基于动态感受野的自适应多尺度信息融合的图像转换

尹梦晓[①②] 林振峰[①] 杨 锋*[①②]

[①](广西大学计算机与电子信息学院 南宁 530004)

[②](广西多媒体通信与网络技术重点实验室 南宁 530004)

摘 要：为提高图像转换模型生成图像的质量，该文针对转换模型中的生成器进行改进，同时探究多样化的图像转换，拓展转换模型的生成能力。在生成器的改进方面，利用选择性(卷积)核模块(SKBlock)的动态感受野机制获取和融合生成器中每个上采样特征的多尺度信息，借助特征的多尺度信息和动态感受野构造选择性(卷积)核的生成式对抗网络(SK-GAN)。与传统生成器相比，SK-GAN以动态感受野获取多尺度信息的生成结构提高了生成图像的质量。在多样化图像转换方面，基于SK-GAN在草图合成真实图像任务提出带引导图像的选择性(卷积)核的生成式对抗网络(GSK-GAN)。该模型利用引导图像指导源图像的转换，通过引导图像编码器提取引导图像特征，然后由参数生成器(PG)和特征转换层(FT)将引导图像特征的信息传递至生成器。此外，该文还提出双分支引导图像编码器以提高转换模型的编辑能力，以及利用引导图像的隐变量分布实现随机样式的图像生成。实验表明，改进后的生成器有助于提高生成图像质量，SK-GAN在多个数据集中获得合理的生成结果。GSK-GAN不仅保证了生成图像的质量，还能生成更多样式的图像。

关键词：图像转换；多尺度信息；动态感受野；自适应特征选择

中图分类号：TN911.73; TP391 **文献标识码**：A **文章编号**：1009-5896(2021)08-2386-09

DOI: 10.11999/JEIT200675

Adaptive Multi-scale Information Fusion Based on Dynamic Receptive Field for Image-to-image Translation

YIN Mengxiao[①②] LIN Zhenfeng[①] YANG Feng[①②]

[①](*School of Computer and Electronics Information*, *Guangxi University*, *Nanning* 530004, *China*)

[②](*Guangxi Key Laboratory of Multimedia Communications and Network Technology*, *Guangxi University*, *Nanning* 530004, *China*)

Abstract: In order to improve the quality of the generated images by the image translation model, the generator in the translation model to obtain high-quality generated images is improved, the diversified image translation is explored and the generation ability of the translation model is expanded. In terms of generator improvement, the dynamic receptive field mechanism of Selective Kernel Block (SKBlock) is used to obtain and fuse the multi-scale information of each up sampling feature in the generator. With the help of multi-scale information of features and dynamic receptive field, the Selective Kernel Generative Adversarial Network (SK-GAN) is constructed. Compared with the traditional generator, SK-GAN improves the quality of the generated image by using dynamic receptive field to obtain multi-scale information. In terms of diversified image translation, the Selective Kernel Generative Adversarial Network with Guide (GSK-GAN) is proposed based on SK-GAN in sketch synthesis realistic image task. GSK-GAN uses the guided image to guide the source image translation and extracts the guide image features through the guided image encoder. Then transmits information of the guided image features to the generator by Parameter Generator (PG) and Feature Transformation (FT). In addition, a dual branch guided image encoder is proposed to improve the editing ability of the translation model. The random style image generation is realized by using the latent variable distribution of the guide image. The experimental results show that the improved generator is helpful to improve the quality of the generated images, and SK-GAN can obtain reasonable results in multiple datasets. GSK-GAN no only ensures the quality of the generated images, but also generates more styles of images

Key words: Image translation; Multi-scale information; Dynamic receptive field; Adaptive feature selection

收稿日期：2020-08-04；改回日期：2021-01-04；网络出版：2021-01-10

*通信作者： 杨锋 yf@gxu.edu.cn

基金项目：国家自然科学基金(61762007, 61861004)，广西自然科学基金(2017GXNSFAA198269, 2017GXNSFAA198267)

Foundation Items: The National Natural Science Foundation of China (61762007, 61861004), The Natural Science Foundation of Guangxi (2017GXNSFAA198269, 2017GXNSFAA198267)

1　引言

图像转换[1]的本质是条件图像生成，目标是将源图像转换成目标图像，如草图生成真实图[2]。由于源图像和目标图像之间存在很大差异，因此需要复杂的变化来完成转换。本文提出一种有效地完成不同类型图像之间转换的方法。

深度神经网络为图像生成提供了有效方法[3–5]，其中深度卷积生成式对抗网络(Deep Convolutional Generative Adversarial Networks, DCGAN)[5]自动学习上下采样以避免信息丢失，提高生成图像质量。在图像转换方面，Pix2pix[1]基于DCGAN，增加编码器实现不同域图像的转换，同时以跳跃连接使编码器的特征绕过瓶颈层直接传至生成器，这些方式提高了Pix2pix对不同转换任务的兼容性以及生成图像质量。后续的研究工作更多关注损失函数的设计[6]、修改生成机制[7]和拓展生成目标[8–10]等，对生成器的研究较少，而生成器作为直接生成图像的部分有较大的探索空间。本文通过改进生成器结构提出选择性(卷积)核的生成式对抗网络(Selective Kernel Generative Adversarial Network, SK-GAN)，避免引入额外的损失函数和超参数，获得高质量的生成图像。

