

基于生成对抗网络的漫画草稿图简化

卢倩雯¹ 陶青川¹ 赵娅琳¹ 刘蔓霄¹

摘要 在漫画绘制的过程中,按草稿绘制出线条干净的线稿是很重要的一环.现有的草图简化方法已经具有一定的线条简化能力,然而由于草图的绘制方式的多样性以及画面复杂程度的不同,这些方法适用范围有限且效果不理想.本文提出了一种新颖的草图简化方法,利用条件随机场(Conditional random field, CRF)和最小二乘生成式对抗网络(Least squares generative adversarial networks, LSGAN)理论搭建了深度卷积神经网络的草图简化模型,通过该网络生成器与判别器之间的零和博弈与条件约束,得到更加接近真实的简化线稿图.同时,为了训练对抗模型的草图简化能力,本文建立了包含更多绘制方式与不同内容的草图与简化线稿图对的训练数据集.实验表明,本文算法对于复杂情况下的草图,相比于目前的方法,具有更好的简化效果.

关键词 草图简化,最小二乘生成式对抗网络,深度学习,条件随机场

引用格式 卢倩雯,陶青川,赵娅琳,刘蔓霄.基于生成对抗网络的漫画草稿图简化.自动化学报,2018,44(5):840–854

DOI 10.16383/j.aas.2018.c170486

Sketch Simplification Using Generative Adversarial Networks

LU Qian-Wen¹ TAO Qing-Chuan¹ ZHAO Ya-Lin¹ LIU Man-Xiao¹

Abstract Sketch simplification is a critical part of cartoon drawing. To some extent, the existing approaches already have basic ability of sketch cleanup, but still have limitation in some situations because of the diversity of sketch drawing methods and complexity of sketch contents. In this paper, we present a novel approach of building a model for sketch simplification, which is based on the conditional random field (CRF) and least squares generative adversarial networks (LSGAN). Through the zero-sum game of generator and discriminator in the model and the learning restriction of conditional random field, we can obtain simplified images more similar to standard clean images. At the same time, we build a dataset containing a large number of pairs of sketches and clean images in different painting ways and contents. Finally, experiments show that our approach can obtain better results than the state of the art approach for sketch simplification.

Key words Sketch simplification, least squares generative adversarial network (LSGAN), deep learning, conditional random field (CRF)

Citation Lu Qian-Wen, Tao Qing-Chuan, Zhao Ya-Lin, Liu Man-Xiao. Sketch simplification using generative adversarial networks. *Acta Automatica Sinica*, 2018, 44(5): 840–854

漫画中的线条是构成整张图的重要部分,如何合理的将草稿中凌乱的线条进行增减、组合,是计算机草图简化的关键.大多数漫画家在绘制草稿时,主要目的是快速地在图纸上确定画面的整体结构、人物的大致形象和动作,这使得草稿图上的线条更加凌乱与不确定.在草稿图完成之后,漫画家们需要在凌乱的草稿图的基础上,勾勒出线条准确、干净的线稿图.在这个过程中,随着草稿图的复杂度的增加,漫画工作者的工作量也会成倍的提升,付出更多的时间和精力.因此,通过训练好的神经网络,自动合理地简化潦草的线条,生成漫画工作者们所需要的线稿图,将大大提升漫画绘制的工作效率,极大地减

轻工作者的负担.

在本文之前,已经出现了一些草稿简化的算法以及基于深度学习的图像处理算法^[1–10].文献[3–5]提出的草稿简化算法,核心在于删减掉画面中多余的线条;文献[6]提出的基于距离和拓扑理论合并线条的算法,文献[7]提出的基于线条与线条间的多个几何关系的约束的简化算法,都是通过修改线条的形状来达到草图简化的目的,文献[8]利用训练图中的线条之间的几何关系进行监督线条聚类来获得简化图;文献[9]提出基于封闭感知的草图简化方法,通过语义分析线条之间产生的封闭区域,再对区域进行处理来达到简化草图的目的.草图简化最新的方法是 Simo-Serra 等^[10]提出的基于卷积神经网络的线稿图生成方法,通过训练卷积神经网络内部的参数,学习标准草稿图与线稿图之间复杂的映射关系,模仿线稿图的生成过程,使卷积神经网络在输入草图后能生成对应的简化图,较以前的方法具有

收稿日期 2017-08-31 录用日期 2018-01-09
Manuscript received August 31, 2017; accepted January 9, 2018
本文责任编辑 王坤峰
Recommended by Associate Editor WANG Kun-Feng
1. 四川大学电子信息学院 成都 610065
1. Collage of Electronic and Information Engineering, Sichuan University, Chengdu 610065

自主学习、输入草图大小不固定、简化效果较以往方法更好的优点,但要获得较好的简化效果,则需要人工调整输入草图的大小,并且对于线条过于凌乱的画面,其简化能力有限. Simo-Serra^[10]的草图简化算法的实验结果证实了卷积神经网络能通过学习获得较好的草图简化能力,在此基础上,本文通过将生成对抗式的卷积神经网络替换基本的卷积神经网络,提升网络模型的草图简化能力. 在网络搭建中,借鉴了 Goodfellow 在机器自主进行图像生成学习的网络构建上提出的一个新颖的理论——生成式对抗网络(Generative adversarial networks, GAN)模型^[11-12]. 该网络由生成器和判别器组成,在迭代学习的过程中,生成器的目的是生成更加趋近于真实图像的伪造图,判别器的目的是正确区分生成器生成的伪造图和真实图像,二者在对抗中逐渐达到纳什均衡,即判别器判断伪造图和真实图,均有相等的可能性为真,且模型中的生成器已经具有优秀的仿真能力. 文献[13-16]在原始GAN上进行改进的算法,其中, Mao等^[16]在 Goodfellow等^[11]的基础上,针对解决生成对抗网络在迭代学习中的梯度消失等问题,提出了最小二乘生成式对抗网络(Least squares generative adversarial networks, LSGAN). 该方法通过改变生成器与判别器的误差函数,使网络整体更加稳定,收敛更快速,具有更强大的无监督图像处理能力.

另一方面,条件随机场(Conditional random field, CRF)作为无向图模型,已经广泛地被用于优化图像处理结果. 文献[17]引入分层的条件随机场处理街景图像的分割与物体分类问题,文献[18]利用分层条件随机场建立模型对人造图和真实图进行分类,文献[19]使用全连接的条件随机场优化卷积网络的语义分割结果.

本文将条件随机场与最小二乘生成式对抗网络

理论相结合,提出了基于条件随机场与最小二乘生成式对抗网络的草图简化方法. 网络中的生成器和判别器均由卷积单元构成,因此,与卷积神经网络相类似,生成器与判别器也能通过迭代学习,自动更新内部卷积核的参数,同时,在生成器的损失函数中,添加了条件随机场的约束以提高生成模型的精度与网络整体捕捉图像细节的能力. 本文网络能通过学习数据集中草稿图与线稿图之间的映射以及线稿图的特征,生成更加接近于真实线稿图的简化图. 本文的方法具有以下优点: 1) 模型构建基于条件随机场与最小二乘生成式对抗网络,在损失函数中添加条件随机场能够增加伪造图的生成约束,使网络向生成理想的伪造图的方向收敛更加迅速,且伪造图中线条内和背景内像素的像素值相差更小,线条与背景交界处的像素点的像素值相差更大,有利于更快速的生成更加优质的线稿图,同时,能够确保模型在通过学习后,能够生成更加接近真实线稿图的草稿简化图. 2) 训练完成后,生成器可以独立使用,本文模型对输入的草稿图的大小没有要求,能够自动生成与输入图相应图片大小的优秀简化图,不再需要判别器部分. 3) 本文首次将生成式对抗网络与条件随机场结合使用在草图简化工作上,将条件随机场作为生成对抗网络损失函数的一部分,加入了先验知识,大大提升了网络的收敛速度和简化效果,具有一定的创新性. 图1是草图和利用模型得到的对应简化图.

