

引文格式:李雪,张力,王庆栋,等.多时相遥感影像语义分割色彩一致性对抗网络[J].测绘学报,2020,49(11):1473-1484. DOI:10.11947/j.AGCS.2020.20190439.
LI Xue,ZHANG Li,WANG Qingdong,et al.Multi-temporal remote sensing imagery semantic segmentation color consistency adversarial network[J]. Acta Geodaetica et Cartographica Sinica,2020,49(11):1473-1484. DOI:10.11947/j.AGCS.2020.20190439.

多时相遥感影像语义分割色彩一致性对抗网络

李雪,张力,王庆栋,艾海滨

中国测绘科学研究院,北京 100830

Multi-temporal remote sensing imagery semantic segmentation color consistency adversarial network

LI Xue,ZHANG Li,WANG Qingdong,AI Haibin

Chinese Academy of Surveying and Mapping, Beijing 100830, China

Abstract: Using deep convolutional neural network (CNN) to intelligently extract buildings from remote sensing images is of great significance for digital city construction, disaster detection and land management. The color difference between multi-temporal remote sensing images will lead to the decrease of generalization ability of building semantic segmentation model. In view of this, this paper proposes the attention-guided color consistency adversarial network (ACGAN). The algorithm takes the reference color style images and the images to be corrected in the same area and different phases as the training set and adopts the consistency adversarial network with the U-shaped attention mechanism to train the color consistency model. In the prediction stage, this model converts the hue of the images to that of the reference color style image, which is based on the reasoning ability of the deep learning model, instead of the corresponding reference color style image. This model transforms the hue of the images to be corrected into that of the reference color style images. This stage is based on the reasoning ability of the deep learning model, and the corresponding reference color style image is no longer needed. In order to verify the effectiveness of the algorithm, firstly, we compare the algorithm of this paper with the traditional image processing algorithm and other consistency adversarial network. The results show that the images after ACGAN color consistency processing are more similar to that of the reference color style images. Secondly, we carried out the building semantic segmentation experiment on the images processed by the above different color consistency algorithms, which proved that the method in this paper is more conducive to the improvement of the generalization ability of multi-temporal remote sensing image semantic segmentation model.

Key words: multi-temporal remote sensing imagery; color consistency; generative adversarial networks; semantic segmentation; attention mechanism

Foundation support: The National Key Research and Development Project (No. 2019YFB1405600); The Basic Scientific Research Project of Chinese Academy of Surveying and Mapping (No. AR1902)

摘要:利用深度卷积神经网络智能化地提取遥感图像中的建筑物对于数字城市构建、灾害侦查、土地管理等具有重要意义。多时相遥感图像之间的色彩差异会导致建筑物语义分割模型泛化能力下降。针对此,本文提出了注意力引导的色彩一致生成对抗网络(attention-guided color consistency adversarial network,ACGAN)。该算法以参考色彩风格图像及相同区域、不同时相的待纠正图像作为训练集,采用加入了U型注意力机制的循环一致生成对抗网络训练得到色彩一致模型。在预测阶段,该模型将待纠正图像的色调转换为参考色彩风格图像的色调,这一阶段基于深度学习模型的推理能力,而不再需要相应的参考色彩风格图像。为了验证算法的有效性,首先,将本文算法与传统的图像处理算法及其他循环一致生成对抗网络做了对比试验。结果表明,ACGAN色彩一致后的图像与参考色彩风

格图像的色调更加相似。其次,将以上不同的色彩一致性算法处理后的结果图像进行建筑物语义分割试验,证明本文方法更加有利于多时相遥感图像语义分割模型泛化能力的提升。

关键词:多时相遥感图像;色彩一致性;生成对抗网络;注意力机制;语义分割

中图分类号:P237

文献标识码:A

文章编号:1001-1595(2020)11-1473-12

基金项目:国家重点研发计划(2019YFB1405600);中国测绘科学研究院基本科研业务项目(AR1902)

利用卷积神经网络(convolutional neural network, CNN)对遥感图像中的建筑物进行智能化地语义分割^[1-2],对于数字城市构建、灾害侦查、土地管理等具有重要意义。基于深度学习的语义分割模型通过卷积操作提取图像局部区域的特征,周围像素的变化会对中心像素造成一定的影响。因此,多时相的遥感图像中植被、水体等显著的色彩差异以及建筑物本身的色彩不一致等会导致建筑物语义分割精度下降,这会给遥感图像语义分割模型的泛化能力带来消极影响。具体来说,若存在有建筑物语义分割标签的遥感数据集 A,称之为源域,与 A 区域相同、时相不同、无语义分割标签的遥感数据集 B,称之为目标域。本文用源域图像训练的建筑物语义分割权重直接预测目标域图像,由于两个域的色彩分布差异,其分割精度会下降。若目标域的数据重新制作标签进行训练,则需要耗费大量的人力与时间,效率低、成本高。

针对以上问题,可以提高源域 A 建筑物语义分割模型的泛化能力,使其能够直接迁移到目标域 B 使用。通过拼接、翻转等方式^[3-5]模拟源域数据来扩充目标域数据多样性的手段获得的数据仍然与源域保持较高的相关性,对源域语义分割模型泛化能力的提升帮助有限。源域与目标域之间的色彩一致性可作为另一种途径,前者为参考色彩风格图像,后者为待纠正图像,缩小两者的色彩分布差异。传统的色彩一致性算法主要分为参数化方法和非参数化方法。参数化方法通过建立参数回归模型恢复图像间的色彩关系。基于伪不变特征点(pseudo-invariant features, PIF)^[6-7]的参数化方法不适用于没有重叠区域的图像。基于图像统计信息的参数化方法包括 Wallis 滤波算法^[8]、直方图匹配(histogram matching, HM)算法^[9]以及线性 Monge-Kantorovitch(linear monge-kantorovitch linear colour mapping, MKL)算法^[10]等,前两种应用较广泛但是效果不稳定,后一种建立了较为通用的线性变换基础但是实现较复杂,效率较低。非参数化方法是根据一定的规则如相似度量^[11]、全局凸问题^[12]、迭代关系^[13]等设计

图像之间的颜色映射模型。以上传统算法虽取得可观的效果,但其实现均是基于图像的底层低级特征,当图像覆盖范围较广、颜色和纹理较为复杂时,色彩一致结果较容易出现色偏现象。

基于深度学习的风格迁移技术^[14]可作为图像色彩一致性处理的新兴手段。与传统算法相比,卷积神经网络可以层次化地提取图像的低级与高级特征,分离并独立处理图像的内容与风格抽象特征^[15],更好地对图像进行深层表征,有利于不同时相遥感图像间的色彩映射。并且,上述诸多传统算法均需要底图作为参考图像,但对于风格迁移技术,参考色彩风格图像只在训练阶段需要,在模型预测阶段,将只包含待纠正图像的预测集输入风格迁移模型即可完成色彩一致性转换,一定程度上避免了传统算法依赖参考底图的限制,应用场景更加广泛。风格迁移技术主要分为两类,一类是基于图片迭代的描述性方法^[15-19],这类方法的结果图像稳健性较好,但是计算时间较长,过于依赖预训练模型。另一类是基于模型迭代的生成式神经网络方法^[20],它比描述性方法速度更快,应用更广泛。循环一致生成对抗网络(cycle-consistent adversarial networks, CycleGAN)^[21]是其中的代表性网络之一,它通过生成器与鉴别器的博弈将源域色彩迁移至目标域,并利用循环一致损失函数约束两者的内容保持一致。但是, CycleGAN 属于全局风格迁移算法,图像局部和细节容易失真。一些学者通过在 CycleGAN 中加入注意力机制(attention mechanism, AM)^[22]从而将其优化为局部描述算法。其中, U-GAT-IT^[23]利用类激活图(class activation map, CAM)找出对类别判断更重要的位置,对此分配更多的关注度;GANimation^[24]算法在生成器中嵌入注意力机制,确保经过变换处理的图像与原始图像之间平滑过渡;注意力引导的生成对抗网络(attention-guided adversarial networks, AGGAN)^[25]的注意力机制独立于生成器,由串联的卷积层及 ResNet 网络^[26]构成,以局部之间的差异化等级为标准生成注意力图,为图