Sun等人[11]提出空间金字塔注意力池(Spatial Pyramid Attentive Pooling, SPAP)模块，利用多级不同的感受野和像素级自适应特征选择从某一个上采样的特征中获取图像由粗到细的变化信息。SPAP在DCGAN[5]和循环生成式对抗网络(Cycle Generative Adversarial Networks, CycleGAN)[12]中发挥了良好性能，一定程度上提高了生成图像质量。本文针对上采样过程每层上采样特征，利用选择性(卷积)核模块(Selective Kernel Block, SKBlock)[13]中的动态感受野机制融合该特征的多尺度信息，这样不仅适应了特征尺度的变化，同时改善了传统生成器以固定感受野解码特征的形式。本文将SKBlock引入生成器，并尝试了不同的结合方式。

在诸如草图转换至真实图像等转换任务中，由引导图像指导的图像生成更具现实意义，该任务根据引导图像的信息生成指定的图像。如何有效利用引导图像的信息是处理此类任务的关键[10]，文献[10]定义参数生成器(Parameter Generator, PG)和特征转换层(Feature Transformation, FT)，通过两个编码器之间的双向特征传递实现源图像和引导图像的信息融合。该方式以特征的局部信息生成传递参数，避免全局一致的变化，然后通过仿射变换融合图像信息。本文结合PG和FT，基于SK-GAN在草图合成真实图像任务提出带引导图像的选择性(卷积)核的生成式对抗网络(Guided SK-GAN, GSK-GAN)，该模型将引导图像信息传至生成器并由动态感受野获取对应的多尺度信息。此外，本文还提出双分支引导图像编码器，用于实现不同引导图像对应生成图像之间的插值。同时还以变分推断[3]学习引导图像的隐变量分布，使GSK-GAN在预测时能采样更多指导信息，实现多样化生成。实验表明，GSK-GAN不仅能够根据引导图像生成指定的图像，还能生成连续变化和引导图像信息之外的图像，同时保证图像质量。

本文主要贡献如下：

(1) 提出动态感受野的自适应多尺度信息融合的生成器结构，使用SKBlock根据上采样特征大小自适应调整感受野，获取特征多尺度信息，改进了传统生成器对特征多尺度信息的忽略和感受野的固定形式。基于此生成器提出图像转换模型SK-GAN。

(2) 基于SK-GAN在草图合成真实图像任务提出GSK-GAN，该模型将引导图像信息直接传至生成器，借助SKBlock获取对应多尺度信息，避免影响源图像编码，保证了图像质量，更利于模型的拓展。

(3) 在GSK-GAN中提出双分支引导图像编码器，通过权重控制每个分支信息的转换程度，实现不同引导图像对应生成图像之间的插值。同时使用额外的生成器，用于生成引导图像信息之外的图像。双分支引导图像编码器学习引导图像的隐变量分布，生成器从该分布中采样隐变量以获得更多指导信息。

2　相关工作

本节简要介绍本文转换模型所密切相关的图像转换(image-to-image translation)、多分支卷积结构和多模态图像转换等工作。

2.1　图像转换

图像生成模型主要包括变分自动编码器[3](Variational AutoEncoder, VAE)和生成式对抗网络[4](Generative Adversarial Networks, GAN)两种类型，其中GAN的对抗学习方式使生成图像更清晰且应用更广泛。图像转换模型以源图像为条件，利用编码器将源图像映射成潜在编码，生成器将潜在编码转换成对应目标图像。Isola等人[1]最早提出同时兼容图像着色、草图合成真实图像和图像补全等多种转换任务的图像转换模型。后续工作分别从增加损失函数[6]、修改生成机制[7]和拓展生成目标[8–10]等方面提升转换模型的处理能力，其中文献[9,10]以引导图像控制目标图像的生成，实现多模态图像转换。现有通用图像转换模型的改进缺少对生成器的关注，而生成器对图像质量的影响更直接。本文

从生成器入手，改进图像转换模型，提高图像生成质量。

2.2 多分支卷积结构

InceptionNets[14]将多分支卷积用于图像分类，以获取特征的多尺度信息。柳长源等人[15]使用类似的结构取得了较好的实验结果。在超分辨率方面，Li等人[16]提出基于残差块的多尺度残差模块(Multi-Scale Residual Block, MSRB)，该结构融合特征的多尺度信息，提高了重建图像的质量。选择性(卷积)核网络(Selective Kernel Network, SKNet)[13]利用两个不同感受野的分支，让网络自适应地从某一个分支中获取信息，增强了网络对目标的适应性。本文将SKNet中的SKBlock引入生成器，增强转换模型自适应调节和提取特征的能力。

