doi: 10.6046/zrzyyg.2021411

引用格式: 付昱凯,杨树文,闫恒,等. 耦合非局部相似性与散度的 SAR 与光学影像融合[J]. 自然资源遥感,2023,35(1):99-106. (Fu Y K, Yang S W, Yan H. An SAR and optical image fusion algorithm coupling non - local self - similarity and divergence [J]. Remote Sensing for Natural Resources 2023.35(1):99 – 106.)

耦合非局部自相似性与散度的 SAR 与光学影像融合

付昱凯1,杨树文1,2,3,闫恒1,薛庆1,洪卫丽1,苏航1

(1. 兰州交通大学测绘与地理信息学院,兰州 730070; 2. 地理国情监测技术应用国家地方联合工程 研究中心,兰州 730070;3.甘肃省地理国情监测工程实验室,兰州 730070)

摘要: SAR 与光学影像的高质量融合应用是目前研究的热点问题之一,然而二者间辐射差异大、灰度相关性弱等 问题,严重影响了融合质量。为此,该文提出一种耦合非局部自相似性与散度的 SAR 与光学影像融合算法。首先 在频率域将影像分解,然后使用非局部方向熵和散度作为特征量分别指导低频和高频分量进行融合,最后将融合 分量重建,得到兼具清晰结构特征和丰富光谱信息的融合影像。通过对比实验,证明所提算法在融合 SAR 与光学 影像方面的有效性,及其在保持结构特征和减小光谱扭曲方面的优越性。

关键词: SAR 与光学影像; 非下采样轮廓波变换; 全局特征; 影像融合; 超球面色彩空间 中图法分类号: TP 751 文献标志码: A 文章编号: 2097-035X(2023)01-0099-08

引言

0

SAR 穿透力强,成像不受天时、天候的限制和 影响,但SAR影像中光谱信息少。光学影像富含光 谱信息,但在多云等地区成像易受限制^[1]。因此, 集成二者的优势能够有效提高遥感监测质量,图像 融合是耦合二者优势的必要手段^[2]。

文献分析表明,SAR 与光学影像像素级融合方 法可分为2类:一类是基于空间域的融合方法,常 见如调制法^[3];另一类是基于频率域的融合方法, 经典的方法有基于降采样小波变换[4]、脊波变换的 融合^[5]等。由于 SAR 和光学影像间存在较大的辐 射差异,基于空间域的方法会损失大量纹理特征和 光谱信息,而基于频率域的融合方法在频率域将图 像分解到多个方向,不仅能够保留大量纹理特征,而 且可以针对各方向分解系数的特点设计不同的融合 策略,从而减少光谱扭曲。因此,基于频率域多分辨 率分析是 SAR 和光学影像融合的主要方向。

基于频率域的融合方法在频率域计算融合系数 时,注射权重是决定融合质量的关键。高文涛等^[6] 在计算注射权重时使用对应像元绝对值计算融合权 重实现了 SAR 与光学影像的融合,该方法计算量 小、易于实现,但由于只利用对应像元的像元值计算 注射权重,而未考虑周围像元的影响,因此融合图像 中会存在明显的光谱扭曲。为更好地将源图像的细 节特征融入到新图像中, Anandhi 等^[7]引入最小似 然比、最大边缘强度和局部梯度等特征量来描述像 元的特征表达,该方法在一定程度上提高了融合质 量;在近期研究中,易维等^[8]提出了非下采样轮廓 波变化(nonsubsampled contourlet, NSCT)平均和平 均 NSCT 这 2 种方法,该方法不受特定融合策略限 制,能够减少光谱扭曲。然而,由于所有特征量都是 基于图像矩形邻域计算的,未能充分考虑影像中各 像元间的相似性及相似像元的影响,导致影像边缘 处出现光谱扭曲和方块效应^[9]。

综上所述,目前 SAR 与光学遥感影像融合面临 的主要问题有:①SAR 与光学影像的成像机理不同 与影像间的非线性辐射差异引起的严重的光谱扭 曲;②在频率域制定低、高频融合策略时,未考虑像 元间的自相似性,抑制在影像边缘处出现光谱扭曲。 鉴于此,本研究提出一种耦合非局部自相似性与散

收稿日期: 2021-11-30;修订日期: 2022-03-07

基金项目:国家自然科学基金项目"基于高分辨率卫星影像的彩钢板建筑与城市空间结构演变关系研究"(编号:41761082)、国家自 然科学基金项目"西北重点城市彩钢板建筑群与产业园区时空关联关系"(编号:42161069)、国家自然科学基金项目"基于 脉冲耦合神经网络的高光谱遥感图像融合方法研究"(编号:41861055)及兰州交通大学优秀平台项目(编号:201806)共同 资助。

第一作者: 付昱凯(1996 -),男,硕士研究生,主要研究方向为遥感图像处理与分析。Email: 736173353@qq.com。

通信作者:杨树文(1975-),男,博士,教授,博士生导师,主要从事遥感数字图像处理和遥感信息识别及提取方面的研究。Email: ysw040966@163.com

度的 SAR 与光学影像融合方法,充分利用影像像元间的自相似性和图像场中的散度信息,有效提高了 SAR 与光学影像的融合质量。

1 基础理论

1.1 HCS 变换

HCS 是 Padwick C 在文献[10]中提出的一种超 球面色彩空间,并将其应用于 WorldView - 2 影像的 多光谱与全色影像融合中,实验证明 HCS 在影像融 合中对源图像空间和光谱信息具有较高的保持度。 相较于其他经典的色彩空间变换方法,如: IHS^[11], HCS 具有能够不受影像通道数限制地提取亮度分 量和亮度分量更加清晰的优势,因此本文通过将影 像转换到 HCS 色彩空间提取亮度分量,以期能够实 现 SAR 与光学遥感影像的高质量融合。