像的不同位置分配不同的权重。以上算法中的注意力机制均是引导生成器关注图像中对于结果影响更大的局部位置,其结构一般较为复杂,不易控制。并且上述网络使用的数据集为人脸、动物等纹理、颜色、场景较为单一的照片,将上述网络应用于颜色、纹理、场景复杂的遥感图像,迁移后的目标域图像与源域图像色彩差异较大,不利于多时相遥感图像建筑物语义分割模型泛化能力的提升。

针对上述问题,本文设计了更加稳健的 U 型注意力机制,并将其加入 CycleGAN 构成了注意力引导的色彩一致生成对抗网络 (attention-guided color consistency adversarial network, ACGAN)。该网络不仅能提取遥感图像的高级特征,而且能通过 U 型注意力机制聚合多尺度的语义信息,减少有效低级特征的损失,更好地缩小源域与目标域之间的色彩分布差异。将该网络色彩一致性处理后的目标域图像直接用源域建筑物

语义分割模型进行预测,分割精度较对比算法有明显的提升,有助于源域建筑物语义分割模型泛化能力的提升。

1 方法

1.1 注意力引导的色彩一致生成对抗网络 (ACGAN)

在遥感应用中,内容完全相同的成对图像很难获取。CycleGAN^[21]能够在没有成对训练数据的情况下,将源域图像的色彩迁移到目标域,而图像的内容保持不变。其由成对的生成器(generator, G)与鉴别器(discriminator, D)构成,如图 1 所示。该算法首先通过生成器产生与源域色彩一致的目标域图像,然后利用鉴别器判断两者的相似性,再将色彩一致后的图像重新转换回原始目标域,这一过程使用循环一致损失函数约束,使得生成图片与输入图片的内容始终保持一致。

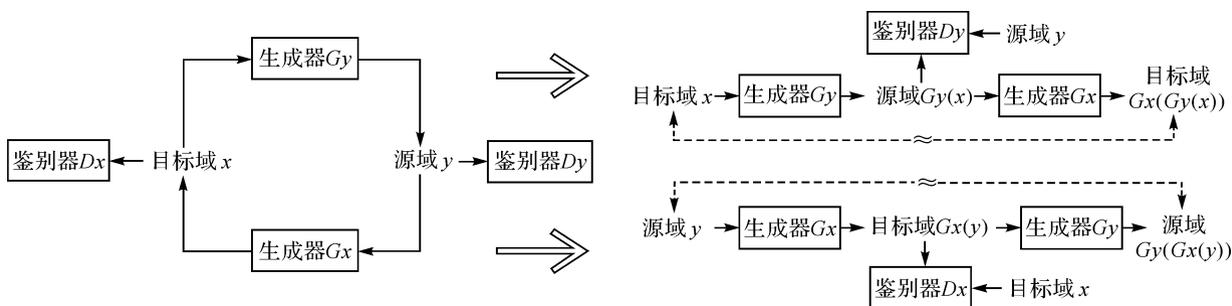


图 1 CycleGAN 网络的基本结构

Fig.1 The basic structure of CycleGAN

由于遥感图像覆盖范围较广,地物类型多样, CycleGAN 等全局色彩一致算法的处理效果不佳。一方面,遥感图像纹理、颜色及场景等均较为复杂,不同时相的遥感图像中河流、植被等地物的色彩差异较为明显,建筑物本身也存在一定的色彩不一致性,即使 CycleGAN 中加入了 PatchGAN 鉴别器^[27],网络对其所处理的图像所有位置赋予同样大的权重,在参考色彩风格图像与待纠正图像之间相似度较高的位置容易过度迁移,而它们相似性较低的区域容易迁移不足,从而导致色彩一致结果图像失真。另一方面,一组神经网络存储信息的容量和神经元的数量以及网络的复杂程度成正比^[28],遥感图像比一般照片蕴含更加丰富的信息量,所以存储遥感图像所需的神经元数量更多,这导致能充分表达遥感图像信息的神经网络参数量相较于一般网络的参数量成倍地

增加。

针对此,本文将聚合多尺度特征的 U 型注意力机制引入到 CycleGAN 网络中,构成注意力引导的色彩一致生成对抗网络 (ACGAN)。在色彩一致的过程中,ACGAN 网络利用鉴别器所认为的参考色彩风格图像与待纠正图像之间最具差异性的区域生成注意力图。其能够引导生成器在整幅图像的不同位置分配不同的关注度,在差异性大的位置分配更多的关注度,差异性小的位置分配较少的关注度,约束生成器在图像对应区域间的映射,最小化参考色彩风格图像和待纠正图像底层分布之间的差异,改善色彩一致过程中的失真与几何畸变。因此,ACGAN 网络将全局风格迁移网络 CycleGAN 优化为局部差异化的色彩一致性处理方法。此外,通过引入注意力机制,能够使网络关注更加重要的信息,减轻其对次要信

息的关注度,在不过度增加模型参数的情况下提高模型的表达能力,从而扩充神经网络的容量^[28],提高神经网络处理信息的能力。

1.2 注意力引导的色彩一致生成对抗网络(ACGAN)的基本原理

与 CycleGAN 类似,ACGAN 仍是具有成对

的生成器与鉴别器的环形结构。利用 ACGAN 网络将目标域 B 的数据分布 x 转换为源域 A 数据分布 y 的过程称为 $x \rightarrow y$ 方向的转换;将源域 A 的数据分布 y 转换为目标域 B 数据分布 x 的过程称为 $y \rightarrow x$ 方向的转换。对于 $x \rightarrow y$ 方向的转换,如图 2 所示。

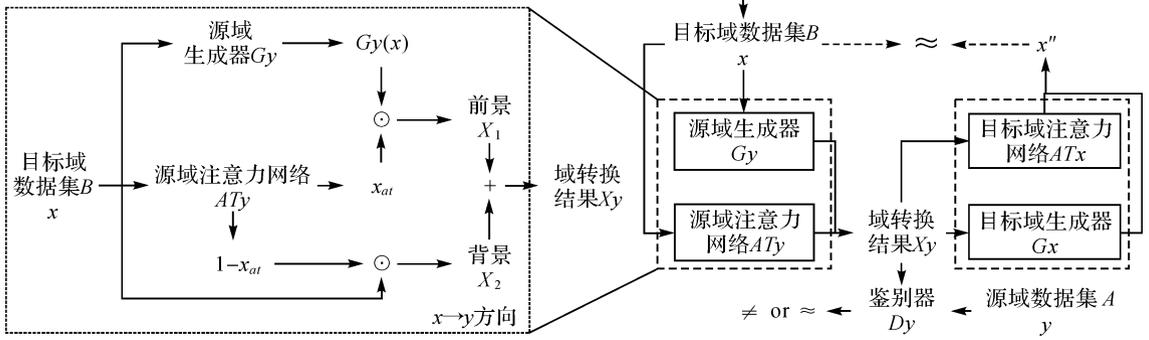


图 2 注意力引导的色彩一致生成对抗网络的基本结构

Fig.2 The basic structure of attention-guided color consistency adversarial network

