基于图像分割及修复的数据生成

Data Generation Based on Image Segmentation and Inpainting

苏海军，薛任谦谦（中讯邮电咨询设计院有限公司，北京 100048）
Su Haijun，Xue Renqianqian（China Information Technology Designing & Consulting Institute Co.，Ltd.，Beijing 100048，China）

摘　要：
随着硬件及技术的发展，深度学习技术得到广泛的研究及应用，而数据作为驱动深度学习发展的重要输入，变得越来越重要。但是针对一些特殊场景，存在数据少、隐私风险等问题。针对上述问题，结合任意物体分割技术和图像修复技术，消除图像中隐私信息，生成不包含隐私数据的新图像；或结合生成模型，在去除指定目标同时生成包含其他正样本的图像；亦或者保留指定正样本，更换背景，生成不同场景数据。实验结果表明该种数据生成的有效性。

关键词：
图像分割；图像修复；数据生成
doi：10.12045/j.issn.1007-3043.2023.07.009
文章编号：1007-3043（2023）07-0049-05
中图分类号：TN919
文献标识码：A
开放科学（资源服务）标识码（OSID）：

Abstract：
With the development of hardware and technology，deep learning has been widely studied and applied. Data，as the key inputs driving the development of deep learning，is becoming more and more important. However，for some special scenarios，there are problems such as lack of data and privacy risks. In view of the above problems，it combines segmentation anything and image inpainting techniques to eliminate the private information in the image and generate a new image without private data. Also，combined with the generatitve model，a new image of different positive sample is generated while removing the specified target，or it keeps the specified target and changes the background to generate different scene image. Experimental results show the effectiveness of this kind of data generation.

Keywords：
Image segment；Image inpainting；Data generation

引用格式：苏海军，薛任谦谦. 基于图像分割及修复的数据生成[J]. 邮电设计技术，2023（7）：49-53.

1 概述

随着深度学习及大模型技术的发展，数据变得愈发重要。但特殊场景的数据存在数据量少、收集困难等问题，此外在已收集的数据中往往包含车牌、人脸等隐私数据以及一些其他的敏感数据。在以往的研究中，对于小样本任务的处理，往往采用迁移学习进行微调参数学习；对于敏感及隐私数据源，往往在数据上进行马赛克处理，遮挡敏感数据。

本文提出一种基于图像分割[1]和图像修复[2]技术的数据生成流程结构。该方法在用户输入的指导下，基于最新的任意物体分割模型自动选取合适的目标图像区域，接着采用前沿的图像填充技术，将选中的目标区域结合周围像素生成目标图像，进而生成一组新的数字图像。同时，该结构还可以结合图像生成大模型，基于现有图像，在指定区域生成新的目标样本或者保留现有目标，更换背景生成新的场景图像。

2 技术方案

本文重点介绍指定物体消除式数据生成结构，该

收稿日期：2023-05-26

架构包含以下3个部分：交互式区域选择、图像自动分割、图像自动修复，同时也可以扩展应用到目标替换和背景替换。数据生成结构如图1所示。

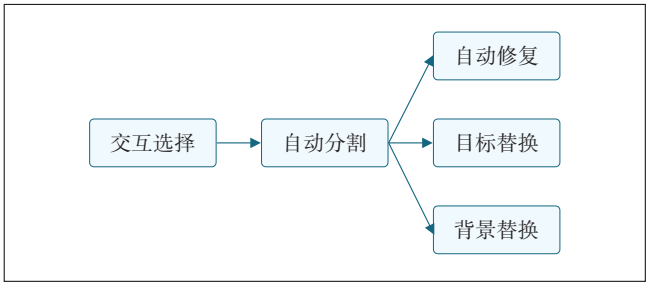

图1　数据生成结构

2.1 交互选择

交互选择是选择给定图中的目标物区域，包括点选以及框选2种方式。点选是指单次点击目标区域，分割模型根据点击位置自动分割出目标的掩码；框选是预先指定一个目标区域，然后分割模型自动在目标区域内分割出目标掩码。