1 技术细节

1.1 网络搭建

本文通过建立一个基于条件随机场和最小二乘生成式对抗网络的卷积网络进行草图的简化学习. 在学习的过程中,草图经过生成器生成对应的伪造



图1 草稿图和与之对应的模型结果图

Fig. 1 Original sketches and output images of model

图, 再将生成的伪造图和真实的线稿图一起, 输入判别器进行分类. 判别器的输出代表对每张输入图片的真假判定, 判定结果与标签求误差后, 再根据梯度下降法对模型参数进行调整, 达到模型自主学习的目的. 模型的构建参考了 Simo-Serra^[10] 搭建的卷积神经网络和最小二乘生成式对抗网络. 其中, 生成器模型内部包含了卷积层、平卷积层和上卷积层, 分别具有不同的步长, 在迭代学习中完成特征提取与图像重构的工作. 判别器模型包括卷积层和全连层, 在迭代中学习准确地区分出标准线稿图和生成器伪造图. 为了提高网络整体的稳定性, 除了生成器的输出卷积层和判别器的输入卷积层, 其他卷积层之后都添加归一化处理层 (Batch normalization)^[20]. 为了加快模型的收敛速度, 模型内使用卷积层代替了池化层. 本文的草稿图简化模型如图 2 所示, 包括生成器模型和判别器模型. 在训练过程中, 生成器模型学习生成与线稿图像相似的伪造图, 判别器模型学习更准确地区分出伪造图与标准线稿图. 模型将生成式对抗网络与深度卷积网络相结合, 遵从文献 [15] 提出的深度卷积生成式对抗网络的构建约束: 1) 在判别网络中使用步幅卷积层替代池化层, 在生成网络中使用微步幅卷积层; 2) 除了生成网络的输出卷积层和判别网络的输入卷积层, 其他卷积层后都添加 Batchnorm 层; 3) 不使用任何全连接隐含层; 4) 生成网络中, 输出层使用 Tanh 激活函数, 其他层使用 Relu 激活函数; 5) 判别网络中, 所有层都使用 Leakyrelu 激活函数.

模型中的生成网络在搭建中参考了 Simo-Serra 等^[10] 提出的卷积神经网络. 网络内部包含了三种

卷积层: 步幅卷积层, 平卷积层和微步幅卷积层. 其中, 步幅卷积层内卷积步幅为 2, 将输出特征图的大小减小为输入特征图大小的一半, 确保生成网络在学习中获得提取输入草图的从像素级特征到内容级特征的能力. 同时, 为了在一些特征级上提取更多特征图, 模型中加入了步幅为 1 的平卷积层. 而为了使生成网络的生成图具有和输入草图相同的尺寸, 模型中使用了步幅为 1/2 的微步幅卷积, 使输出图片的大小为输入图片的两倍. 另外, 为了增加生成网络应对线条复杂程度更高的输入图, 增加了模型中平卷积层的数量提高了对高层特征的提取能力.

草稿简化模型的判别网络, 由卷积层与全连接层构成, 具有提取输入图片特征以及根据提取特征图对输入图像分类的能力. 模型结构的细节如表 1 所示. 1~25 层为生成器内卷积层, 26~36 层为判别器内卷积层. 除了生成器和判别器的输出卷积单元和判别器的输入卷积单元, 每个卷积单元都添加了 Batchnorm 层和一个 Relu 层, 在生成网络的输出层使用了 Tanh 激活函数. 为了保持输入图片和输出图片的大小一致, 每个卷积层都进行补零. 在训练完成后, 生成器模型可以单独使用, 处理任何尺寸的图片.

在模型中, Batch normalization 在除了生成网络的输出卷积层和判别网络的输入卷积层之外的卷积层后使用, 通过在每个 Mini-batch 中计算得到均值和方差后, 将卷积层输出信号进行规范化. Batch normalization 的使用能够增加模型的稳定性, 提高模型的收敛速度.

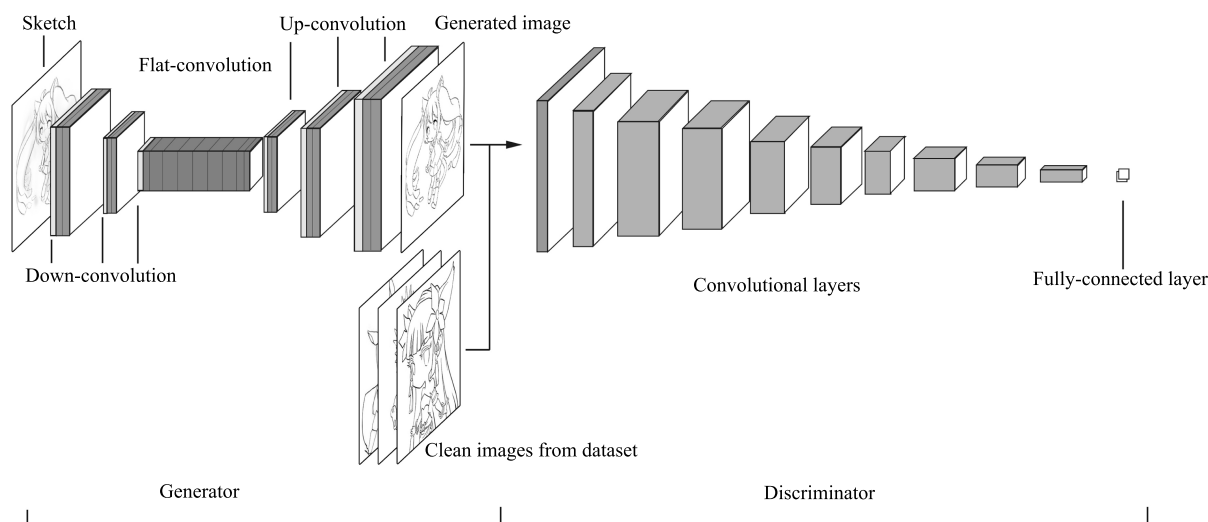


图 2 模型示意图

Fig. 2 Schematic figure of model

表 1 基于条件随机场和最小二乘生成式对抗网络的草图简化模型
Table 1 Sketch simplification model based on the conditional random field and least squares generative adversarial networks

序号	层类型	卷积核	步长	深度	序号	层类型	卷积核	步长	深度
0	Input	—	—	1	20	Up-convolution	4 × 4	1/2 × 1/2	128
1	Down-convolution	5 × 5	2 × 2	48	21	Flat-convolution	3 × 3	1 × 1	128
2	Flat-convolution	3 × 3	1 × 1	128	22	Flat-convolution	3 × 3	1 × 1	48
3	Flat-convolution	3 × 3	1 × 1	128	23	Up-convolution	4 × 4	1/2 × 1/2	48
4	Down-convolution	3 × 3	2 × 2	256	24	Flat-convolution	3 × 3	1 × 1	24
5	Flat-convolution	3 × 3	1 × 1	256	25	Flat-convolution	3 × 3	1 × 1	1
6	Flat-convolution	3 × 3	1 × 1	256	26	Convolution	5 × 5	1 × 1	20
7	Down-convolution	3 × 3	2 × 2	256	27	Convolution	5 × 5	1 × 1	50
8	Flat-convolution	3 × 3	1 × 1	512	28	Convolution	4 × 4	1 × 1	500
9	Flat-convolution	3 × 3	1 × 1	1 024	29	Convolution	5 × 5	2 × 2	500
10	Flat-convolution	3 × 3	1 × 1	1 024	30	Convolution	5 × 5	2 × 2	500
11	Flat-convolution	3 × 3	1 × 1	1 024	31	Convolution	5 × 5	2 × 2	500
12	Flat-convolution	3 × 3	1 × 1	1 024	32	Convolution	5 × 5	2 × 2	500
13	Flat-convolution	3 × 3	1 × 1	1 024	33	Convolution	5 × 5	2 × 2	1 000
14	Flat-convolution	3 × 3	1 × 1	1 024	34	Convolution	5 × 5	2 × 2	1 000
15	Flat-convolution	3 × 3	1 × 1	512	35	Convolution	4 × 4	2 × 2	1 000
16	Flat-convolution	3 × 3	1 × 1	256	36	Fully-connected	1 × 1	1 × 1	2
17	Up-convolution	4 × 4	1/2 × 1/2	256	37	Dloss	—	—	—
18	Flat-convolution	3 × 3	1 × 1	256	38	Gloss	—	—	—
19	Flat-convolution	3 × 3	1 × 1	128					

在生成网络中,除了输出卷积层,每个卷积单元都加入了 Relu 激活函数.

