DOI:10.16515/j.cnki.32-1722/n.2023.03.001

基于U-Net网络的动漫草图生成方法

赵海峰[1,2,3],高梓玉[1,3,4],张 燕[1,3*],高顺祥[1,3]

(1.金陵科技学院软件工程学院,江苏 南京 211169;2.江苏润和软件股份有限公司,江苏 南京 210012;
3.江苏省信息分析工程研究中心,江苏 南京 211169;4.南京师范大学计算机与电子信息学院,江苏 南京 210023)

摘 要:提出了一种基于U-Net网络的动漫草图生成方法,用于解决草图着色时难以获取成对的草图与RGB图的问题。该方法整体结构基于U-Net网络,通过添加残差块,有效利用了草图抽象信息和RGB图颜色信息,从而实现草图线稿的有效提取。实验结果表明,该方法在标准草图数据集上的FID值为74.56,超过了同类边缘提取方法40%,能够更好地提取草图,并且与人工绘制的草图风格接近。

关键词:U-Net网络;动漫草图;提取方法;深度学习;残差块

中图分类号:TP181 **文献标识码**:A **文章编号**:1672-755X(2023)03-0001-07

An Approach to Generating Animation Sketch Based on U-Net Network

ZHAO Hai-feng[1,2,3], GAO Zi-yu[1,3,4], ZHANG Yan[1,3*], GAO Shun-xiang[1,3]

(1. Jinling Institute of Technology, Nanjing 211169, China; 2. Jiangsu Hoperun Software Co., Ltd., Nanjing 210012, China; 3. Information Analysis Engineering Research Center of Jiangsu Province, Nanjing 211169, China; 4. Nanjing Normal University, Nanjing 210023, China)

Abstract: An animation sketch generation method based on U-Net network is proposed to solve the problem of difficulty in obtaining paired sketches and RGB images during sketch coloring. The overall structure of the method is based on the U-Net network, and by adding the residual module, it effectively utilizes the sketch abstraction information and the color information of RGB images to achieve the effective extraction of sketch line drawings. The experimental results show that the FID value of this method on the standard sketch dataset is 74.56, which outperforms edge detection methods by 40%. It can better extract the sketches and it is similar in the style to the manually drawn sketches.

Key words: U-Net network; animation sketch; extraction method; deep learning; residual blocks

动漫以动画、漫画为主要表现形式,是文化创意产业的主要组成部分,在人们的生活中扮演着重要角色。在动漫创作过程中,动漫设计人员首先根据剧本绘制出较为抽象的动漫草图,即动漫清稿。然后在此基础上绘制细节,形成线条草图,即动漫精稿。接着,进行草图着色和后期处理,最终形成一套完整的动漫故事。整个动漫创作过程由多人协作完成,需要花费大量的时间和精力。随着人工智能技术的不断发展,

收稿日期:2023-09-10

基金项目:江苏省国际科技合作项目(BZ2020069);江苏省高校自然科学研究重大项目(21KJA520001);江苏省体育局重大体育科研课题(ST221108)

作者简介:赵海峰(1984—),男,河南三门峡人,副教授,博士,主要从事计算机视觉研究。

通信作者:张燕(1969—),女,河南商丘人,教授,博士,金陵科技学院副校长,主要从事软件工程、模式识别研究。

利用人工智能辅助加快动漫制作过程成为动漫领域的重要需求，也是计算机视觉与模式识别领域的研究热点，这涉及抽象的动漫草图与彩色的动漫图像之间的相互转换问题。

在当前互联网上，存在大量已经制作好的动漫资源，包括动漫图像和视频。对这些已经制作完成的动漫资源进行处理，还原其制作过程，对支撑动漫的自动化制作有重要意义。具体来说，就是根据已经制作完成的彩色动漫图像，还原出原始的线条草图，从而建立两者之间的对应关系，为后续从线条草图自动生成彩色动漫图像建立预处理数据集。从彩色动漫图像提取出线稿动漫草图，主要是要保留原始艺术风格和提高准确性。传统的图像边缘提取算法容易受到噪声的干扰，无法准确反映出草图的结构信息和语义信息，同时缺乏直接对颜色信息的建模。

