文章编号：2095-2163(2022)08-0136-06 中图分类号：TP 183 文献标志码：A

基于 U-net 变体和分类器的动漫线稿风格迁移

冯煜颋，李志伟
（上海工程技术大学 电子电气工程学院，上海 201620）

摘 要： 近年来，随着一些突破性的神经风格迁移方法的出现，一张动漫线稿和一张匹配的风格图像可以通过风格迁移的方法，生成一张彩色图像。但是，当需要将这幅图像的风格具体应用到某张动漫线稿的时候，这些方法都只是将线稿的素描线随即上色作为输出，并且无法得到想要的具体风格类型迁移。在本文中，利用一种改进的残差增强 U-net 变体结合辅助分类器组成辅助分类器生成对抗网络模型（AC-GAN）应用于神经风格迁移动漫线稿上色中。实验结果表明，该方法能够将辅助图像的颜色风格应用到线稿当中，同时生成的彩色图像具有较高的质量。

关键词： 动漫线稿上色；残差连接；风格迁移；生成对抗网络

Style transfer of animation sketch based on U-net variant and the classifier

FENG Yuting, LI Zhiwei
(School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

【Abstract】 In recent years, with the emergence of some breakthrough neural style transfer methods, an animation line draft and a matching style image can generate a color image through the method of style transfer. However, when the users need to apply the style of the style image to a specific animation line draft, these methods just color the sketch line of the line draft as the output, and it is difficult to get the wanted specific style type transfer. In this paper, combined with am auxiliary classifier, an improved residual enhanced U-net variant is used to form an auxiliary classifier generative adversal network model (AC-GAN), which could be applied into neural style transfer animation line coloring. Experimental results show that this method can apply the color style of auxiliary images to the line draft, and the generated colorful images have high quality.

【Key words】 animation sketch coloring; residual connection; style transfer; generative adversarial network

0 引 言

素描或线稿艺术上色是一个有着巨大市场需求的研究领域。与强烈依赖纹理信息的普通照片上色不同，草图上色更具挑战性，因为草图可能没有纹理。在动漫、游戏这些产业中，大部分的作品都是通过素描或线稿来进行创作的，这就会耗费大量的时间和精力，因为需要人工去给这些线稿上色来达到人们想要的状态。如果尝试将某种绘画风格应用到半成品的动漫线稿上，那么就可以省去不少多余的工作，比如用一个动漫的特定人物的某张图片作为风格参考图像，并将这种颜色风格应用到人物的素描上。而图像上色一般分为 2 种：有引导上色和无引导上色。其中，无引导指的是全交由算法进行自动化上色，而有引导是在上色过程中有人为（其它参照）干预，比如给出一幅风格参考图像或指定某一区域为特定颜色。本文提出的上色方法属于一种有引导上色。

神经风格算法可以结合线稿图和风格图生成优秀的图像，但是却缺乏处理素描线稿的能力，生成的图像远未达到人工上色的预期效果。事实上，U-net 和生成对抗网络已获证明在图像上色方面有着很好的效果。Zhang 等人[1]提出了一种二阶段的线稿上色方法：第一阶段是草稿阶段，根据人工输入颜色提示或是提供参考图生成模拟合成草图；第二阶段是精修阶段，通过 Inception V1 网络提取草图的颜色特征和预先提示的颜色特征，来控制最终生成图像的颜色风格。Zhang 等人[2]开发了线稿风格迁移工具 Style2paints，实现了线稿到彩色图像的风格迁移上色，根据自带一些颜色风格特征或是参考图像可以进行快速的线稿上色。于是改进了一种残差增强的 U-net 来增加生成网络学习特征图的能力，结合辅助分类器对抗生成网络（ACGAN）作为解决方案[3-5]。这种前馈网络能够快速地合成绘画，节省

基金项目： 国家自然科学基金（61705127）。

作者简介： 冯煜颋（1997-），男，硕士研究生，主要研究方向：图像处理与机器视觉；李志伟（1982-），男，博士，副教授，主要研究方向：光电子、图像处理与机器视觉。