$$\sigma_{\text{Relu}}(x) = \max(0, x) \quad (1)$$

在判别网络中,所有的卷积单元都加入了 Leaky relu 激活函数.

$$\sigma_{\text{Leaky relu}}(x) = \begin{cases} x, & \text{若 } x > 0 \\ 0.01x, & \text{否则} \end{cases} \quad (2)$$

在生成器的输出卷积层,添加了 Tanh 激活函数代替 Relu 激活函数,将生成器输出归一化至区间 $[-1, 1]$.

$$\sigma_{\text{Tanh}}(x) = \frac{\sinh x}{\cosh x} \quad (3)$$

1.2 网络损失函数

在确定模型的损失时,引入 Goodfellow 等^[11]提出的生成式对抗网络的相关理论.为从标准数据 x 中学习生成器目标分布,生成式对抗网络的输入定义为 $p_z(z)$,具有多层感知器功能的生成器定义

为 $G(z; \theta_g)$,其中 θ_g 为生成器内的参数, $G(z)$ 代表生成器输出的伪造图.具有分类功能的判别器定义为 $D(x; \theta_d)$,其中 θ_d 代表判别器内的参数.判别器的输出 $D(x)$ 代表对应的输入被判定为真的比率.在训练过程中,生成器学习更好的欺骗判别器的能力,使判别器对伪造图的评分更高,即学习最小化 $\log(D(x)) + \log(1 - D(G(z)))$.同时,判别器通过学习,不断提升区分伪造图与真实图的能力,使其对伪造图的评分更低,真实图的评分更高,即学习最大化 $\log(D(x)) + \log(1 - D(G(z)))$.生成器和判别器在博弈的过程如下所示:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (4)$$

当生成器固定, $V(D, G)$ 形如函数 $y \rightarrow a \log(y) + b \log(1 - y)$,其中 $(a, b) \in \mathbb{R}^2 \setminus \{0, 0\}$,求导可得 $[0, 1]$ 区间上的最大值在点 $a/(a + b)$ 处,即 $V(D, G)$ 要获得最大值必须满足条件如下:

$$D_G^*(x) = \frac{bp_{\text{data}}(x) + ap_g(x)}{p_{\text{data}}(x) + p_g(x)} = \frac{p_{\text{data}}(x)}{p_{\text{data}}(x) + p_g(x)} \quad (5)$$

同时, 为了使该网络达到纳什均衡, 生成器需要在 $\max_D V(G, D)$ 取到最小值, 函数表示如下:

$$C(G) = \max_D V(G, D) = \mathbb{E}_{X \sim p_{\text{data}}} \left[\log \frac{p_{\text{data}}(x)}{p_{\text{data}}(x) + p_g(x)} \right] + \mathbb{E}_{X \sim p_g} \left[\log \frac{p_g(x)}{p_{\text{data}}(x) + p_g(x)} \right] \quad (6)$$

为了使 $p_{\text{data}}(x) = p_g(x)$, 此时 $D_G^*(x) = 1/2$, $C(G)$ 取得最小值.

$$C(G) = V(D_G^*, G) = -\log(4) + 2 \times JSD_{\text{data}} \|p_g \quad (7)$$

此时生成器生成的伪造图与真实图几乎一致, 香农散度为 0, $C(G)$ 最终取得值为 $-\log(4)$ 的全局最小值.

需要注意, 原始生成对抗网络在训练中, 判别器的 Sigmoid 交叉熵损失只关注输入的图片是否被正确区分, 对于伪造图分类正确的情况不进行惩罚, 使得该损失函数能提供的损失误差较小, 容易在学习时产生梯度消失的情况. 文献 [16] 在 Goodfellow^[11] 的基础上提出的最小二乘生成式对抗网络, 优化了生成器和判别器的损失函数.

生成器损失函数如下:

$$\min_D V_{\text{LSGAN}}(D) = \frac{1}{2} \mathbb{E}_{x \sim p_{\text{data}}(x)} [(D(x) - b)^2] + \frac{1}{2} \mathbb{E}_{z \sim p_z(z)} [(D(G(z)) - a)^2] \quad (8)$$

判别器损失函数如下:

$$\min_G V_{\text{LSGAN}}(G) = \frac{1}{2} \mathbb{E}_{z \sim p_z(z)} [(D(G(z)) - c)^2] \quad (9)$$

其中, 设定 $a = 0$, $b = c = 1$, 且 a , b , c 分别代表伪造图的标签、真实图的标签和生成器期望判别器对伪造图判定的标签. 与原始的生成式对抗网络一致, 在 $p_{\text{data}}(x) = p_g(x)$ 时, 达到网络内部的纳什均衡.

为了进一步优化对抗网络的精度, 本文在最小二乘生成式对抗网络的损失函数的基础上, 在生成器模型中添加了条件随机场的能量函数作为损失函

数的一部分. 通过生成器的仿造图中像素点与其四邻域点之间的关系约束, 在迭代学习中不需要进行人为的干预, 网络整体能够自主地向像素间能量更低、更稳定的方向收敛, 生成更加符合要求的仿造图.

在本文的模型中, 生成器和判别器分别具有不同的损失函数. 生成器的损失函数主要包括内容损失、对抗损失和条件随机场能量损失. 内容损失计算生成器生成图与对应标准线稿图的逐像素均方误差, 在学习过程中保证生成图与对应线稿图的像素分布相似. 对抗损失计算判别器对伪造图判定的标签与真实图标签的误差, 确保生成网络在学习过程中增加欺骗判别网络的能力. 另外, 图像中的每个像素点的灰度值, 都与其领域内的其他点存在相关关系. 同为背景点, 线条内部点之间像素值相差很小, 线条边缘点像素值则相差较大. 因此, 本文参考文献 [19] 的算法, 利用条件随机场的能量损失对伪造图像素点与其四邻域像素点之间的关系进行约束, 使简化网络在收敛的过程中, 学习增加像素标签不同的伪造图像素点之间的灰度差, 减小像素标签相同的像素点之间的灰度差, 使生成的伪造图边缘对比更加明显、背景与线条的颜色更加稳定.

生成网络的总误差定义如下:

$$l_G(x, z, \theta_g) = k_1 \|G(z) - x\|_{\text{FRO}}^2 + k_2 \|D(G(z)) - 1\|_{\text{FRO}}^2 + k_3 \sum_{ij} E_{ij}(z_i, z_j) \quad (10)$$

其中, $\|\cdot\|_{\text{FRO}}$ 代表 Frobenius 范数, 系数 $k_1 = 1$, $k_2 = 0.1$, $k_3 = 0.5$. $\sum_{ij} E_{ij}(z_i, z_j)$ 计算如下:

$$\sum_{ij} E_{ij}(z_i, z_j) = \mu(l_i, l_j) \exp\left(-\frac{\|I_{z_i} - I_{z_j}\|^2}{2\sigma_\alpha^2}\right) + (1 - \mu(l_i, l_j)) \exp\left(-\frac{\|1 - |I_{z_i} - I_{z_j}|\|^2}{2\sigma_\beta^2}\right) \quad (11)$$

其中, z_i, z_j 分别代表在生成器伪造图内选取的像素点和像素点的四邻域点, l_i, l_j 代表像素点和四邻域点对应的像素标签. 像素的像素标签由生成器伪造图对应的标准线稿图上对应的像素点的位置决定, 本文规定标准图内位于背景的像素点的像素标签为 1, 位于线条的像素点的像素标签为 0, 且当 $l_i = l_j$ 时, $\mu(l_i, l_j) = 0$, 否则 $\mu(l_i, l_j) = 1$. I_{z_i}, I_{z_j} 分别代表 z_i, z_j 像素点的灰度值. 经过验证图集验证, 参数 σ_α 和 σ_β 通常取值为 1.

判别网络作为分类器, 目的是学习正确地生

成器伪造图与真实图进行区分. 判别器的损失函数由最小二乘生成式对抗网络误差函数定义如下:

$$l_D(x, G(z), \theta_d) = \frac{1}{2}(D(x) - 1)^2 + \frac{1}{2}(D(G(z)))^2 \tag{12}$$

利用最小二乘的损失函数, 使网络的损失函数在判别器分类错误的情况下提供学习误差, 在判别器分类正确的情况下, 仍然可以对远离决策边界的点进行惩罚, 并将它们拉向决策边界. 使判别器具有更强的辨别能力, 能够对生成器学习提供更多贡献, 避免了模型崩溃情况的出现, 大大提高了网络的图像处理精度.