第 1 阶段,通过源域生成器 G_y 将目标域分布 x 转换为与源域基本一致的数据分布 $G_y(x)$;同时,将 x 输入源域的注意力网络 AT_y ,得到注意力图 x_{at} 。 $G_y(x)$ 与 x_{at} 相乘的结果为目标域图像转换后的前景 X_1 。 x 与反注意力图 $(1-x_{at})$ 相乘的结果为目标域图像未转换的背景 X_2 。将 X_1 与 X_2 相加,即得到如下目标域到源域的转换结果

$$X_y = x_{at} \odot G_y(x) + (1-x_{at}) \odot x \quad (1)$$

式中, $x_{at} = AT_y(x)$ 。 x_{at} 的像素值取值范围为 $[0, 1]$,当像素值均等于 1 时,即注意力机制为目标域整幅图像的所有区域分配数值为 1 的权重时,ACGAN 网络退化为 CycleGAN 网络;当像素值均等于 0 时,即注意力机制为目标域整幅图像的所有区域分配数值为 0 的权重时,目标域图像并未进行色彩一致性转换,此时,若将 X_y 与源域 A 的真实数据一起输入鉴别器 D_y ,它会准确地将两者区分开。通过加入注意力机制 AT_y ,ACGAN 在 CycleGAN 的基础上构成了局部差异化的色彩一致性处理算法。并且,只有当 AT_y 利用 D_y 认为的源域与目标域之间最具差异性的区域生成注意力图 x_{at} 时, X_y 才可以“欺骗” D_y ,使其接受目标域色彩一致性转换后的结果。

上述 $x \rightarrow y$ 方向转换的对抗损失可用如下交叉熵表示

$$L_y(G_y, AT_y, D_y) = E_{y \sim A} [\log D_y(y)] +$$

$$E_{x \sim B} [\log(1 - D_y(X_y))] \quad (2)$$

式中, $E(*)$ 为分布函数的期望值; $E_{y \sim A} [\log D_y(y)]$ 为源域数据对应的损失; $\log D_y(1 - D_y(X_y))$ 为目标域到源域的转换结果 X_y 对应的损失。

第 2 阶段,为避免色彩一致过程中图像内容失真,ACGAN 利用目标域的注意力网络 AT_x 、生成器 G_x ,将 X_y 再转换回与目标域 B 一致的数据分布 x'' ,此时应有 $x'' \approx x$,两者的循环一致损失定义如下

$$L_{cyc}^y(x, x'') = \|x'' - x\|_1 \quad (3)$$

式中, $x'' = AT_x(X_y) \odot G(X_y) + (1 - AT_x(X_y)) \odot X_y$ 。

$y \rightarrow x$ 方向的转换与上述过程同理。将 $x \rightarrow y$ 、 $y \rightarrow x$ 方向的对抗损失 L_y 、 L_x 、循环一致损失 L_{cyc}^y 、 L_{cyc}^x 按一定比例相加,作为 ACGAN 网络的总损失函数

$$L(G_x, G_y, AT_y, AT_x, D_x, D_y) = L_x + L_y + \lambda(L_{cyc}^x + L_{cyc}^y) \quad (4)$$

式中, λ 为可调整的常数。

对于 $x \rightarrow y$ 方向的转换,生成器 G_y 与注意力机制 AT_y 共同产生的分布 X_y 应该尽量接近源域 A 的数据分布 y ,即最小化 X_y 与 y 之间的交叉熵。相反,鉴别器 D_y 应该尽可能区分 X_y 与 y 的不同,即最大化 X_y 与 y 之间的交叉熵。 $y \rightarrow x$ 方向的转换同理。故,ACGAN 包含的生成器、注意力机制与鉴别器之间相互博弈的过程

可表示如下

$$G_x^*, G_y^*, AT_x^*, AT_y^*, D_x^*, D_y^* = \underset{G_x, G_y, AT_x, AT_y, D_x, D_y}{\operatorname{argmin}} (\underset{D_x, D_y}{\operatorname{argmax}} L(G_x, G_y, AT_x, AT_y, D_x, D_y)) \quad (5)$$

在模型训练过程中,生成对抗网络一般通过依次优化鉴别器、生成器的方式来求得最优解。先固定生成器(G_x, G_y)和注意力机制(AT_x, AT_y)的参数,采用随机梯度下降(stochastic gradient descent, SGD)优化鉴别器(D_x, D_y);再固定鉴别器(D_x, D_y),采用 SGD 优化生成器(G_x, G_y)和注意力(AT_x, AT_y)的参数,此过程交替进行。随着训练轮次的推进,生成器、注意力机制、鉴别器三者逐渐达到纳什均衡^[29]状态。

1.3 U型注意力机制

针对纹理、颜色、场景均较为复杂的遥感图像,本文设计了一种更加稳健的 U 型注意力机制,如图 3 所示, K, H, W 分别代表输入图片的长、宽、通道数。整个 U 型注意力网络分为两部分,左侧主要负责特征提取,右侧主要负责上采样,可以将其看作是一对编码器与解码器。注意力网络以任意尺寸的图像作为输入,首先经过由卷积层构成的编码部分,第 1 卷积层,滤波器个数为 32,卷积核为 3×3 ,步长为 1,第 2 卷积层,滤波器个数为 64,卷积核尺寸为 7×7 ,步长为 2,第 3 卷积层,滤波器个数为 128,卷积核为 7×7 ,步长为 2。这部分的作用是将图像尺寸逐层缩小一倍、通道数逐层增加至 128 个,在尽可能减小计算量的基础上层次化地提取输入图像的高维特征。然后,将上述过程得到的特征图输入解码部分,其先经过第 1 次上采样,通道数不变的情况下,特征图尺寸扩大一倍;再经过第 4 卷积层(滤波器个数为 64,卷积核尺寸为 3×3 ,步长为 1)提取尺寸为 $K \times H \times 64$ 的特征图,然后利用跳跃连接(图 3 中空箭头①所示)将其与编码部分对应的特征图拼接,通道数增加一倍;将上述结果再次上采样后输入第 5 卷积层(滤波器个数为 32,卷积核尺寸为 3×3 ,步长为 1)提取尺寸为 $K \times H \times 32$ 的特征图,并利用跳跃连接(图 3 中空箭头②所示)将其与编码部分对应的特征图拼接;最后利用第 6 卷积层(滤波器个数为 1,卷积核尺寸为 3×3 ,步长为 1)得到尺寸为 $K \times H \times 1$ 的注意力图。这部分的作用除了提取图像特征外,主要是利用反卷积操作将特征图通道数逐渐收缩变小,图像