收稿日期： 2021-11-10

时间。另外，U-net 和条件 GAN（CGAN）在没有成对输入和输出信息的均衡量时性能会相对下降。因此，本文在原有的生成网络基础上附加了 2 个指引解码器来实现额外的损失。网络的整体结构如图 1 所示。

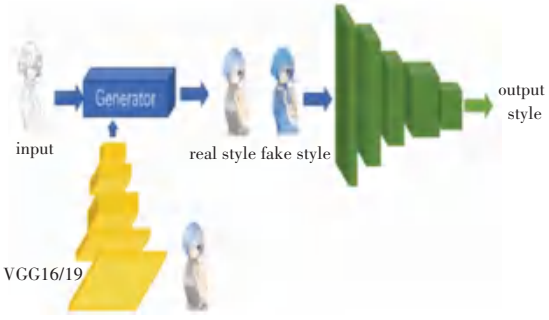


图 1　生成对抗网络的整体结构

Fig. 1　The overall structure of the generative adversarial network

生成网络由残差增强的 U-net、分类器和指引解码器组成；判别网络由 AC-GAN 进行改进，经判别器处理后会输出一个 2 048 或 4 096 维的特征向量对应 VGG 的特征向量（颜色风格），而不再是原始 GAN 的二分类值。全局的颜色风格提示可以被看作是一个具有 2 048 或 4 096 个类的低级分类结果。

1　相关工作

生成对抗网络（GANS）的出现，在深度学习领域中日益受到广大学者的关注[6]。生成对抗模型通常是由一个生成器和一个判别器组成，其中生成器捕捉真实样本的潜在分布，并且生成新的数据样本；判别器往往是一个二值分类器，通过训练可以尽可能正确地从生成样本中区分出真实样本。利用判别器来引导生成器的训练，通过 2 个模型之间的交替训练不断进行对抗，最终使得生成模型能够更好地完成生成任务。而随着越来越多 GANS 变体的出现，GANS 在图像的各个领域都取得了不错的成果。在图像上色领域中，GANS 同样在主流算法中占据着至关重要的地位。目前，基于深度学习的自动着色模型大多采用 GANS 体系结构。

Lee 等人[7]提出的动漫线稿自动上色算法是基于二次规划图匹配，但是这种基于参考图的自动上色方法难度较大，原因在于线稿中人物姿态的变化，使得参考图和线稿的一些区域无法对应起来，给图匹配算法带来了极大挑战。这种上色方法生成的彩色图像因参考图和线稿有些区域并不匹配，一些区域只能随机上色，导致图像的质量很差。GAN 的出现使线稿基于参考图上色逐渐变得可靠，在生成器和判别器的对抗式训练中，模型不断学习并将线稿到对应彩色图像间的映射关系做了进一步优化。

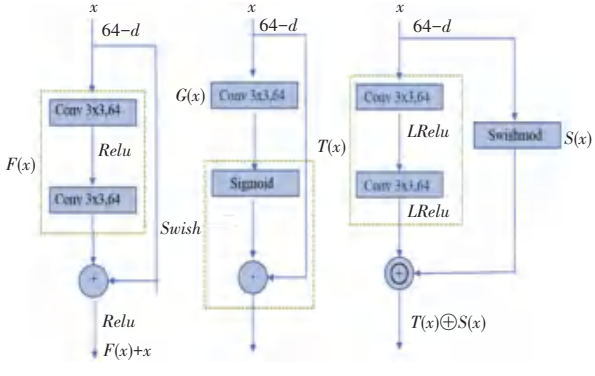
神经风格迁移，是通过基于最小化深度卷积层的格拉姆矩阵的差异的算法，可以将一张普通的照片赋予另外一种艺术作品风格[8-10]。然而，本文的目标是将风格图像和草图相结合。事实上，从风格图像到草图的神经风格迁移得到的最终图像远不是一幅正常的图像，往往和风格图像有很大差异。

