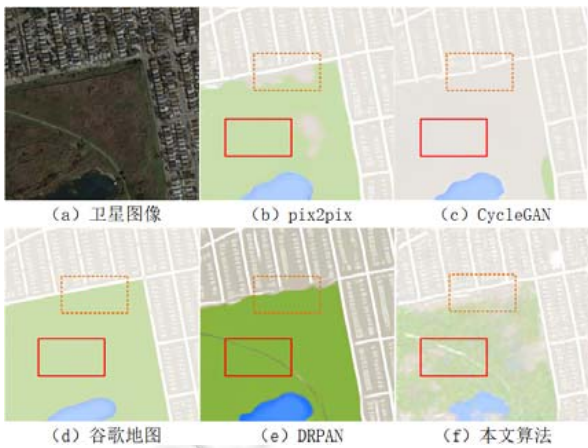


Fig.9 Comparison of map synthesis results

图 9 地图合成结果对比

3.2.4 白天到黑夜转换

在本节中,本文对白天到黑夜的转换进行了训练和测试,如图 10 所示.训练集同样来自 Pix2pix[1].同时,在训练测试过程中发现:本文算法在该训练集下,判决器个数设置为 1 时效果更好.考虑到是由于 input 和 label 并不完全匹配,数据集的图像虽然为同一地点同一位置,但仍有不同:一是拍摄时间不同,二是其他内容不同,比如马路上的车辆个数和位置、季节变化等的差异,而采用多个多尺度判决器原本目的是对于图像细节进行矫正,但在白天黑夜转换中则应该忽略掉小的细节上的差异,并对这些差异有所保留.

从图 10 中可见,本文算法生成图像的内容基本不变,但对于天空区域则均变为黑色,埃菲尔铁塔则亮起了灯.视觉效果上,pix2pix 和 DRPAN 算法生成的黑夜图更接近真实的夜晚图像,即色彩和亮度上的相似性,但是图很明显地出现了块效应,尤其是在天空与地面的交界处.图 CycleGAN 生成的图像下沿即建筑物区域一片模糊,很明显没有转换成功.而本文提出的算法虽然在色彩和亮度上没有更接近与真实夜晚图,但是生成的图片减少了块效应,很好地保持了纹理结构,且更具真实感.

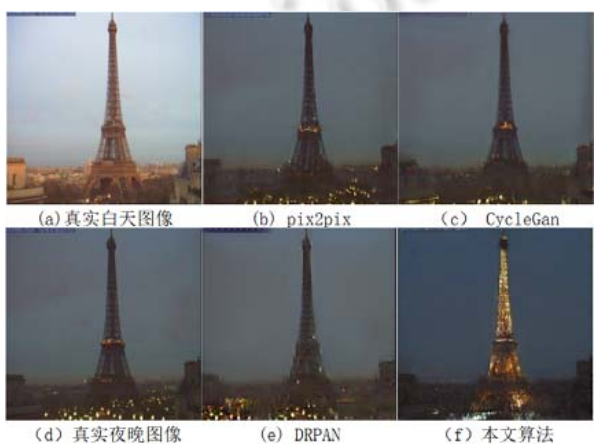


Fig.10 Comparison of night image synthesis results

图 10 夜晚场景合成结果对比

3.3 客观指标分析

3.3.1 图像加雾客观指标分析

首先,Choi 等人[30]计算雾浓度用了算法 FADE(fog aware density evaluator),利用该算法对本文 40 张测试集分别求出雾浓度指标.雾浓度结果见表 1.指标越高,说明雾浓度越大.计算了 40 张图的平均值及均方差,并对比无雾图像、CycleGAN、Pix2pix、DRPAN 和软件加雾效果.

Table 1 Comparison of FADE indicators

表 1 FADE 指标对比

图像	无雾图像	CycleGAN	Pix2pix	DRPAN	软件加雾	本文算法
平均指标	0.230±0.116	0.736±0.431	0.689±0.348	0.459±0.250	0.670±0.412	0.634±0.410

由表 1 可见:在利用 Pix2pix、CycleGAN、DRPAN、软件加雾和本文算法处理后,图像的 fog density 指标明显上升,相对 DRPAN 加雾程度最低,CycleGAN 程度最高.Pix2pix 和本文算法以及软件加雾的加雾程度相近.这也与主观效果相似.CycleGAN 之所以该项指标更高,也是由于图像色彩较少,整体图像色彩偏黄绿色.