尺寸逐渐扩大,端到端地输出注意力图。

与 AGGAN^[25]等模型中串联型的注意力网络相比,本文的 U 型注意力机制中间共有 2 个跳跃连接结构。这是因为遥感图像上下文语义信息较为复杂,且低级特征对于色彩一致性处理也具有重要的意义,这个结构能够使低级特征与高级特征聚合起来,减少卷积过程中低级特征的损失,最大限度地保证注意力图中融合了有效的低级特征,有利于遥感图像多尺度的特征提取,从而避免了直接在高级语义特征上进行梯度反传。本文 U 型注意力机制中包含的两次上采样操作使生成的结果图边缘等信息更加精确。此外,ACGAN 网络的总参数量比 AGGAN 网络^[25]少了近 500 万个,学习效率更高,且 U 型结构更加稳健,对于具有丰富信息量的遥感图像来说,能够在网络参数量没有大量增加的情况下,提高色彩一致模型的表达能力。

2 试验

2.1 试验数据概况

为了更全面地分析本文算法的可靠性,此次试验选取了不同源的遥感图像,包括覆盖国外地区的多时相航空图像以及覆盖国内地区的多时相卫星图像。

航空图像从 WHU 建筑物数据集^[30]中获取,包含新西兰基督城($S43^\circ, E172^\circ$)2012 年和 2016 年变化检测图像,下文称其为数据集 1。该区域于 2011 年 2 月发生了 6.3 级地震,在接下来的年份中一直处于地物重建状态,生态环境变化较为明显。两个时相的图像地面分辨率均为 0.2 m,大小均为 $32\ 507 \times 15\ 354$ 像素,覆盖了总面积为 $20.5\ \text{km}^2$ 的区域,并且它们均有 RGB 3 个波段。本次试验将数据集 1 无重叠裁剪为 512×512 像素的图片,如图 4 所示,2016 年为源域的参考色彩风格图像,2012 年为目标域的待纠正图像,即 2016 年图像的色彩迁移到 2012 年的图像。

卫星图像来自 GF-2 卫星拍摄的武汉市($E114.30^\circ, N30.6^\circ$)2015 年和 2016 年变化检测遥感图像,覆盖了 $196.0\ \text{km}^2$ 的区域,其中包含建筑物密集的城区、河流、农田及工业区,地物类型丰富,下文称其为数据集 2。两个时相的图像地面分辨率均为 1 m,大小均为 $13\ 999 \times 14\ 000$ 像素,并且均有 RGB 3 个波段。本次试验将数据集 2 无重叠裁剪为 512×512 像素的图片,如图 5 所示,2015 年为源域的参考色彩风格图像,2016 年

为目标域的待纠正图像,即 2015 年图像的色彩迁

移到 2016 年的图像。

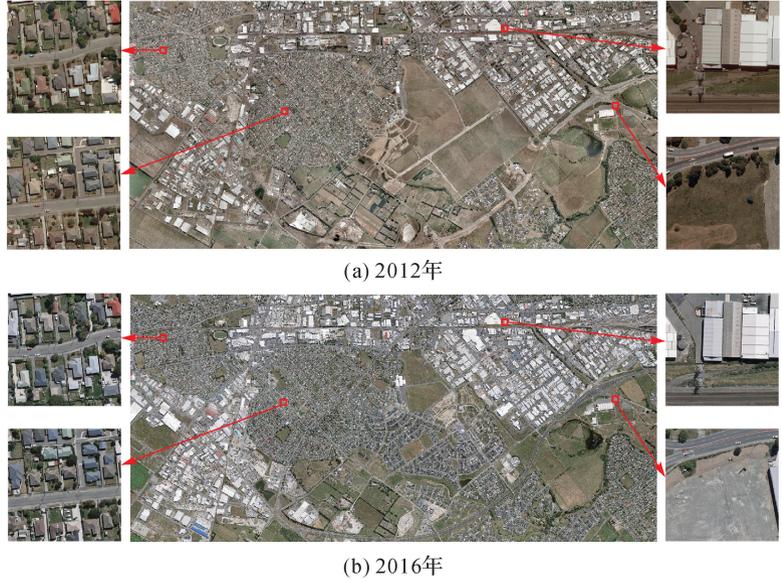
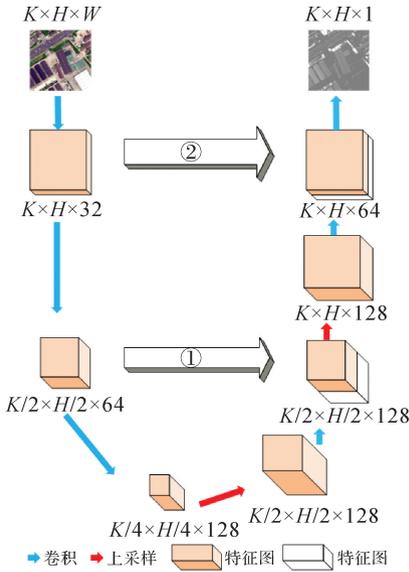


图3 ACGAN中的注意力网络结构
Fig.3 Attention network structure in ACGAN

图4 数据集1中2012年和2016年航空图像
Fig.4 Aerial images of 2012 and 2016 in dataset 1

2.2 试验设计

此次试验中的 AGGAN、ACGAN 为无监督的生成对抗网络,需要训练集和测试集,不需要验证集。利用有监督的神经网络 U-Net 网络^[31]和 FCN 网络^[32]验证色彩一致性前后对于语义分割模型泛化能力的影响,判断此次色彩一致试验的可靠性,即利用参考色彩风格图像(即航空图像 2016 年的数据,卫星图像 2015 年的数据)去训练权重模型,来预测待纠正图像(即航空图像 2012 年的数据,卫星图像 2016 年的数据)色彩一致前后其上建筑物的分割效果。此过程需要训练集、验证集以及测试集。色彩一致网络 AGGAN、ACGAN 与分割网络 U-Net、FCN 的训练集图片数量相同。整个试验所用数据集图片数量见表 1。

本文利用待纠正的原始图像、传统的 MKL、HM 算法及 AGGAN 模型色彩一致后的图像与 ACGAN 网络的试验结果进行对比,整体试验设计如图 6 所示:①利用源域的参考色彩风格图像训练集得到建筑物语义分割模型,直接分割参考色彩风格图像测试集以及相应的待纠正图像测试集。②利用①中的训练集得到 AGGAN、ACGAN 网络的色彩一致性权重,对只包含待纠正图像的测试集进行色彩转换。③利用①得到的建筑物语义分割模型直接分割 HM 算法、MKL 算法及②中色彩一致性处理后的待纠正图像测试

集,计算建筑物分割精度。试验所用 GPU 为 NVIDIA TITAN XP。

表 1 数据集图像数量分布

Tab.1 The number of images in the dataset

数据集	年份	训练集	验证集	测试集	总数
数据集 1	2016	899	299	300	1498
	2012	899	—	300	1199
数据集 2	2015	400	167	162	729
	2016	400	—	162	562

2.3 试验结果与分析

2.3.1 多时相遥感图像色彩一致性结果

数据集 1 色彩一致性结果如图 7 所示,两种传统算法生成的结果图像比 2016 年参考色彩风格图像亮度低,且部分图像出现与参考色彩风格图像相差很大的色块。AGGAN 网络与 ACGAN 网络的色彩映射较为稳定,色偏现象少。但是,与 AGGAN 网络相比,本文算法的结果与 2016 年原始图像在亮度上更加接近。所以对于多时相的航空影像,本文算法有较好的色彩一致性效果。