针对以上问题，本文提出了基于 U-Net 网络[1]的草图提取方法，通过引入残差块，提取彩色动漫图像中的结构信息与颜色信息，还原出原始的线条草图。通过建立线稿形式的动漫草图与彩色动漫图像之间的对应关系，构建成对的动漫草图与彩色动漫图像训练数据集，对于动漫制作中的动漫草图着色具有重要意义。通过设计与不同算子的边缘提取算法的对比实验与用户调研实验，验证本文方法的可行性。

1 相关工作

根据生成对象类型不同，草图生成可以分为从草图生成图像、从草图生成三维形状、从图像生成草图以及从三维形状生成草图等。

当草图生成的对象为图像时，早期的方法包括 Sketch2Photo[2]和 PhotoSketcher[3]，主要使用图像匹配和图像合成的思路进行生成。随着深度学习[4]的快速发展，研究人员提出了基于生成式对抗神经网络(GAN)的草图生成模型[5]。Chen 等[6]提出了 SketchyGAN 方法，该方法利用数据增强技术来增强训练数据，通过引入新的生成式对抗网络模块来提升图像合成的质量，从而得到更加真实的图像。ContextualGAN[7]则利用联合图像学习草图与图像的联合分布，将草图作为弱监督信息，实现无需草图与图像对齐的生成结果。Sarvadevabhatla 等[8]使用草图解译的方式来分析草图，将对象姿态预测作为一种草图分析的辅助任务，提高草图生成的整体性能。

除了一些特定用于草图生成的网络结构，还有类似 Pix2Pix[9]的适用于许多场景的解决方案，同样可以用于解决草图到图像的生成问题。SketchyCOCO[10]实现了从场景级手绘草图到场景级图像的生成，将场景级草图分为前景和背景两个部分，通过生成前景物体，再以前景物体为导向辅助生成背景的方式生成草图。Lei 等[11]结合多尺度卷积神经网络和注意机制的优点，实现人脸的草图到图像的转化。

在图像生成草图方面，CLIPasso[12]将草图看作是多个 B 样条曲线的组合，利用 CLIP 的能力提取图像语义来生成对应的草图。CLIPascene[13]通过对草图进行不同精准度的刻画，从场景级的图像中生成不同类型和不同抽象尺度的草图。

与二维草图的生成研究不同，关于草图直接生成三维形状的研究相对较少。一种典型的方法是将二维形状作为中间过程来生成三维模型[14]，另一类方法是将生成的过程进行分解，逐步形成可用的三维形状[15-16]。通过添加文本引导，Wu 等[17]使用了扩散模型生成三维形状。此外，在三维形状生成对应的二维草图方面，Ye 等[18]采用训练生成式对抗网络，仅使用解码器来生成二维草图。

与之前基于 GAN 方法生成草图不同，本文基于 U-Net 方法根据原始图像直接生成草图。2015 年，Ronneberger 等[1]提出一种网络结构左右对称、形状类似于 U 形的新型卷积神经网络，即 U-Net。从网络结构上来看，U-Net 网络包含左侧的特征提取网络和右侧的特征融合网络，网络整体上与自动编码器类似，都拥有编码部分和解码部分。不同的是 U-Net 网络将编码与解码严格对应起来，特征提取网络每一步得到的特征图都有与之对应的上采样特征图，更多地保留了输入端的图像特征信息。本文通过添加残差块，更好地提取了图像的结构信息，取得了较好的草图生成效果。

2　方法框架

2.1　模型整体结构

本文模型的整体结构基于 U-Net 网络，在网络中引入残差块。残差块是残差网络 ResNet 的组成元素，有效地解决了深度神经网络的退化问题。残差块将深度神经网络原本 $X \rightarrow Y$ 的学习变成了 $X \rightarrow Y \rightarrow X$ 的学习，从而阻止了网络对简单“图像搬运”的学习，转而学习输入与输出之间的区别，达到更优的效果。残差块结构可以充分利用网络的深度，使网络能够学习到更加复杂和深层次的特征。由于残差块的设计，网络在训练过程中可以更好地保留原始输入信息的细节和结构，从而提高了特征的表示能力。图 1 为残差块示意图。