本文网络的损失函数中未添加文献 [10] 中使用的 Lossmap. 添加 Lossmap 虽有利于加速网络收敛速度, 且使网络在学习的过程中不因注重于较粗的线条而忽略较细的线条, 但增加 Lossmap 进行训练的网络在训练完成之后有个明显的缺陷: 在处理较粗的线条时, 网络的生成图更倾向于将组线条分离成两个独立的线条, 或出现线条消失的情况, 如图 3 所示. 为避免上述情况出现, 本文网络使用 CRF 代替 Lossmap 加速网络的收敛.

2 实验结果及分析

2.1 数据集构成

本文构建了由 240 对在纸和手绘板上完成的草图与线稿图组成的原始大图数据集. 数据集中的原始大图分辨率不一致, 但所有图片都大于 500 像素 × 500 像素. 在实验中, 为了使本文中构建的草图简化网络具有更好的简化能力, 原始的图片数据集被分成训练图集、验证图集和测试图集三部分.

训练图集中的大图在经过变换截取后获得了

1 000 张以上, 大小均为 128 像素 × 128 像素的成对的小图, 作为简化模型在训练时的输入与标准图. 这些训练图中的小图对是通过从原始大图对数据集抽取部分图片后, 经过旋转随机角度, 随机地选取坐标截取获得, 如图 4 所示. 同时, 为了增加模型的稳定性, 针对纸质草图在扫描后产生的画面变脏、画质更低、噪声点更多等问题, 本文将图片对中的草稿图使用 Adobe photoshop 随机进行了 4 种变换: 质感变化, 图像模糊, 增加图像噪声, 增加斑点. 在图 5 中列举出了本文中一部分用于训练简化模型的数据集图片, 分类为动植物、建筑与场景、日漫风格漫画形象和迪士尼风格漫画形象, 且一部分在实际训练前进行了加噪处理. 此外, 在构建训练图集之前, 由于纸质漫画与电脑绘制的漫画在线条上存在区别, 如图 6 所示, 在两幅图中, 由图形板描绘的草图更凌乱, 且使用文献 [10] 中提供的模型更难清理干净. 而由于越来越多的漫画家和插图师选择使用手绘板作图, 在进行网络训练之前, 本文收集了一些由手绘板在电脑上进行绘制的图片对, 加入到网络训练中, 增强模型应对复杂情况时的草图简化能力. 验证图集包含 400 对大小为 128 像素 × 128 像素的图片, 主要用于在模型训练时, 通过验证模型训练图对中的草稿图的简化效果确定模型的超参数. 同时, 本文构建了图片对数量为 400、大小为 128 像素 × 128 像素的测试图集, 用于测试已训练好的模型的草图简化能力, 通过与最新的草图简化算法的输出结果对比, 证明本文提供的草图简化模型确实具有更优秀的简化能力. 需要注意的是, 验证图集和测试图集不参与模型内部的参数更新.

2.2 训练过程

在实验中, 本文使用了卷积生成式对抗网络^[11]、



图 3 本文网络输出图与使用了 Lossmap 的 Simo-Serra^[10] 的结果图对比

Fig. 3 The comparison of our model outputs and results from Simo-Serra algorithm^[10] with the usage of lossmaps

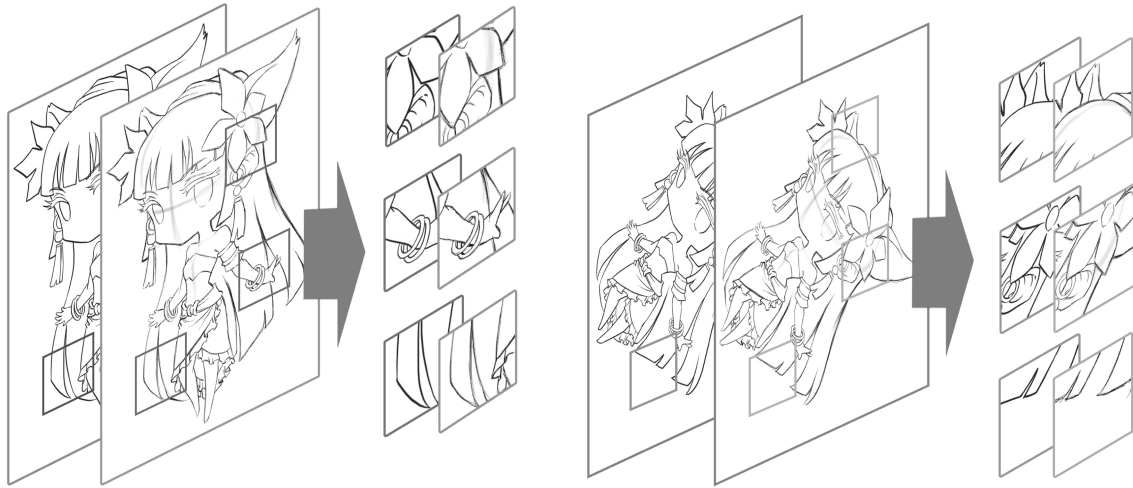


图 4 在大图中截取用于训练的小图示意

Fig. 4 Extracting small images from the original images for training



(a) 从互联网下载的原始的线稿与草稿的图片对

(a) Some original images and sketches downloaded from the internet



(b) 通过 Photoshop CS6 进行噪声添加后的训练图

(b) Images after adding with noise using Photoshop CS6

图 5 图片集构成

Fig. 5 Images dataset construction

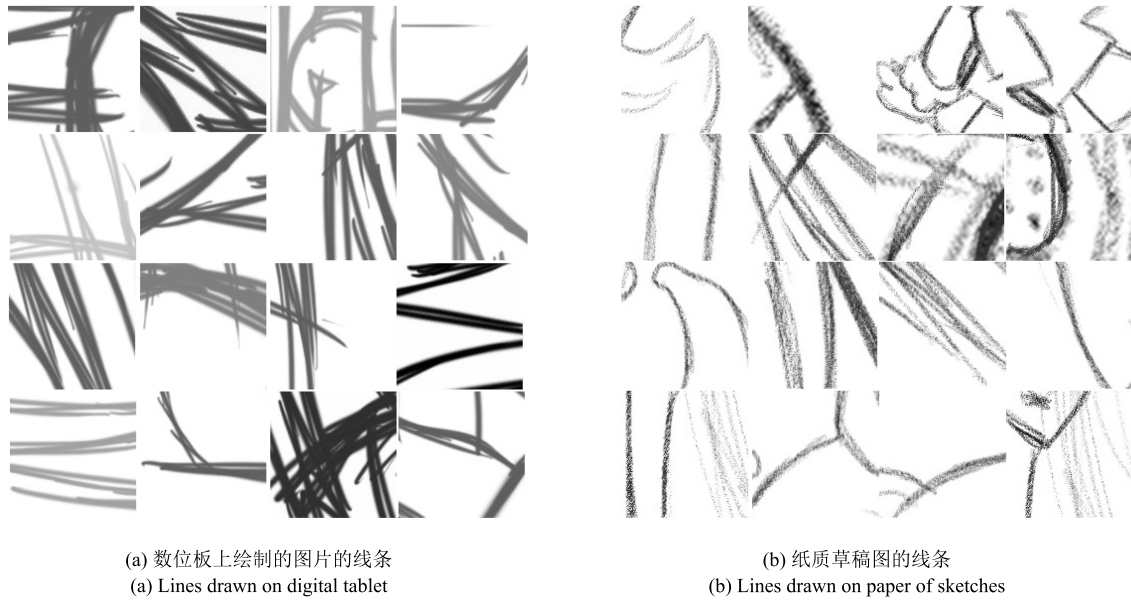


图 6 数位板与纸质草稿图中的线条对比

Fig. 6 The comparison of lines in sketches drawn on digital tablet and real paper

最小二乘生成式对抗网络和添加条件随机场能量损失的最小二乘生成式对抗网络^[16]三种方法进行对比训练. 同时, 还将本文算法和未添加条件随机场能量函数的最小二乘生成式对抗网络与全连接条件随机场优化相结合的算法进行了对比.