Pix2pix 是基于条件生成对抗网络 CGAN 的风格迁移模型之一，在成对数据集的情况下，可以完成很多任务。如：将素描画轮廓转换成图片，将黑夜场景转换成白天场景，自动上色等等[11]。但在实验中发现，网络的输出的质量最终取决于输入信息和输出信息的差距程度。实际上，条件判别器很容易导致生成器过于关注草图和绘画之间的关系，因此，在某种程度上，忽略了绘画的组成，导致不可避免的过拟合。

2　本文方法

2.1　增强型的残差连接

本文提出的残差连接 Enhanced residual connection，是对 ResNet 中残差模块的一种改进。这种连接方式是 SwishMod 集成残差模块的连接方式，SwishMod 包含了卷积层和 *Swish* 激活函数。其残差连接结构如图 2 所示。



（a）ResNet connection（b）SwishMod（c）Enhanced residual connection

图 2　增强残差连接方式

Fig. 2　Enhanced residual connection mode

本文图 2 中，x 表示输入数据，$F(x)$ 表示残差，$F(x)+x$ 是残差连接后的输出，“+”表示像素点对应相加；$G(x)$ 表示 SwishMod 中卷积层的输出，“·”表示像素点对应相乘；$T(x)$ 表示卷积层经过非线性 *LReLU* 函数后的输出，$S(x)$ 是 SwishMod 的

输出，“⊕”表示特征图之间进行拼接，“$T(x) \oplus S(x)$”是 *Enhanced residual connection* 的最终输出。

在残差连接中，输入数据 x 没有经过处理就直接和残差相加；而在 SwishMod 中，对 x 进行了处理，使用了 *Sigmoid* 函数，该种设计优势就在于能够控制数值的幅度，在深层网络中可以保持数据幅度不会出现大的变化。此外，对 Enhanced residual connection 中的卷积层使用了非线性 *LReLU*，对于生成类的任务比 *ReLU* 有着更好的效果。

采用 SwishMod 滤波输入数据 x，控制了输入数据 x 从底层到高层之间通过一个捷径的特征图传输，得到更精细和准确的颜色特征。

SwishMod 定义为：

$$S(x) = x \cdot \sigma(G(x)) \tag{1}$$

SwishGatedBlock 的输出为：

$$y = T(x) \oplus S(x) \tag{2}$$

其中，$T(x)$ 是模块中的残差部分，$S(x)$ 是 SwishMod 滤波后的信息，两者拼接在一起输出得到更精细的颜色特征。

2.2 生成网络结构和损失函数

U-net 虽然在图像合成领域有着出色的表现，能够提取每个层次的特征图像，一旦 U-net 具备了能够在低级层中处理问题的能力后，那么高级层就不会再去学习任何东西。如果训练一个 U-net 来做一项简单的工作、即复制图像（如图 3 所示），当输入和输出相同时，损失值将立即降至 0。因为第一层编码器发现，可以简单地经由跳过连接，将所有特征图直接传输到解码器的最后一层，来最小化损失。在这种情况下，无论训练多少次，中间层都不会得到任何梯度。对于 U-net 的解码层来说，每一层的特征图都是由更高层或是跳接层中获得。在训练过程的每次迭代中，这些层选择了经过非线性激活其它层的输出来最小化损失。当 U-net 用高斯随机数初始化网络时，编码器中第一层的输出具有足够的信息来表达完整的输入映射，而解码器中第二到最后一层的输出似乎存在噪声。因此，“懒惰的”U-net 放弃了相对来说有噪声的特征图。