2.3 多模态图像转换

传统转换模型仅以源图像为输入，只能产生确定的输出，但实际应用中常存在一对多的转换情况。Zhu等人[8]针对上述问题提出双向循环生成式对抗网络(Bidirectional cycle Generative Adversarial Networks, BicycleGAN)，通过成对图像中目标图像的隐变量改变生成图像的样式，但预测时从正态分布中采样的隐变量无法获取指定样式，只能生成随机样式的图像。纹理生成式对抗网络(Texture Generative Adversarial Networks, TextureGAN)[9]以引导图像提供额外信息，通过风格迁移中常用的内容和样式损失函数将引导图像信息迁移至生成图像。文献[10]提出参数生成器和特征转换层，将引导图像信息的迁移过程加入转换模型，避免过多的损失函数使转换模型的训练变得复杂。以引导图像指导源图像的转换只能生成与引导图像相关的图像，限制了多样性生成。本文将使用隐变量和引导图像提供额外信息的方式结合，不仅能够获得指定的生成图像，还能通过隐变量产生更多不同的结果。此外，本文还提出双分支引导图像编码器，实现在已有的引导图像中编辑生成图像，进一步增强了转换模型的处理能力。

3 主要方法

本文目标是将源图像x转换成目标图像y，即$T_{\mathrm{SK}}:(x)\to y$，其中T_{SK}表示SK-GAN的编码器和生成器。多模态图像转换任务增加了引导图像和双分支引导图像编码器，对应的转换过程描述为$T_{\mathrm{GSK}}:(x,c_1,c_2,\omega)\to y$，其中$T_{\mathrm{GSK}}$表示GSK-GAN的源图像编码器、双分支引导图像编码器、与引导图像信息对应的生成器和与隐变量对应的生成器，c_1和c_2分别表示不同的引导图像，ω和$(1-\omega)$分别表示双分支引导图像编码器中不同分支的权重。3.1节和3.2节将分别介绍SK-GAN和GSK-GAN的实现。

3.1 基于动态感受野的自适应多尺度信息融合的转换模型SK-GAN

本文使用Pix2pix[1]结构实现SK-GAN，如图1所示，该模型主要包括编码器E、生成器G和判别器D。编码器和生成器将源图像映射至目标图像，判别器通过判断输入图像的真假优化转换过程。

生成器上采样阶段由多个转置卷积组成，如图2中模式1。本文在生成器中引入SKBlock，使生成器获得动态感受野机制。SKBlock的结构如图3(a)所示，提取和融合多尺度信息的步骤包括：(1)使用文献[13]中SKBlock感受野的设置，以3×3和5×5的感受野分支获取特征I_F的多尺度信息；(2)将2个分支获取的特征相加并以全局平均池化GAP统计全局信息；(3)通过全连接FC将全局信息的特征降维并增加归一化层和激活函数，提高模块的学习能力，然后再次经过全连接恢复至原维度；(4)利用激活函数和全连接层输出的特征学习选择权重，该权重与每个分支输出的特征相乘以控制多尺度信息的转换；(5)通过像素级和融合每个分支的转换信息输出特征O_F。图3(b)展示了多尺度信息的动态选择过程，不同感受野获取的特征$F_{3\times3}$和$F_{5\times5}$分别由对应的权重$W_{3\times3}$和$W_{5\times5}$控制转换程度，通过权重变化改变固定感受野的特征提取方式。

图2模式2和模式3分别展示不同的SKBlock与生成器的结合方式。模式2简单地将SKBlock加入每个上采样层之间，而模式3以残差模式将SKBlock与生成器结合。本文主要基于模式3进行实验，并在4.4.1节讨论每个模式的生成效果。