1.2 NSCT 变换

多分辨率分析方法的性能是影响融合质量的关 键因素之一。常见的多分辨率分析方法都是采用降 采样的方式获取图像的多尺度特征,但降采样的方 式会造成信息损失和像元移位,导致融合结果中存 在像元失真和轻微划痕等问题。对此本文采用 NSCT的方法对影像进行多分辨率分析,并在 NSCT 变换的基础上,对影像的高频和低频子带的融合过 程做出改进^[12]。NSCT 是在小波变换和轮廓波变换 的基础上发展而来的,相比于降采样的分解方法, NSCT 集合了多方向、多尺度、平移不变性等优势, NSCT 变换方法的结构如图 1 所示。





2 本文融合方法

在基于频率域的多尺度影像融合过程中,描述 影像低、高频子带像元特征显著度的特征量和对应 的融合规则是决定融合影像质量的关键因素。故本 文针对影像低、高频子带的特点,设计使用不同的特 征量和融合规则对影像进行融合。

2.1 低频融合规则

低频是图像的近似,包含图像大部分能量和轮 廓信息。在计算融合权重时,常采用规则的矩形邻 域作为窗口,用局部窗口的统计特征作为特征量。 此类方法考虑相邻像元间的影响,但未顾及像元间 的相似性,这将导致融合结果中出现方块效应及边 缘处的光谱扭曲。

为此,本文在设计低频子带融合策略时,在考虑 相邻像素影响的同时,顾及像元间的相似性,引入非 局部自相似性^[13]的思想,设计非局部方向熵作为特 征量辅助计算低频融合系数。

非局部方向熵是基于非局部自相似性对图像方 向熵的扩展。图像方向熵是处在方向场中图像的局 部邻域二维熵,能够反映图像局部邻域灰度的聚集 特征和空间特征^[14],计算方法为:

$$G_{pq} = \frac{f(p,q)}{mn}, 0 \le p, q \ge 255$$
 , (1)

$$DE(i,j) = -\sum_{p=0}^{255} \sum_{q=0}^{255} G_{pq} \lg G_{pq} \quad , \qquad (2)$$

式中:p,q 为图像的灰度值; G_{pq} 和f(p,q)是由像元 灰度和邻域灰度均值组成的特征二元组(p,q)的出现的概率和频数;m,n为邻域的行数和列数; DE(i,j)为图像(i,j)处像元的方向熵。

非局部方向熵在计算过程中,首先设定一个搜 索窗口,然后通过在搜索窗口内滑动窗口,按照一定 的相似性测度计算滑动窗口与中心窗口的相似性, 并依据相似性确定权重,再以搜索窗口各像元方向 熵的加权平均作为中心像元的非局部方向熵。考虑 到像元间的结构相似性,本文计算非局部相似性时, 设定的相似性测度为图像块间的结构相似性(structural similarity,SSIM)^[15],计算方法为:

$$SSIM = \frac{(2\mu_w\mu_c + C_1)(2\delta_{wc} + C_2)}{(\mu_w^2 + \mu_c^2 + C_1)(\delta_w^2 + \delta_c^2 + C_2)} , (3)$$

式中:c,w 为核心像元的邻域窗口以及搜索范围的 滑动窗口; $\mu_w,\mu_e,\delta_w,\delta_c,\delta_we$ 分别为它们的均值、方差 以及协方差; C_1,C_2 为2个极小的常数,防止分母为 0导致无法计算。

依据式(1)可计算搜索窗口 *se* 内 (*i*,*j*) 处像元 的权值 *P* 为:

$$P(i,j) = \frac{SSIM(i,j)}{\sum_{(i,j) \in S^{e}} SSIM(i,j)} , \qquad (4)$$

由式(1)—(4)可得图像 R 中各像元的非局部加权 熵计算公式为:

$$NLDE(i,j) = \sum_{(i,j) \in R} P(i,j) \times DE(i,j) \circ (5)$$

在使用非局部方向熵作为特征量计算低频子带融合系数时,首先计算光学影像亮度分量和 SAR 影像各低频子带像元对应的非局部方向熵,然后依据非局部方向熵最大规则选择用于融合的高频系数。同时为抑制噪声的表达,在使用 SAR 影像低频子带的系数时,使用搜索窗口所有像元的非局部均值代替。本文融合方法低频部分位置(*i*,*j*)处融合像元计算公式为:

$$C_{\rm F} = \begin{cases} C_{\rm I} & NLDE(C_{\rm I}) \ge NLDE(C_{\rm SAR}) \\ P_{ij} w_{ij} & \ddagger \& \end{cases} , (6)$$

式中: C_F , C_I , C_{SAR} 分别代表融合低频、I 分量低频和 SAR 低频; $P_{ij}w_{ij}$ 为 SAR 低频中以 (i,j) 为中心像元 的搜索域和权值的点积; $NLDE_I$ 和 $NLDE_{SAR}$ 分别为 I 和 SAR 对应像元的非局部方向熵。

2.2 高频融合规则

影像的高频子带中包含大量小尺度纹理和细节,需要应用对图像细节和边缘敏感的特征量。由此本文引入场论中的图像散度作为高频融合规则的特征量。图像散度场是对图像梯度场中像元聚散程度的描述,某一点的散度精确表述了该点在梯度场中的聚散程度,散度的值越大,则该点的发散程度越大,即该点处在图像边缘等结构特征上的可能越大^[16]。因此,在高频融合中用散度描述特征的显著性,可以对源图像的结构特征具有较高的保持度。