网络生成器网络整体结构是基于残差增强 U-net的变体（如图 4 所示），每个蓝色模块都是一个 Enhanced residual connection，在下一个分辨率提取特征时，通过残差增强可以得到更精细的特征。随着等级的提高，分辨率也逐渐降低。该网络也可以看做是左、右两个分支，但是把同一个分辨率等级的左、右分支之间嵌入一个 SwishMod，来滤波编码路径传递到解码路径的信息，而不再是原来的跳接。因此，SwishMod 在提高网络收敛速度的同时，还能提高网络的性能。在左侧分支中，每个 Enhanced residual connection 的输出由残差部分输出的特征图和经过 SwishMod 滤波的特征图组成；而在右侧分支中，每个 Enhanced residual connection 的输出是由残差部分输出的特征图、经过 SwishMod 滤波的特征图、以及对应左侧分支通过 SwishMod 滤波的特征图三部分组成。由此可见，经过残差增强的 U-net，完全解决了 U-net 在训练时中间层不会得到任何梯度的问题。

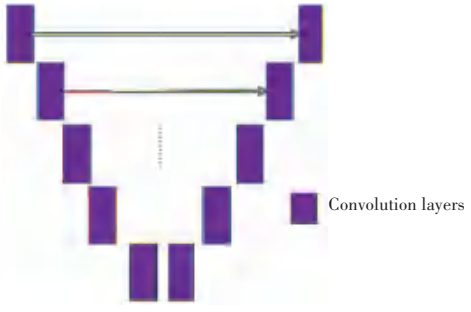


图 3 U-net 的跳接方式

Fig. 3 Skipped connection between U-net layers

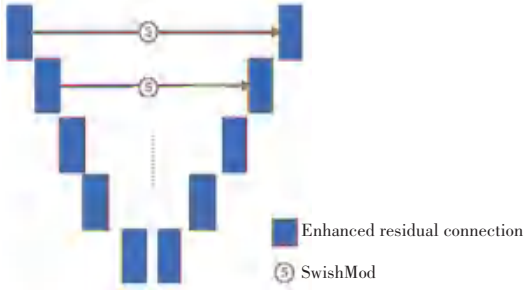


图 4 残差增强 U-net 层与层之间的连接

Fig. 4 The connections between residual enhanced U-net layers

此外，本文在生成器的结构中增加了一个分类器，如图 5 所示。相对来说，1×1×256 的风格提示不能够满足动漫线稿的颜色风格，所以在 VGG19 全连接层的输出中不再使用 *ReLu* 激活函数，则会得到更多的 1×1×4 096 的颜色风格提示。然而，对于一个新初始化的 U-net，如果将 4 096 维的特征向量直接添加到该层中，中间层的输出噪声可能会非常大。由于有噪声的中间层会被 U-net 放弃，因此这些层不能接收到任何梯度。

为了解决上述问题，本文在原有的生成网络中附加了 2 个解码器（见图 5）。如果给每一层附加额外的损失，无论中间一层的输出有多嘈杂，该层将永远不会被 U-Net 放弃，不会出现梯度消失的情况，

从而会得到稳定的梯度。通过向中间层添加一个有信息量和有具体内容的噪声提示，解决了原本网络传递特征信息跳过中间层而导致训练时中间层梯度消失的问题。通过在“指引解码器 1”和“指引解码器 2”中实现了 2 个额外的损失，因此就避免了中间层的梯度消失。