3.2 多模态转换模型GSK-GAN

GSK-GAN基于SK-GAN并增加双分支引导图像编码器和额外的生成器，如图4所示，GSK-GAN

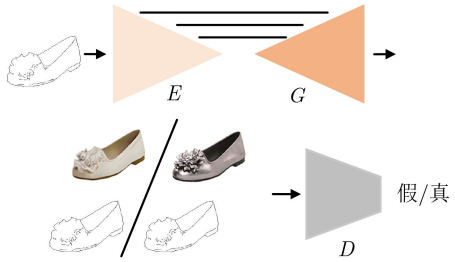


图 1 转换模型结构

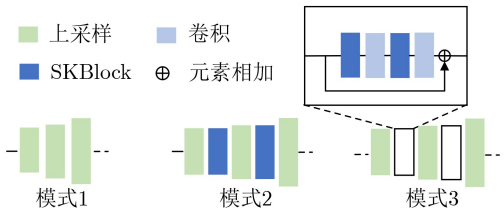
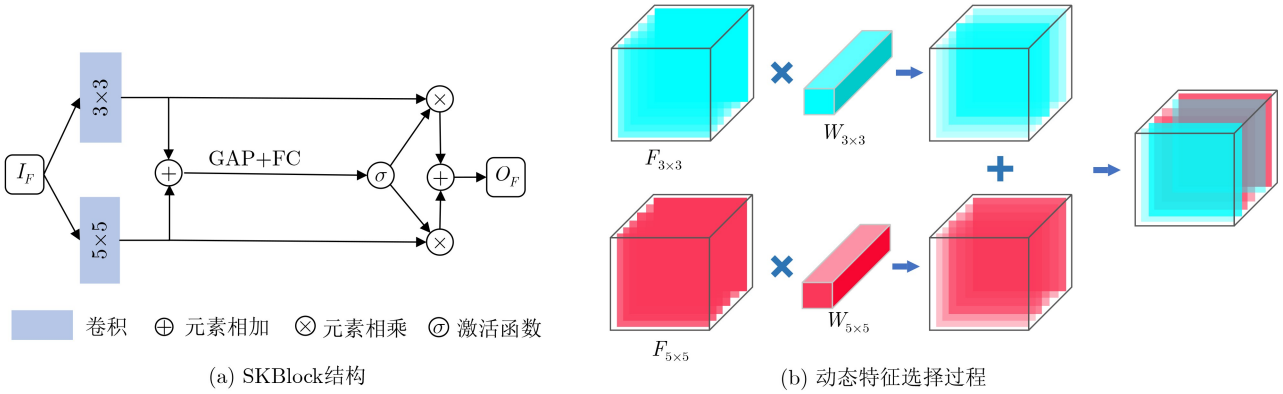
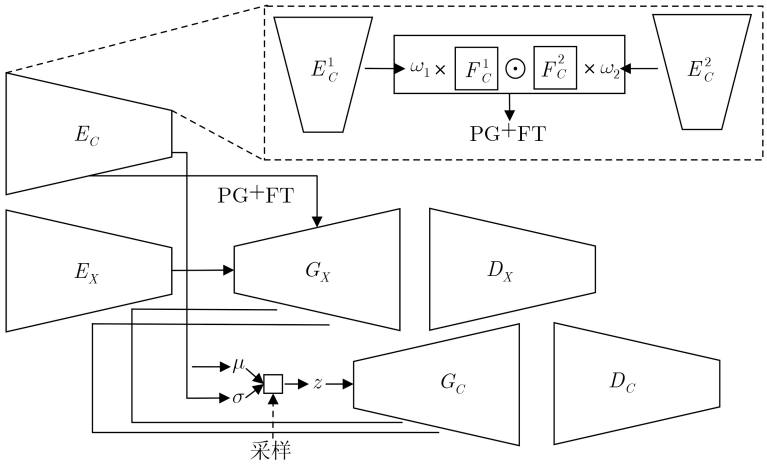


图 2 生成器中的上采样过程



(a) SKBlock结构　　(b) 动态特征选择过程

图 3 SKBlock的结构和动态特征选择过程



μ和σ分别为引导图像隐变量分布均值和标准差，z为隐变量，⊙表示沿通道方向拼接特征。

图 4 GSK-GAN模型结构

包括源图像编码器E_X、双分支引导图像编码器E_C、对应的生成器G_X和G_C及判别器D_X和D_C。其中E_C用于提取引导图像特征，通过参数生成器PG和特征转换层FT将引导图像信息传至生成器，同时学习引导图像的隐变量分布。G_C利用采样的隐变量生成引导图像信息之外的目标图像。

图4还展示了双分支引导图像编码器的结构，分支编码器E_C^1和E_C^2采用原编码器的编码形式，最大特征数量为原编码器的1/2。每个分支中网络层的特征F_C^1和F_C^2分别与对应的权重ω_1和ω_2相乘，然后沿特征通道拼接作为参数生成器的输入。GSK-GAN训练时使用随机权重值，且$\omega_1+\omega_2=1$，以学习每个分支中不同程度的信息转换，测试时通过改变权重获得不同引导图像对应生成图像之间的插值。

在引导图像信息融合方面，GSK-GAN基于SK-GAN中生成器的结构进行多级信息融合，如图5(b)所示，参数生成器利用引导图像编码器输出的特征生成转换参数，生成器中每个SKBlock前包含特征转换层，用于转换引导图像的信息，然后由SKBlock获取多尺度信息。相比文献[10]中编码器之间双向信息传递的方式(图5(a))，GSK-GAN将引导图像信息直接传递至生成器，避免了对源图像编码的影响，不仅有利于模型的拓展还减少了1/2参数生成器的使用，同时也保证了生成图像的质量。

3.3 损失函数

本文沿用Pix2pix中的对抗损失函数和L1损失函数，其中对抗损失函数使用LSGAN[17]。优化的转换模型包括SK-GAN和GSK-GAN, GSK-GAN中还使用KL散度学习引导图像的隐变量分布。

SK-GAN的优化目标包括转换模型T_{SK}和判别

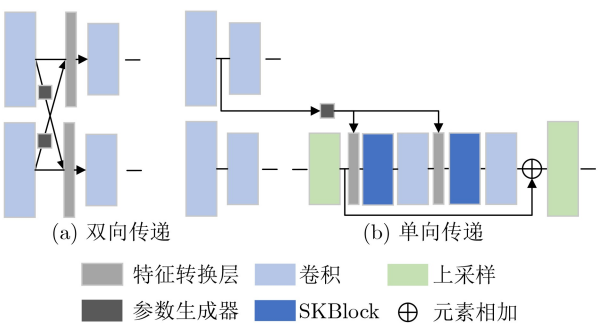


图 5 引导图像信息的传递方式

器D_{SK}，对应的损失函数分别如式(1)和式(2)

$$L(T_{\mathrm{SK}})=\lambda_1 L_{\mathrm{LSGAN}}(T_{\mathrm{SK}})+\lambda_2 L_{\mathrm{L1}}(T_{\mathrm{SK}}) \tag{1}$$

$$\begin{aligned}L(D_{\mathrm{SK}})&=L_{\mathrm{LSGAN}}(D_{\mathrm{SK}})\\&=0.5(D_{\mathrm{SK}}(x,y)-1)^2+0.5(D_{\mathrm{SK}}(x,T_{\mathrm{SK}}(x)))^2\end{aligned} \tag{2}$$