由图 7 可知,2012 年与 2016 年两期图像除了对应位置的建筑物、车辆等具体地物存在少许差别外,最主要的是两期图像的整体色彩存在明显不同,这是对数据集 1 进行色彩一致性转换的重要关注点。其对应的注意力图如图 8 所示,亮

度越高代表注意力网络分配的权重越大。由图 8 可知,AGGAN 网络与 ACGAN 网络的注意力图都可以关注到影像的总体色彩。但是,AGGAN 网络的注意力图较为粗糙,不同地物之间边界模糊,图像色彩等低频信息没有形成整体的关注度,部分建筑物顶部分配的权重过少,一定程度上会导致建筑物顶部色彩一致效果不佳。ACGAN 网络的注意力图中不同地物之间的边界较为清晰,并且图像整体蒙上白色,因此它对图像的整体色彩等低频信息的关注度较为一致,基本不会出现权重分配不均的情况。

数据集 2 色彩一致性结果如图 9 所示,两种传统算法的结果在河流处均出现了较为明显的色偏现象,且对于 HM 算法来讲,其结果图像中的植被色彩饱和度较高、建筑物顶部并没有完全脱离 2016 年原始图像的紫色调。两种深度学习算法的色彩一致性效果较稳定,但是对于河流来讲,两种深度学习方法均带来了一定的噪声。对于植被和建筑物来讲,AGGAN 网络的结果图像整体泛白,ACGAN 网络的结果与 2015 年原始图像的色调更加接近。所以,对于多时相卫星图像的色彩一致性处理,本文算法具有一定的优越性。

由图 9 可知,2016 年与 2015 年影像最显著的差异是植被之间的色彩以及图像整体色调等低频信息,这是对数据集 2 进行色彩一致性转换的重要关注点。其对应的注意力图如图 10 所示,亮度越高代表注意力网络分配的权重越大。从图 10 可以看出,AGGAN 网络与 ACGAN 网络的注意力图都可以关注到影像的总体色彩。但是,AGGAN 网络生成的注意力图质量较差,除了图像整体色调等低频信息没有形成均匀的关注度外,色彩差异较显著的植被没有被分配较高的权重。ACGAN 网络的注意力图质量较好,它对整幅图像中色调等低频信息的关注度较为均匀,并且为图像之间差异最显著的植被分配了更高的权重。因此,对于数据集 1 这类整体色调存在差异的图片,U 型注意力机制能分配较为平均的权重;对于数据集 2 这类植被等具体地物色调差异较显著的图片,U 型注意力机制能进一步为其分配更大的权重。因此,ACGAN 网络的注意力机制更加灵活,可以针对图像的具体差异制定不同的注意力分配策略。

此次试验利用峰值信噪比 (peak signal to noise ratio, PSNR) 和结构相似性 (structural

similarity index, SSIM) 定量评价 4 种方法处理后的测试图像,结果见表 2,粗体表示本文的方法。对于数据集 1, MKL 算法的结果图像 PSNR 和 SSIM 值最高,图像质量最优,但是由前文可知其目视效果不佳。本文算法色彩一致后的图像质量次高。对于数据集 2, ACGAN 网络生成的图像 PSNR 值和 SSIM 值均最高,说明其图像质量最高,且由前文可知其色彩一致性效果也较好。综上所述,ACGAN 网络的结果比传统算法在目视效果上更接近参考色彩风格图像,比 AGGAN 网络的图像质量更好、噪声更少。

表 2 各算法结果图像的 PSNR 及 SSIM 值

Tab.2 The PSNR and SSIM values of each algorithm result images

参数	数据集 1				数据集 2			
	MKL	HM	AGGAN	ACGAN	MKL	HM	AGGAN	ACGAN
PSNR	25.223	23.612	21.225	24.889	21.882	21.761	20.834	22.394
SSIM	0.968	0.937	0.877	0.957	0.820	0.829	0.821	0.856

2.3.2 多时相遥感图像色彩一致性前后建筑物分割结果与分析

为了验证色彩一致后的图像有利于建筑物语义分割模型泛化能力的提升,从而设计了该试验。建筑物分割的精度评价指标为交并比 (intersection over union, IoU)、召回率 Recall、精确度 Precision、Recall 和 Precision 的调和平均数 F1 score 以及总体像素精度 (overall accuracy, OA)。

数据集 1 的建筑物分割精度见表 3,粗体表示本文方法。由表 3 可以看出,使用 2016 年航空图像训练的语义分割模型直接测试 2012 年的原始数据,精度下降较为明显。利用 4 种色彩一致性方法处理后建筑物语义分割精度均有不同程度的提升。与两种传统方法中表现较好的算法相比,对于 U-Net 网络,ACGAN 网络的分割结果 IoU 上升了 11 个百分点;Precision 上升了 5 个百分点;F1 Score 上升了 6 个百分点;OA 上升了 5 个百分点。对于 FCN 网络,ACGAN 网络色彩一致后的结果也略有提升。这是由于 ACGAN 网络通过卷积神经网络逐层提取航空图像更高级的色彩特征进行映射,色偏现象比传统算法少。与 AGGAN 网络相比,ACGAN 网络的结果图像建筑物分割精度也有一定的提升,这是由于本文的 U 型注意力机制能将低级特征与高级特征结合,有利于多尺度的特征提取。并且,在模型训练

