

文章编号: 2095-2163(2022)11-0087-05

中图分类号: TP391

文献标志码: A

语义兴趣点检测算法在图像分割中的研究

陆宏菊

(济南市技师学院 信息工程学院, 济南 250115)

摘要: 针对不同类型的图像,不同模态的特征能够对图像分割提供不同的帮助。本文提出一种基于无监督学习策略的多模态特征映射模型,通过对多模态特征在解空间的投影将图像分割问题转化为最小化像素相似度距离的优化问题。本文提出算法通过无监督学习策略得到特征映射矩阵,实现将高维度的多模态特征进行投影得到低维的特征解平面,用来进行图像的前景-背景分割。最后,在 BSD500 数据集中进行了对比,验证了本文算法的有效性和先进性。

关键词: 图像自动分割; 多模态特征; 特征映射

Research on semantic interest points detection algorithm in images segmentation

LU Hongju

(School of Information Engineering, Jinan Technician College, Jinan 250115, China)

[Abstract] For different types of images, the features of different modalities can provide different help for images segmentation. In this paper, a multimodal feature mapping model based on the unsupervised learning strategy is proposed, which transforms the image segmentation problem into an optimization problem of minimizing the pixel similarity distance by projecting the multimodal features into the solution space. The algorithm in this paper obtains the feature mapping matrix through the unsupervised learning strategy, and realizes the projection of the high-dimensional multimodal features to obtain the low-dimensional feature solution plane for the foreground-background segmentation of the image. Finally, a comparison is made in the BSD500 dataset to verify the effectiveness and advancement of the algorithm proposed in this paper.

[Key words] automatic images segmentation; multimodal features; feature mapping

0 引言

图像分割是计算机视觉中最基础的研究问题之一,其目的是根据相关性原则将图像分成若干互不重叠的区域,并将感兴趣的区域与背景标记出来^[1]。图像分割算法中,交互式分割研究相对比较成熟,包括基于边界的和基于区域的方法^[2]。其中,基于边界的方法通过用户获得边界信息,例如 Snake 模型^[3]、GAC 模型^[4]和 LSM 模型^[5]。基于区域的方法预定义或交互标记不同的目标和背景,通过相似性计算完成分割,如 Intelligent Paint^[6]和 Marker Drawing^[7]。为了降低用户干预,Dambreville 等人^[8]提出 KPCA 算法在 GAC 中引入形状先验模型。Liao 等人^[2]提出结合显著性特征和星型形状先验进行前景特征点定位。尽管已经获得较好的分割结果,交互式分割算法仍存在时间长、操作繁琐和结果易受标记点位置影响的问题。

在上述方法中,局部特征相似性计算在像素级分

割算法中起着关键的作用,但也随即带来了较高的计算代价。对于某些参数依赖和模型依赖的算法,数据标注化成本也很高,且只适用于特定种类和指定的数据集。针对这些问题,本文提出一种无监督学习的多模态特征语义分割模型。受 FCN 模型的启发,多种像素级特征(颜色、显著性、空间、深度等)联合进行图像表示能够获得更加完整的图像描述。针对不同类型的图像,多种特征在图像分割解空间的贡献度不同。本文提出特征选择矩阵 S , 能够将多模态特征在解空间进行映射,从而获得低维、平滑的解平面用于图像分割,通过引入语义检索模型简化求解难度。

1 本文方法

1.1 多模态图像特征描述

目前,出现了众多针对图像特征的描述子,例如显著性特征、景深特征、空间特征和颜色特征等。这些特征针对图像的不同侧面进行了描述和表示。在图像分割中,采用显著性特征表示对于包含有明显

基金项目: 国家自然科学基金(61902225)。

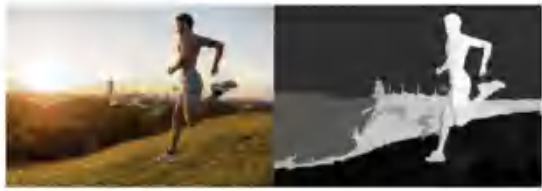
作者简介: 陆宏菊(1982-),女,硕士,高级讲师、多媒体二级技师、高级电子商务师,主要研究方向:图像识别、图像处理、人工智能等。