卷积生成式对抗网络、最小二乘生成式对抗网络和本文网络的损失函数值随迭代次数的变化如图 7 所示. 从图 7 可以看出, 本文提供的网络具有最快的收敛速度, 且最快达到较平稳的状态. 在训练的过程中, 卷积生成式对抗网络无法经过训练学习复杂的草图简化, 由于判别器起初不具有正确分类的能力, 并且当 Sigmoid 交叉熵作为损失函数时, 只关注模型是否正确分类而不关心偏离正确分类的程度, 使得判别模型对标准图和伪造图持续判为错, Fake loss 持续为 0, 失去一部分对生成器提供残差的能力, 最终训练失败. 在最小二乘生成式对抗网络与添加条件随机场能量损失的最小二乘生成式对抗网络的对比中, 可以看到添加条件随机场能量损失的最小二乘生成式对抗网络中的生成器具有最快的收敛效果. 网络整体在迭代次数 5000 时, 生成器与判别器误差都已趋于稳定, 已经具有较好的生成伪造图的能力, 判别器也能够持续为生成器提供稳定的残差, 有助于判别器与生成器内部的参数微调. 学习过程中, 三个网络的输出图片随迭代次数变化的对比如图 8 所示. 从图 8 可以看出, 本文添加了条件随机场能量函数的网络明显较未添加的网络具有明显的收敛速度更快、生成图更清晰、更贴近标准线稿图的优势, 同时体现了条件随机场能量函数对于提

升简化效果的重要性.

为了对比文献 [19] 提出的先通过卷积神经网络 (Convolutional neural network, CNN) 分类, 再经过条件随机场进行迭代优化分类结果的方法, 本文将相同迭代次数下未添加条件随机场的最小二乘生成式对抗网络生成图再经过条件随机场优化的结果图与添加了条件随机场能量方程的最小二乘生成式对抗网络的结果图进行了对比实验. 由于未添加在模型内部的条件随机场优化不能使用网络生成图相应的标准线稿图进行参数的训练, 因此不能使用与本文添加在网络内部的条件随机场能量方程一致的能量函数, 本文利用与文献 [19] 一致的全连接条件随机场能量方程, 如下:

$$E(x) = \sum_i \theta_i(x_i) + \sum_{ij} \theta_{ij}(x_i, x_j) \quad (13)$$

其中, x 为输入图中的像素, $\theta_i(x_i) = -\log P(x_i)$, $P(x_i)$ 为 CNN 网络输出的标签分类概率, $\theta_{ij}(x_i, x_j) = \mu(x_i, x_j) \sum_{m=1}^K \omega_m \times k^m(f_i, f_j)$, 当 $x_i \neq x_j$ 时, $\mu(x_i, x_j) = 1$, 否则 $\mu(x_i, x_j) = 0$. k^m 代表基于像素点 i 与像素点 j 的特征 f 且权值为 ω_m 的高斯核, 其公式如下:

$$\omega_1 \exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma_\alpha^2} - \frac{\|I_i - I_j\|^2}{2\sigma_\beta^2}\right) + \omega_2 \exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma_\gamma^2}\right) \quad (14)$$

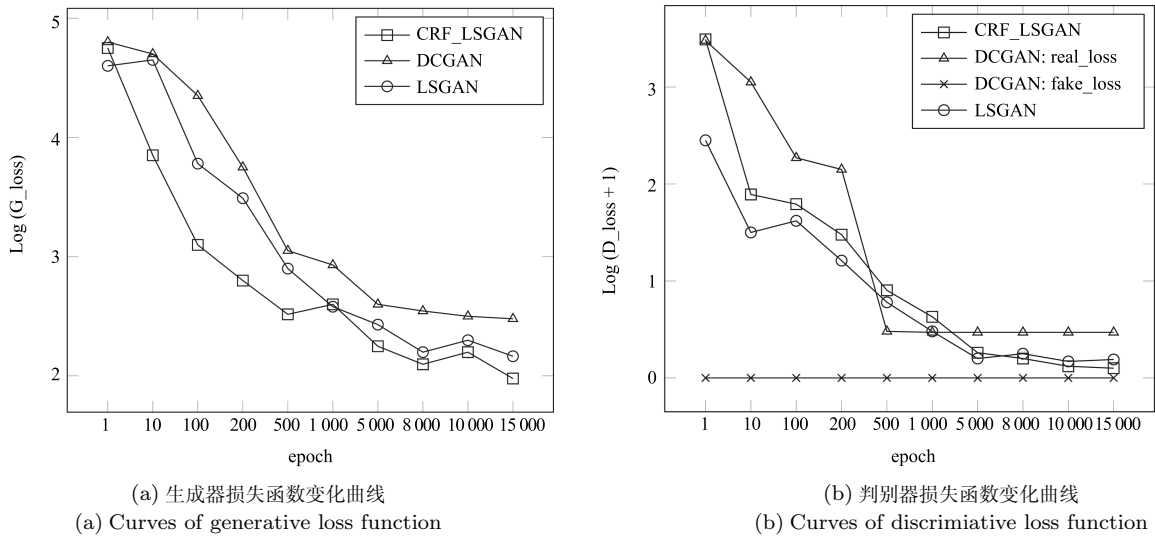


图 7 三网络损失函数变化图

Fig. 7 Curve of loss functions changing with iterations

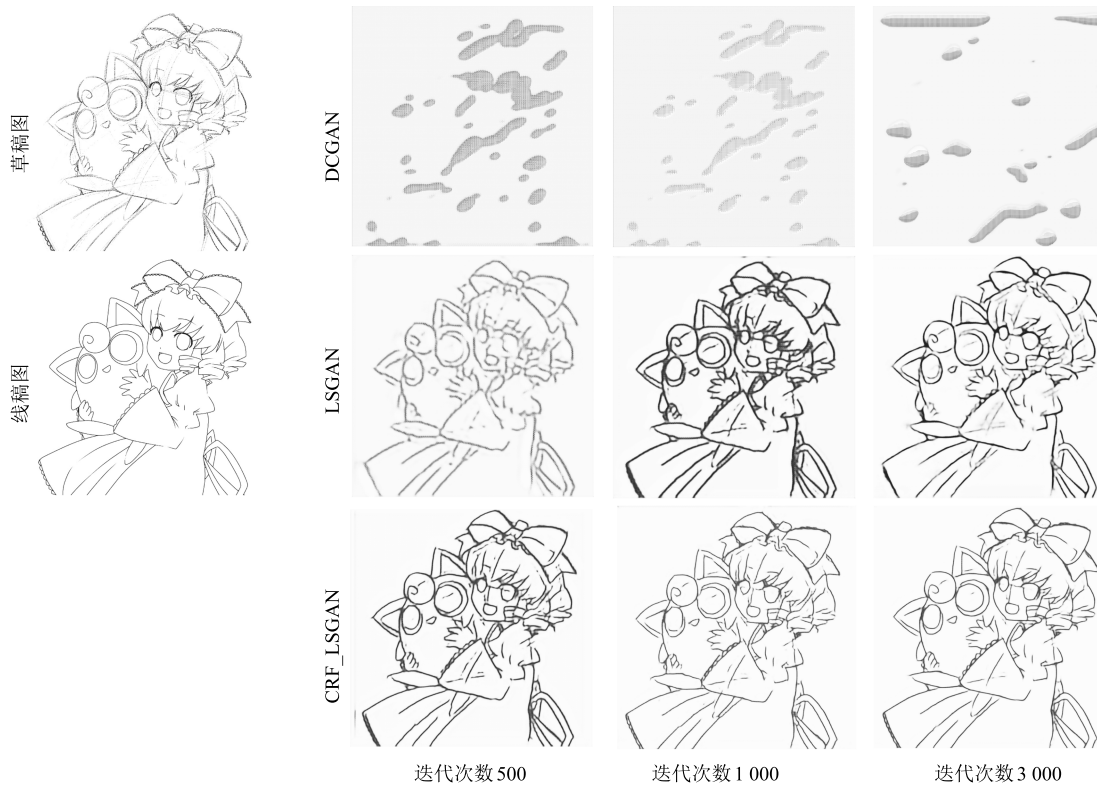


图 8 三个网络的输出图随迭代次数变化对比

Fig. 8 Outputs of three models changing with iterations

在实验中, 为了获得较好的二分类效果, 参数设置为 $\omega_1 = 10, \omega_2 = 0, \sigma_\alpha = 60, \sigma_\beta = 20, \sigma_\gamma = 3$.