过程中,本文算法比 AGGAN 网络节省了近 10 h,这得益于本文方法的参数较少、结构更加稳健。以上结果均表明本文方法对于提升多时相航

空图像的建筑物语义分割模型泛化能力的思路是正确且有效的,并且与对比算法相比,ACGAN 网络的帮助更显著。



图 5 数据集 2 中 2015 年和 2016 年卫星图像

Fig.5 Satellite images of 2015 and 2016 in dataset 2

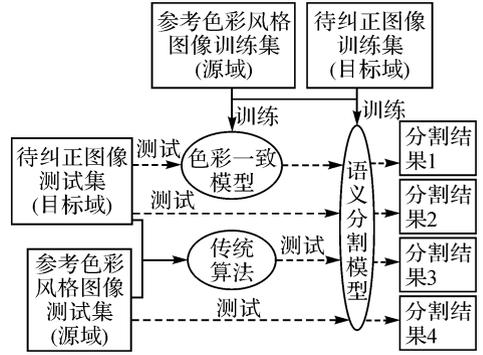


图 6 试验的整体设计

Fig.6 The overall design of the experiment

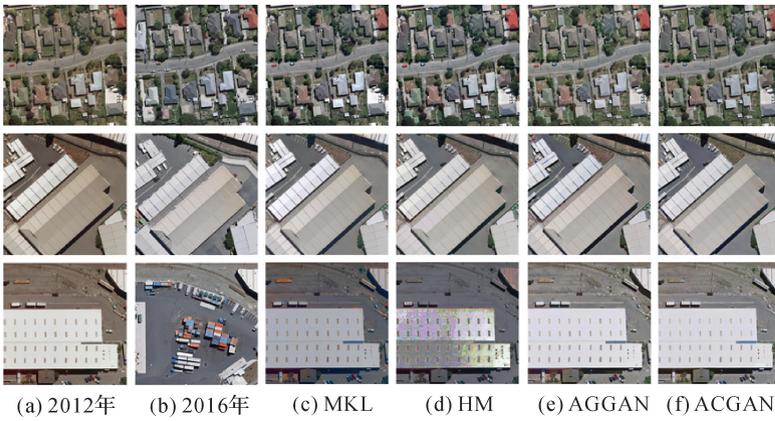


图 7 数据集 1 中 2012 年部分航空图像色彩一致性结果

Fig.7 Color consistency results of some 2012 aerial images in dataset 1

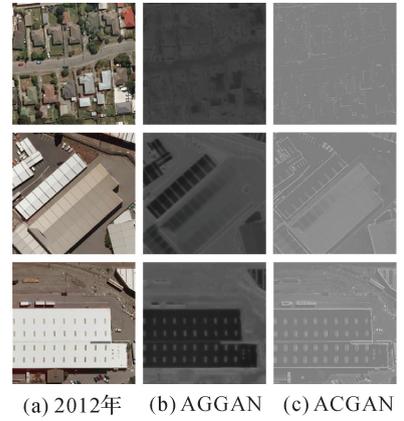


图 8 数据集 1 中 2012 年部分航空图像注意力

Fig.8 Attention map of some 2012 aerial images in dataset 1

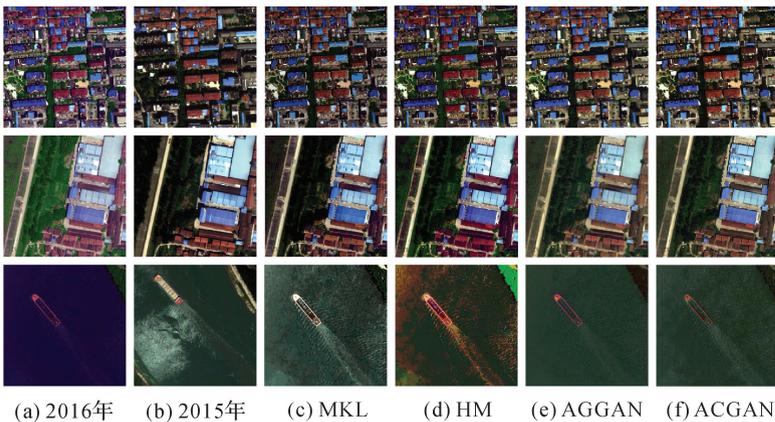


图 9 数据集 2 中 2016 年部分卫星图像色彩一致性结果

Fig.9 Color consistency results of some 2016 aerial images in dataset 2

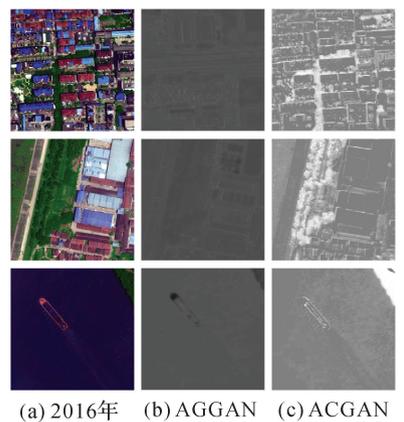


图 10 数据集 2 中 2016 年部分卫星图像注意力

Fig.10 Attention map of some 2016 satellite images in dataset 2

表 3 数据集 1 经各算法色彩一致后建筑物语义分割结果

Tab.3 Dataset 1 building semantic segmentation results after color consistency of each algorithm

年份	算法	U-Net					FCN				
		IoU	Recall	Precision	F1 Score	OA	IoU	Recall	Precision	F1 Score	OA
2016	原始图像	0.819	0.972	0.839	0.901	0.940	0.827	0.967	0.960	0.963	0.945
	MKL	0.705	0.873	0.786	0.827	0.888	0.872	0.937	0.927	0.932	0.965
2012	HM	0.631	0.961	0.647	0.774	0.854	0.866	0.941	0.917	0.929	0.962
	AGGAN	0.777	0.942	0.816	0.874	0.930	0.876	0.940	0.923	0.934	0.964
	ACGAN	0.784	0.941	0.824	0.875	0.932	0.879	0.943	0.929	0.936	0.966

数据集 1 的建筑物语义分割结果如图 11 所示,前两行为 U-Net 网络分割结果,后两行为 FCN 网络的分割结果。2012 年的原图、MKL 算法及 HM 算法生成的图像中建筑物均存在较为明显的漏检、错检等情况。AGGAN 网络的结果图像中建筑物边缘噪声多,轮廓较为模糊。ACGAN 网络的结果图像分割后建筑物的轮廓更清晰,噪声更少,检测的建筑物个数更全。

数据集 2 的建筑物分割精度见表 4,粗体表示本文方法。从中可知,使用 2015 年卫星图像训练的语义分割模型直接测试 2016 年的数据,精度下降较为明显。通过 4 种色彩一致性方法处理后,两种语义分割网络建筑物提取精度均有不同程度的上升。与两种传统方法中表现较好的算法相比,对于 U-Net 网络,ACGAN 网络迁移后的结果 IoU 上升了 7 个百分点,Precision 上升了 10 个百分点,F1 Score 上升了 5 个百分

点;对于 FCN 网络,ACGAN 网络迁移后的结果也略有提升。这是因为传统算法利用卫星图像的低级特征,但本文算法可以提取卫星图像更高级的色彩特征进行映射。与 AGGAN 网络相比,对于 U-Net 网络,ACGAN 网络迁移后的结果 IoU、Recall 上升了 4 个百分点,F1 Score 上升了 3 个百分点;对于 FCN 网络,ACGAN 网络迁移后的结果 IoU 上升了 3 个百分点,Recall 上升了 14 个百分点,F1 Score 上升了 2 个百分点。因为 ACGAN 网络的 U 型注意力机制能结合上下文语义信息,避免有效低级特征的损失。并且,由于本文方法参数较少,它的训练时间比 AGGAN 网络节约了 8 h。以上结果均表明,ACGAN 网络的结果图像对于多时相卫星图像建筑物语义分割模型泛化能力的提升帮助更加明显。

表 4 数据集 2 经各算法色彩一致后建筑物语义分割结果

Tab.4 Dataset 2 building semantic segmentation results after color consistency of each algorithm

年份	算法	U-Net					FCN				
		IoU	Recall	Precision	F1 Score	OA	IoU	Recall	Precision	F1 Score	OA
2015	原始图像	0.587	0.726	0.754	0.740	0.913	0.617	0.813	0.719	0.763	0.915
	原始图像	0.334	0.471	0.535	0.501	0.847	0.486	0.786	0.536	0.637	0.854
	MKL	0.459	0.595	0.669	0.630	0.886	0.563	0.732	0.711	0.721	0.902
2016	HM	0.393	0.491	0.664	0.564	0.896	0.546	0.767	0.655	0.706	0.896
	AGGAN	0.474	0.578	0.726	0.643	0.895	0.547	0.688	0.728	0.707	0.907
	ACGAN	0.492	0.600	0.733	0.660	0.900	0.565	0.781	0.672	0.723	0.907

数据集 2 的建筑物分割结果如图 12 所示,前两行为 U-Net 的分割结果,后两行为 FCN 的分割结果。对于面积较小的建筑物,2016 年原始图像、MKL 算法及 HM 算法的结果图像均存在明显的漏检、错检。对于面积较大的建筑物,AGGAN 的结果图像错检的情况较多,且建筑物边缘轮廓不清晰。ACGAN 的结果图像建筑