通讯作者: 陆宏菊 Email: 112578687@qq.com

收稿日期: 2022-03-13

主题物体的图像、采用景深特征表示对于具有距离深度的图像非常有效。每一种特征单独表示能对于相关类型图像具有良好的分割效果。本文提出假设:采用多模态特征融合的方法进行图像分割,相对于单一特征描述,分割效果则会更佳。因此,本文提出基于多模态特征的无监督学习算法进行图像前景和背景的分割。

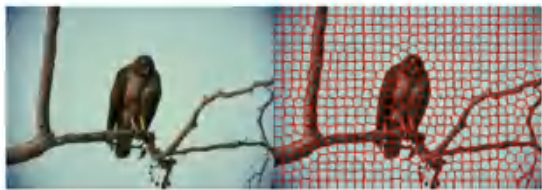
特征合并是多特征学习算法常用的策略,即将不同模态的特征通过合并(多维特征组成一维特征)进行分割训练。这种方法简单易行,但是未考虑不同模态特征对于最终分割问题解平面的贡献度不同。比如说在空旷风景照中景深特征会对于区分前景和背景贡献度较大;对于人的肖像照,显著性特征和颜色特征贡献度较大。图像及其多模态特征表示如图1所示。因此本文提出通过学习策略获得特征选择矩阵 S 。针对不同类型的图像,矩阵 S 能够合理地将不同模态特征在分割解空间内进行投影,使得贡献度大的特征能够在投影中占据主导。



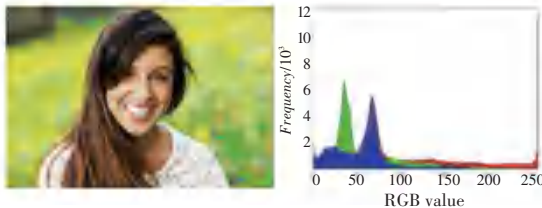
(a) 显著性特征



(b) 景深特征



(c) 空间特征



(d) 颜色直方图

图1 图像及其多模态特征表示

1.2 多模态特征映射

在规模为 n 的邻域内,像素多模态特征 Ω 可以表示为所有特征组成的高维向量, $\Omega = \{Cr_{i,j}, Cg_{i,j}, Cb_{i,j}, X_{position}, Y_{position}, f_{texture}, f_{lightness}, f_{saliency}, \dots\}$ 。每个像素的特征向量表示为 n 邻域范围内所有像素的特征集合。本文中提出的“语义”与 FCN 中的语义有所不同,本文中主要是指在高维特征空间内,针对目标函数的特征的投影,也就是本文提出的特征语义。因此,可以将分割问题表示为:

$$\begin{aligned} \operatorname{argmin}_{\Omega} \sum_{n=1}^k \|\Omega_{pi} - \Omega_{pj}\|^2 \omega_{i,j} = \\ \operatorname{argmin} \operatorname{tr} \left(\sum_{n=1}^k \Omega_{pi} L_n \Omega_{pj}^T \right) \end{aligned} \quad (1)$$

其中, $\operatorname{tr}(\cdot)$ 为迹运算。经过化简, $L \in \mathbb{R}^{(n+1) \times (n+1)}$ 体现了在 n 维空间内的相似度量目标函数。对于 L 的估算仍旧很难进行,经过分析发现,虽然不同的特征对于不同类型图像的贡献度不同,但是可以认为其贡献度皆为非负的。本文研究提出存在一个潜在的特征选择矩阵 S ,通过 S 可以将多模态特征进行全局映射,即: $\Omega_{pi} = \Omega S_n$ 。

因此,公式(1)可以进一步化简为:

$$\begin{aligned} \operatorname{argmin}_{\Omega} \sum_{n=1}^k \operatorname{tr}(\Omega S_n L_n S_n^T \Omega^T) = \\ \operatorname{argmin}_{\Omega} \sum_{n=1}^k \operatorname{tr}(\Omega L'_n \Omega^T), L'_n = S_n L_n S_n^T \end{aligned} \quad (2)$$