本文也采用 PSNR 和 SSIM[31]指标对进行真实图像定性对比分析.PNSR 值越高,说明生成图像与原图更加相似,失真越少.当 SSIM 值越接近 1 时,则生成图像与原图的结构越相近,表明生成图像效果越好.对比结果见表 2,将原始图像做基准图像,分别对本文算法加雾结果,将本文算法得到的加雾图像利用 DCP 算法得到去雾图像求 PSNR 和 SSIM 指标.在表 2 中,对 40 张测试图的 PSNR 和 SSIM 进行统计,其平均值和均方差结果如下.

Table 2 Comparison of PSNR and SSIM indicators

表 2 PSNR 和 SSIM 指标对比

图像	DCP 去雾	CycleGAN	Pix2pix	DRPAN	软件加雾	本文算法
PSNR	19.231±2.157	12.703±1.708	13.830±0.690	17.211±2.541	13.611±1.147	14.760±0.733
SSIM	0.789±0.052	0.349±0.084	0.752±0.074	0.809±0.082	0.782±0.075	0.725±0.062

由表 2 分析 PSNR 和 SSIM.首先,本文算法在进行去雾后,图像的 *PSNR* 值明显上升,其中,红色为本文加雾后结果,深蓝色为本文算法的结果进行去雾后的效果.同时,本文算法和软件加雾以及 Pix2pix 算法在对图像进行加雾后,图像的 PSNR 值基本维持在一定范围内波动较小,而 DRPAN 和 CycleGAN 波动较大,则证明 DRPAN 和 CycleGAN 算法加雾效果较差.这一点,从主观效果中也能够明显体现.而 CycleGAN 在几个加雾算法中 *PSNR* 值基本最低,也是由于该算法生成的图像内容出现误差和缺失.而 DRPAN 的 PSNR 值整体偏高,并非该算法的加雾效果好,而是证明该算法更与无雾图像接近,也即加雾效果并不明显.

SSIM 指标只能做参考,因为去雾算法并不能够实现完全去雾,且任何去雾算法或多或少都会出现雾的残留、块效应、光晕、天空色彩失真等现象.而本文算法的结果在进行去雾之后,SSIM 指标有所上升.CycleGAN 算法的 SSIM 指标则十分低,说明生成图像的内容与输入图像差别较大;其余 4 个方法的加雾效果则基本维持在同一个水平.相比之下,DRPAN 的 SSIM 值较高,因为其与输入图像相近.

3.3.2 SAR 图像转换客观指标分析

在 Pix2pix[1]算法中,对于图像转换后生成的图像采取人为进行观察评价的方法.于是,本文对 SAR 图像转换以及地图转换进行调查问卷调查,对多种算法进行对比,并对图像的真实程度评分(未给出真实图像),当生成图像越真实,真假难辨评分越高.最高 5 分,最终在随机人群中回收到的 21 份问卷评分统计结果见表 3.

Table 3 Comparison of SAR image conversion scores

表 3 SAR 图像转换得分对比

算法	CycleGAN	MUNIT	Pix2pix	本文算法
平均得分	2.964±1.142	1.205±0.462	3.181±0.926	**4.193±0.634**

由结果可见:MUNIT 得分最低,主要是由于其图像的边缘基本都难以分辨;CycleGAN 得分仅比 Pix2pix 略低,原因在于 CycleGAN 能够保留较为清晰和准确的图像边缘,图像整体也呈现灰白色彩,但其对于 SAR 图像的

特征转换的不够准确;而本文算法得分略高,原因在于不论色彩还是边缘,本文算法效果均突出,且没有块效应的出现.

3.3.3 地图转换客观指标分析

同样采取 SAR 图像转换一样的问卷调查,对于 21 份回收结果进行分析——平均值及均方差,结果见表 4.