单通道的图像是处在二维平面中的,以二维直 角坐标空间为例,对于一个二维场 *F*(*x*,*y*)在 (*x*,*y*)处的梯度可以定义为:

$$gradF(x,y) = \nabla F(x,y) = \left[\frac{\partial F}{\partial x}, \frac{\partial F}{\partial y}\right] \quad (7)$$

而对于一个二维矢量场的 V(x,y) 在 (x,y) 处

的散度表示为:

$$divV(x,y) = \nabla V(x,y) = \frac{\partial V}{\partial x} + \frac{\partial V}{\partial y} \quad (8)$$

由式(7)(8)可知,单通道图像散度的计算公式 为:

$$div(G(x,y)) = \nabla(\nabla G(x,y)) = \frac{\partial^2 G}{\partial x^2} + \frac{\partial^2 G}{\partial y^2} \circ \quad (9)$$

在使用散度作为特征量计算高频融合系数时, 首先计算光学影像亮度分量和 SAR 影像各高频子 带像元对应的散度,然后再依据散度最大规则选择 用于融合的高频系数。但由于 SAR 影像的高频子 带中含有大量噪声,因此在使用 SAR 影像高频子带 系数时选用邻域窗口的中值以抑制 SAR 影像中噪 声的影响。本文融合方法高频部分融合像元计算公 式为:

2.3 融合流程

基于上述研究,依据提出的低、高频融合策略, 得本文融合方法的主要步骤为:①对多通道光学影 像进行 HCS 变换,获取到光学影像的亮度分量 I; ②参照 I 分量的灰度范围,对 SAR 分量的像元值进 行调整,统一灰度范围;③对 I 分量和 SAR 影像进 行 NSCT 变换,获取各自的低频和高频分量;④按 照上文的融合规则分别计算高频和低频融合系数; ⑤对高频和低频的融合系数进行 NSCT 逆变换,获 取融合后的亮度分量 I';⑥用 I'替换原始分量 I,并 进行 HCS 逆变换到原始色彩空间,得到融合结果。

本文设计的针对 SAR 与光学遥感影像的融合 框架如图 2 所示。



图 2 本文融合算法流程结构图 Fig. 2 Flow chart of the proposed fusion algorithm

3 实验结果与分析

3.1 实验数据源

实验选用两组 SAR 与光学影像对算法性能进行测试。第一组为海南省万宁市的 3 通道谷歌影像

和机载 SAR 影像,主要场景为农田; 第二组为河南 省郑州市大暴雨期间的高分一号多光谱影像和高分 三号影像,主要场景为城市,且有云层遮挡; 第三组 为新疆伊犁的 Landsat8 假彩色合成(NIR, SWIR1, Red)影像和哨兵一号 SAR 影像,主要场景为山地和 湖泊。影像详细信息如表1 所示。

(10)

表 1 实验数据信息 Tab.1 Data information of the experiment

实验	区域	场景	影像类型	尺寸 (像素×像素)	分辨 率/m
实验1	海南 万宁市	农田	机载影像 (SAR)	378 × 404	0.25
			谷歌影像(3通 道)	192 × 205	0.5
实验2	河南 郑州市	城市	高分三号 (SAR)	1 159 ×1 211	3
			高分一号(多 光谱)	435 × 454	8
实验3	新疆 伊犁市	山地 +	哨 兵 一 号 (SAR)	875 ×1 576	20
		湖泊	Landsat8(假 彩色3通道)	583 ×1 051	30

3.2 融合质量评价标准

为定量评价融合算法的性能,本研究引入 SSIM、光谱映射(spectral angle mappig,SAM)、失真 度(degree of distortion,DD)、均方根误差(root - me an - square error,RMSE)和相对无量纲全局误差 (relative dimensionless global error in synthesis,ER-GAS)^[17-20]作为客观评价指标对融合质量评价。其 中 SSIM 用以表征结构特征保存完整度,其值越大, 融合影像质量越高;SAM 可以衡量融合结果的光谱 扭曲程度,值越小则光谱扭曲越小;DD 用以衡量融 合影像相对于原始光学影像的光谱失真程度,DD 越 小,失真度越小,融合质量越高;RMSE 可以衡量全局 光谱失真,值越小则全局光谱失真越小;ERGAS 可以 反映影像整体光谱质量,融合影像的 ERGAS 值越小则融合影像的光谱质量越高。除 SSIM 外的其他4 种评价指标的数学描述如式(11) - (14)所示:

$$SAM(\boldsymbol{v}_F, \boldsymbol{v}_O) = \arccos(\frac{\langle \boldsymbol{v}_F, \boldsymbol{v}_O \rangle}{\|\boldsymbol{v}_F\|_2 \| \|\boldsymbol{v}_F\|_2}) , (11)$$

$$DD(O,F) = \frac{1}{MN} \| vec(O) - vec(F) \|_{1}, \quad (12)$$

$$RMSE(O,F) = \sqrt{\frac{\sum_{i=1}^{M} \sum_{j=1}^{N} (O(i,j) - F(i,j))^{2}}{MN}},$$
(13)

 $ERGAS = 100 \times \frac{R_F}{R_0} \sqrt{\frac{1}{B} \sum_{i=1}^{B} \left(\frac{RMSE(i)}{\mu_i}\right)^2}, \quad (14)$

式中: $f_o(i,j)$, $f_F(i,j)$ 为源图像 O 与融合图像 F 对 应 (i,j) 位置的像元值; M,N 为图像的行、列数; $vec(\cdot)$ 表示将矩阵展成向量; v_F,v_o 为融合图像和 源图像的光谱向量; R_F,R_o 为源图像和融合图像的 空间分辨率; B 为源图像的通道数; μ_i 是源图像经 过 IHS 变换后的第 i 个分量的平均值。