将公式(1)带入到公式(2)中可以得到:

$$L_n = D_n - W_n \quad (3)$$

其中, $W_n \in \mathbb{R}^{n \times n}$;当像素 i 和像素 j 非常相似的时候, $(W_n)_{i,j} = \exp\left(-\frac{\|x_i^n - x_j^n\|^2}{\sigma}\right)$;当像素 i 和 j 非常不相似时, $(W_n)_{i,j} = 0$; D_n 为对角线形矩阵, $D_n = \sum_n [W_{i,j}^n]$ 。因此, L_n 可以看作是一个 Normalized Laplacian matrix^[9]。

本文采用文献[9]中的方法对式(3)进行求解。学习出的模型 $(\Phi_k(x), \lambda_k)$ 可以用来将高维的特征矩阵 Ω 映射到由用户指定相似距离阈值 θ 的解空间上。

1.3 基于多模态特征映射的图像分割算法

经过式(3)可以针对目标图像学习出其相应的特征模式(modes),并寻找到相似度最小的像素集,如图2所示。

针对目标兴趣点集完成自动分割,首先利用检测出的兴趣点产生种子区域。通过图像的颜色距离 d_c 和位置距离 d_s ^[10],将与兴趣点像素“同质”的像素计算出来。因此,本文提出的算法代码的设计表

述如下。



(a) 目标图像 (b) 目标兴趣点集

图 2 目标兴趣点集

Fig. 2 Object interesting pixels set

输入: 图像 I , 特征 n , 距离阈值 θ

输出: 分割结果

- 1: 提取特征矩阵 $\Omega_{i,j}^n$;
- 2: 基于 $\Omega_{i,j}^n$, 通过式(3)计算 L_n ;
- 3: [特征向量, 特征值] $\leftarrow L_n$;
- 4: 训练学习出特征模式, 从而得到兴趣点;
- 5: 通过式(3)和距离阈值, 根据兴趣点得到兴趣区域;
- 6: 根据兴趣区域获得分割结果。

2 实验结果与分析

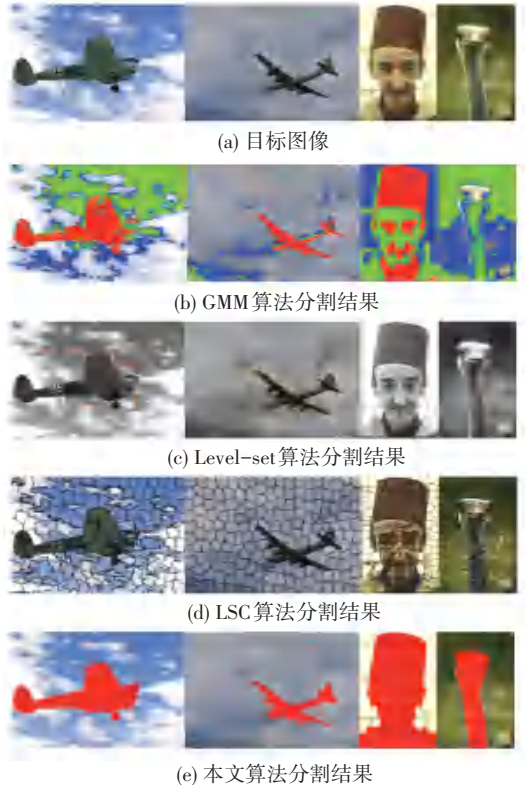
2.1 实验内容

本文实验环境为 Windows 10, I9CPU, 32GB 内存, 英伟达 GTX 1080Ti 11G。实验选取图像的显著性^[11]、Focal 长度^[12]、空间特征^[13]和颜色直方图^[14]四种不同模态的图像特征进行对比实验。实验选用的是 BSD 数据集。与 3 种最新方法, 即高斯混合模型 (GMM)^[15]、Level-set 算法^[16]和 LSC 超像素分割算法^[17]进行定性比较。