其中，L_{LSGAN}和L_{L1}分别表示对抗损失函数和L1损失函数，$L_{\mathrm{LSGAN}}(T_{\mathrm{SK}})=0.5(D_{\mathrm{SK}}(x,T_{\mathrm{SK}}(x))-1)^2$, $L_{\mathrm{L1}}(T_{\mathrm{SK}})=\|y-T_{\mathrm{SK}}(x)\|_1$, λ_1和λ_2表示用于平衡损失函数的超参数。

GSK-GAN的优化目标包括转换模型T_{GSK}和T_{GSK}中生成器G_X和G_C对应的判别器D_X和D_C。G_X的损失函数为

$$\begin{aligned}&L_{\mathrm{LSGAN}}(G_C,E_C,E_X)\\&\quad=0.5(D_X(x,G_X(E_X(x),E_C(c_1,c_2,\omega)))-1)^2\end{aligned} \tag{3}$$

$$L_{\mathrm{L1}}(G_X)=\|G_X(E_X(x),E_C(c_1,c_2,\omega))-y\|_1 \tag{4}$$

G_C的优化过程与G_X一致，包含损失函数$L_{\mathrm{LSGAN}}(G_C,E_X,E_C)$和$L_{\mathrm{L1}}(G_C)$。此外，$T_{\mathrm{GSK}}$中引入双分支引导图像编码器$E_C$学习引导图像的隐变量分布，使用的损失函数为

$$L_{\mathrm{KL}}(E_C)=-D_{\mathrm{KL}}(q_\phi(z|x)||p(z)) \tag{5}$$

其中，$q_\phi(z|x)$表示引导图像的隐变量分布，$p(z)$表示正态分布。

T_{GSK}总的损失函数为

$$\begin{aligned}&L_{\mathrm{GSK}}(E_C,E_X,G_C,G_X)\\&\quad=\lambda_1(L_{\mathrm{LSGAN}}(G_X,E_X,E_C)+L_{\mathrm{LSGAN}}(G_C,E_X,E_C))\\&\qquad+\lambda_2(L_{\mathrm{L1}}(G_X)+L_{\mathrm{L1}}(G_C))+\lambda_3 L_{\mathrm{KL}}(E_C)\end{aligned} \tag{6}$$

其中，λ_3表示用于平衡损失函数L_{KL}的超参数。

GSK-GAN中判别器D_X的损失函数为

$$\begin{aligned}L_{\mathrm{LSGAN}}(D_X)=&0.5(D_X(x,y)-1)^2\\&+0.5(D_X(x,T_{\mathrm{GSK}}(x,c_1,c_1,\omega)))^2\end{aligned} \tag{7}$$

D_C的优化过程与D_X一致，表示为$L_{\mathrm{LSGAN}}(D_C)$。

4 实验

本节详细介绍实验使用的设备参数、生成图像的评价指标、与现有方法的对比结果和实验分析。

4.1 实验设置

所有实验在NVIDIA Tesla V100 GPU上运行，训练过程中转换模型和判别器的学习率均为0.0002并使用beta1为0.5的Adam优化器。超参数$\lambda_1=2$, $\lambda_2=100$和$\lambda_3=0.01$。GSK-GAN和SK-GAN分别使用128×128和256×256的图像分辨率，所有任务的数据批大小均在1～10。

4.2 评价指标

本文采用文献[7]所使用的结构相似性(Structural SIMilarity, SSIM)和峰值信噪比(Peak Signal to Noise Ratio, PSNR)来评价目标图像和生成图像的相似性，两者越相似，SSIM和PSNR的评分越高。此外，利用深度神经网络来评价生成图像质量也是常用的评价方法，该类评分包括弗雷歇Inception距离(Fréchet Inception Distance, FID)[18]，学习的感知图像块相似性(Learned Perceptual Image Patch Similarity, LPIPS)[19]和全卷积网络评分(FCN Score, FCNS)[1]等。FID和LPIPS评分越低表示生成图像质量越高。FCNS以语义分割模型分割cityscapes[20]数据集的生成结果并计算相应的分割精确度，其值越高表明生成图像越接近目标图像。