物语义分割更加准确,漏检、错检的情况更少。综上所述,经本文算法处理后,多时相遥感图像目视效果更加接近于参考色彩风格图像,且图像中建筑物的语义分割精度提升较为明显,证明该算法是一种有效地提升多时相遥感图像深度学习语义分割模型泛化能力的途径。

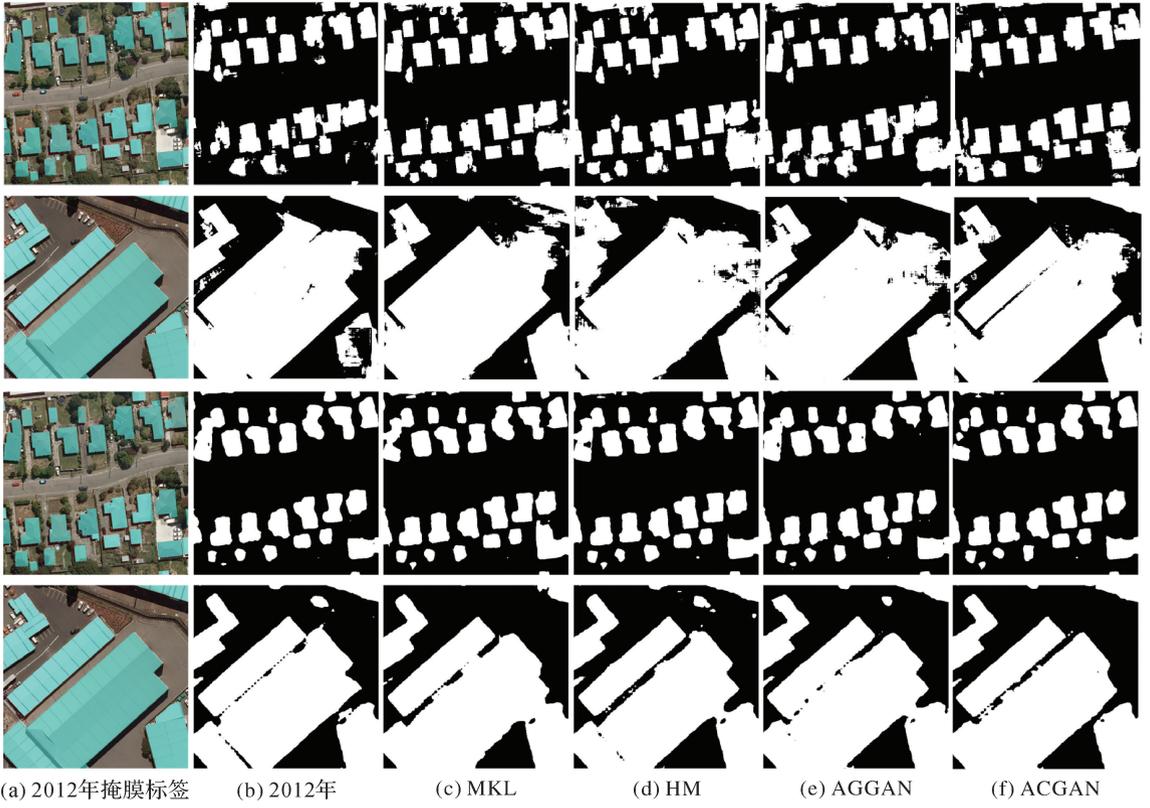


图 11 数据集 1 建筑物语义分割的结果图像

Fig.11 Dataset 1 results of building semantic segmentation images

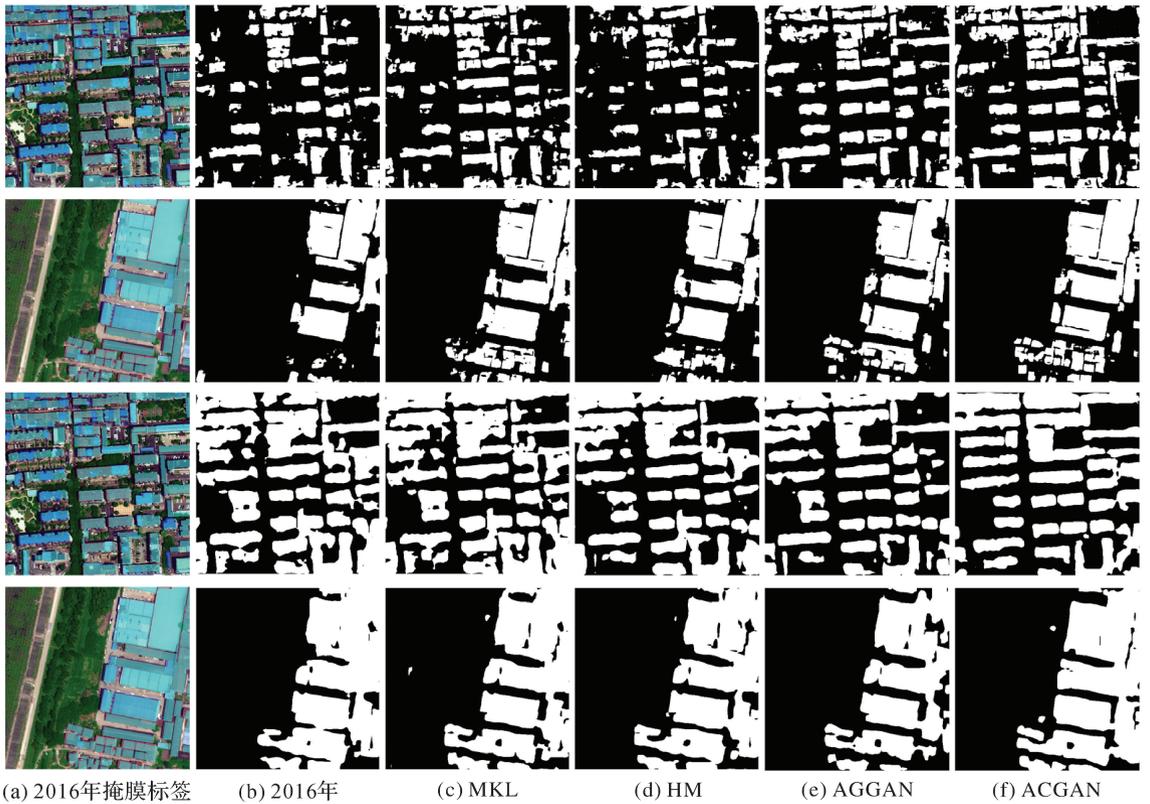


图 12 数据集 2 建筑物语义分割的结果图像

Fig.12 Dataset 2 results of building semantic segmentation images

3 结束语

针对多时相遥感图像色彩不一致导致的建筑物语义分割模型泛化能力差等问题,本文结合遥感图像颜色、纹理及场景等较为复杂的特点,设计了一种注意力引导的色彩一致生成对抗网络(ACGAN)。该网络能够层次化地提取遥感图像的高级特征,缩小了多时相遥感图像之间的数据分布差异,通过聚合多尺度特征的 U 型注意力机制生成更加精细的注意力图,改善了传统方法的偏色现象。试验证明,第一,经本文算法处理后的待纠正图像与参考色彩风格图像的整体色调更加接近,目视效果以及纹理特征更加真实,且该算法的测试阶段是基于深度学习模型的推理能力,一定程度上避免了传统算法依赖参考图像的限制,应用场景更加广泛。第二,经本文算法处理后的待纠正图像建筑物语义分割精度提升较为明显,增强了多时相遥感图像建筑物语义分割模型的泛化能力,节省了为待纠正图像人工手绘建筑物标签的时间,提高了生产效率,降低了生产成本。本文算法可以应用于遥感图像变化检测等诸多场景,并且对于其他地物类型的提取也具有一定的借鉴作用。

参考文献:

[1] JI Shunping, WEI Shiqing, LU Meng. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(1): 574-586.