本文分别选用了迭代了 5000 次的添加了条件随机场的最小二乘生成式对抗网络的生成图和未添加条件随机场的最小二乘生成式对抗网络的生成图, 再经过全连接的条件随机场迭代 10 次后的结果图

进行了对比, 如图 9 所示. 需要注意, 条件随机场迭代 10 次以上, 结果图几乎不再发生改变. 根据图片显示, 由于缺少条件随机场能量方程对内部的参数训练提供的有力的帮助, 未添加条件随机场的网络在同等迭代次数的情况下, 学习效果远不如添加了条件随机场约束的网络. 并且, 虽然通过全连接的条件



图9 本文网络输出图、最小二乘生成式对抗网络输出图、最小二乘生成式对抗网络与全连接条件随机场优化结合的输出图对比

Fig.9 The comparison of outputs from our model, LSGAN model and the model combined with LSGAN and CRF

随机场迭代优化, 网络的生成图边缘确实较之前分界明显, 但是远不足以改善生成图的整体情况, 仍然存在大面积黑块、细节缺失等情况, 生成图质量较差. 另外, 由于条件随机场添加在外部, 则即使在网络训练好后, 每次输入图片都需要进行额外的迭代优化的步骤, 比本文网络更加耗时, 效率较低.

本文网络在学习的过程中, 生成网络的参数 θ_g 更新, 通过计算生成网络误差 $l_G(x, z, \theta_g)$ 与 θ_{g_t} 的一阶导数实现, 如下:

$$\theta_{g_{t+1}} = \theta_{g_t} + \Delta\theta_{g_t} = \theta_{g_t} - \frac{RMS[\Delta\theta_g]_{t-1}}{RMS[\delta\theta_g]_t} \delta\theta_{g_t} \quad (15)$$

其中, $\Delta\theta_{g_t}$ 为参数 θ_g 的更新量, $\delta\theta_{g_t}$ 代表参数在迭代 t 次时的梯度.

判别网络也利用了梯度下降法更新网络内参数 θ_d , 为了获得误差函数 $l_D(x, G(z), \theta_d)$ 的最小值, 参数的更新方法如下:

$$\theta_{d_{j+1}} = \theta_{d_j} - \alpha \frac{\partial}{\partial \theta_{d_j}} l_D(x, G(z), \theta_d) \quad (16)$$

其中, α 表示判别网络的参数更新速率.

2.3 生成图向量化

本文提供的网络通过训练事实上已经能够生成逼真的草稿简化图, 但由于图像线条的边缘在像素级别上仍然具有不够圆滑、边界对比不够强烈的问题, 这些问题虽不影响美观, 但对绘画工作者后续的上色工作将产生一定的影响. 因此, 本文使用

Potrace software^[21] 对网络生成图进行向量化, 使用其默认值, 并且在向量化过程中不需要人为干预. 在图 10 中展示了向量化的输入与输出结果. 需要注意, 文献 [10] 已经证明直接对草稿图进行向量化不能得到较好的简化结果.

2.4 实验结果

通过学习现有的简化草图方法并结合对 LSGAN 模型的分析, 经过大量实验, 证明与目前最新的 Simo-Serra 等提出的方法^[10] 相比, 本文提供的生成对抗网络模型更适合于简化草图. 实验中, 搭建模型使用 Python3.5 和 Tensorflow 1.1.0, 并在 GPU 为 Nvidia gtx980ti 的 Windows 平台上训练 4 周. 需要注意, 模型输入不需要任何预处理, 并且在训练完成后, 生成网络可以输入任意大小的草图, 如图 11 所示, 其中生成图 2 均为以原始大小的左侧草图为输入图获得的网络生成图, 生成图 1 为以扩展为原始长宽的 2 倍的草图为输入图获得的网络生成图, 生成图 3 为以缩减为原始长宽的 1/2 倍的草图为输入图获得的网络生成图. 从图 11 可以看出, 当输入图尺寸较大时, 结果图获得的细节相对较多, 输入图尺寸较小时, 结果图相对简化, 但输入草图无论如何变化大小, 本文提供的简化算法都能得到相应优秀的简化图, 可直接被绘画工作者使用. 利用本文提供的测试图集, 基于条件随机场与最小二乘生成式对抗网络的草图简化模型的输出图和原始草稿图的对比如图 12 所示. 从图 12 可以看出, 在各种内容、风格、绘制方式的草图输入下, 本文网络均能获得较好的简



模型输出

向量化处理结果

图 10 模型输出与向量化结果图对比

Fig. 10 Comparison between output of our model and image after vectorization



草图

生成图 1

生成图 2

生成图 3

图 11 多尺寸草图获得的算法生成图对比

Fig. 11 The comparison of output images obtained by different sizes of input images

化结果. 另外, 在 GPU 为 Nvidia gtx980ti 的 Windows 平台上, 草图简化模型在训练完成后, 生成器网络生成简化图所需时间如表 2 所示.

2.5 对比现有技术

本文在算法验证阶段使用了测试图集, 并且利

用了宏查全率、宏查准率以及相应的宏 F1 的相关原理, 对本文算法的简化图与 Simo-Serra^[10] 提供的例子的简化结果做了详细的计算与对比分析, 通过结果验证, 本文的简化模型与最新的草图简化模型相比, 确实具有更加优秀的草图简化能力.

在二分类问题中, 查全率与查准率常用来共同衡

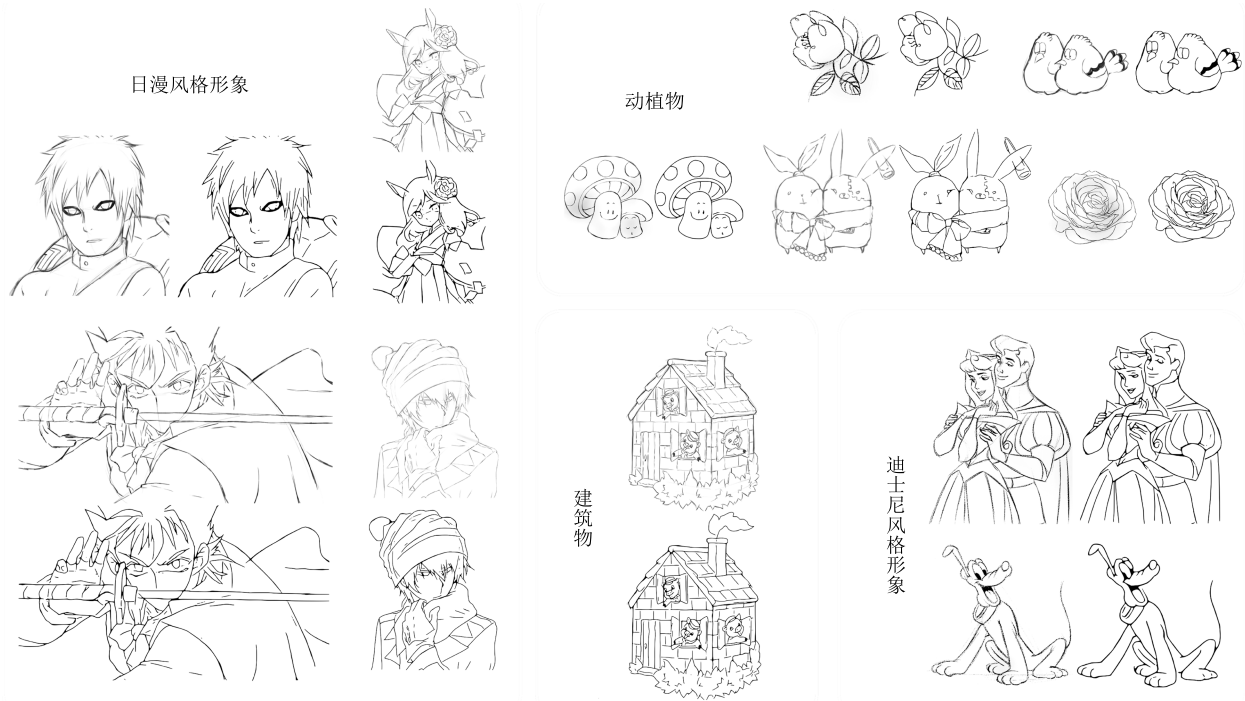


图 12 不同草图与本文模型对应的生成图

Fig.12 Fake images generated by our generator using different sketches

表 2 生成器网络生成简化图所需时间
Table 2 Time needed for generating a simplified image by the generator

图片大小 (像素)	像素个数	GPU (Nvidia gtx980ti) (s)
500 × 500	250 000	1.377
1500 × 1500	2250 000	11.632
2000 × 2000	4000 000	21.431

量一个模型的输出结果. 其中, 可将样本根据其真实的类别与分类器预测的类别的组合划分成真正例 (TP)、假正例 (FP)、真反例 (TN)、假反例 (FN) 四种情况.