2.2 实验结果定性分析

本文进行了分割结果的可视化定性比较, 对比结果如图 3 所示。图 3(a)是从 BSD500 数据集选定的待分割图像。图 3(b)受复杂背景图像的影响, 有部分区域被误分类为前景区域, 对于边界保持也并不令人满意。图 3(c)受初始位置的选择影响, 当初始位置选择合理时, 分割结果较好, 例如“鸵鸟”图像; 反之则不然, 例如“飞机”图像。图 3(d)为 LSC 超像素算法的结果, 在前景物体分割上出现错误。图 3(e)为本文算法, 采用了多模态的特征, 对于 2 幅“飞机”图像, 景深特征的作用可以确保前景边界信息的完整。“老人”图像中, 显著性特征能够有效分割出老人与椅子背景区域。本节实验呈现结果并不是在相同模式下, 所有代码均来自于作者本人自主研发。经过视觉对比可以明显看出, 本文方法在

场景图像、人物图像、动物图像等不同类型的图像前景背景分割中均获得了理想的结果。



(e) 本文算法分割结果

图 3 定性对比

Fig. 3 The qualitative comparison

2.3 实验结果定量分析

在定量对比方面, 本文主要设置 2 组实验: 采用参数 PRI 、 GCE 和 VOI 指标的对比; 采用 $F - measure$ 与 BSD500 数据集中的标定数据进行对比。对此拟展开探讨分述如下。

(1) 采用 PRI 、 GCE 、 VOI 指标的对比。在 BSD500 图库中抽取 30 幅图像进行定量分析对比实验。采用 PRI ^[18] (probabilistic rand index)、 GCE ^[19] (Global Consistency Error) 和 VOI (Variation of Information) 评价指标进行评估^[20]。其中, PRI 计算分割结果与真实标记相一致的像素数的比例, 其值越大、分割结果与参考值间的属性共生一致性也就越好。 VOI 是信息差异指标, 计算像素点分割产生的信息熵的变化程度。在概率论和信息论中, 信息或共享信息距离的变化是对 2 个聚类 (元素分区) 间距离的度量, 且与相互信息密切相关; 实际上, 这就是一个包含相互信息的简单线性表达式, 其值越小越好。 GCE 衡量分割结果之间互相包括的概率, 其值越小越好。本次研究中, 计算 50 次的平均值作为最终结果, 见表 1。由表 1 中可以看出, 本文算法在 3 项指标上皆高于文中选择的对比方法。

表1 PRI、GCE 和 VOI 的平均值

Tab. 1 The average value of PRI, GCE and VOI

指标	GMM	Level-set	LSC	The proposed
PRI	0.853 5	0.853 5	0.939 0	0.976 0
GCE	0.156 7	0.176 4	0.224 3	0.014 4
VOI	0.146 4	0.137 5	0.554 6	0.133 3

(2) 采用 $F - measure$ 指标的对比。采用 $F - measure$ 衡量本文分割方法与 BSD500 中提供的 6 个人工标注数据之间的性能比较, 数学定义式具体如下:

$$F_{\beta} = \frac{(1 + \beta^2) Precision \times Recall}{\beta^2 Precision + Recall} \quad (4)$$

其中, $\beta^2 = 0.3^{[21]}$ 。

与 BSD 标定数据性能对比结果见表 2。在表 2 中, 本文从 BSD500 数据集中选取 20 个图像, 求其平均准确率、召回率和 $F - measure$ 数值。通过实验可以看出, 这些指标大多超过 95%。实验结果表明本文方法的效果接近 BSD500 所提供的 6 个人工标注分割结果。

表 2 与 BSD 标定数据性能对比

Tab. 2 The quantitative comparison with BSD labels

Bench	precision	recall	$F - measure$
Benchmark1	0.985 5	0.999 9	0.982 7
Benchmark2	0.975 9	0.999 9	0.982 8
Benchmark3	0.967 7	0.999 5	0.983 5
Benchmark4	0.989 5	0.999 5	0.991 1
Benchmark5	0.972 2	0.998 9	0.923 8
Benchmark6	0.976 4	0.999 8	0.988 0