3.3 对比实验

实验将本文算法与6种经典的融合方法进行对 比。包括 IHS^[21], PCA^[22], HCS^[10]、基于小波分解和 IHS 变换的融合算法^[23]、NSCT 平均^[8]及平均 NSCT^[8]融合方法。

3.3.1 主观评价

 (a) 原始 SAR 影像
 (b) 原始光学影像
 (c) HS 方法

 (a) 原始 SAR 影像
 (b) 原始光学影像
 (c) HS 方法

 (d) PCA 方法
 (e) 基于小波变换的方法
 (f) HCS 方法

 (g) 平均 NSCT
 (h) NSCT 平均
 (i) 本文算法

第一组实验的结果如图3所示,主要场景为农

图 3 实验一融合结果 Fig. 3 Fusion result of experiment 1

• 102 •

田,实验通过将机载 SAR 影像与天地图 3 通道影像 融合,增强或补全天地图影像的结构特征。在结构 特征完整度方面,除平均 NSCT 算法以外,其余方法 均能够将 SAR 影像中的结构特征注射到融合影像 中,其中 IHS 方法、HCS 方法 PCA 方法以及本文算 法融合结果中的结构特征更加清晰。在光谱保持度 方面,IHS 和 HCS 出现了较为严重的光谱扭曲,基 于小波变换的方法和 NSCT 平均 2 种算法相较于原 始光学影像都出现了整体光谱值降低的问题,PCA 方法和平均 NSCT 方法在局部区域存在较大的光谱 扭曲,相比之下,本文算法融合结果的光谱扭曲较 小,光谱特征更贴近源影像。

第二组实验结果如图 4 所示,主要场景为城市, 且由云层遮挡,实验通过将 GF - 3 的 SAR 影像与 GF-1B的多光谱影像融合,填补光学影像中被云 层遮挡部分的结构信息,降低云层遮挡对光学影像 目视解译的影响。在结构特征完整度方面,各算法 皆能在光学影像中的云层遮挡处填补上 SAR 影像 的结构特征,不同的是成分替换类的 IHS,HCS 以及 PCA 方法的融合结果中将云层完全去除,这是由算 法原理所决定的,成分替换的方法直接使用无云层 遮挡的 SAR 影像替换光学影像的亮度分量,因此基 于成分替换的方法可以完全去除光学影像中的云层 部分。在光谱保持度方面,基于成分替换的 3 种方 法以及平均 NSCT 方法均出现明显的光谱扭曲, NSCT 平均和基于小波变换的方法分别存在不同程 度的全局光谱失真,而本文算法能够更加完整地保 持源影像的光谱特征。



图4 实验二融合结果

Fig. 4 Fusion result of experiment 2

第三组实验结果如图 5 所示,主要场景为山地 和湖泊,实验通过将 Landsat8 和哨兵一号 SAR 影像 融合,将 SAR 影像中清晰的特征注射到光学影像, 得到具有丰富的结构和光谱信息的融合影像。在结 构特征完整度方面,各算法均能将实现特征注射,其 中 IHS 和 HCS 方法的注射效果最完整、立体。其 次,NSCT 平均和本文方法效果相近,相较于其他方 法结构特征保持的更为完整; 在光谱保持度方面, 成分替换类的方法的融合结果中均存在较大的光谱 扭曲。混合方法中, NSCT 平均和基于小波变换的 方法的效果相近,均存在全局光谱值降低的问题,平 均 NSCT 和本文算法效果相近,光谱信息更加贴近 原始影像。综合来看,实验三中,本文方法融合结果 的综合质量最高。



图 5 实验三融合结果

Fig. 5 Fusion result of experiment 3

3.3.2 客观评价

对几种方法的定量评价结果如表 2 所示,综合 3 组实验的指标评价结果来看,在结构特征保持度 方面,IHS 和 HCS 两方法的 SSIM 值最大,即结构特 征保持度最高,本文方法的 SSIM 值次之,也具有较 高的结构特征保持度。在光谱特征保持度方面, SAM,DD,RMSE 和 ERGAS 从不同角度反映了融合

	表 2	2 组实	<u></u> {验客观评(介指标	
Tab. 2	Obj	jective	evaluation	indicators	of
	2 g	roups	of experim	ents	

实验	파 스 → ン+	SSIM †	SAM	$DD(\times$	DMCE	ERGAS
组	融合力法		(°)↓	10 $^{-2}$) \downarrow	KMSE ↓	$(10^{-2})\downarrow$
	IHS	0.989 5	0.635 4	14.233 8	26.991 0	8.9794
	PCA	0.449 9	0.003 5	5.940 5	13.701 5	7.8624
实	HCS	0.908 3	0.668 6	13.239 6	25.321 2	8.415 0
验	Wavelet – IHS	0.682 0	0.188 3	7.098 1	13.5077	4.4894
	平均 NSCT	0.480 0	0.162 8	1.646 1	5.223 0	1.749 3
	NSCT 平均	0.628 9	0.056 3	3.913 0	7.648 6	2.585 0
	本文算法	0.729 8	0.041 1	1.457 1	3.532 9	1.238 4
	IHS	0.973 3	1.054 0	21.344 4	47.6397	15.477 8
	PCA	0.400 8	0.050 5	19.038 5	42.555 9	25.060 3
实	HCS	0.968 8	1.027 9	21.305 0	47.6187	15.488 6
验	Wavelet – IHS	0.5264	0.289 2	10.682 4	23.854 3	6.862 1
	平均 NSCT	0.3527	0.418 6	8.085 0	25.089 3	7.668 2
	NSCT 平均	0.472 3	0.368 8	6.9977	15.800 9	5.061 3
	本文算法	0.672 2	0.1099	1.7797	3.852 9	1.232 6
	IHS	0.998 9	5.732 0	41.8869	48.6669	46.6727
实验三	PCA	0.665 3	0.018 2	33.972 9	41.683 4	57.306 0
	HCS	0.989 5	5.6793	42.464 5	49.058 4	47.048 8
	Wavelet – IHS	0.6178	2.415 3	21.682 1	24.993 8	23.970 3
	平均 NSCT	0.389 6	0.0647	8.600 4	15.922 6	16.336 1
	NSCT 平均	0.637 9	0.509 2	13.193 3	19.6109	19.237 5
	本文算法	0.622 9	0.481 0	5.463 6	10.775 7	11.708 9