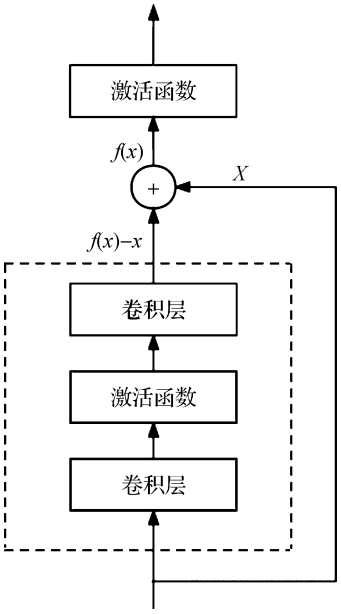

图 1　残差块示意图

由于保留了原始的输入信息，因此随着深度的增加，可以获取更高的精度，较浅的网络因具有更多的特征信息而获得更好的效果。在本文任务中，需要模型学习 $I_{\mathrm{RGB}} \rightarrow I_{\mathrm{sketch}}$ 的变化。从本质上来说，RGB 图像与其对应的草图之间具有如下关系：

$$I_{\mathrm{sketch}} = I_{\mathrm{RGB}} - I_{\mathrm{color}} \tag{1}$$

变换可得：

$$I_{\mathrm{RGB}} = I_{\mathrm{sketch}} + I_{\mathrm{color}} \tag{2}$$

可以通过残差块使模型学习到 $I_{\mathrm{RGB}} \rightarrow I_{\mathrm{sketch}}\ (\rightarrow I_{\mathrm{color}})$ 的变化，即获得 RGB 图像的多余颜色信息，进而通过数值计算去除多余的颜色信息，得到想要的漫画草图。

图 2 为本文模型的具体结构，该结构总体基于 U-Net，浅灰色方块为下采样残差块，深灰色方块为上采样残差块。输入为 $M \times N$ 的 RGB 图像，经过一次下采样残差块，即使用 3×3 的卷积层进行卷积计算，增加通道数，减少图像尺寸。以中间的六层残差块为轴，下采样阶段与上采样阶段分别包含两组残差块以及输入层与输出层。每一个残差块内部网络如图 3 所示。

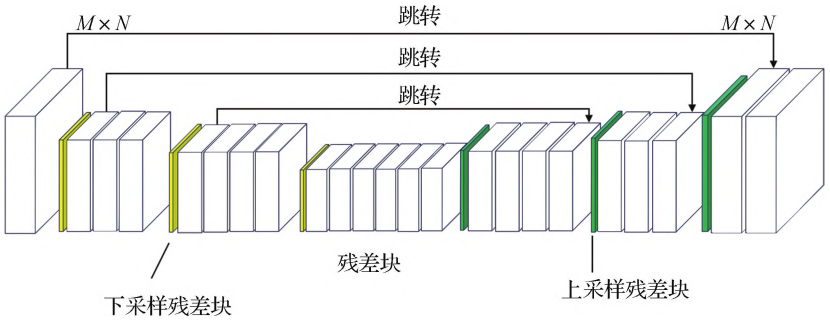

图 2　漫画草图提取方法的整体网络结构

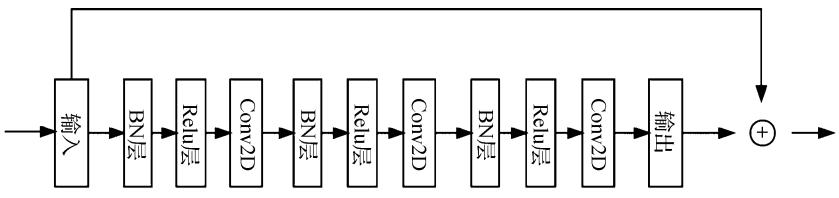

图 3　残差块内部网络结构示意图

2.2　损失函数

本文的目标是从 RGB 漫画图像中生成与之对应的漫画草图，其损失函数可以定义为生成草图 $I_{\mathrm{sketch}}^{\mathrm{g}}$ 与真实草图 $I_{\mathrm{sketch}}^{\mathrm{real}}$ 之间的误差。由于生成草图与真实草图均为线条图，图像内容相对简单，学习的目标是

尽可能降低两者之间的距离。L1 损失对于减小这种差距具有很好的效果，因为它对小的差距和大的差距都给予同样的重视，可以促进模型更好地拟合数据，生成更准确、更接近真实草图的线条图。综上所述，选用 L1 损失作为损失函数是合适的。具体来说，L1 损失是目标值与预测值之差绝对值的和，表示预测值的平均误差幅度，不需要考虑误差的方向。式(3)是本文模型的损失函数：

$$L=E(\| I_{\text{sketch}}^{\text{real}}-I_{\text{sketch}}^{\text{g}} \|_1) \tag{3}$$