4.3 实验结果

实验内容包括模型SK-GAN和模型GSK-GAN的实验结果与分析，通过定性和定量的对比展示了SK-GAN和GSK-GAN的优势。

4.3.1 SK-GAN的实验对比

SK-GAN在草图合成真实图像和语义图像合成真实图像任务中进行实验对比，使用的数据集分别为Edges2handbags[1]和Edges2shoes[1], Facades[21]和Cityscapes[20]。图6和图7分别展示了SK-GAN在草图合成真实图像任务中与Pix2pix和判别区域对抗网络(Discriminative Region Proposal Adversarial Networks, DRPAN)定性对比结果、在语义图像合成真实图像任务中与Pix2pix定性对比结果，这些结果表明SK-GAN生成的图像伪影较少，细节较丰富。两种任务的定量对比分别如表1和表2，其中表2引用文献[7]的对比结果，包含级联优化网络(Cascaded Refinement Network, CRN)[22]的实验对比。本文方法对生成器的改善增强了图像特征的提取，在小样本数据中也能获得更多细节，保持较完整的图像结构，如图7中Facades数据集。此外，本文在Cityscapes数据集中获得更高的FCNS评分(表2)，这表明SK-GAN的生成结构优于CRN和DPRAN。

(a) 源图像 (b) Pix2pix (c) DRPAN (d)SK-GAN (e) 真实图像

图 6 草图合成真实图像实验结果对比

4.3.2 GSK-GAN的实验对比

本文使用TextureGAN[9]中的采样方式从目标图像中采样纹理来替换对应源图像中的信息作为引导图像，实验数据集为包含对象掩码的Edges2shoes[9]和Edges2handbags[9]。本文通过文献[10]提供的模型获取统一的纹理图像并随机计算10次生成结果，GSK-GAN中双分支引导图像编码器采用同一输入且ω= 0.5，与TextureGAN和文献[10]定性对比结果如图8，该结果表明GSK-GAN生成的图像与对应引导图像的纹理更接近，更光滑和精细。与文献[10]一致，本文使用FID和LPIPS评估GSK-GAN生成图像的质量，对应结果如表3，相比文献[10]，GSK-GAN在FID评分中获得较大程度提升且在可视化效果中更接近真实图像。

GSK-GAN中还包含双分支引导图像编码器和以隐变量获得多样性生成效果的生成器，本文在Edges2shoes数据集中展示这两部分生成图像的效果，分别如图9和图10所示，GSK-GAN能够利用已有的引导图像c_1和c_2产生样式连续变化的生成图

(a) 源图像　(b) Pix2pix　(c) SK-GAN　(d) 真实图像

图 7 语义图像合成真实图像实验结果对比

表 1　Edges2shoes和Edges2handbags数据集中定量对比结果

	Edges2shoes			Edges2handbags		
	Pix2pix[1]	DRPAN[7]	SK-GAN	Pix2pix[1]	DRPAN[7]	SK-GAN
SSIM	0.749	0.764	**0.788**	0.641	0.671	**0.676**
PSNR	20.001	19.739	**20.606**	16.475	**17.384**	17.171
FID	69.213	**43.883**	45.168	73.675	69.606	**68.957**
LPIPS	0.183	0.176	**0.161**	0.267	0.260	**0.254**

表 2　Cityscapes数据集中定量对比结果

	Per-pixel acc	Per-class acc	Class IOU
L1+CGAN[1]	0.63	0.21	0.16
CRN[22]	0.69	0.21	**0.20**
DPRAN[7]	0.73	0.24	0.19
SK-GAN	**0.76**	**0.25**	**0.20**

(a)引导图像　(b) TextureGAN　(c) 文献[10]　(d) GSK-GAN　(e) 真实图像

图 8 多模态图像转换生成的结果对比

表 3　多模态图像转换Edges2shoes和Edges2handbags数据集中定量对比结果

	Edges2shoes			Edges2handbags		
	TextureGAN[9]	文献[10]	GSK-GAN	TextureGAN[9]	文献[10]	GSK-GAN
FID	**44.190**	118.988	45.041	61.068	73.290	**60.753**
LPIPS	0.123	0.123	**0.119**	0.171	0.162	**0.154**

像，以及利用隐变量产生更多引导图像信息之外的生成图像，同时图像的整体质量能够很好地保持。

此外，本文在融合源图像和纹理图像的过程中对纹理图像进行随机翻转，增强了模型对纹理图像的泛化。如图11，当引导图像纹理和目标图像不匹配时，GSK-GAN仍能够产生较为合理的生成图像。

4.4 模型分析

本节分析SK-GAN的模型结构，包括SKBlock与生成器结合的方式、不同对抗损失函数和上采样过程不同感受野对转换模型的影响，以及GSK-GAN中引导图像信息融合方式对多样性生成的影响。

图 9 Edges2shoes数据集中使用双分支引导图像编码器的生成结果

图 10 Edges2shoes数据集中使用隐变量的生成结果

(a) 引导图像 (b) TextureGAN (c) 文献[10] (d) GSK-GAN

图 11 Edges2shoes数据集中纹理不匹配的生成结果

4.4.1 SK-GAN的模型结构

图2展示生成器的3种模式，其中模式1为常用结构，模式2和模式3为SKBlock和生成器结合的方式。每种模式在Facades数据集上的结果对比如表4所示，其中模式2和模式3相对模式1都有一定程度的提升，这表明SKBlock的加入有助于改善转换模型的生成器，从而提高图像的生成质量(模式2)，而模式3的结合方式进一步提升了生成器的性能。模式1、模式2和模式3均使用LSGAN。在模式3中对比原始GAN和LSGAN对转换模型的影响，如表4所示，LSGAN在SSIM, PSNR和FID评分中都有一定程度提升，相比原始GAN, LSGAN对转换模型的优化使生成图像更接近真实图像。