点选包含正样本点选择和负样本点选择，正样本点明确当前位置是目标，负样本点明确当前位置为背景区域。点选过程可以是一次性输入所有正负样本点，进而通过模型直接获得分割结果；也可以是渐进式选择，即在上一步点选及分割结果的基础上，根据分割效果，选择加入正样本或者负样本点，以达到进一步优化分割效果的目标。综上所述，点选的优点是选择快速，不限制选择范围，比如显著的大目标，仅仅通过一次点击，就可完成大目标分割；而点选的缺点是在相对复杂的背景中，如果仅通过单次点击选择不能精确锁定目标，需要根据分割结果进行多次调整选择，而每次调整都会进行模型的推理，耗时较多。总体来说，在实际应用中，相对于背景突出的简单目标推荐选择该种方式。

框选是指在目标周围提前选定好范围，相比于点选，该方式明确了目标主要集中的区域；对于模型来说，输入的引导信息更丰富。具体地，框选包括四边形以及任意多边形（通常大于四边形）。在实际使用中常见的方式是四边形，该种方式规定了目标在图像中的左上及右下位置。另一种更精确的方式是任意多边形，任意多边形构成了一个不规则的封闭区域，更好地约束了目标所在的区域范围。一个极端的多边形可以直接是目标区域，相当于直接指定了目标的所有边界点，并不需要分割。总体来说，框选方式能为模型提供更多的引导信息，适用于背景复杂的场景以及精确选择目标分割的场景。

2.2 图像分割

与检测和识别任务类似，目标分割任务兴起时也预先定义了多个目标类别，并准备大量标注数据，因此早期目标分割的研究方向集中在特定目标分割领域，而具有交互输入的任意物体分割算法研究相对较少。

特定目标分割或者语义分割往往只关注预定义好的目标类别，并且不需要交互输入，能自动在图像中分割出特定的目标区域，比如只关注人像的分割称为人像分割[3]，只关注天空的分割称为天空分割[4]等。该种分割算法的特点是往往只关注某一大类目标，并不关注图像中的其他类别目标。通常语义分割只关注指定类别，而不对同一类别的不同个体做区分，比如人像语义分割，最终输出结果是所有人像的掩码区域，并不区分人像个体。在语义分割基础上进一步区分每个目标的个体则被称为实例分割，比如人像实例分割，需要分割出每个人的掩码区域。

任意目标分割或者交互式分割是预先不指定目标类别，通过用户交互式提示，分割出用户想要的目标区域。因此不论是点选还是框选交互，都作为分割模型的一个输入，旨在获得目标的精确掩码。与特定目标分割相比，交互式分割具有以下优点。

a）简单交互便能够获得较好的分割结果。

b）比语义分割和全景分割更具有针对性，对单个物体的分割效果也较优。

c）由于分割一般只集中在用户需要分割的部分图像，所以计算复杂度不高，运行时间短。

d）训练过程中多数使用的是类别不可知方式，因此对没有出现的类别有一定的泛化能力。

在深度学习兴起之前，交互式分割的主流方法是优化算法，交互的主要形式是划线，然后利用图论的方法建立能量方程，通过最小化能量方程的方式迭代得到分割结果。由于是基于优化的实现过程，这类方法的分割效果一般，但这种优化的思路可以作为基于深度学习的一种后处理，进一步提升分割精度。Xu[5]基于2015年提出的FCN网络，在2016年CVPR中首次将深度学习方法引入到交互式分割的研究中，通过对FCN进行少量的改动和微调，得到了第一篇关于深度学习的交互式分割方法的论文。同时，论文中提出了正负点的采样方式和生成距离特征图的方法，打开了