Table 4 Comparison of map conversion scores

表 4 地图转换得分对比

算法	Pix2pix	CycleGAN	DRPAN	本文算法
平均得分	2.672±0.908	2.475±0.977	3.754±0.888	**3.984±1.041**

由图像的得分可见:Pix2pix 和 CycleGAN 得分相近,均为中等分数,图像转换效果可圈可点,道路、草地等转换也都较为准确;本文算法和 DRPAN 得分相近且略高,DRPAN 的色彩鲜艳,且分割完整,线条流畅;而本文算法色彩较暗,分割完整,但在线条上更加笔直.

3.3.4 白天夜晚转换客观指标分析

同样采取 SAR 图像转换一样的问卷调查,对于 21 份回收结果进行分析——平均值及均方差,结果见表 5.

Table 5 Comparison of day-night conversion scores

表 5 白天夜晚转换得分对比

算法	Pix2pix	CycleGAN	DRPAN	本文算法
平均得分	3.572±0.858	3.268±0.624	3.682±0.988	**3.768±0.909**

由图像的得分可见:Pix2pix,DRPAN 和本文算法得分相近,且波动范围较大,夜晚图像转换效果可圈可点;Pix2pix 和 DRPAN 在色彩与饱和度方面做得更好,本文算法则在纹理结构以及图像失真方面做得更好;CycleGAN 算法在这一转换任务上表现较差,色彩上较为暗淡,图像的纹理结构也大多丢失.

4 结 论

本文介绍了基于生成对抗网络的图像场景转换算法的具体内容.首先介绍了算法的跨层连接生成器网络设计、多尺度判决器网络设计以及损失函数的 4 种组合;接着对网络模块性能进行分析,从实验证明本文算法设计的合理性;接着介绍实验的平台、硬件软件等,之后分别从主观效果和客观指标进行分析.在主观效果上,分析了本文算法与其他基于生成对抗网络的图像转换算法对于场景转换的效果,包括雾霾场景转换、SAR 图像转换、谷歌地图转换以及白天黑夜转换.在客观指标上,本文算法效果表现也略为突出.

References:

[1] Isola P, Zhu JY, Zhou T, *et al*. Image-to-image translation with conditional adversarial networks. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2017. 1125−1134.

[2] Ledig C, Theis L, Huszár F, *et al*. Photo-realistic single image super-resolution using a generative adversarial network. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2017. 4681−4690.

[3] Xiao JS, Liu EY, Zhu L, *et al*. Single image dehazing algorithm based on the learning of hazy layers. Acta Electronica Sinica, 2019,47(10):2142−2148 (in Chinese with English abstract).

[4] Zhang R, Isola P, Efros AA. Colorful image colorization. In: Leibe B, *et al*., eds. Proc. of the European Conf. on Computer Vision. Cham: Springer, 2016. 649−666.

[5] Goodfellow IJ, Pouget-Abadie J, Mirza M, *et al*. Generative adversarial nets. In: Advances in Neural Information Processing Systems. 2014. 2672−2680.

[6] Wang C, Zheng H, Yu Z, *et al*. Discriminative region proposal adversarial networks for high-quality image-to-image translation. In: Proc. of the European Conf. on Computer Vision. 2018. 770−785.

[7] Wang KF, Zuo WM, Tan Y, *et al*. Generative adversarial networks: From generating data to creating intelligence. Acta Automatica Sinica, 2018,44(5):769−774 (in Chinese with English abstract).

[8] Wang WL, LI ZR. Advances in generative adversarial network. Journal on Communications, 2018,39(2):135−148 (in Chinese with English abstract).

[9] Xiao J, Tian H, Zhang Y, *et al*. Blind video denoising via texture-aware noise estimation. Computer Vision and Image Understanding, 2018,169(4):1−13.

[10] Freedman D, Kisilev P. Object-to-object color transfer: Optimal flows and smsp transformations. In: Proc. of the 2010 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. IEEE, 2010. 287−294.

[11] Laffont PY, Ren Z, Tao X, *et al*. Transient attributes for high-level understanding and editing of outdoor scenes. ACM Trans. on Graphics, 2014,33(4):149.

[12] Tsai YH, Shen X, Lin Z, *et al*. Sky is not the limit: Semantic-aware sky replacement. ACM Trans. on Graphics, 2016,35(4): 149:1−149:11.

[13] Gatys LA, Ecker AS, Bethge M. Image style transfer using convolutional neural networks. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2016. 2414−2423.

