激光与光电子学进展 先进成像

基于生成对抗网络的轻量级图像卡通风格化方法

孙劲光，王伟*

辽宁工程技术大学电子与信息工程学院，辽宁 葫芦岛 125105

摘要 艺术家创作卡通是一项具有挑战性和耗时的任务。将真实照片自动转换为高质量卡通风格图像的自动技术具有很高价值。提出一种基于生成对抗网络的轻量级图像卡通风格化方法。通过观察卡通绘画行为，将卡通图像风格解耦为平滑表面、稀疏色块、高频纹理3种表示方法。运用生成对抗网络框架学习提取的表示，进而学习卡通图像风格，在生成网络中采用深度可分离卷积和反向残差块来减少网络的参数量与计算成本。为验证所提方法的有效性，进行定性比较和定量分析。结果显示，所提方法能够快速地将真实世界的照片转换为高质量的卡通图像，在时间效率和转换质量方面与已有方法相比有一定优势。

关键词 生成对抗网络；解耦表征；轻量级网络；风格迁移

中图分类号 TP389.1 **文献标志码** A **DOI**：10.3788/LOP213143

Lightweight Cartoonlization Method Based on Generative Adversarial Network

Sun Jinguang, Wang Wei*

School of Electronic and Information Engineering, Liaoning Technical University, Huludao 125105, *Liaoning, China*

Abstract Creating cartoon is a challenging and time-consuming task for artists. However, automated technology that converts real photos into high-quality cartoon-style images is significantly valued. Therefore, based on a generative adversarial network, this study proposes a lightweight image cartoon stylization method. By observing the cartoon drawing behavior, the cartoon image style is decoupled into three representations, including smooth surface, sparse color block, and high frequency texture. A generative adversarial network framework is used to learn the extracted representation and the style of cartoon images. Furthermore, deep detachable convolution and reverse residual blocks are used in generative networks to reduce number of network parameters and computational costs. Qualitative comparison and quantitative analysis are conducted in this study to evaluate the proposed method's effectiveness. The results show that the proposed method can quickly convert real-world photos into high-quality cartoon images and is superior to the existing methods.

Key words generative adversarial network; disentangled representation; lightweight network; style transfer

1 引 言

卡通是一种流行的艺术形式，常用于数字娱乐、游戏甚至是人们的社交账户简介中。目前，卡通制作主要依靠手工实现。由于对创新的强烈要求，要精心设计布局和纹理，艺术家创作卡通图片是一项具有挑战性和耗时的任务。因此，将真实照片自动转换为高质量卡通风格图像的自动技术是非常有价值的。

Gatys等[1]率先提出了一种基于神经网络的图片风格迁移方法，该方法一经提出便风靡一时，解决了传统方法手动建模所存在的复杂过程问题。接着Chen等[2]提出了一种改进漫画风格的深度卷积神经网络方法。该方法使用内容图像的灰度图像初始化输出图像，合成了更好的漫画图像，但是漫画风格图像中的样式和内容分离存在不可避免的模糊性，导致生成的漫画图片出现了现实照片中并不存在的结构。沈瑜等[3]

收稿日期：2021-12-02；**修回日期**：2021-12-20；**录用日期**：2022-01-05；**网络首发日期**：2022-01-15

基金项目：国家重点研发计划(2018YFB1403303)

通信作者：*1803671965@qq.com

提出了目标边缘清晰化的图像迁移算法，该算法细化了图像风格化的结果。由于卡通图像是由稀疏色块组成的，并且具有清晰边缘和光滑表面，该算法并不能生成干净清晰的卡通图像。

随着生成对抗网络（GAN）应用于风格迁移，许多研究者提出了基于生成对抗网络的风格迁移方法。Zhu 等[4]提出的 CycleGAN 使用一个循环框架，该框架是两对生成器和判别器组成的循环框架，第一对学习从源到目标的映射，第二对学习反向映射，实现了在没有成对示例的情况下将图像从源域 X 映射到目标域 Y 的目标。但由于卡通图像的高度抽象和边缘清晰等特点，此方法生成的卡通图像呈现暗沉效果并会产生高频伪影，由于要训练两个生成器与两个判别器，网络参数量巨大。Chen 等[5]提出了一种具有新颖边缘损失的 GAN 框架 CartoonGAN，来代替 CycleGAN 的循环结构，某些情况下取得了良好的效果。但是 CartoonGAN 会扭曲色彩，这并不符合图像卡通风格化的预期效果，并且此方法使用标准卷积，导致网络参数量较大。

综上所述，现有方法并不能准确地学习卡通图像风格，并且多数方法网络参数量巨大，计算成本高。因此本文提出了一种基于生成对抗网络的轻量级框架用于图像卡通风格化。通过对不同风格的人类绘画行为和卡通形象进行观察，发现卡通图像有着一定的共同特征：由稀疏色块组成的全局结构；清晰的边缘勾勒出的细节；平整光滑的表面。

将图像分解为三种卡通表示。表面表示：表示卡通图像的光滑表面。对图像进行边缘保持滤波，忽略边缘和纹理细节的同时保留原有真实图像的颜色和光滑表面。结构表示：表示卡通图像中的整体结构和稀疏色块。对真实图像进行像素分割，之后对每个分割区域运用自适应着色算法，继而生成结构表示，此模块提取的特征具有清晰的边界和稀疏的色块。纹理表示：表示卡通图像绘制的细节和边缘。将真实图像转换为单通道图，去除颜色和亮度，保留图像的高频纹理细节。此外，为了解决图像转换网络框架拥有大量的网络参数和更多的内存容量的问题，本文运用深度可分离卷积和反向残差结构，提出了一种用具有较少网络参数量的轻量级生成网络架构进行图像卡通风格化的方法。实验结果表明，所提方法在主观感受和客观指标方面均取得了较大的进步。