在感受野分析方面，对卷积核为1×1和3×3, 3×3和5×5以及5×5和7×7 3种组合分支进行实验，分别以K13, K35和K57表示，3种组合的卷积核对应的感受野依次增大。如表5所示，K13和K57分别获得较优的LPIPS和PSNR评分，K35对应的这些评分只有较小的差距，综合性能更具优势。SK-GAN中使用K35的卷积分支组合，感受野的变化通过选择特征的权重控制。图12展示多个数据集上采样过程不同特征的选择权重，图中“3×3”和“5×5”分别表示不同的感受野，对不同上采样层特征，生成器能通过学习选择权重动态获取不同尺度信息，从而控制感受野变化。

4.4.2 引导图像信息融合方式对多样性生成的影响

为验证引导图像信息的融合方式对多样性生成的影响，本文的GSK-GAN使用单一输入的引导图像编码器，简化模型训练。编码器之间双向传递信息使源图像编码器包含引导图像信息，这相对隐变量对生成器G_C的影响更大，因此生成器G_C无法通过隐变量改变生成图像的样式。本文提出的单向信息传递的方式中仅以隐变量作为指导源图像转换的信息，有效地产生了多样化的生成结果。如图13所

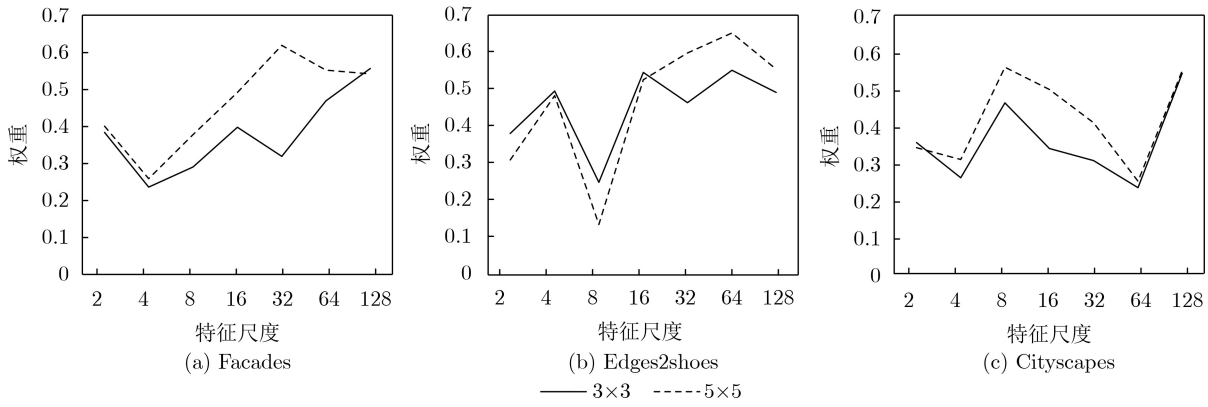


图 12 多个数据集中上采样层的特征对应的多尺度信息的选择权重

(a) 引导图像　　(b) 双向传递方式(第1行)和单向传递方式(第2行)生成的图像

图 13 不同引导图像信息传递方式对应的多样性生成结果

表 4　生成器中不同的上采样过程生成的图像质量对比结果

	SSIM	PSNR	FID	LPIPS
模式1	0.267	12.821	102.771	0.415
模式2	0.267	12.853	92.608	0.404
模式3	**0.284**	**12.981**	**89.718**	0.405
模式3 (GAN)	0.262	12.568	97.828	**0.399**

表 5　SKBlock中不同感受野分支组合对应的图像质量对比结果

	SSIM	PSNR	FID	LPIPS
K13	0.276	12.961	100.532	**0.398**
K35	**0.284**	12.981	**89.718**	0.405
K57	0.268	**13.007**	98.132	0.400

示，双向信息传递的方式在生成器中G_C只能产生与引导图像相关的图像，单向信息传递能够通过隐变量获得更多样式。此外，两种方式都生成了细节丰富的清晰图像，但单向信息传递减少了一半参数生成器，降低了模型参数。单向信息传递更有利于转换模型对多样性生成的拓展，同时保证了图像质量。

5　结论

本文通过实验验证了SK-GAN以动态感受野获取生成器上采样过程中特征的多尺度信息有助于提高生成器的性能，从而获得高质量的生成图像。在GSK-GAN中，本文提出了双分支引导图像编码器和新的引导图像信息融合的方式，同时以隐变量提高转换模型的多样性生成能力。实验表明，GSK-GAN不仅实现了可控的图像生成，还能获得更多引导图像信息之外的多样性生成结果，且保证了图像质量。

参 考 文 献

[1] ISOLA P, ZHU Junyan, ZHOU Tinghui, *et al*. Image-to-image translation with conditional adversarial networks[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, USA, 2017: 5967–5976. doi: 10.1109/CVPR.2017.632.