影像的光谱扭曲程度。在3组实验中,IHS和HCS 方法对应的各指标的结果值均远高于其他方法,即 IHS和HCS方法的融合影像中存在极为严重的光 谱扭曲。而本文方法在3组实验中的DD,RMSE和 ERGAS指标值均为7个方法中最小,SAM值仅次于 PCA方法。由此可知,本文方法在SAM,DD,RMSE 和ERGAS这4个指标上的综合表现最优,即本文方 法的融合影像中光谱扭曲最小,对光谱信息保持的 完整度最高。综合3组实验的所有评价结果来看, 相较于其他几种融合方法,本文方法的融合结果不 仅保留了丰富的空间纹理特征,且光谱扭曲最小,证 明了本文算法在SAR与光学遥感影像融合中的有 效性与优越性。

4 结论

为了提高 SAR 与光学遥感影像融合的质量,减 少光谱扭曲,本文在 SAR 与光学遥感影像融合过程 进行了2方面的改进:

1)在计算注射权重时,在低频子带融合中,充 分考虑影像中具有相似结构特征的像元间的影响, 设计非局部方向熵作为特征量指导融合低频系数的 选择;在高频子带融合中,选用对图像边缘纹理敏 感的散度作为特征量,指导融合高频系数的选择,减 少了融合结果中,尤其是边缘处的光谱扭曲。

2) 在决定使用 SAR 影像的高频或低频子带系数时,选用高频子带系数的邻域中值和低频子带的

非局部均值作为融合系数替代单一像元值,对 SAR 影像中的噪声进行抑制同时减小了光谱扭曲。

实验证明,本文算法的融合结果光谱扭曲小、结构特征保全完整,应用在多源、多尺度、多场景的 SAR 与光学遥感影像融合中行之有效。

参考文献(References):

- 王梦瑶,孟祥超,邵枫,等.基于深度学习的 SAR 辅助下光学 遥感图像去云方法[J].光学学报,2021,41(12):243-251.
 Wang M Y, Meng X C, Shao F, et al. SAR - assisted optical remote sensing image cloud removal method based on deep learning [J]. Acta Optica Sinica,2021,41(12):243-251.
- [2] 李卫国,蒋 楠,熊世为.基于 ARSIS 策略的 SAR 影像与多光谱 遥感小波融合[J].农业工程学报,2012,28(s1):158-163.
 Li W G, Jiang N, Xiong S W. Multi - spectral and SAR wavelet fusion based on ARSIS strategy[J]. Transactions of the Chinese Society of Agricultural Engineering,2012,28(s1):158-163.
- [3] 陈东,李 飚,沈振康. SAR 与可见光图像融合算法的研究
 [J]. 系统工程与电子技术,2000(9):5-7,43.
 Chen D,Li B,Shen Z K. Research on data fusion algorithm of SAR and optical images[J]. Systems Engineering and Electronics,2000 (9):5-7,43.
- [4] 徐 赣,尤红建.小波的 SAR 和光学图像融合方法比较研究
 [J]. 测绘科学,2008(1):109-112,249.
 Xu G,You H J. Comparative study on wavelet based SAR and optical image fusion methods [J]. Science of Surveying and Mapping,2008(1):109-112,249.
- [5] 李晖晖,郭 雷,李国新. 基于脊波变换的 SAR 与可见光图像融合研究[J]. 西北工业大学学报,2006(4):418-422.
 Li H H,Guo L,Li G X. Is Ridgelet transform better than wavelet transform in SAR and optical image fusion[J]. Journal of Northwestern Polytechnical University,2006(4):418-422.
- [6] 高文涛,汪小钦,凌飞龙,等. 基于纹理的雷达与多光谱遥感数 据小波融合研究[J]. 中国图象图形学报,2008(7):1341-1346.

Gao W T, Wang X Q, Ling F L, et al. Fusion algorithm research based on texture for SAR and multispectral images with wavelet transform [J]. Journal of Image and Graphics, 2008(7):1341 - 1346.

- [7] Anandhi D, Valli S. An algorithm for multi sensor image fusion using maximum a posteriori and nonsubsampled contourlet transform
 [J]. Computers and Electrical Engineering, 2018, 65(1):139 – 152.
- [8] 易维,曾湧,原征.基于NSCT变换的高分三号SAR与光学 图像融合[J].光学学报,2018,38(11):76-85.
 Yi W,Zeng Y,Yuan Z. Fusion of GF-3 SAR and optical images based on the nonsubsampled contourlet transform[J]. Acta Optica Sinica,2018,38(11):76-85.
- [9] 李小军,闫浩文,杨树文,等.一种多光谱遥感影像与航拍影像融合算法[J].遥感信息,2019,34(4):11-15. Li X J, Yan H W, Yang S W, et al. A fusion algorithm of multispectral remote sensing image and aerial image[J]. Remote Sensing Information,2019,34(4):11-15.
- [10] Padwick C, Deskevich M, Pacifici F, et al. WorldView 2 pan sharpening[C]//Proceedings of the ASPRS 2010 Annual Confer-

ence, San Diego, CA, USA. 2010, 2630:1-14.