2　所提方法内容

提出了一种用基于生成对抗网络的轻量级框架进行图像卡通风格化的方法，网络模型架构如图 1 所示。将卡通图像分解为表面、结构和纹理 3 种表示方法，并引入三个独立的模块来提取相应的特征。提出一种具有发生器 G 和两个判别器 D_s 和 D_t 的 GAN 框架，其中 D_s 用于区分从网络生成图像和卡通图像中提取的表面特征，D_t 用于区分从网络生成图像和卡通图像中提取的纹理特征。使用预先训练的 VGG-19 网络[6]提取高级特征，并在提取的结构特征与生成图像之间、输入照片与生成图像之间的整体内容中加入空间约束。为了有效地减少生成器参数的数量，在生成网络中运用

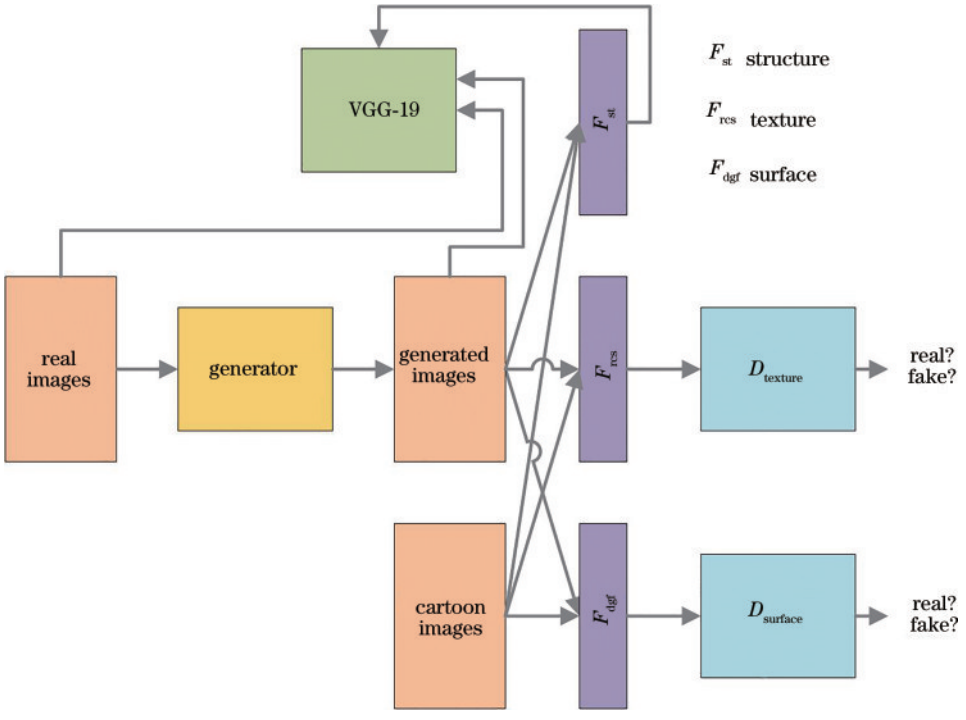

图 1　图像卡通风格化框架

Fig. 1　Image animation stylization framework

深度可分离卷积和反向残差块，构建了轻量级的生成网络架构。

2.1　解耦表征

1)表面表示

表面表示模仿艺术家颜色草图的绘画过程。为了平滑图像的同时保持图像的整体语义结构，采用可微的引导滤波器[7]进行边缘保留滤波。该滤波器整合了由2个具有1×1卷积核的卷积层组成的完全卷积神经网络和1个由线性变换矩阵组成的计算图。此计算图包含均值滤波器和局部线性模型。令输入高分辨图为I_{h}，低分辨图为I_{l}，输出的低分辨图为O_{l}，高分辨图为O_{h}，将引导滤波学习的线性变换系数分别设为A_{l}和b_{l}，通过最小化输入和输出的损失可以计算得到这些系数，则有

$$O_{\mathrm{l}}{}^{i}=A_{\mathrm{l}}^{k}I_{\mathrm{l}}^{i}+b_{\mathrm{l}}^{k},\tag{1}$$

式中：i是像素的标号；k是引导滤波中局部窗口的标号。A_{h}和b_{h}是对A_{l}和b_{l}进行上采样得到的，输出的高分辨图通过线性公式获得：

$$O_{\mathrm{h}}=A_{\mathrm{h}}*I_{\mathrm{h}}+b_{\mathrm{h}}。\tag{2}$$

I_{l}和O_{l}通过一个均值滤波器和局部线性模型，得到A_{l}和b_{l}；然后通过上采样，获得A_{h}和b_{h}；再线性组合A_{h}、b_{h}、I_{h}，得到输出O_{h}；O_{h}可以反向传播到A_{h}、b_{h}及I_{h}，进而通过引导滤波层直接训练得到O_{l}。将图像I作为输入，自身作为引导图，返回提取的表面表示，结果表示为F_{dgf}。经过表面表示，处理的图片结果如图2所示。

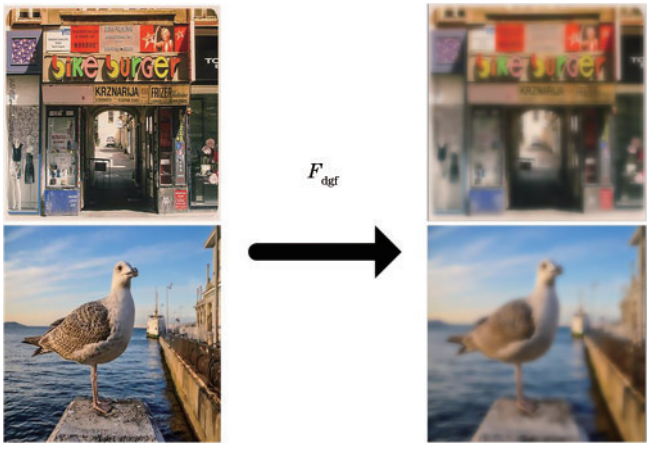

图2　表面表示

Fig. 2　Surface representation

2)结构表示

结构表示模拟卡通图像中的全局内容、稀疏色块和清晰边界。首先使用Felzenszwalb算法[8]将图像分割成不同的区域。之后根据图像的语义信息运用选择性搜索[9]合并分割区域，进而生成图像的稀疏分割图。最后使用自适应着色算法为合并后的分割区域着色，公式为

$$\begin{cases}S_{i,j}=(\theta_1\times\overline{S}+\theta_2\times\tilde{S})^{\mu}\\(\theta_1,\theta_2)=\begin{cases}(0,1), & \sigma(S)<\gamma_1\\(0.5,0.5), & \gamma_1<\sigma(S)<\gamma_2\\(1,0), & \gamma_2<\sigma(S)\end{cases}\end{cases}。\tag{3}$$