式中，E 是期望函数。

3 实验结果与分析

3.1 数据准备与预处理

3.1.1 数据集

本文的数据集以 Danbooru2019[19] 数据集为基础。首先，通过联合标签“1girl 1boy”从整个数据集中随机选取约 500 张动漫人物 RGB 图像。然后，进行人工筛选，筛选标准包括：图像内容干净，无多余噪点和色块；图像以人物为主，背景不过于复杂；内容和谐，没有不良信息。最终，经筛选得到 200 张可用于实验的动漫人物 RGB 图像。图 4 为部分使用数据的展示。

图 4 数据集部分使用的数据

3.1.2 实验设置

本文实验在安装了 64 位 Ubuntu 16.04 系统的服务器上运行，服务器装有 2 块 NVIDIA Tesla V100 显卡，每张内存为 32 G，采用 CUDA 11.0 库进行加速。实验使用深度学习框架 PyTorch 实现。

所有的实验采用 Adam 优化器进行训练，实验的批大小(batch size)为 24，共训练 15 个 epoch。初始学习率设为 0.000 2，随后采用 PyTorch 官方学习率衰减策略，在每一个 epoch 结束后对学习率进行衰减。

3.2 结果与分析

3.2.1 效果展示

随机选取 100 张 RGB 图像，使用训练好的模型进行草图提取实验，最终验证结果如图 5 所示，每组图的左边为 RGB 原始图像，右边为经过模型计算得到的草图提取图。可以看出，本文提出的网络模型基本能够满足草图提取的要求，整体效果保持较好，并且与人工绘制的草图风格接近。对于复杂线条的图像，本文方法能够较好地适应，并且不会产生影响画面的噪点。

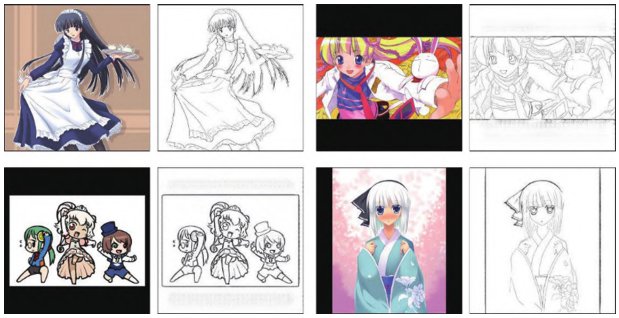

图 5 漫画草图提取效果展示

3.2.2 定性对比实验

为了验证本文算法的有效性，实验对比了常用的经典边缘提取方法，结果如图 6 所示。从图 6 可以看出，Sobel 算子与 Prewitt 算子对边缘的提取更注重对象的外轮廓，对明显的线条边界更加敏感，这使得对象最外的边缘轮廓粗而内部的线条浅。Laplacian 算子与 Canny 算子得到的边缘图都存在较多的噪声点。Laplacian 算子在人物脸部的噪声点较多，可能是与该部分线条密集有关；Canny 算子提取的线条信息相对完整，但同样在线条密集的地方存在不必要的噪声点。相对来说，Roberts 算子得到的边缘图线条之间强度相差不大，更接近人工绘制的草图结构，但线条整体强度较弱。将 Roberts 算子的边缘图进行线条强化处理，效果如图 7 所示，强化后效果变好。