- [11] Chong X J. Comparative analysis of different fusion rules for SAR and multispectral image fusion based on NSCT and IHS transform [C]//International Conference on Computer and Computational Sciences(ICCCS), IEEE:2015,271 - 274.
- [12] Maryam G, Mohammad S H, Habibollah D. Nonsubsampled contourlet transform – based conditional random field for SAR images segmentation[J]. Signal Processing, 2020, 174(9):107623.
- [13] 卜丽静,赵 爽,张正鹏. NLM 与比率图像的多时相 SAR 图像 去噪方法[J]. 遥感信息,2021,36(3):17-24.
 Pu L J,Zhao S,Zhang Z P. A multi temporal SAR image denoising method based on NLM and ratio image[J]. Remote Sensing Information,2021,36(3):17-24.
- [14] 胡学敏,余 进,邓重阳,等. 基于时空立方体的人群异常行为 检测与定位[J]. 武汉大学学报(信息科学版),2019,44(10): 1530-1537.

Hu X M, Yu J, Deng C Y, et al. Abnormal crowd behavior detection and location based on spatial – temporal cube [J]. Geomatics and Information Science of Wuhan University, 2019, 44 (10): 1530 – 1537.

- [15] Samadhan C, Kulkarni, Priti P. Rege, Pixel level fusion techniques for SAR and optical images: A review [J]. Information Fusion (2020), doi:https://doi.org/10.1016/j.inffus.2020.01.003.
- [16] 郑 胜,田 岩,柳 健,等. 基于散度的不同焦点图像融合方法
 [J]. 华中科技大学学报(自然科学版),2007(4):7-10.
 Zheng S, Tian Y, Liu J, et al. Divergence based multifocuses image fusion [J]. Journal of Huazhong University of Science and Technology(Natural Science Edition),2007(4):7-10.
- [17] 尹 峰,孟祥超,梁 鹏.一种国产高分卫星遥感影像变分融合 方法[J].国土资源遥感,2018,30(2):100-106.doi:10.6046/ gtzyyg.2018.02.14.
 Yin F, Meng X C, Liang P. A variational fusion method for remote sensing images of China's domestic high - resolution satellites
 [J]. Remote Sensing for Land and Resources,2018,30(2):100-106. doi:10.6046/gtzyg.2018.02.14.
- [18] Lichun S, Jonathan L. Fusion of hyperspectral and multispectral images based on a Bayesian nonparametric approach [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2019, 12(2):1205-1218.
- [19] Zhang Q, Maldague X. An adaptive fusion approach for infrared and visible images based on NSCT and compressed sensing[J]. Infrared Physics & Technology, 2016, 74(1):11 - 20.
- [20] Li X J, Yan H W, Xie W Y, et al. An improved pulse coupled neural network model for pan sharpening [J]. Sensors, 2020, 20 (13),2764.
- [21] Tu T M, Su S C, Syu H C, et al. A new look at HIS like image fusion methods[J]. Information Fusion, 2001, 2(3):177 – 186.
- [22] Psjr C, Sides S C, Anderson J A. Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic [J]. Photogrammetric Engineering & Remote Sensing, 1991, 57(3):256 - 303.
- [23] Hong G, Zhang Y, Mercer B. A wavelet and HIS integration method to fuse high resolution sar with moderate resolution multispectral images [J]. Photogrammetric Engineering & Remote Sensing, 2009,75(10):1213 - 1223.

An SAR and optical image fusion algorithm coupling non – local self – similarity and divergence

FU Yukai¹, YANG Shuwen^{1,2,3}, YAN Heng¹, XUE Qing¹, HONG Weili¹, SU Hang¹

 Faculty of Geomatics, Lanzhou Jiaotong University, Lanzhou 730070, China; 2. National – Local Joint Engineering Research Center of Technologies and Applications for National Geographic State Monitoring, Lanzhou 730070, China;
 Gansu Provincial Engineering Laboratory for National Geographic State Monitoring, Lanzhou 730070, China)

Abstract: Currently, the high – quality fusion of SAR and optical images is a hot research topic. However, the significant radiation difference and weak gray correlation between SAR and optical images greatly reduce the fusion quality. In this regard, this study proposed a SAR and optical remote sensing image fusion algorithm that coupled non – local self – similarity and divergence. First, images were decomposed in the frequency domain. Then, the non – local directional entropy and divergence were used as characteristic parameters to guide the fusion of low – and high – frequency components, respectively. Finally, the fusion components were reconstructed to obtain fusion images with clear structural features and rich spectral information. The comparative experiments verified the effectiveness of the proposed algorithm in fusing SAR with optical images and its superiority in maintaining structural features and reducing spectral distortion.

Keywords: SAR and optical image; non – subsampled contourlet transform; global feature; image fusion; hyper – spherical color space

(责任编辑:李瑜)

doi: 10.6046/zrzyyg.2021460

引用格式:赵凌虎,袁希平,甘淑,等.改进 Deeplabv3 + 的高分辨率遥感影像道路提取模型[J].自然资源遥感,2023,35(1): 107 – 114. (Zhao L H, Yuan X P, Gan S, et al. An information extraction model of roads from high – resolution remote sensing images based on improved Deeplabv3 + [J]. Remote Sensing for Natural Resources,2023,35(1):107 – 114.)