当$\gamma_1=20$、$\gamma_2=40$、$\mu=1.2$时，可以有效地增加图像的对比度并减少模糊效果。结构表示处理的结果如图3所示。

图3　结构表示

Fig. 3　Structure representation

3)纹理表示

卡通图像的高频特征是图像卡通风格化过程中的主要学习目标，但亮度和颜色信息的影响使卡通图像和真实照片易于区分。本文使用随机颜色偏移算法将三通道RGB彩色图像生成单通道图像，在保留图像高频纹理的同时减少了真实图像中颜色和亮度的影响。

$$F_{\mathrm{rcs}}(I_{\mathrm{rgb}})=(1-\alpha)(\beta_1\times I_{\mathrm{r}}+\beta_2\times I_{\mathrm{g}}+\beta_3\times I_{\mathrm{b}})+\alpha\times Y,\tag{4}$$

式中：I_{rgb}表示3通道RGB彩色图像；I_{r}、I_{g}和I_{b}表示三个彩色通道；Y表示从RGB彩色图像转换而来的标准灰度图像。设置$\alpha=0.8$，β_1、β_2、$\beta_3\sim U(-1,1)$。如图4所示，随机颜色偏移算法可以随机生成去除亮度和颜色信息的不同强度的图像。

2.2　生成网络

所提方法的生成网络的主体结构如图5所示，其中c表示每个模块的输出尺寸。该网络借鉴了U-net的设计思路，是一个对称的编码-解码网络，在编码和解码的对应阶段使用长跳跃连接，叠加低层次的特征使解码网络可以融合不同规模的特征，增强图像卡通风格化结果的细节表现。生成网络各模块结构主要由标准卷积模块(Conv_Block)、深度可分离卷积模块(DSConv)、反向残差模块(IRB)、下采样卷积模块(Down-Conv)、上采样卷积模块(Up-Conv)组成，最后一层为具有1×1卷积核的卷积层，后接Tanh非线性激活函数。

标准卷积模块(Conv_Block)、深度可分离卷积模

图 4　纹理表示

Fig. 4　Texture representation

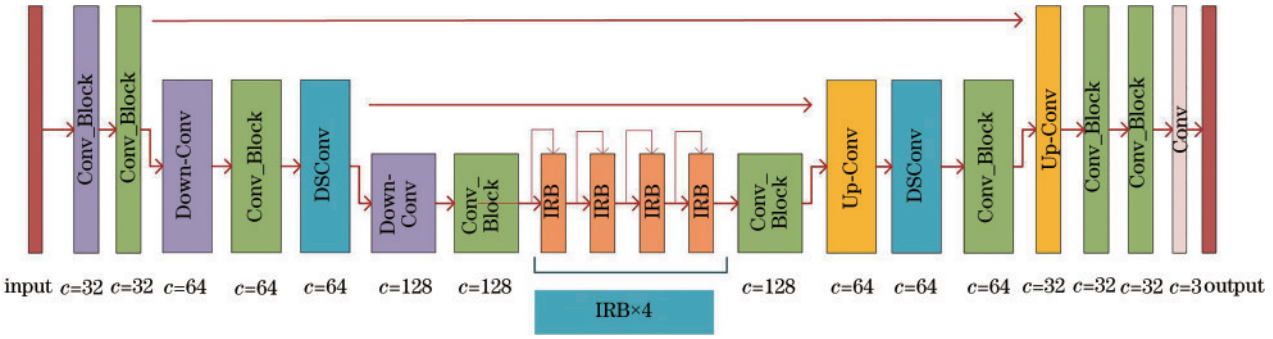

图 5　生成网络的主体结构

Fig. 5　Main structure of the generative network

块(DSConv)、反向残差模块(IRB)、下采样卷积模块(Down-Conv)、上采样卷积模块(Up-Conv)的具体结构如图 6 所示，k 表示卷积核大小，s 表示步长，H 和 W 分别表示图像的像素高和宽。对输入尺寸进行填充，保证步长为 1 时卷积层的输入维度和输出维度一致。Conv_Block 由具有 3×3 卷积核的标准卷积和 Leaky ReLU(LReLU)激活函数组成。DSConv 由具有 3×3 卷积核的深度可分离卷积和 Leaky ReLU 激活函数组成。IRB 由标准卷积和深度可分离卷积[10]组成，并以反向残差方式连接。

假设输入图像尺寸为 $H\times W\times c_1$，卷积核大小为 $h_1\times w_1\times c_1$，输出图像尺寸为 $H\times W\times c_2$，标准卷积对图像的 g 个通道同时进行卷积，则标准卷积的参数量为 $h_1\times w_1\times c_1\times c_2$；而深度可分离卷积用 g 个卷积对图像的 g 个通道分别进行卷积，一次卷积操作之后输出 g 个中间数值，再通过一个 $1\times1\times g$ 的卷积核得到最后卷积结果，在卷积操作后，标准卷积的参数量为 $h_1\times w_1\times c_1\times c_2\times g$，深度可分离卷积产生的参数量为 $h_1\times w_1\times c_1\times c_2\times 1/g$。相比之下，深度可分离卷积构造的网络参数量明显较少。申毫等[11]在网络中引入倒置残差结构，有效地减少了参数量。为了有效地减少生成器的参数量，本文引入反向残差结构[12]来构造轻量级网络模型，设计了 DSConv 与 IRB，网络的中间使用了 4 个连续相同的 IRB。IRB 使用深度可分离