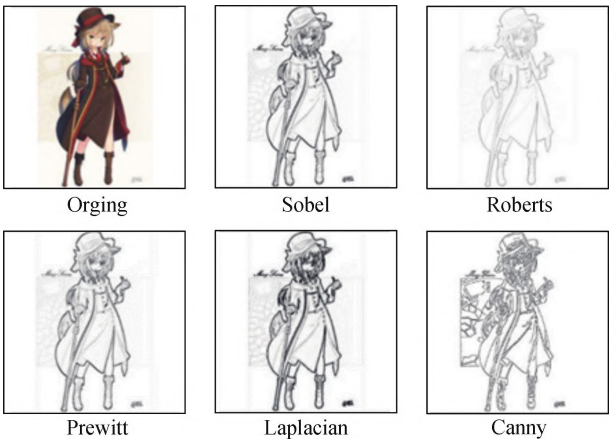

图 6　不同边缘提取算法的效果对比

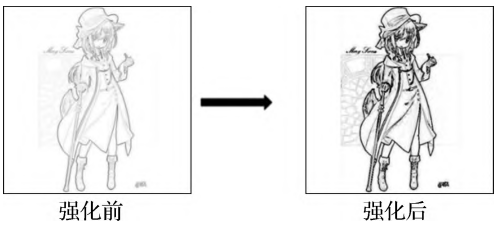

图 7　基于 Roberts 算子的边缘图线条强化处理效果

3.2.3　定量对比实验

本文方法生成的漫画草图，其本质可以看作只包含结构不包含风格内容的结构化图像，因此可以使用 FID(fréchet inception distance)作为生成效果的度量。FID 值用于衡量生成图像与真实图像分布之间的距离。FID 值越低，表明两组图像的相似度越高。表 1 为不同方法的 FID 值对比。可以看出，本文方法的 FID 值比其他方法至少低 40%，对草图提取有更优的效果，能够得到清晰、合理、保留大部分底层信息的漫画草图。

表 1　不同方法 FID 值对比

方法	FID 值
Canny[0,20]	202.89±0.21
Laplacian	151.70±0.26
Prewitt	145.65±0.18
Sobel	144.95±0.23
Roberts	135.63±0.22
Canny [50,100]	129.04±0.18
Canny [50,200]	125.88±0.20
本文	74.56±0.19

值得一提的是，从表 1 可以看出，使用 Canny 算子进行草图提取时，上下限的影响较大。虽然可以通过调整 Canny 算子的参数得到较好的 FID 值，但是在相同条件下，本文提出的方法仍然更加高效。

3.2.4　用户评估实验

为了评估本文方法和其他方法各方面的效果，邀请了 20 名参与者来评价不同方法生成图的效果。评价指标包括完整度、清晰度、精细度、结构合理性以及视觉美感 5 个维度。完整度表示生成的线条是否包括实际的线条，清晰度表示生成的线条整体是否清晰，精细度表示生成图是否包括了应有的细节，结构合理性表示是否在宏观和细节部分分配合理，视觉美感表示生成的线条整体是否较为舒服。针对某种方法，用户首先对其每个维度从 1 星到 5 星给出评分，1 星为最差，5 星为最好。然后，对评分的平均值进行归一化操作，将其归一化到[0,1]。式(4)是维度 j 的评分计算公式。

$$W_j = \frac{\left(\sum_{i=1}^{n} X_{i,j}\right)/n - S_{\min,j}}{S_{\max,j} - S_{\min,j}} \tag{4}$$

式中，n 为参与者人数；$X_{i,j}$ 为针对维度 j 在[1,5]范围的评分；$S_{\max,j}$ 为评分上边界，即 5；$S_{\min,j}$ 为评分下边界，即 1。

图 8 为 Canny、Laplacian、Prewitt、Roberts、Sobel 和本文方法的用户评估图。可以看出，本文方法在完整度、清晰度、视觉美感 3 个维度上超过了所有方法，表明本文方法能够提取更多的草图线条信息；在结构合理性与精细度两个维度比 Sobel 算子略低，但高于其他 4 种方法。综合来看本文提出的方法相较其他方法效果更好。