改进 Deeplabv3 + 的高分辨率遥感影像道路提取模型

赵凌虎1,袁希平2,3,甘淑1,2,胡琳1,丘鸣语1

(1.昆明理工大学国土资源工程学院,昆明 650093;2.云南省高校高原山区空间信息测绘技术应用工程 研究中心,昆明 650093;3.滇西应用技术大学地球科学与工程学院,大理 671000)

摘要:针对传统的道路提取方法在高分辨率遥感影像中存在提取效果差和提取速度慢的问题,提出了改进 Deeplabv3 + 的高分辨率遥感影像道路提取模型。采用 MobileNetv2 主干特征提取网络与 Dice Loss 函数相结合,较好地 平衡了高分辨率遥感影像道路提取精度与速度的矛盾,实现较高提取精度的同时减少了模型参数,满足了时效性 的要求。基于开源道路提取数据集的实验结果表明:①该文提出的道路提取模型在高分辨率遥感影像上具有可行 性,提取道路的整体精度达到 98.71%,具有较高的提取精度;②在提取道路的速度方面该方法平均帧数达到 120.05,模型参数量仅为 5.81 M,总体上比原模型更加轻量化,表明该方法满足了时效性的要求。该方法在大幅减 少参数量、满足时效性的同时保证了提取的精确度,为提高基于高分辨率影像的道路提取精度和速度提供了一种 新的改进思路和方法。

关键词: 遥感影像; 道路提取; 深度学习; 语义分割; DeepLabv3 + 模型 中图法分类号: P 2 文献标志码: A 文章编号: 2097 - 035X(2023)01 - 0107 - 08

0 引言

道路是人类在生活与工作中不可或缺的元素, 如改善民生生活条件及配套设施,提供出行便利; 提高城市各项工作的运行效率,加快城市化建设; 优化城市资源配置,便于各项特色资源的输入输出 等。道路信息也是重要的地理信息,是城市地理信 息数据库的重要组成部分。随着遥感技术的快速发 展,遥感图像的空间分辨率也在逐步提高,目前已经 达到亚米级别,并且发展趋势呈上升态。随之而来 的是地物特征的细节信息更加丰富,非道路信息 (植被阴影、车辆流动、高层建筑物遮挡和人流流 动)非常复杂,因其作为噪声使得提取道路目标信 息更加困难。此外,具有不同质地材料的道路在识 别的过程中将具有不同的光谱特征^[1],它们表现为 不同道路具有相同光谱或同一道路具有不同光谱。 目前,绝大多数道路提取采用半自动化方式,半自动 化方法是需要人工给定道路的起始点,按照特定的 规则和逻辑提取道路信息, Kass 等^[2]提出的 Snake 算法;罗庆洲等^[3]利用光谱特征和形状特征识别道路的方法;Ghaziani等^[4]提出利用二值图像分割提取道路的方法;Sirmaçek等^[5]利用边缘检测和投票机制进行道路的提取。半自动化道路提取方式具有算法鲁棒性差、识别精度低、流程繁琐等一系列问题,一般适用于中低空间分辨率的遥感影像,而不太适用于高空间分辨率的遥感影像^[6]。因此,研究一种自动的针对高分辨率遥感影像的道路提取方法具有非常重大的意义。

深度学习是机器学习的一个分支,传统机器学 习需要人工设计特殊的特征量,如尺度不变特征转 换(scale – invariant feature transform,SIFT)和方向梯 度直方图(histogram of oriented gradient,HOG)等,而 深度学习直接学习数据本身,数据中包含的所有特 征都由机器来学习。深度学习网络利用训练数据的 损失函数进行误差的反向传播从而实现参数的更 新,这样就能达到直接从数据中学习的目的^[7]。近 些年来随着深度学习在计算机视觉中的兴起^[8],其 在图像的识别、分类和分割中都取得了较大的突破。 最近几年,也有越来越多的学者将深度学习应用于

收稿日期: 2021-12-27;修订日期: 2022-04-12

基金项目: 国家自然科学基金项目"滇中星云湖高原湖泊流域聚落空间格局演化研究"(编号:41561083)和"东川小江泥石流迹地的 多尺度遥感探测试验分析研究"(编号:41861054)共同资助。

第一作者:赵凌虎(1998-),男,硕士研究生,研究方向为遥感图像处理。Email: 2919404153@qq. com。

通信作者: 甘 淑(1964 -),女,教授,博士生导师,研究方向为摄影测量与遥感技术。Email: gs@kust.edu.cn。

遥感影像的道路分割。Mnih 等^[9]最先提出使用卷 积神经网络(convolutional neural network, CNN)提取 遥感影像的道路,并采取条件随机场(conditional random field, CRF) 进行后处理细化分割效果, 但分 割精度并不高,存在明显的椒盐噪声;叶雪娜^[10]比 较了基于小批量梯度下降(mini - batch gradient descent, MBGD) 和基于拟牛顿法(BFGS method) 的 CNN 在道路提取上的效果,结果表明 BFGS 算法不 需要计算梯度,不用考虑学习率参数设置问题,不会 出现训练过程中梯度下降缓慢或发生震荡的现象: Long 等^[11]提出了全卷积网络(fully convolution network, FCN),其旨在于对图像进行像素级的分类,即 对每个像素都产生了一个预测,从而解决了语义级 别的图像分割问题。FCN 就是将 CNN 最后的全连 接层替换为卷积层,称为反卷积,采用反卷积层对最 后一个卷积层的特征图进行上采样, 使它恢复到输 入图像相同的尺寸,在预测每个像素的同时还保留 了空间信息。但全卷积的结果比较模糊,对图像中 的细节不够敏感,且只考虑了单个像素的预测,没有 关注像素与像素之间的关系,缺乏空间统一性。 Chen 等^[12]提出 Deeplabv1 模型,采用空洞卷积进行 特征提取并利用条件随机场恢复边界信息。Deeplabv2 在 Deeplabv1 基础上引入了空洞空间金字塔池 化(atrous spatial pyramid pooling, ASPP)并采用残差 网络(residual neural network, Resnet)作为主干网 络。Deeplabv3 改进了空洞空间金字塔池化模块,加 入了批量归一化层(batch normalization, BN),加快 了网络收敛速度避免了网络过拟合。Deeplabv3+^[13] 引入编码-解码结构且再次修改了主干网络,将残 差网络升级为 Xception 网络,使物体边界分割效果 更好。