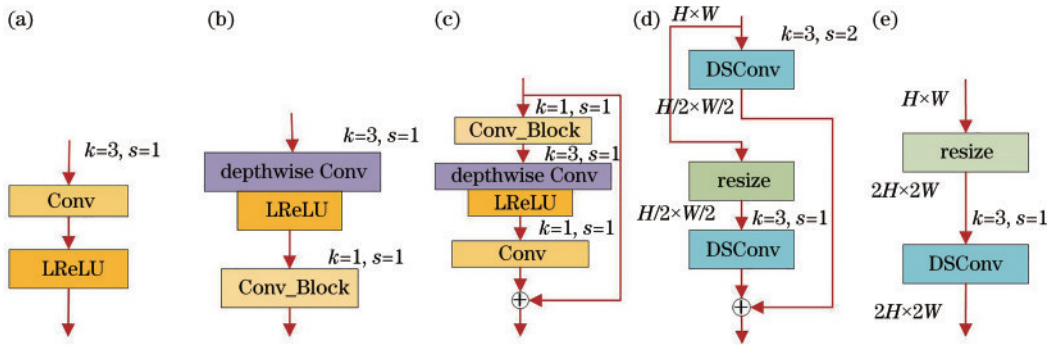

图 6　生成网络的各模块结构。(a)标准卷积模块；(b)深度可分离卷积模块；(c)反向残差模块；(d)下采样卷积模块；(e)上采样卷积模块

Fig. 6　Each module structure of the generative network. (a) Conv_Block; (b) DSConv; (c) IRB; (d) Down-Conv; (e) Up-Conv

卷积和反向残差结构，与标准残差块相比，IRB可显著减少网络的参数量和计算工作量。

使用Down-Conv来降低特征映射的分辨率，此模块可以解决最大池化造成特征信息丢失的问题。先使用步长为2的DSConv将特征映射的大小调整为输入特征映射的1/2，后连接步长为1的DSConv，最后的输出是步长为2的DSConv和步长为1的DSConv的输出之和。为了避免生成的图像出现棋盘伪影而影响生成的卡通图像质量，使用Up-Conv提高特征映射的分辨率，此模块输出的特征映射的大小为输入尺寸的2倍。

2.3　判别网络

使用PatchGAN[13]中的判别网络结构，如图7所示，在每个3×3卷积层（最后一层除外）之后放置谱归一化层[14]，可以在训练网络时强化Lipschitz约束并稳定训练，其最后一层是1×1卷积层。此判别网络输出特征图中的每个元素对应于输入图像中的一个图像块，图像块的大小等于感受野大小，用于判断图像块属于卡通图像还是生成图像，对每个图像块进行有差别的判别，实现了对局部图像特征的提取和表征。最终整个图像的判别结果是对所有图像块的判别结果求均值得到的。此判别网络结构增强了对细节的辨别能力，加快了训练速度。

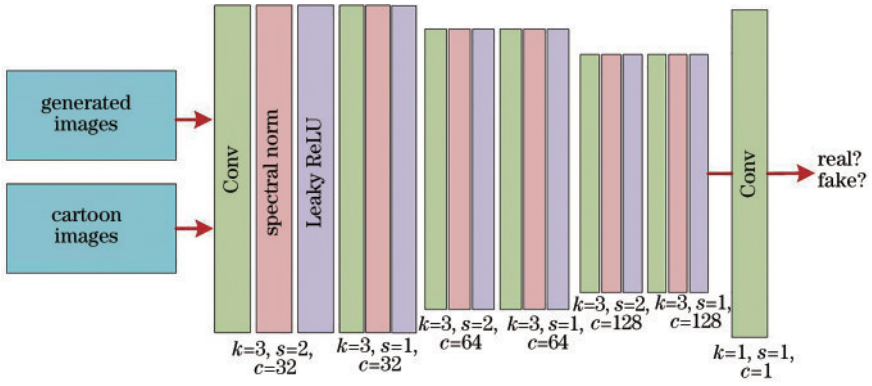

图7　判别网络结构

Fig. 7　Discriminant network structure

2.4　损失函数

使用判别器D_s来判断生成网络输出和卡通图像是否具有相似的曲面，引导生成网络G学习存储在提取的表面表示中的信息。令$\boldsymbol{I}_p$表示输入照片，$\boldsymbol{I}_c$表示参考卡通图像，表面损失L_{sur}的公式为

$$L_{sur}(G,D_s)=\log D_s\left[F_{dgf}(\boldsymbol{I}_c,\boldsymbol{I}_c)\right]+\log\left\{1-D_s\left\{F_{dgf}\left[G(\boldsymbol{I}_p),G(\boldsymbol{I}_p)\right]\right\}\right\}。\tag{5}$$

使用预先训练过的VGG-19网络提取高级特征，来加强结果和提取的结构之间的空间约束。令F_{st}表示结构提取，结构损失L_{str}可表示为

$$L_{str}=\left\|\mathrm{VGG}_n\left[G(\boldsymbol{I}_p)\right]-\mathrm{VGG}\left\{F_{st}\left[G(\boldsymbol{I}_p)\right]\right\}\right\|,\tag{6}$$

式中：$\mathrm{VGG}_n(\cdot)$表示VGG-19网络中第n层的特征图，n层表示VGG-19网络中的Conv4－4。

引入判别器D_t来区分从生成网络输出和卡通图像中提取的纹理，引导生成网络G学习卡通图像的清晰轮廓和精细纹理，纹理损失L_{tex}表示为

$$L_{tex}(G,D_t)=\log D_t\left[F_{rcs}(\boldsymbol{I}_c)\right]+\log\left\{1-D_t\left\{F_{rcs}\left[G(\boldsymbol{I}_p)\right]\right\}\right\}。\tag{7}$$

为了减少生成图像中的高频噪声，使用总变差损失L_{tv}来给生成图像施加空间平滑度。

$$L_{tv}=\frac{1}{H\times W\times C}\left\|\nabla_x\left[G(\boldsymbol{I}_p)\right]+\nabla_y\left[G(\boldsymbol{I}_p)\right]\right\|。\tag{8}$$

为了使卡通化图像与输入照片的语义保持不变，在预先训练的VGG-19特征空间上计算内容丢失L_{con}，公式为

$$L_{con}=\left\|\mathrm{VGG}_n\left[G(\boldsymbol{I}_p)\right]-\mathrm{VGG}_n(\boldsymbol{I}_p)\right\|。\tag{9}$$