基于深度学习的交互式分割研究的大门。虽然该方法实现过程较为繁琐，实现难度也较大，但提出的思路在后期得到广泛关注及应用，在之后的研究中，提升分割精度也多以网络结构更新、交互输入形式的改进为主。虽然这些交互式分割能根据用户提示分割出目标，但精度仍然有提升空间。2023年，由Meta AI研究中心提出了一个创新性大模型，号称能分割一切的模型Segment Anything Model（SAM）[6]。虽然没有ChatGPT的应用广泛，但SAM的出现也被称为分割领域的GPT。

SAM定义了一种新的图像分割任务及模型结构，在1 100万个图像中训练学习了超过10亿个掩码。由于该模型的设计出发点是可提示输入，因此该模型不仅能分割参与训练的已知目标类型，还能分割未在训练中出现的目标类型，即可以在零样本（zero-shot）数据中实现目标分割。与特定目标分割不同，可提示的分割任务对模型架构施加了约束，使模型在结构上必须支持灵活的提示，并对输入提示做到模糊感知；在性能上，需要摊销实时计算掩码耗时以允许交互式使用。基于该思路，SAM结构设计为3个部分（见图2）：1个强大的图像编码器，用于提取图像中的特征；1个提示编码器，用于捕获交互输入信息；2个信息源在轻量级的掩码解码器中组合，预测最终的分割掩码。其中，SAM的图像编码器允许对输入的图像一次性提取特征，在之后的提示输入中，可以重用已提取的图像特征，达到摊销时间成本的目标。通过不同的提示编码器和图像特征融合，并且通过轻量级的掩码解码器，可最终得到基于该提示输入的不同分割结果。

2.3 图像修复

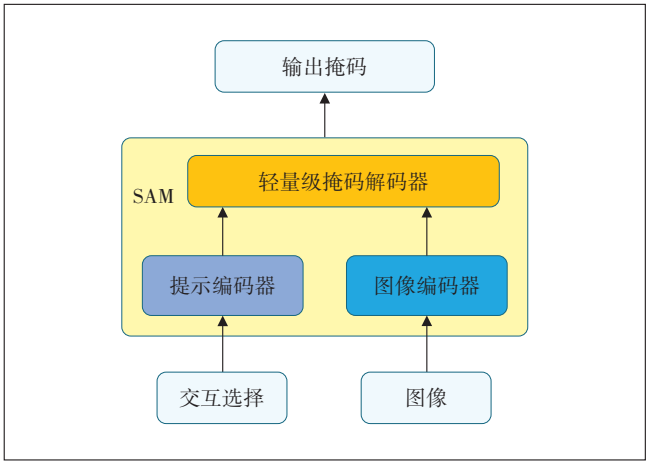

图2　Segment Anything模型结构

图像修复是一种重要的图像处理技术，通过恢复被损坏或缺失的图像信息来改善图像质量，主要的图像修复方法包括基于插值的方法和基于深度学习的方法[2]。基于插值的方法是一种传统的图像修复方法，该方法根据缺失图像位置周边信息，对缺失图像从边缘到中心进行像素插值，进而实现图像的修复。同时，为获得较高的修复精度，会采用图像金字塔的方式，由粗到细逐层修复，常用的方法有patch match[7]。该种方法不需要标注数据及训练模型，实现思路及原理简单，操作简单易行，在处理局部小面积修复时效果很好，但在处理大面积或者复杂纹理的场景时，修复效果略差。由于采用金字塔式逐层修复，随着修复面积增大，耗时也会增多。

基于深度学习的方法是近年来得到广泛关注的一种图像修复方法[8-10]，该类方法利用深度学习模型学习图像的特征，从而实现图像的修复和重建。在实现过程中，通过设计一个专用的卷积神经网络以及对应的损失函数，在指定的数据集上进行训练，进而获得一个图像修复的模型。由于经过大量数据学习训练，该方法鲁棒性更强，可以有效地修复不同程度、不同场景的图像。