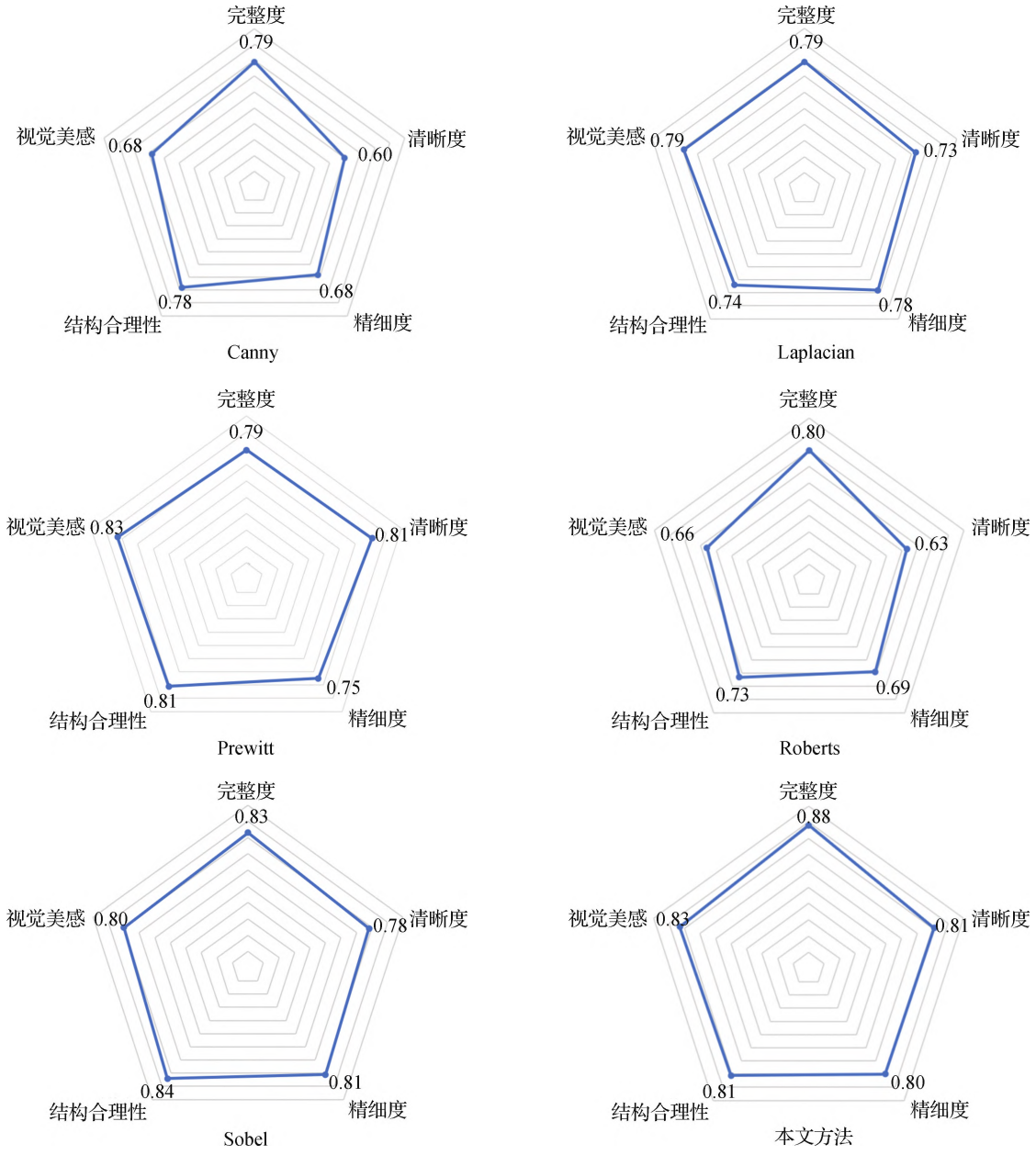

图 8 用户评估图

3.2.5 训练损失分析

图 9 展示的是本文模型的 L1 损失图。可以看出 L1 损失在训练过程中并非持续下降，而是在不断振荡中下降。本文实验中，在每个 epoch 结束后都对学习率进行一次调整，确保损失函数逐渐逼近全局最优解。在经过 2 000 次迭代后，最终损失值在 0.08～0.10 范围内波动。

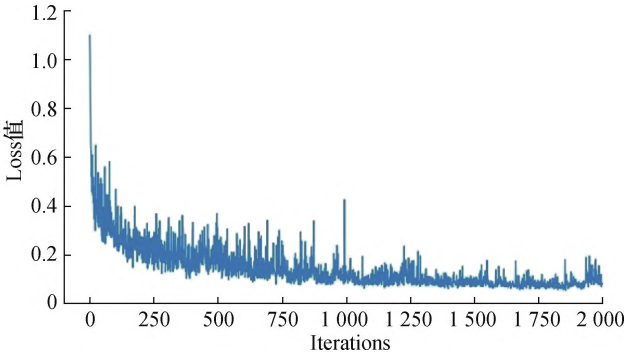

图 9 训练过程中的 L1 损失变化

4 结 语

本文提出了一种基于 U-Net 网络的漫画草图提取方法，通过在 U-Net 结构中引入残差块，增加了网络的结构信息提取能力。通过与不同边缘算法进

行对比，本文方法的 FID 值比其他算法至少低 40%，验证了本文方法的有效性。本文方法可以用于从动漫 RGB 图像中提取动漫线稿草图，作为后续动漫图像的预处理。本文提出的模型虽然在实验中得到了较好的结果，但仍有改进的空间。除了残差块之外，注意机制也具有较强的信息提取能力，特别在线条高度相似和背景复杂的情况下，采用基于注意机制的 Transformer 等技术有助于进一步改进本文的方法。