为了能够对局部特征进行卡通化，使用具有稀疏性的L_1范数进行正则化。所提模型是具有1个生成网络和2个判别网络的轻量级GAN框架，结合从表面、结构、纹理三种表示中学习到的特征，通过调整来优化风格化结果，可表示为

$$L_{total}=\lambda_1\times L_{sur}+\lambda_2\times L_{tex}+\lambda_3\times L_{str}+\lambda_4\times L_{con}+\lambda_5\times L_{tv}。\tag{10}$$

通过调节损失参数λ_1、λ_2、λ_3、λ_4来学习不同的卡通风格。使用Adam算法[15]优化生成网络和判别网络。训练时的学习率设置为2×10^{-4}，批次大小设置为8。

3　实验结果与分析

实验平台为Ubuntu18.04、Tensorflow以及Python3.6。采用易于获取的非配对真实图像与卡通图像作为训练数据。训练数据中的真实世界的照片是从FFHQ数据集[16]中收集的5000张人脸图像，从文献[17]中收集的5000张风景图像。对于卡通图像，从动漫中收集5000张人脸图像和三种不同卡通风格的15320张风景图像。收集的卡通风格包括新海诚、细田

直人和宫崎骏三种风格。另外在DIV2K数据集[18]中收集了2932张卡通图像和1533张真实照片作为验证集。在开始训练之前，对生成器执行初始化训练，以使GAN的训练更容易、更稳定。首先对包含内容损失的生成网络进行50000次的预训练，然后共同优化基于GAN的框架。10万次迭代或算法收敛后停止训练。在训练过程中，所有图像的大小调整为256×256。

生成网络和判别网络是互相促进的。生成网络的训练目标是使生成图像更趋近于真实卡通图像，进而迷惑判别网络使其分辨不出生成图像和卡通图像，判别网络的目的是准确分辨真假卡通图像。所以生成网络生成的图像会越来越真，判别网络的辨别能力会越来越强，最终达到一个纳什平衡。损失函数收敛情况如图8所示。可以看出：D_s和D_t的损失d_loss_blur和d_loss_gray逐渐收敛，它们的总损失d_loss_total也逐渐收敛，表明判别网络的辨别能力越来越强并逐步稳定；总变差损失tv_loss逐渐收敛并逐渐稳定，避免了图像过度风格化；生成网络中的各部分损失以及总损失函数g_loss_total都逐渐增大，这表明卡通风格化效果越来越明显，和真实图片相似度就会越来越小，与卡通图片的相似度越来越大，模型的卡通风格化能力逐渐提高。