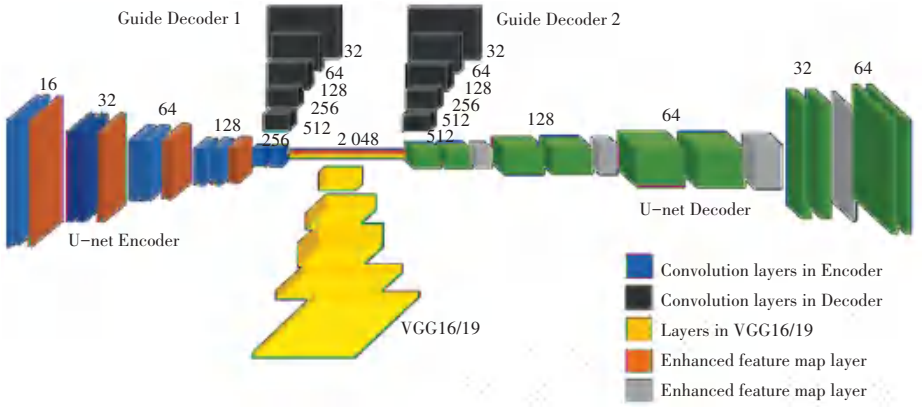


图 5　生成器的网络结构

Fig. 5　The network structure of the generator

损失函数定义为：

$$L_{l1}(V,G_{f,g_1,g_2}) = \mathrm{E}_{x,y\sim P_{data}(x,y)}\left[\|y - G_f(x,V(y))\|_1 + \alpha\|y - G_{g_1}(x)\|_1 + \beta\|y - G_{g_2}(x,V(y))\|_1\right] \quad (3)$$

其中，x、y 是线稿和参考图的配对域；$V(x)$ 是不经过 *ReLu* 的 VGG19 全连接层输出；$Gf(x, V(x))$ 是 U-net 的最终输出；$G_{g_1}(x)$ and $G_{g_2}(x, V(x))$ 是 2 个指引解码器在中间层的入口和出口的输出。这里，α 和 β 的推荐参数值为 0.3 和 0.9。

此外，通过用灰度图输入位于中间层入口的指引解码器，可以改善颜色的分布，让颜色分布不会特别单一，因此最终的损失如下：

$$L_{l1}(V,G_{f,g_1,g_2}) = \mathrm{E}_{x,y\sim P_{data}(x,y)}\left[\|y - G_f(x,V(y))\|_1 + \alpha\|T(y) - G_{g_1}(x)\|_1 + \beta\|y - G_{g_2}(x,V(y))\|_1\right] \quad (4)$$

其中，$T(y)$ 可以将 y 转换为灰度图像。

2.3　判别网络结构和损失函数

绘画是一项复杂的过程，需要人类考虑到色彩的选择、构图和微调，所有这些都需要一个艺术家专注于绘画的整体方式。然而，条件鉴别器总是更倾向于关注素描线和颜色之间的关系，而不是全局信息。比如在 Pixtopix 中使用的是条件鉴别器，生成器会产生强烈的抵抗，这就导致了最终的彩色图像会出现颜色溢出和颜色混淆的结果。

在进行风格迁移时，需要判别器具有判断图像的颜色风格、并在风格转移时相应地提供梯度的能力，因此选用了集成 AC-GAN 的判别器。与 AC-GAN 判别器相比，本文判别器输出不含 *real/fake* 二分类，只包含生成图像的类标签。具体而言，判别器的输出为一个 4 096 维的特征向量，与 VGG 输出特征向量的意义基本相同，可视为色彩风格类别的分类结果。当判别器的输入图像为 *fake* 时，输出向量接近于全为 0；当判别器的输入图像为 *real* 时，输出向量接近于 VGG19 的全连接层输出的特征向量。

最终的损失函数定义为：

$$L_{GAN}(V,G_f,D) = E_{y\sim P_{data}(y)}[Lb(D(y) + (1 - norm(V(y))))] + E_{x\sim P_{data}(x)}[Lb(1 - D(G_f(x,V(y))))] \quad (5)$$

其中，$norm(x)$ 是归一化函数，用于对 VGG 的输出进行归一化，来拟合经过线性激活的特征之间的对数似然距离；$E_{y\sim P_{data}(y)}$ 是对真实样本的判别，$E_{x\sim P_{data}(x)}$ 是对生成样本的判别。在此，$D(y)$ 越接近 $norm(V(y))$ 的值越好，$D(G_f(x, V(y)))$ 越接近 0 越好。$norm(x)$ 归一化函数非常适合本文的判别器网络结构，可使得 VGG 输出的三维向量能够成为判别器标签。

本文使用的归一化函数如下：

$$Norm_{t\ sigmoid}(x) = 2sigmoid(relu(x)) - 1 \quad (6)$$

最终目标函数为：

$$G^* = \arg\min_{G_f}\max_{D} L_{GAN}(V,G_f,D) + \lambda L_{l1}(V, G_{f,g_1,g_2}) \quad (7)$$