参考文献：

[1] RONNEBERGER O，FISCHER P，BROX T. U-net：convolutional networks for biomedical image segmentation[M]// Lecture Notes in Computer Science. Cham：Springer International Publishing，2015：234－241

[2] CHEN T，CHENG M M，TAN P，et al. Sketch2photo：Internet image montage[J]. ACM Transactions on Graphics (TOG)，2009，28(5)：1－10

[3] EITZ M，RICHTER R，HILDEBRAND K，et al. Photosketcher：interactive sketch-based image synthesis[J]. IEEE Computer Graphics and Applications，2011，31(6)：56－66

[4] HE K M，ZHANG X Y，REN S Q，et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Las Vegas：IEEE，2016：770－778

[5] GOODFELLOW I，POUGET-ABADIE J，MIRZA M，et al. Generative adversarial nets[EB/OL]. [2023－08－20]. https：//arxiv. org/abs/1406. 2661

[6] CHEN W L，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：IEEE，2018：9416－9425

[7] LU Y Y，WU S Z，TAI Y W，et al. Image generation from sketch constraint using contextual GAN[C]//European Conference on Computer Vision(ECCV). Cham：Springer International Publishing，2018：213－228

[8] SARVADEVABHATLA R K，DWIVEDI I，BISWAS A，et al. SketchParse：towards rich descriptions for poorly drawn sketches using multi-task hierarchical deep networks[C]//Proceedings of the 25th ACM International Conference on Multimedia. New York：ACM，2017：10－18

[9] 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(CVPR). Honolulu：IEEE，2017：5967－5976

[10] GAO C Y，LIU Q，XU Q，et al. SketchyCOCO：image generation from freehand scene sketches[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle：IEEE，2020：5173－5182

[11] LEI Y T，DU W W，HU Q H. Face sketch-to-photo transformation with multi-scale self-attention GAN[J]. Neurocomputing，2020，396：13－23

[12] VINKER Y，PAJOUHESHGAR E，BO J Y，et al. CLIPasso：semantically-aware object sketching[J]. ACM Transactions on Graphics，2022，41(4)：1－11

[13] VINKER Y，ALALUF Y，COHEN D，et al. CLIPascene：scene sketching with different types and levels of abstraction[C]//2023 IEEE/CVF International Conference on Computer Vision. Paris：IEEE，2023：4146－4156

[14] WANG L J，QIAN C，WANG J F，et al. Unsupervised learning of 3D model reconstruction from hand-drawn sketches[C]//Proceedings of the 26th ACM International Conference on Multimedia. New York：ACM，2018：1820－1828

[15] HUANG H B，KALOGERAKIS E，YUMER E，et al. Shape synthesis from sketches via procedural models and convolutional networks[J]. IEEE Transactions on Visualization and Computer Graphics，2017，23(8)：2003－2013

[16] CHAKRABARTY S，JOHNSON R F，RASHMI M，et al. Generating abstract art from hand-drawn sketches using GAN models[M]//UDDIN M S，BANSAL J C. Algorithms for Intelligent Systems. Singapore：Springer Nature Singapore，2023：539－552

[17] WU Z J，WANG Y N，FENG M T，et al. Sketch and text guided diffusion model for colored point cloud generation[C]//2023 IEEE/CVF International Conference on Computer Vision. Paris：IEEE，2023：8929－8939

[18] YE M J，ZHOU S Z，FU H B. DeepShapeSketch：generating hand drawing sketches from 3D objects[C]//2019 International Joint Conference on Neural Networks(IJCNN). Budapest：IEEE，2019：1－8

[19] BRANWEN G，ANONYMOUS，COMMUNITY D. Danbooru2019 portraits：a large-scale anime head illustration dataset[EB/OL]. (2020－08－05)[2023－09－05]. https：//gwern. net/crop＃danbooru2019-portraits

（责任编辑：湛　江　马金玉）