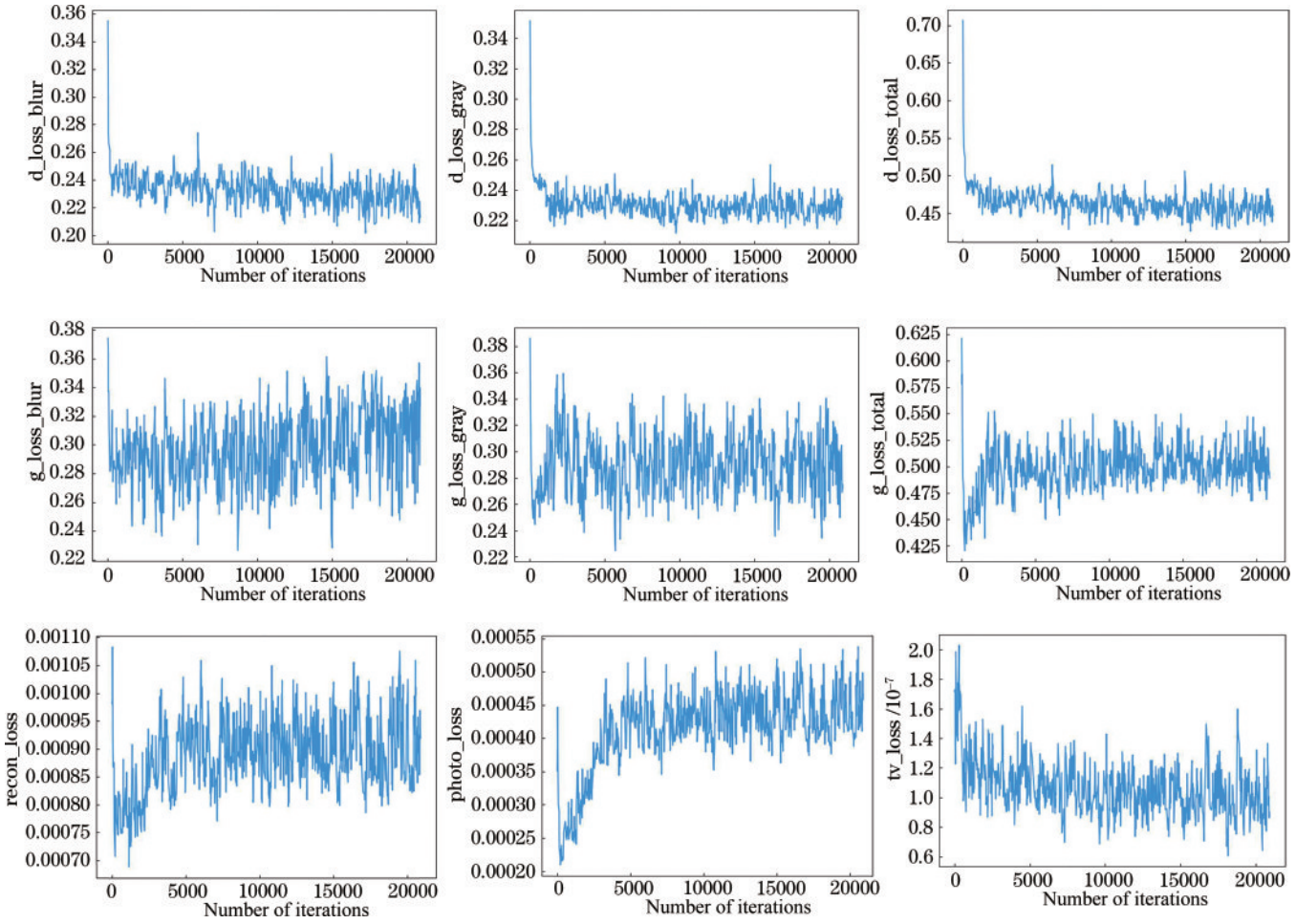

图8　损失函数收敛情况

Fig. 8　Convergence of loss functions

3.1　定性比较

图9显示了所提方法与其他方法的定性比较。可以清楚地看出：所提方法有助于生成具有干净轮廓的卡通图像。由于卡通图像具有清晰的边缘、平滑的色彩底纹和相对简单的纹理，快速风格迁移方法并不能有效地生成具有卡通风格的图像。CartoonGAN和CycleGAN都能有效地模仿卡通风格。CycleGAN可以很好地应用于非配对的图像风格转换，但是由于其只是粗暴地对抗学习大致的卡通风格，忽略了局部细节，导致生成的图像颜色暗沉，并且很容易在局部区域过度风格化，从而导致生成的图像丢失原始照片的内容。虽然CartoonGAN提出了新颖的边缘损失，但最终并没有排除颜色对卡通风格迁移的影响，因此生成的图像在局部区域产生了明显的彩色伪影，并失去了原始内容图像的颜色。所提方法将卡通图像的风格学习分为了表面、结构、纹理三部分，可以通过调节相应的损失来更好地学习不同的卡通风格，在有效地减少了伪影的同时保留了原真实照片的精细细节。所提方法在生成颜色和谐、边界清晰、细节精细和噪声较少的图像方面优于其他方法。

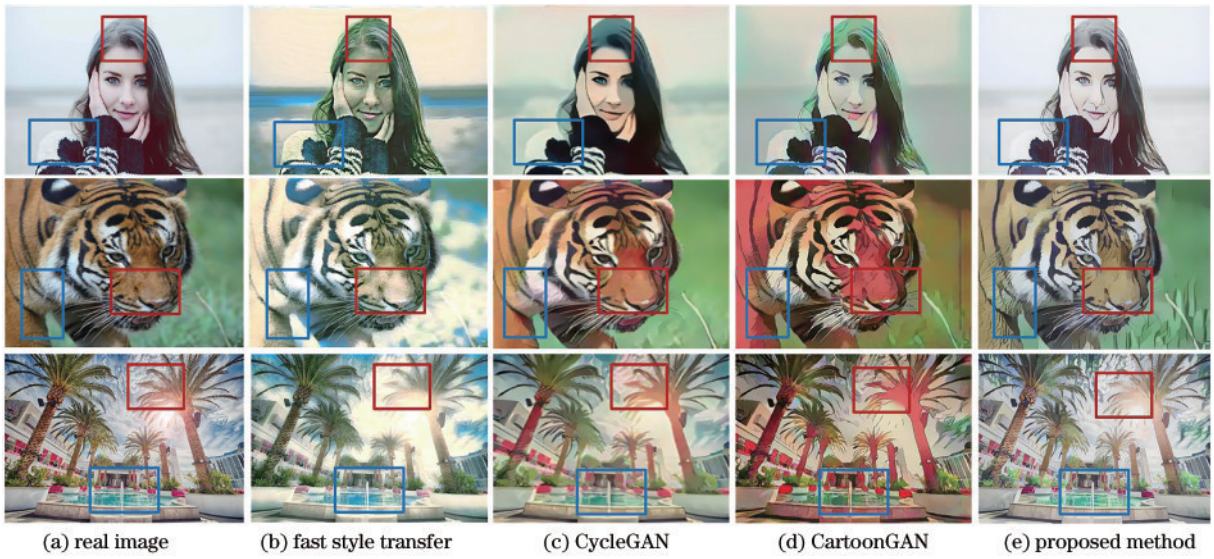

图 9　图像卡通风格化主观效果对比

Fig. 9　Comparison of subjective effects of image animation stylization

为了充分验证所提方法的图像卡通风格化的效果，在收集的 3 个卡通风格数据集上进行大量实验，并与 CartoonGAN 方法进行比较，实验结果如图 10 所示。可以看出：CartoonGAN 方法生成的图像整体呈现暗沉效果，并不能很好地模仿对应风格的卡通图像，尤其是宫崎骏这种颜色明艳的卡通风格，生成图像的局部区域出现了色彩扭曲和高频伪影，比如暗红色的天空和建筑；与 CartoonGAN 相比，所提方法由于纹理模块，有效地减小了图像生成过程中颜色和亮度的影响，并且结构模块使得真实照片原本的精细细节得到保留，表面模块学习到了卡通的光滑表面，最后呈现的生成图像成功地学习了三种不同的卡通风格，并生成了高质量的卡通视觉图像。

图 10　三种卡通风格实验结果与所提方法和 CartoonGAN 的比较

Fig. 10　Experimental results of three animation styles and comparison between proposed method and CartoonGAN

为了验证所提方法的适用性，进一步在各种真实场景进行图像生成效果验证，包括自然景观、城市景观、建筑物、人、动物和植物，生成结果如图 11 所示。可以看到：所提方法可以适用于不同场景的照片，生成颜色和谐、边界清晰、细节精细和噪声较少的卡通风格图像。