[2] CHEN Wengling and HAYS J. SketchyGAN: Towards diverse and realistic sketch to image synthesis[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 9416–9425. doi: 10.1109/CVPR.2018.00981.

[3] KINGMA D P and WELLING M. Auto-encoding variational Bayes[EB/OL]. https://arxiv.org/abs/1312.6114, 2013.

[4] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, *et al*. Generative adversarial nets[C]. The 27th International Conference on Neural Information Processing Systems, Montreal, Canada, 2014: 2672–2680.

[5] RADFORD A, METZ L, and CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks[EB/OL]. https://arxiv.org/abs/1511.06434, 2015.

[6] SUNG T L and LEE H J. Image-to-image translation using identical-pair adversarial networks[J]. *Applied Sciences*, 2019, 9(13): 2668. doi: 10.3390/app9132668.

[7] WANG Chao, ZHENG Haiyong, YU Zhibin, *et al*. Discriminative region proposal adversarial networks for high-quality image-to-image translation[C]. The 15th European Conference on Computer Vision, Munich, Germany, 2018: 796–812. doi: 10.1007/978-3-030-01246-5_47.

[8] ZHU Junyan, ZHANG R, PATHAK D, *et al*. Toward multimodal image-to-image translation[C]. The 31st International Conference on Neural Information Processing Systems, Long Beach, USA, 2017: 465–476.

[9] XIAN Wenqi, SANGKLOY P, AGRAWAL V, *et al*. TextureGAN: Controlling deep image synthesis with texture patches[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 8456–8465. doi: 10.1109/CVPR.2018.00882.

[10] ALBAHAR B and HUANG Jiabin. Guided image-to-image translation with bi-directional feature transformation[C]. The 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019: 9015–9024. doi: 10.1109/ICCV.2019.00911.

[11] SUN Wei and WU Tianfu. Learning spatial pyramid attentive pooling in image synthesis and image-to-image translation[EB/OL]. https://arxiv.org/abs/1901.06322, 2019.

[12] ZHU Junyan, PARK T, ISOLA P, *et al*. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]. 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017: 2242–2251. doi: 10.1109/ICCV.2017.244.

[13] LI Xiang, WANG Wenhai, HU Xiaolin, *et al*. Selective kernel networks[C]. 2019 IEEE/CVF Conference on

Computer Vision and Pattern Recognition (CVPR), Long Beach, USA, 2019: 510–519. doi: 10.1109/CVPR.2019.00060.

[14] SZEGEDY C, IOFFE S, VANHOUCKE V, *et al*. Inception-v4, inception-ResNet and the impact of residual connections on learning[EB/OL]. https://arxiv.org/abs/1602.07261, 2016.

[15] 柳长源, 王琪, 毕晓君. 基于多通道多尺度卷积神经网络的单幅图像去雨方法[J]. 电子与信息学报, 2020, 42(9): 2285–2292. doi: 10.11999/JEIT190755.
LIU Changyuan, WANG Qi, and BI Xiaojun. Research on Rain Removal Method for Single Image Based on Multi-channel and Multi-scale CNN[J]. *Journal of Electronics & Information Technology*, 2020, 42(9): 2285–2292. doi: 10.11999/JEIT190755.

[16] LI Juncheng, FANG Faming, MEI Kangfu, *et al*. Multi-scale residual network for image super-resolution[C]. The 15th European Conference on Computer Vision, Munich, Germany, 2018: 527–542. doi: 10.1007/978-3-030-01237-3_32.

[17] MAO Xudong, LI Qing, XIE Haoran, *et al*. Least squares generative adversarial networks[C]. 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017: 2813–2821. doi: 10.1109/ICCV.2017.304.

[18] HEUSEL M, RAMSAUER H, UNTERTHINER T, *et al*. Gans trained by a two time-scale update rule converge to a local nash equilibrium[C]. The 31st International Conference on Neural Information Processing Systems, Long Beach, USA, 2017: 6629–6640.

[19] ZHANG R, ISOLA P, EFROS A A, *et al*. The unreasonable effectiveness of deep features as a perceptual metric[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 586–595. doi: 10.1109/CVPR.2018.00068.

[20] CORDTS M, OMRAN M, RAMOS S, *et al*. The cityscapes dataset for semantic urban scene understanding[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 2016: 3213–3223. doi: 10.1109/CVPR.2016.350.

[21] TYLEČEK R and ŠÁRA R. Spatial pattern templates for recognition of objects with regular structure[C]. The 35th German Conference on Pattern Recognition, Saarbrücken, Germany, 2013: 364–374. doi: 10.1007/978-3-642-40602-7_39.

[22] CHEN Qifeng and KOLTUN V. Photographic image synthesis with cascaded refinement networks[C]. 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017: 1520–1529. doi: 10.1109/ICCV.2017.168.

尹梦晓：女，1978年生，博士，副教授，CCF会员，研究方向为计算机图形学与虚拟现实、数字几何处理、图像与视频编辑.

林振峰：男，1996年生，硕士生，研究方向为图像生成、图像转换.

杨 锋：男，1979年生，博士，副教授，CCF会员，研究方向为人工智能、网络信息安全、大数据与高性能计算、精准医学.

责任编辑：余 蓉