3　实验分析

3.1　数据集

研究指出，由于目前还没有一个动漫线稿和参考图配对的数据集，本文使用的是训练好的 VGG 网络-ImageNet 图片分类数据集。由于本文的生成网

络使用的是 U-Net 网络,可以对任意形状大小的图片进行卷积操作,特别是任意大的图片。因此,在图像上色任务中,就可以对任意分辨率的灰度图像进行上色。实验数据随机截取了 ImageNet 图片分类数据集中的 5 000 幅匹配图像进行训练,并将所有图像分辨率都调整为 256×256。

3.2 实验结果

为了证明本文采用的 2 个指引解码器能够解决训练时中间层梯度消失的问题,实验对象分别采用 2 个指引解码器和无指引解码器的上色模型;目标函数分别采用指引解码器的额外损失和原 GAN 的生成对抗损失。Style2color-Guide 使用了 2 个指引解码器的生成模型,生成模型的目标函数选择了 2 个指引解码器的额外损失作为目标函数;Style2color-GAN不使用指引解码器的生成模式,生成模型的目标函数采用原始 GAN 的生成对抗损失。实验结果如图 6 所示。由图 6 可见,Style2color-Guide 上色模型生成的彩色图像有着更多的颜色层次,彩色图像质量优于 Style2color-GAN。此外,Style2color-Guide 生成的彩色图像颜色风格在细节上更接近于风格图像,而 Style2color-GAN 在细节上的表现依然欠佳(如图 6 中眼睛的颜色部分)。

(a) 线稿

(b) 参考图

(c) Style2color-Guide

(d) Style2color-GAN

图 6 2 种方法结果图对比

Fig. 6 Results of two coloring methods

为了验证本文方法中判别器和生成器都能学习到深层次的颜色特征、同时训练时不会有中间梯度消失,将本文方法与 Style2paints 方法进行了对比,对比结果如图 7 所示。由图 7 可以看出,Style2paints 过于追求风格迁移,在很多区域的颜色都出现了溢出现象,而本文方法生成的图像在视觉上更符合审美观念,同时也能生成更精细的颜色特征,颜色分布不会混淆,算法生成的图像有着更高的视觉质量和更加自然的色彩梯度。

(a) 线稿

(b) 参考图

(c) Style2paints

(d) 本文方法

图 7 本文方法和 Style2paints 生成图像对比

Fig. 7 Results of the proposed method and Style2paints

3.3 定量分析

从上色结果可以直观地看出,Style2color-Guide 的上色效果相比 Style2color-GAN 的上色效果更加细腻连贯。因为 2 个上色模型结构几乎一致,进一步采用 *FID* (Frechet Inception Distance) 指标[12-13] 来评价最终生成的彩色图像的质量。*FID* 指标的实验结果见表 1。由此可见,Style2color-Guide 生成的彩色图像质量略优于 Style2color-GAN 生成的彩色图像。对于动漫线稿颜色迁移来说,2 种指引解码器额外实现的损失函数效果不仅比传统 GAN 生成的对抗损失效果更好,还能避免网络训练时梯度消失的问题。

表 1　Style2color-Guide 和 Style2color-GAN 的实验结果

Tab. 1　Results of Style2color-Guide and Style2color-GAN

上色模型	*FID*
Style2color-Guide	**98.326**
Style2color-GAN	99.663

此外，为了进一步证明本文方法的优越性，采用峰值信噪比(*PSNR*)、相似结构性(*SSIM*)、特征相似度(*FSIM*)[14-15] 三种常规评价图像质量的方法，评价本文算法和现在流行的 Style2paints 算法生成的彩色图像质量(清晰度)和色彩多样性，结果见表 2。由表 2 可见，本文方法在所有指标上都获得了较好的表现，说明这种残差增强型的生成网络能够解决U-net训练时中间层梯度消失的问题。