2.4 与 FCN 模型的分割对比分析

深度学习模型是当下研究热点, 本文思路也是受深度模型多层次、多特征加工启发产生, 因此, 本文针对目前广泛应用的 FCN^[22] 进行对比实验。FCN 通过梯度累积、正则化 $loss$ 函数和标准化动量方式进行训练, 比较结果如图 4 所示。本次研究中, 选取了 FCN8s (2 stream, 8 pixel prediction)、FCN16s (2 stream, 16 pixel prediction) 和 FCN32s (1 stream, 32 pixel prediction) 模型进行对比实验。

由图 4 可知, FCN8s 模型效果明显好于其他 2 个。但是由于 FCN 模型的主要目的是进行语义分割 (Semantic segmentation), 即将具有相同语义的物体进行分割和标注。因此, 在对于边缘分割准确率方面也并不精确。在这一方面, 本文算法的结果较 FCN 更加准确。

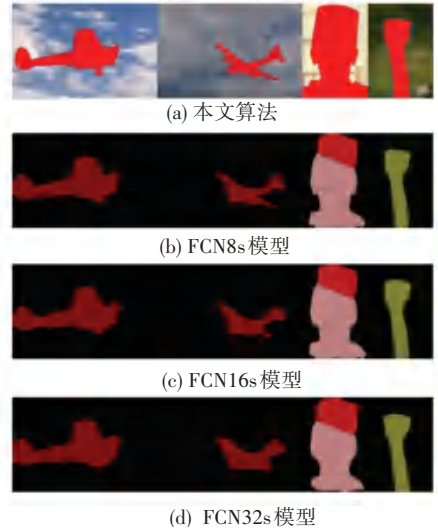


图 4 FCN 模型分割对比

Fig. 4 FCN models segmentation comparison

在 1.3 节提到的本文算法中, 相似度距离阈值 θ 为超参, 表示相似度距离, 也就是当相似度距离小于 θ 就表示 2 个像素同质, 否则表示 2 个像素异质。

参数分析结果如图 5 所示。图 5 中, 图 5(a) ~ (c) 中的各子图从左至右的 θ 取值依次为 8, 9, 10, 11, 12。从图 5 中可以看出, 当 θ 取值趋向于 0, 检测到的兴趣点像素就非常多, 但包含大量的噪声, 从而出现过分割。当 θ 取值很大时, 相似度距离会比较严格, 检测出的兴趣点会减少, 从而出现欠分割现象。总之, 合适的相似度距离阈值 θ 对于结果是非常重要的, 本文中选取的 θ 在 [9, 11] 之间, 一般情况下 $\theta = 10$ 。

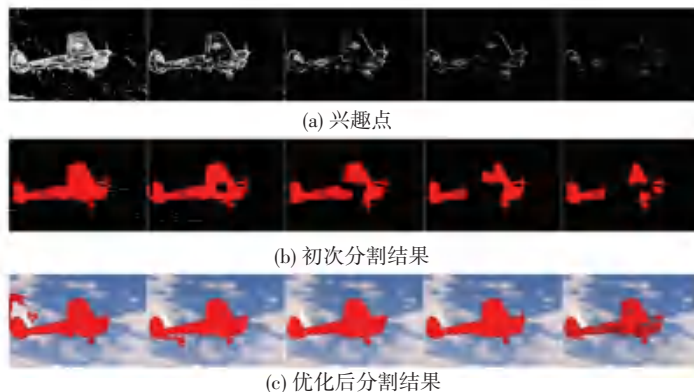


图 5 参数分析结果

Fig. 5 The parameters analysis results

3 结束语

本文提出一种基于无监督的多模态特征映射策略,将传统的图像分割问题转化为在全局范围内寻找相似距离最小的最优化问题。通过引入特征选择矩阵和语义哈希模型,简化了计算复杂度,降低了计算代价,实现了对不同种类图像的前景-背景分割。通过与其他无监督算法和深度学习模型进行对比,验证了本文方法的可行性。

本文方法中还有很多需要改进的地方。首先,虽然多模态特征映射采用的是无监督学习策略,但相似度距离阈值 θ 是由用户指定的。如何能够在学习过程中将阈值参数加入到目标函数中是未来工作亟待解决的问题。通过引入显著性、景深、空间和颜色特征,本文方法展现出对于多类型图像分割的鲁棒性。在当前框架下,如何引入新的特征能够适应更加泛化的图像类型将成为本文后续工作的重点。