查全率 (P) 和查准率 (R) 定义如下:

$$P = \frac{TP}{TP + FP} \tag{17}$$

$$R = \frac{TP}{TP + FN} \tag{18}$$

F1 为查全率和查准率的加权调和平均. 但在有多个混淆矩阵的情况下, 常常使用宏查全率和宏查准率来计算得出相应的宏 F1, 以代替 F1. 宏查全率、宏查准率和宏 F1 的公式下:

$$\text{macro-}P = \frac{1}{n} \sum_{i=1}^n P_i \tag{19}$$

$$\text{macro-}R = \frac{1}{n} \sum_{i=1}^n R_i \tag{20}$$

$$\text{macro-F1} = \frac{2 \times \text{macro-}P \times \text{macro-}R}{\text{macro-}P + \text{macro-}R} \tag{21}$$

本文中的简化模型也可以看成一个二分类的分类器, 规定标测试集图片对中的每张标准线稿图的所有像素点分别为每次计算时的总样本, 其中线条部分的像素点是实际为“真”的样本, 背景的像素点则是实际为“假”的样本. 同时, 在对应的输出简化图中, 规定线条部分的像素点是分类器预测为“真”的样本, 而背景部分的像素点则是分类器预测为“假”的样本. 同理, 例子的结果也可以作为 Simo-Serra^[10] 草图简化模型的分类结果.

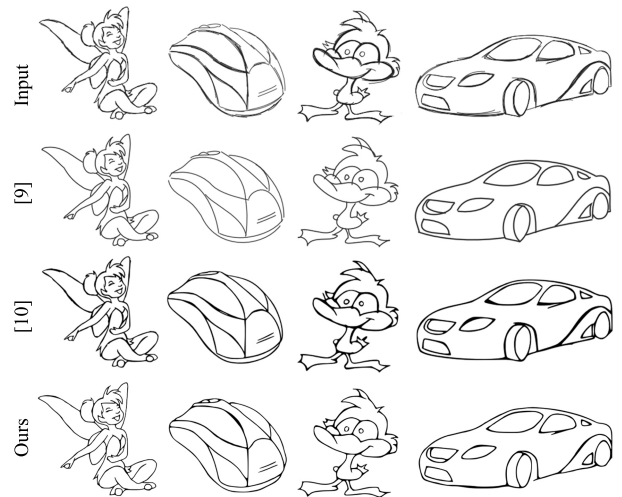
测试的过程中, 将测试图集的草稿图输入本文的简化网络, 得到本文算法的结果图, 同时, 将同样的草稿图输入 Simo-Serra^[10] 的例子中, 获得 Simo-Serra^[10] 算法的结果图. 上述步骤完成之后, 再利用测试图集的标准线稿图和查全率和查准率的公式获得对比测试结果, 如表 3. 需要注意, 由于经过卷积神经网络后输出图中的线条会有些许像素位移, 不能对标准图和两个对比图进行像素间的一一对应判定, 而应该对计算设置一定的范围. 为确定相对合理的范围, 本文对范围的变化对测试的结果 macro-F1 与 macro-F1_s 的影响进行了实验, 结果如

表 3 利用测试集获得的 Simo-Serra^[10] 简化模型输出图和本文简化模型输出图的宏查全率、宏查准率和宏 F1 的测试结果
 Table 3 Test results of macro-*R*, macro-*P* and macro-F1 of output images of Simo-Serra^[10] simplification model and output images of our model using test dataset

测试项目	Simo-Serra 模型结果图	本文模型结果图
宏查全率 (范围为 5 像素 × 5 像素)	0.6660	0.7278
宏查准率 (范围为 5 像素 × 5 像素)	0.7105	0.7078
宏 F1 (范围为 5 像素 × 5 像素)	0.6875	0.7181

图 13 所示. 本文算法结果图的测试结果 macro-F1 与 Simo-Serra^[10] 算法结果图的测试结果 macro-F1_s. 在范围增大时, 均能获得稳定提升. 另外, 在范围为 3 像素 × 3 像素时, macro-F1 与 macro-F1_s 达到最大的差值, 在小于该范围时, 由于网络输出的像素位移对二者均有显著影响, 二者结果下降的同时, 差值也相对下降, 而在大于该范围时, 二者由于容忍条件的放宽, 测试结果的可信程度相对降低, 虽然二者的结果均有提升且差异减小, 但过大的范围获得的测试结果已经不具有实际意义. 最后, 本文选择将范围为 5 像素 × 5 像素时的 macro-F1 与 macro-F1_s 的数值作为对比测试的结果. 需要注意, 范围变化的过程中, macro-F1 始终大于 macro-F1_s, 因此, 测试证明了本文提供的基于条件随机场与最小二乘生成式对抗网络的草图简化模型确实优于最新的、由 Simo-Serra 等^[10] 提出的基于卷积神经网络的草图简化模型. 另外, 在输入同一内容和大小的草稿图的情况下, 本文算法得到的结果图与文献 [9-10] 得到的结果图对比如图 14 所示. 从图 14 可以看出, 本文提供的算法确实具有比最新的算法^[9-10] 更好的草图简化能力, 输入相同草图得到的简化结果图细节更加的丰富, 画面中的线条虽有粗细的变

化, 但既体现了作者对线条的强调或忽略的作画意图, 也保持了线条的粗细均匀, 具有很强的实用性.



(a) 本文结果图与文献 [9-10] 的结果图对比
 (a) Compared with outputs of our method, [9] and [10]

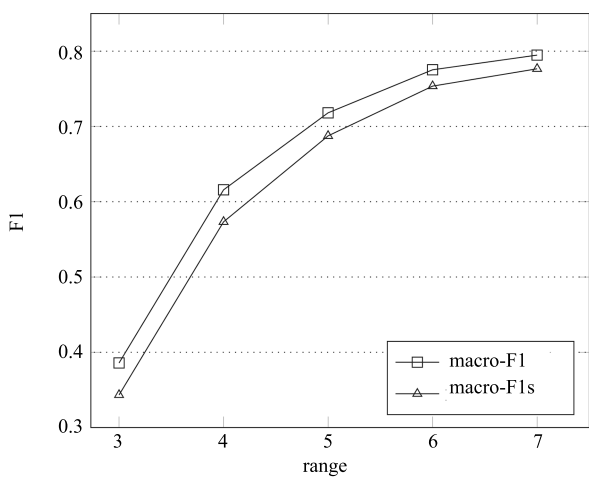
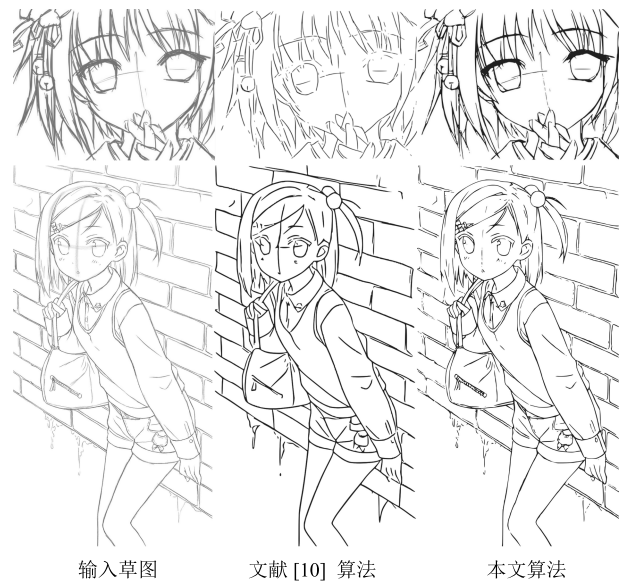


图 13 macro-F1 与 macro-F1_s 对比图

Fig. 13 Experiment on the influence of range change on macro-F1 and macro-F1_s.