[14] Li Y, Fang C, Yang J, *et al*. Diversified texture synthesis with feed-forward networks. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2017. 3920−3928.

[15] Chen D, Yuan L, Liao J, *et al*. Stylebank: An explicit representation for neural image style transfer. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2017. 1897−1906.

[16] Huang X, Belongie S. Arbitrary style transfer in real-time with adaptive instance normalization. In: Proc. of the IEEE Int'l Conf. on Computer Vision. 2017. 1501−1510.

[17] Li S, Xu X, Nie L, *et al*. Laplacian-steered neural style transfer. In: Proc. of the 25th ACM Int'l Conf. on Multimedia. ACM, 2017. 1716−1724.

[18] Wang TC, Liu MY, Zhu JY, *et al*. High-resolution image synthesis and semantic manipulation with conditional GANs. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2018. 8798−8807.

[19] Liu MY, Tuzel O. Coupled generative adversarial networks. In: Proc. of the 30th Conf. on Neural Information Processing Systems. Barcelona, 2016. 469−477.

[20] Shrivastava A, Pfister T, Tuzel O, *et al*. Learning from simulated and unsupervised images through adversarial training. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2017. 2107−2116.

[21] Zhu JY, Park T, Isola P, *et al*. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proc. of the IEEE Int'l Conf. on Computer Vision. 2017. 2223−2232.

[22] Huang X, Liu MY, Belongie S, *et al*. Multimodal unsupervised image-to-image translation. In: Proc. of the European Conf. on Computer Vision. 2018. 172−189.

[23] Luan F, Paris S, Shechtman E, *et al*. Deep photo style transfer. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2017. 4990−4998.

[24] Mirza M, Osindero S. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784, 2014.

[25] Bousmalis K, Silberman N, Dohan D, *et al*. Unsupervised pixel-level domain adaptation with generative adversarial networks. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2017. 3722−3731.

[26] Chen Q, Koltun V. Photographic image synthesis with cascaded refinement networks. In: Proc. of the IEEE Int'l Conf. on Computer Vision. 2017. 1511−1520.

[27] Dosovitskiy A, Brox T. Generating images with perceptual similarity metrics based on deep networks. In: Proc. of the 30th Conf. on Neural Information Processing Systems. Barcelona, 2016. 658−666.

[28] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2015. 3431−3440.

[29] Chen Y, Lai YK, Liu YJ. CartoonGAN: Generative adversarial networks for photo cartoonization. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2018. 9465−9474.

[30] Scharstein D, Hirschmüller H, Kitajima Y, *et al*. High-resolution stereo datasets with subpixel-accurate ground truth. In: Proc. of the German Conf. on Pattern Recognition. Cham: Springer-Verlag, 2014. 31–42.

[31] Choi LK, You J, Bovik AC. Referenceless prediction of perceptual fog density and perceptual image defogging. IEEE Trans. on Image Processing, 2015,24(11):3888–3901.

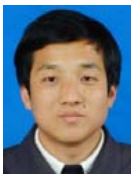
[32] Xiao J, liu E, Zhao L, *et al*. Detail enhancement of image super-resolution based on detail synthesis. Signal Processing: Image Communication, 2017,50(1):21–33.

附中文参考文献:

[3] 肖进胜,周景龙,雷俊锋,等.基于霾层学习的单幅图像去雾算法.电子学报,2019,47(10):2142–2148.

[7] 王坤峰,左旺孟,谭营,等.生成式对抗网络:从生成数据到创造智能.自动化学报,2018,44(5):769–774.

[8] 王万良,李卓蓉.生成式对抗网络研究进展.通信学报,2018,39(2):135–148.



肖进胜(1975－),男,博士,副教授,CCF 专业会员,主要研究领域为计算机视觉,图像处理与分析.

李亮(1995－),男,硕士,主要研究领域为图像处理.

周景龙(1996－),男,硕士,主要研究领域为图像处理与分析.

丁玲(1979－),男,博士,副教授,主要研究领域为人工智能,计算机视觉,图像处理.

雷俊锋(1975－),男,博士,副教授,主要研究领域为计算机视觉,图像处理与分析.

杜治一(1997－),男,学士,主要研究领域为计算机视觉,图像处理与分析.