参考文献

- [1] FARMER M E, JAIN A K. A wrapper-based approach to image segmentation and classification [J]. IEEE Transactions on Image Processing, 2005, 14(12): 2060-2072.
- [2] LIAO Xiangli, XU Hongbo, ZHOU Yicong, et al. Automatic image segmentation using salient key point extraction and star shape prior [J]. Signal Processing, 2014, 105(12): 122-136.
- [3] PATIL B M, BURKPALLI V. Segmentation of cotton leaf images using a modified chan vese method [J]. Multimedia Tools and Applications, 2022, 81: 15419-15437.
- [4] CHANG Y T, SHIH T K, LI Y H, et al. Effectiveness evaluation of iris segmentation by using geodesic active contour (GAC) [J]. The Journal of Supercomputing, 2021, 76(3): 1628-1641.
- [5] LIU Cheng, LIU Weibin, XING Weiwei. An improved edge-based level set method combining local regional fitting information for noisy image segmentation [J]. Signal Processing, 2017, 130: 12-21.
- [6] REESE L J, BARRETT W A. Image editing with intelligent paint [C]// the Proceedings of Eurographics. Aire-la-Ville: Eurographics Association, 2002: 714-724.
- [7] FALCO A X, STOLFI J, DE ALENCAR LOTUFO R. The image foresting transform: Theory, algorithms, and applications [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(1): 19-29.
- [8] DAMBREVILLE S, RATHI Y, TANNENBAUM A. A shape-based approach to robust image segmentation [M]//CAMPILHO A, KAMEL M S. Image analysis and recognition. ICIAR 2006. Lecture Notes in Computer Science. Berlin/Heidelberg: Springer, 2006, 4041: 173-183.
- [9] WEISS Y, TORRALBA A, FERGUS R. Spectral hashing [C]// Proceedings of International Conference on Neural Information Processing Systems. NY, USA: Curran Associates Inc., 2008: 1753-1760.
- [10] ACHANTA R, SHAJI A, SMITH K, et al. SLIC superpixels compared to state-of-the-art superpixel methods [J]. IEEE transactions on pattern analysis and machine intelligence, 2012, 34(11): 2274-82.
- [11] YAN Qiong, LI Xu, SHI Jianping, et al. Hierarchical saliency detection [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2013: 1155-1162.
- [12] SHI Jianping, LI Xu, JIA Jiaya. Just noticeable defocus blur detection and estimation [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2015: 657-665.
- [13] ACHANTA R, SHAJI A, SMITH K, et al. SLIC superpixels compared to state-of-the-art superpixel methods [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2012: 2274-2282.
- [14] HAFNER J, SAWHNEY H S, EQUITZ W, et al. Efficient color histogram indexing for quadratic form distance functions [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 1995: 729-736.
- [15] RAGOTHAMAN S, NARASIMHAN S, BASAVARAJ M G, et al. Unsupervised segmentation of cervical cell images using gaussian mixture model [C]// Proceedings of International Conference on Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 1347-1376.
- [16] WU Yongfei, HE Chuanjiang. A convex variational level set model for image segmentation [J]. Signal Processing, 2015, 106(C): 123-133.
- [17] CHEN Jiansheng, LI Zhengqin, HUANG Bo. Linear spectral clustering superpixel [J]. IEEE Transactions on Image Processing, 2017, 26(7): 3317-3330.
- [18] UNNIKRISSNAN R, PANTOFARU C, HEBERT M. Toward objective evaluation of image segmentation algorithms [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6): 929-944.
- [19] MARTIN D, FOWLKES C, TAL D, et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics [C]// Proceedings of IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2002: 416-423.
- [20] HANBURY A, STOTTINGER J. On segmentation evaluation metrics and region counts [C]// Proceedings of International Conference on Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2012: 1-4.
- [21] CHENG Mingming, MITRA N J, HUANG Xiaolei, et al. Global contrast based salient region detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3): 569-582.
- [22] ZHAO Lixin. Image semantic segmentation method based on GAN network and FCN model [J]. The Journal of Engineering, 2022(1): 1-9.