研究发现，要想获得好的修复效果，需要网络模型和损失函数的计算都具备足够的感受野，Lama[11]提出使用快速傅里叶卷积[12]来增大感受野，进而实现大掩码下的图像修复，与降分辨率获得大感受野相比，快速傅里叶卷积能保留更多图像细节。Lama基于快速傅里叶卷积提出新的图像修复网络，使得网络在浅层阶段即使未经过大的下采样，也能获得整个图片的感受野。此外，基于快速傅里叶卷积的大感知视野，Lama提出使用感知损失函数[13]。同时，在训练过程中提出了遮挡区域掩码生成策略，通过调整输入掩码的大小，使模型对于不同场景、不同大小掩码都具有较好的鲁棒性。因此，快速傅里叶卷积不仅提升了模型的修复质量，还降低了模型参数量，同时兼顾了修复效果和性能。

2.4 扩展应用

随着图像生成技术的发展，图像修复也可以看作是一种局部图像生成技术[14]，不仅可以基于原来的图像做修复生成新的图像，还可以根据新的提示生成完全不同的图像数据。

Stable diffusion[15]是一个基于潜在扩散模型的文图生成模型，相比于普通扩散模型，潜在扩散模型先

将图像进行压缩，生成一个潜在空间特征，然后在潜在空间中进行迭代去噪，最后将新的表示结果解码为一个完整的图像，该过程如图3所示。假如输入图像是分辨率为512×512×3的彩色图像，其像素空间大小为786 432，如果进行8倍下采样，同时生成4维潜在空间，则特征维度大小为16 384，可以看出图像像素空间是潜在空间维度的48倍。由于潜在空间维度远远小于图像像素空间，在潜在空间中做图像生成耗时相对较少，这也使得在单个消费级GPU上用生成模型做图文生成成为现实。同时，潜在空间的生成结果要经过解码网络恢复到图像像素空间，该过程为上采样学习过程，因此该方法同时兼顾了高分辨率图像的生成和耗时问题。

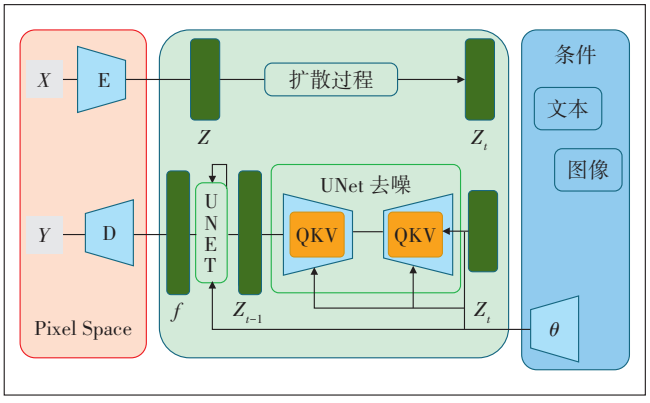

图3　Stable diffusion结构

图3描述了整个模型学习及生成的过程。在训练过程中，给定输入图像X，经过图中模块E进行编码，提取潜在空间特征Z，然后进行加噪声的扩散过程得到噪声图Z_t；接着进行去噪过程，单次去噪过程由UNet[16]结构注意力模型及输入提示θ构成，其中QKV模块代表注意力机制；一次去噪后得到Z_{t-1}，重复迭代去噪过程，最终得到生成的潜在空间特征f，然后经过图像解码器D得到与原始分辨率一致的新输出图像Y。在推理生成阶段，即只给定文本生成图像的过程中，系统首先随机生成潜在空间的噪声图，然后经过上述描述的去噪过程和图像解码D过程，最终得到新图像Y。

在具体使用中，生成模型也可以完成图像修复任务。此外，生成模型可以在已有的掩码上生成新的指定类别目标，进而生成包含不同正样本的图片，也可以保留已有掩码对应的目标，更换背景，生成包含相同正样本但背景不同的样本图片。

3 实验效果

考虑到车牌为隐私数据，因此只展示经过马赛克处理和分割及自动修复后的图像，其中图4(a)为车牌经过马赛克处理后效果图，图4(b)为车牌经过SAM自动分割及修复后的效果图。从图4中可以看出，自动修复后的图像更加平滑，更符合视觉观感。