[2] 左宗成, 张文, 张东映. 融合可变形卷积与条件随机场的遥感影像语义分割方法[J]. *测绘学报*, 2019, 48(6): 718-726. DOI: 10.11947/j.AGCS.2019.20170740.

ZUO Zongcheng, ZHANG Wen, ZHANG Dongying. A remote sensing image semantic segmentation method by combining deformable convolution with conditional random fields [J]. *Acta Geodaetica et Cartographica Sinica*, 2019, 48(6): 718-726. DOI: 10.11947/j.AGCS.2019.20170740.

[3] ZHANG Hongyi, CISSE M, DAUPHIN Y N, et al. Mixup: beyond empirical risk minimization [J]. *arXiv preprint arXiv: 1710.09412*, 2017.

[4] INOUE H. Data augmentation by pairing samples for images classification[J]. *arXiv preprint arXiv: 1801.02929*, 2018.

[5] CUBUK E D, ZOPH B, MANE D, et al. AutoAugment: learning augmentation strategies from data[C]//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision*

and Pattern Recognition. Long Beach, CA: IEEE, 2019: 113-123.

[6] CANTY M J, NIELSEN A A. Automatic radiometric normalization of multitemporal satellite imagery with the iteratively re-weighted MAD transformation [J]. *Remote Sensing of Environment*, 2008, 112(3): 1025-1036.

[7] HUANG T W, CHEN H T. Landmark-based sparse color representations for color transfer [C]//*Proceedings of 2009 IEEE International Conference on Computer Vision*. Kyoto, Japan: IEEE, 2009: 199-204.

[8] 李德仁, 王密, 潘俊. 光学遥感影像的自动匀光处理及应用[J]. *武汉大学学报(信息科学版)*, 2006, 31(9): 753-756.

LI Deren, WANG Mi, PAN Jun. Auto-dodging processing and its application for optical RS images [J]. *Geomatics and Information Science of Wuhan University*, 2006, 31(9): 753-756.

[9] HORN B K P, WOODHAM R J. Destriping Landsat MSS images by histogram modification [J]. *Computer Graphics and Image Processing*, 1979, 10(1): 69-83.

[10] PITIÉ F, KOKARAM A. The linear Monge-Kantorovitch colour mapping for example-based colour transfer [C]//*Proceedings of the 4th European Conference on Visual Media Production*. London: IET, 2007: 27-28.

[11] YOO J D, PARK M K, CHO J H, et al. Local color transfer between images using dominant colors [J]. *Journal of Electronic Imaging*, 2013, 22(3): 033003.

[12] MOULON P, DUISIT B, MONASSE P. Global multiple-view color consistency [J]. *International Conference on Visual Media Production*, 2013.

[13] HACOHEM Y, SHECHTMAN E, GOLDMAN D B, et al. Non-rigid dense correspondence with applications for image enhancement [J]. *ACM Transactions on Graphics*, 2011, 30(4): 70.

[14] GATYS L A, ECKER A S, BETHGE M. A neural algorithm of artistic style [C]. *arXiv: 1508.06576*, 2016: 326.

[15] GATYS L A, ECKER A S, BETHGE M, et al. Image style transfer using convolutional neural networks [C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV: IEEE, 2016: 2414-2423.

[16] LI Chuan, WAND M. Combining markov random fields and convolutional neural networks for image synthesis [C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV: IEEE, 2016: 2479-2486.

[17] LUAN Fujun, PARIS S, SHECHTMAN E, et al. Deep photo style transfer [C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI: IEEE, 2017: 6997-7005.

[18] WANG Li, WANG Zhao, YANG Xiaosong, et al. Photographic style transfer [J]. *The Visual Computer*, 2020, 36

- (2): 317-331.
- [19] LI Yijun, LIU Mingyu, LI Xueting, et al. A closed-form solution to photorealistic image stylization [C] // Proceedings of the 16th European Conference on Computer Vision. Munich: Springer, 2018: 453-468.
- [20] JOHNSON J, ALAHI A, LI Feifei. Perceptual losses for real-time style transfer and super-resolution[C]//Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2016: 694-711.
- [21] ZHU Junyan, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C] // Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017: 2242-2251.
- [22] MNIH V, HEES N, GRAVES A. Recurrent models of visual attention[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal: NIPS, 2014: 2204-2212.
- [23] KIM J, KIM M, KANG H, et al. U-GAT-IT: unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation[J]. arXiv preprint arXiv: 1907.10830, 2019.
- [24] PUMAROLA A, AGUDO A, MARTINEZ A M, et al. Ganimation: anatomically-aware facial animation from a single image[C]//Proceedings of the 16th European Conference on Computer Vision. Munich, Germany: Springer, 2018: 818-833.
- [25] MEJATI Y A, RICHARDT C, TOMPKIN J, et al. Unsupervised attention-guided image-to-image translation[C] // Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montreal: NIPS, 2018: 3693-3703.
- [26] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C] // Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV: IEEE, 2016: 770-778.
- [27] ISOLA P, ZHU Junyan, ZHOU Tinghui, et al. Image-to-image translation with conditional adversarial networks [C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI: IEEE, 2017: 1125-1134.
- [28] 邱锡鹏. 神经网络与深度学习[EB/OL]. [2017-04-21]. <https://nndl.github.io/>.
- QIU Xipeng. Neural network and deep learning[EB/OL]. [2017-04-21]. <https://nndl.github.io/>.
- [29] PAUSANG S M C, FAJARDO-LIM Y. Nash equilibria in a multiple stage inspection game[J]. Electronic Notes in Discrete Mathematics, 2016, 56: 49-57.
- [30] 季顺平, 魏世清. 遥感影像建筑物提取的卷积神经网络与开源数据集方法[J]. 测绘学报, 2019, 48(4): 448-459. DOI: 10.11947/j. AGCS.2019.20180206.
- JI Shunping, WEI Shiqing. Building extraction via convolutional neural networks from an open remote sensing building dataset[J]. Acta Geodaetica et Cartographica Sinica, 2019, 48 (4): 448-459. DOI: 10. 11947/j. AGCS. 2019.20180206.
- [31] BADRINARAYANAN V, KENDALL A, CIPOLLA R. U-Net: a deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39 (12): 2481-2495.
- [32] LONG J, SELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015: 3431-3440.

(责任编辑:张艳玲)

收稿日期: 2019-10-28

修回日期: 2020-07-07

第一作者简介: 李雪(1993—),女,硕士生,研究方向为遥感图像智能解译。

First author: LI Xue (1993—), female, postgraduate, majors in remote sensing image interpretation.

E-mail: amberlixue1229@163.com

通信作者: 张力

Corresponding author: ZHANG Li

E-mail: zhangl@casm.ac.cn