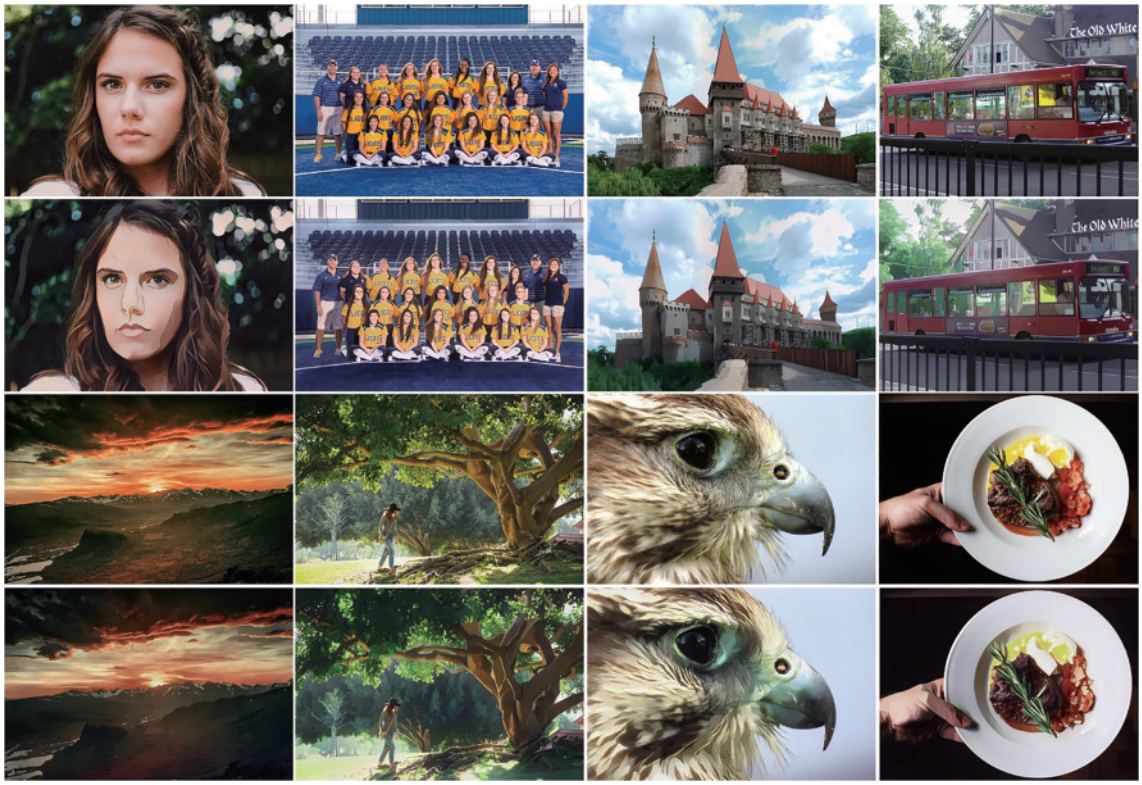

图 11　所提方法在不同场景的结果

Fig. 11　Results of proposed method in different scenarios

3.2　定量分析

为了验证所提方法的轻量级网络性能，分别与图像转换方法 CycleGAN 和卡通风格转换方法 CartoonGAN 在 DIV2K 数据集上进行性能比较。CycleGAN 要同时训练两对生成网络与判别网络。CartoonGAN 只有一对生成网络与判别网络，但是生成网络要使用标准卷积和标准残差块。所提方法的生成网络模型是基于 CartoonGAN 的生成网络模型进行改进的，加入了深度可分离卷积和反向残差块。不同网络模型在模型参数、网络模型尺寸大小、模型计算量、每张图片的推理时间 4 个方面的性能比较如表 1 所示。在表 1 中，每个输入照片的大小为 256×256。可以看出：与其他两种方法相比，所提方法由于使用深度可分离卷积和反向残差块，显著减少了参数量，降低了计算成本，并且对每幅图像的推理速度更快。

表 1　不同网络模型性能的比较

Table 1　Performance comparison of different network models

Network	Number of parameters	Model size /MB	FLOPs	Inference time /ms
Proposed network	2526810	11.33	25.32	16
CartoonGAN	12253152	46.74	108.98	51
CycleGAN	24310250	50.34	155.32	132

Frechet 初始距离（FID）[19]被广泛用于定量评估合成图像的质量。使用预训练的 Inception-V3 模型[20]来提取图像的高级特征，并计算两个图像分布之间的距离。为了验证所提方法对卡通图像进行解耦表征学习的合理性和有效性，在 DIV2K 数据集上进行分类实验和基于 FID 的定量实验，结果如表 2 所示。在训练数据集上训练一个二值分类器来区分真实照片和卡通图像。此分类器是根据在判别网络上添加一个全连接层来设计的。然后在 DIV2K 数据集上评估训练的分类器，以验证每个表示对学习卡通风格的影响。结果显示：提取的图像表示成功地欺骗了训练的分类器；与真实图像相比，三种表示分别提取的卡通表示图像的准确率都较低，并且 FID 都更小，表明解耦表征有助于缩小生成图像与卡通图像之间的差距。

使用 FID 来评估不同方法的性能，并在 DIV2K 数据集进行了对比实验，结果如表 3 所示。与卡通图像

表 2　解耦表征的分类准确度和 FID 评估

Table 2　Classification accuracy and FID evaluation of decoupling characterization

Parameter	Surface	Structure	Texture	Photo
Accuracy	0.821	0.6341	0.8341	0.9481
FID	113.57	112.56	112.41	162.88

表 3 基于 FID 的性能评价

Table 3 Performance evaluation based on FID

Parameter	Photo	Fast style transfer	CycleGAN	CartoonGAN	Proposed method
FID to cartoon	162.89	146.34	141.50	130.76	101.31
FID to photo		103.48	122.12	58.13	28.79

相比，所提方法生成的图像的 FID 数值最小，这证明它生成的结果与卡通图像最相似。与现实图像相比，所提方法的输出具有最小的 FID 数值，表明所提方法较好地保留了图像内容信息。

3.3 消融实验

图 12 显示了所提方法分别对表面、结构、纹理消融的实验结果，其中 W/O 表示消融。消融表面会出现噪波和混乱的细节。图 12(b)显示了不清晰的岩石结构和苔藓上的噪音，原因是引导滤波抑制高频信息并保持表面平滑。消融结构会导致图 12(c)中的高频噪声，苔藓和树干上出现了严重的椒盐噪声，这是因为结构表示将图像展平并删除高频信息。去除纹理会导致生成的图像中出现混乱的细节，如图 12(d)所示，树干和瀑布上的不规则纹理仍然存在。这是由于纹理中存储着高频信息，缺乏纹理会降低模型的卡通风格化能力。作为比较，所提方法的完整结果如图 12(e)所示，具有平滑的特征、清晰的边界和更少的噪声。总之，这三种表示有助于提高所提方法的卡通风格化图片的质量。