(a) 马赛克　　(b) inpainting修复

图4　车牌马赛克与消除对比

以动物数据为例，图5(a)为原图，图5(b)为点选输入时，SAM分割得到的掩码在原图中的显示，图5(c)为采用Lama模型去除SAM掩码生成的新图像，图5(d)再次使用SAM分割得到图5(c)中的目标，并更换新的背景图，生成新的图像。

4 结束语

针对隐私数据及少样本数据在深度学习中的应用，本文提出一种基于图像分割及图像修复技术的数据生成架构。与传统的马赛克处理相比，图像修复在消除敏感隐私信息的同时，尽可能保留了图像的原始结构，更好地保持了数据的一致性；而任意物体分割与图像生成模型的融合则会生成新的多样性场景数据，实现少样本数据扩充。实验证明了该数据生成方法的可行性，为隐私数据应用及样本扩充提供了一种新的思路。

参考文献：

[1] 田萱，王亮，丁琪. 基于深度学习的图像语义分割方法综述[J]. 软件学报，2019，30(2)：440-468.
[2] 强振平，何丽波，陈旭，等. 深度学习图像修复方法综述[J]. 中国图像图形学报，2019，24(3)：447-463.
[3] 杨坚伟，严群，姚剑敏，等. 基于深度神经网络的移动端人像分割[J]. 计算机应用，2020，40(12)：3644-3650.
[4] 王柯俨，胡妍，王怀，等. 结合天空分割和超像素级暗通道的图像去雾算法[J]. 吉林大学学报(工学版)，2019，49(4)：1377-1384.

图5 图像修复与背景更换生成效果

[5] XU N, PRICE B, COHEN S, et al. Deep interactive object selection [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Las Vegas, NV, USA: IEEE, 2016: 373-381.

[6] KIRILLOV A, MINTUN E, RAVI N, et al. Segment anything[DB/OL]. [2023-04-09]. https://arxiv.org/abs/2304.02643.

[7] BARNES C, SHECHTMAN E, FINKELSTEIN A, et al. PatchMatch: a randomized correspondence algorithm for structural image editing [J]. ACM Transactions on Graphics, 2009, 28(3): 1-11.

[8] CRIMINISI A, PEREZ P, TOYAMA K. Region filling and object removal by exemplar-based image inpainting [J]. IEEE Transactions on Image Processing, 2004, 13(9): 1200-1212.

[9] DONG Q L, CAO C J, FU Y W. Incremental transformer structure enhanced image inpainting with masking positional encoding[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE, 2022: 11348-11358.

[10] LI W B, LIN Z, ZHOU K, et al. MAT: mask-aware transformer for large hole image inpainting[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE, 2022: 10748-10758.

[11] SUVOROV R, LOGACHEVA E, MASHIKHIN A, et al. Resolution-robust large mask inpainting with fourier convolutions [C]//2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa, HI, USA: IEEE, 2022: 3172-3182.

[12] CHI L, JIANG B R, MU Y D. Fast fourier convolution[C]//34th Conference on Neural Information Processing Systems (NeurIPS 2020). Online: Curran Associates, Inc., 2020: 4479-4488.

[13] JOHNSON J, ALAHI A, FEI-FEI L. Perceptual losses for real-time style transfer and super-resolution [C]//Computer Vision - ECCV 2016. Cham: Springer, 2016: 694-711.

[14] LUGMAYR A, DANELLJAN M, ROMERO A, et al. RePaint: inpainting using denoising diffusion probabilistic models [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE, 2022: 11451-11461.

[15] ROMBACH R, BLATTMANN A, LORENZ D, et al. High-resolution image synthesis with latent diffusion models [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE, 2022: 10674-10685.

[16] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation [C]//Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015. Cham: Springer, 2015: 234-241.

作者简介：

苏海军，硕士，主要从事计算机视觉算法研发工作；薛任谦谦，学士，主要研究方向为计算机视觉。