表 2　本文方法和 Style2paints 的实验结果

Tab. 2　Results of the proposed method and Style2paints

上色模型	*PSNR*	*SSIM*	*FSIM*
The proposed method	**17.486**	**0.812 323**	**0.819 591**
Style2paints	15.632	0.784 283	0.791 362

4　结束语

本文提出了一种集成 U-net 变体和分类器的线稿风格迁移模型。通过残差增强的 U-net 变体能够更好地传递颜色特征图信息，避免了 U-net 训练时中间层梯度容易消失的问题，生成的彩色图像不会出现颜色混淆和颜色溢出的问题。同时引入 2 个指引解码器来附加 2 个损失，通过这 2 个损失来训练生成网络，取代了原来的生成对抗的训练方式，使得网络模型能够更多地聚焦于全局信息、而不再关注颜色和线条的关系。经过实验证明，本文算法比 Style2paints 在输出的结果上有着更高的颜色质量和更加平滑的颜色梯度，满足了人们的艺术审美需求。

本文的不足在于 VGG 的分类是 ImageNet 分类，只能使用训练好的 VGG，如果可以找到或者制作一个庞大的线稿匹配数据自行训练，则网络的训练效果会更趋完善，甚至于生成的彩色图像会完全接近人工上色的效果。

参考文献

[1] ZHANG Lvmin, LI Chengze, WONG T T, et al. Two-stage sketch colorization[J]. ACM Transactions on Graphics(TOG), 2018, 37(6): 1-14.

[2] ZHANG Lvmin, JI Yi, LIN Xin, et al. Style transfer for anime sketches with enhanced residual U-net and auxiliary classifier GAN[C]// 2017 4th IAPR Asian Conference on Pattern Recognition. Nanjing, China: IEEE, 2017: 506-511.

[3] 吕李娜，刘镇，夏炎. 基于生成对抗网络的灰度照片上色方法[J]. 计算机与数字工程，2021，49(02)：388-391,416.

[4] 蒋文杰，罗晓曙，戴沁璇. 一种改进的生成对抗网络的图像上色方法研究[J]. 计算机技术与发展，2020，30(07)：56-59.

[5] WU Jiong, WANG Kai, TANG Xiaoying. Skip connection U-Net for white matter hyperintensities segmentation from MRI[C]// Proceedings of IEEE Access. United States: Institute of Electrical and Electronics Engineers Inc, 2019: 155194-155202.

[6] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Genera-tive adversarial networks[C]// Proceedings of the 27th In-ternational Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2014: 2672-2680.

[7] LEE J, KIM E, LEE Y, et al. Reference-based sketch image colorization using augmented-self reference and dense semantic correspondence[C]// Proc of the 33rd IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 5801-5810.

[8] 吴子扬，贺丹，李映琴. 基于 VGG-19 神经网络模型的图像风格迁移[J]. 科技与创新，2021(13)：171-173.

[9] 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 (CVPR). Las Vegas, NV, USA: IEEE, 2016, 2414-2423.

[10] 陈淮源，张广驰，陈高等. 基于深度学习的图像风格迁移研究进展[J]. 计算机工程与应用，2021，57(11)：37-45.

[11] ISOLA P, ZHU Junyan, ZHOU Tinghui. Image-to-image translation with conditional adversarial networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 5967-5976.

[12] MIRZA M, OSINDERO S. Conditional Generative Adversarial Nets[J]. arXiv preprint arXiv: 1411.1784, 2014.

[13] ZHANG Lin, ZHANG Lei, MOU Xuanqin, et al. Fsim: A feature similarity index for image quality assessment[J]. IEEE Transactions on Image Processing, 2011, 20(8): 2378-2386.

[14] WU Jiong, ZHANG Yue, WANG Kai. Skip connection U-Net for white matter hyperintensities segmentation from MRI[C]// Proceedings of IEEE Access. United States: Institute of Electrical and Electronics Engineers Inc., 2019: 155194-155202.

[15] WANG Z, BOVIK A C, SHEIKH H R. Image quality assessment: From error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.