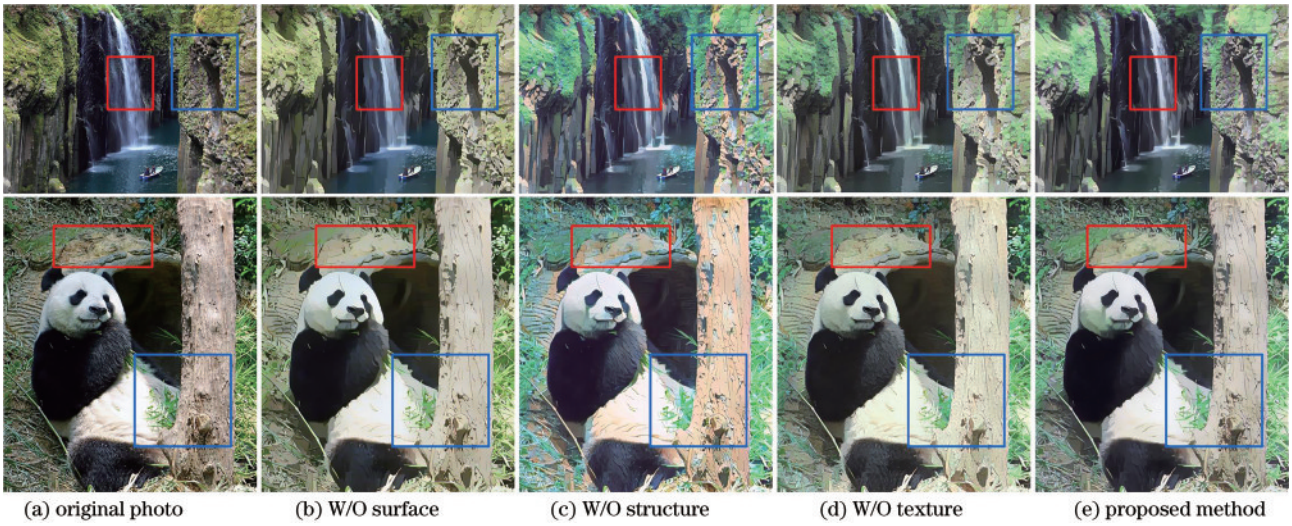

图 12 消融实验

Fig. 12 Ablation study

4 结 论

提出了一种轻量级生成对抗网络架构用于图像卡通风格化。它可以从真实世界的照片中生成高质量的卡通化图像。运用深度可分离卷积与反向残差块，有效减少了网络参数量，降低了计算成本。根据卡通图像的绘制过程，将图像分解为表面、结构和纹理三种表示，使用相应的图像处理模块提取用于网络训练的三种表示。表面模块学习到了卡通的光滑表面，结构模块使得真实照片原本的细节得到保留，纹理模块有效地减小了图像生成过程中颜色和亮度的影响。通过定性比较与定量分析，结果显示所提模型具有更少的参数量和更小的计算成本，并在生成颜色和谐、边界清晰、细节精细和噪声较少的图像方面优于以前的方法。在未来的工作中，将加强对生成图像局部细节的优化，并期望把图像卡通风格迁移网络模型应用在更多其他的转换工作中，如视频等。

参 考 文 献

[1] Gatys L A, Ecker A S, Bethge M. Image style transfer using convolutional neural networks[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 2414-2423.

[2] Chen Y, Lai Y K, Liu Y J. Transforming photos to comics using convolutional neural networks[C]//2017 IEEE International Conference on Image Processing, September 17-20, 2017, Beijing, China. New York: IEEE Press, 2017: 2010-2014.

[3] 沈瑜, 杨倩, 苑玉彬, 等. 目标边缘清晰化的图像风格迁移[J]. 激光与光电子学进展, 2021, 58(12): 1210021.
Shen Y, Yang Q, Yuan Y B, et al. Image style transfer with clear target edges[J]. Laser & Optoelectronics Progress, 2021, 58(12): 1210021.

[4] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//2017 IEEE International Conference on

Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2242-2251.

[5] Chen Y, Lai Y K, Liu Y J. CartoonGAN: generative adversarial networks for photo cartoonization[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 9465-9474.

[6] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014-09-04)[2021-06-06]. https://arxiv.org/abs/1409.1556.

[7] Wu H K, Zheng S, Zhang J G, et al. Fast end-to-end trainable guided filter[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 1838-1847.

[8] Felzenszwalb P F, Huttenlocher D P. Efficient belief propagation for early vision[J]. International Journal of Computer Vision, 2006, 70(1): 41-54.

[9] Uijlings J R R, van de Sande K E A, Gevers T, et al. Selective search for object recognition[J]. International Journal of Computer Vision, 2013, 104(2): 154-171.

[10] Chollet F. Xception: deep learning with depthwise separable convolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 1800-1807.

[11] 申毫, 孟庆浩, 刘胤伯. 基于轻量卷积网络多层特征融合的人脸表情识别[J]. 激光与光电子学进展, 2021, 58(6): 0610005.
Shen H, Meng Q H, Liu Y B. Facial expression recognition by merging multilayer features of lightweight convolutional networks[J]. Laser & Optoelectronics Progress, 2021, 58(6): 0610005.

[12] Sandler M, Howard A, Zhu M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 4510-4520.

[13] Isola P, Zhu J Y, Zhou T H, et al. Image-to-image translation with conditional adversarial networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 5967-5976.

[14] Miyato T, Kataoka T, Koyama M, et al. Spectral normalization for generative adversarial networks[EB/OL]. (2018-02-16)[2021-06-08]. https://arxiv. org/abs/1802.05957.

[15] Kingma D P, Ba J. Adam: a method for stochastic optimization[EB/OL]. (2014-12-22)[2021-08-06]. https://arxiv.org/abs/1412.6980.

[16] Karras T, Laine S, Aila T M. A style-based generator architecture for generative adversarial networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 4396-4405.

[17] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2242-2251.

[18] Agustsson E, Timofte R. NTIRE 2017 challenge on single image super-resolution: dataset and study[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 1122-1131.

[19] 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 30, December 4-9, 2017, Long Beach, CA, USA. New York: Curran Associates, 2017: 6629-6640.

[20] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 2818-2826.