(b) 本文结果图与文献 [10] 结果图对比
 (b) Compared with outputs of our method and Simo-Serra^[10]

图 14 本文算法与其他算法的两组对比图

Fig. 14 Compared with outputs of the latest algorithms

2.6 局限性

本文中的基于条件随机场与最小二乘生成式对抗网络的草图简化模型由于包含生成网络和判别网络两个部分, 拥有更多的隐含层和更高的复杂度, 导致训练草图简化网络的过程更耗时. 此外, 模型生成图像的质量取决于训练数据的质量和数量, 这与由卷积层构建的其他模型相同, 原因在于由于模型内部参数较多, 不充足的训练数据更容易使模型过拟合, 得不到较好的实验效果. 但在提供更多的高质量训练图像以及迭代次数充足的情况下, 能够尽量避免过拟合导致的训练失败, 同时, 基于条件随机场与最小二乘生成式对抗网络的草图简化模型能够为图形艺术家提供更优秀的草稿简化图.

3 结论

本文提出了一种基于条件随机场和最小二乘生成式对抗网络 (LSGAN) 的新颖草图简化方法, 并能获得更高质量的草图简化图. 同时, 构建了新的数据集, 其中包含在真实纸张和数码手绘板上绘制的草稿图与其对应的干净线稿图. 此外, 草图简化模型使用 GPU 进行训练, 并且计划收集更多的图像对, 使模型在后续的训练中生成更好的简化图像.

References

- Sun Xu, Li Xiao-Guang, Li Jia-Feng, Zhuo Li. Review on deep learning based image super-resolution restoration algorithms. *Acta Automatica Sinica*, 2017, **43**(5): 697–709 (孙旭, 李晓光, 李嘉锋, 卓力. 基于深度学习的图像超分辨率复原研究进展. *自动化学报*, 2017, **43**(5): 697–709)
- Zhang Hui, Wang Kun-Feng, Wang Fei-Yue. Advances and perspectives on applications of deep learning in visual object detection *Acta Automatica Sinica*, 2017, **43**(8): 1289–1305 (张慧, 王坤峰, 王飞跃. 深度学习在目标视觉检测中的应用进展与展望. *自动化学报*, 2017, **43**(8): 1289–1305)
- Wilson B, Ma K L. Rendering complexity in computer-generated pen-and-ink illustrations. In: Proceedings of the 3rd International Symposium on Non-Photorealistic Animation and Rendering. New York, NY, USA: ACM, 2004. 129–137
- Grabli S, Durand F, Sillion F X. Density measure for line-drawing simplification. In: Proceedings of the 12th Pacific Conference on Computer Graphics and Applications. Seoul, South Korea: IEEE, 2004. 309–318
- Cole F, DeCarlo D, Finkelstein A, Kin K, Morley K, Santella A. Directing gaze in 3D models with stylized focus. In: Proceedings of the 17th Eurographics Conference on Rendering Techniques. Nicosia, Cyprus: ACM, 2006. 377–387
- Grimm C, Joshi P. Just Drawit: a 3D sketching system. In: Proceedings of the 2012 International Symposium on Sketch-Based Interfaces and Modeling. Annecy, France: ACM, 2012. 121–130
- Fišer J, Asente P, Sýkora D. ShipShape: a drawing beautification assistant. In: Proceedings of the 2015 Workshop on Sketch-Based Interfaces and Modeling. Istanbul, Turkey: ACM, 2015. 49–57
- Orbay G, Kara L B. Beautification of design sketches using trainable stroke clustering and curve fitting. *IEEE Transactions on Visualization and Computer Graphics*, 2011, **17**(5): 694–708
- Liu X T, Wong T T, Heng P A. Closure-aware sketch simplification. *ACM Transactions on Graphics (TOG)*, 2015, **34**(6): Article No. 168
- Simo-Serra E, Iizuka S, Sasaki K, Ishikawa H. Learning to simplify: fully convolutional networks for rough sketch cleanup *ACM Transactions on Graphics (TOG)*, 2016, **35**(4): Article No. 121
- Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. In: Proceedings of the 2014 Advances in Neural Information Processing Systems (NIPS). Montreal, Canada: Curran Associates, Inc., 2014. 2672–2680
- Wang Kun-Feng, Gou Chao, Duan Yan-Jie, Lin Yi-Lun, Zheng Xin-Hu, Wang Fei-Yue. Generative adversarial networks: the state of the art and beyond. *Acta Automatica Sinica*, 2017, **43**(3): 321–332 (王坤峰, 苟超, 段艳杰, 林懿伦, 郑心湖, 王飞跃. 生成式对抗网络 GAN 的研究进展与展望. *自动化学报*, 2017, **43**(3): 321–332)
- Qi G J. Loss-sensitive generative adversarial networks on Lipschitz densities. ArXiv: 1701.06264, 2017.
- Arjovsky M, Chintala S, Bottou L. Wasserstein GAN. ArXiv: 1701.07875, 2017.
- Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. In: Proceedings of the 2015 International Conference on Learning Representations (ICLR). San Diego, CA, USA: ICLR, 2015.
- Mao X D, Li Q, Xie H R, Lau R Y K, Wang Z, Smolley S P. Least squares generative adversarial networks. In: Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017. 2813–2821
- Huang Q X, Han M, Wu B, Ioffe S. A hierarchical conditional random field model for labeling and segmenting images of street scenes. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition. Colorado Springs, CO, USA: IEEE, 2011. 1953–1960
- Yang M Y, Förstner W. A hierarchical conditional random field model for labeling and classifying images of man-made scenes. In: Proceedings of the 2011 IEEE International Conference on Computer Vision Workshops. Barcelona, Spain: IEEE, 2011. 196–203
- Chen L C, Papandreou G, Kokkinos I, Murphy K, Yuille A L. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, **40**(4): 834–848
- Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: Proceedings of the 32nd International Conference on Machine Learning. Lille, France: PMLR, 2015.

- 21 Selinger P. Potrace: a polygon-based tracing algorithm [Online], available: <http://potrace.sourceforge.net/potrace.pdf>, May 10, 2017



卢倩雯 四川大学硕士研究生. 2015 年获得上海大学学士学位. 主要研究方向为机器视觉, 模式识别, 机器学习, 图像处理. E-mail: lqwjk2018@163.com

(**LU Qian-Wen** Master student at Sichuan University. She received her bachelor degree from Shanghai University in 2015. Her research interest covers

computer vision, pattern recognition, machine learning, and image processing.)



陶青川 四川大学电子信息学院副教授. 分别于 1997 年、2000 年和 2005 年获得四川大学学士学位、硕士学位和博士学位. 主要研究方向为机器视觉, 模式识别, 机器学习, 图像处理和三维显微成像技术. 本文通信作者.

E-mail: taoqingchuan@scu.edu.cn

(**TAO Qing-Chuan** Associate professor at the College of Electronics and Information Engineering, Sichuan University. He received his bachelor degree, master degree, and Ph.D. degree from Sichuan University, in 1997, 2000, and 2005, respectively. His research

interest covers computer vision, pattern recognition, machine learning, and image processing.)

interest covers computer vision, pattern recognition, machine learning, image processing, and optical sectioning microscopy 3D imaging technique. Corresponding author of this paper.)



赵娅琳 四川大学硕士研究生. 2014 年获得重庆交通大学学士学位. 主要研究方向为机器视觉, 模式识别, 机器学习, 图像处理. E-mail: caeluszhao@yeah.cn

(**ZHAO Ya-Lin** Master student at Sichuan University. She received her bachelor degree from Chongqing Jiaotong University in 2014. Her research

interest covers computer vision, pattern recognition, machine learning, and image processing.)



刘蔓霄 四川大学硕士研究生. 2016 年获得成都学院学士学位. 主要研究方向为机器视觉, 模式识别, 机器学习, 图像处理. E-mail: 18708179608@163.com

(**LIU Man-Xiao** Master student at Sichuan University. She received her bachelor degree from Chengdu University in 2016. Her research interest covers

computer vision, pattern recognition, machine learning, and image processing.)