随着网络深度和结构的不断加深与复杂,分割 精度在不断提升,但随之而来的是网络参数的规模 也愈发庞大。因此,本文针对深度学习的道路提取 在算力有限的计算机和嵌入式系统等设备上的应 用,研究一种基于 Deeplabv3 + 的道路提取网络,在 满足较高提取精度的前提下实现了实时处理。

方法原理

1.1 整体流程

高分辨率遥感影像道路提取实质上是一种二元 的语义分割问题,语义分割的主旨是为图像中的每 一个像素都分配一个语义标签,即为图像中的每一 个像素都进行分类^[14]。本文提出了基于 Deeplabv3 + 的轻量化高分辨遥感影像道路语义分割模型。整体 流程可以分为3个阶段:首先,遥感影像数据的预 处理阶段包含对图像的标注、裁剪、增强以及训练集 与测试集的划分等;其次,模型的训练阶段包含参 数的设置、主干特征提取网络的替换和损失函数的 改进等方面;最后,遥感影像道路的提取分割,并对 提取结果进行评价,如图1所示。



1.2 Deeplabv3 + 网络模型

Deeplabv3 + 以编码 - 解码为基础结构。其中 编码器负责特征抽取,将输入的图像编程为中间 表达形式,即逐渐减小特征图并提取高层语义信 息。相应的解码器负责将中间表达形式解码为输 出,逐渐恢复图像的空间信息并给每个像素分类, Deeplabv3 + 的整体网络结构如图 2 所示。其编码 部分包括作为主干网络的深度卷积神经网络 (deep convolutional neural networks, DCNN)和ASPP 这2个部分。其解码部分,接收来自主干网络中 间层的低级特征图和来自 ASPP 模块的输出作为 输入。首先,对低级特征图使用1×1卷积进行通 道降维; 然后,对 ASPP 模块的特征图进行 4 倍上 采样得到与低级特征图尺寸相同的特征图;再使 用 Concat 函数将上述 2 种特征图堆叠起来,一起 送入3×3卷积块进行特征融合;最后,进行线性 插值上采样得到与原始图像分辨率大小相等的预 测图。



图 2 Deeplabv3 + 网络结构 Fig. 2 Structure of Deeplabv3 +

1.3 改进 Deeplabv3 + 网络模型

Deeplabv3 + 原文中所用的主干网络是 Xception 特征提取网络,但 Xception 对模型的参数规模和运 行速度控制不佳,所以本文采用 MobileNetv2^[15]特征 提取网络作为主干网络以提升模型的提取效率;并 且针对遥感影像中道路提取类别不均衡的问题,本 文采用 Dice Loss 与交叉熵损失函数相加作为损失 函数来解决样本极度不均衡的情况。

1.3.1 MobileNetv2 网络

MobileNetv2 是 Google 提出的一个轻量化 CNN, 其将 MobileNetv1 中的深度可分离卷积(depthwise separable convolution, DepSep Conv)沿用,与常规卷 积不同的是 DepSep Conv 是将一个完整的卷积分解 成2步进行,分别是深度卷积与逐点卷积,图3为 DepSep Conv 示意图。相比常规的卷积操作,其参 数规模和运算成本比较低。



图 3 DepSep Conv 结构 Fig. 3 Structure of DepSep Conv

首先 MobileNetv2 是由多个瓶颈层构成的,如表 1 所示,其中 t 为输入通道的倍增系数,c 为输出通 道数,n 为该模块重复的次数,s 为该模块第一次重 复时的步长,k 为实际输出的通道数。每个瓶颈层 都包含扩张、卷积和压缩3个部分(图4)。其中扩 张是利用1×1 卷积来增加特征图的通道数;卷积 是用3×3的 DepSep Conv 提取图像每个通道的特 征以此减少模型参数;压缩是利用1×1 卷积来减 少特征图的通道数。通过扩张、卷积、压缩这3个过 程可以在保留空间信息的同时提取更多的通道信息。 在最后的压缩过程后,如果仍使用 ReLU 函数就会有 较大的信息丢失,因此为了减少信息丢失使用 Linear 作为激活函数。其次 MobileNetv2 为每个瓶颈层加入 了残差连接使得网络训练精度有明显的提升。

表1		Mobile	结构		

Tab. 1 Structure of MobileNetv2

输入/(像素 × 通道数)	模块类型	t	с	n	\$
$224^2 \times 3$	Conv2d	-	32	1	2
$112^2 \times 32$	Bottleneck	1	16	1	1
$112^{2} \times 16$	Bottleneck	6	24	2	2
$56^2 \times 24$	Bottleneck	6	32	3	2
$28^2 \times 32$	Bottleneck	6	64	4	2
$14^2 \times 64$	Bottleneck	6	96	3	1
$14^2 \times 96$	Bottleneck	6	160	3	2
$7^2 \times 160$	Bottleneck	6	320	1	1
$7^2 \times 320$	Conv2d 1×1	-	1 280	1	1
$7^2 \times 1\ 280$	Avgpool 7×7	-	-	1	-
$1 \times 1 \times 1$ 280	Conv2d 1×1	-	k	-	-



在一张图像中,如果一个像素只看到它相邻的 像素则无法判断此像素的类别和属性,那就应该增 加观察的视野以将更多的像素点纳入考察范围。在 CNN 中"视野"被称为感受野,通过增大感受野来捕 获图像的上下文信息,增大感受野的方式一般有池 化和增大卷积核,但这2种方法带来的是计算量和 参数量的激增。空洞卷积就是权衡感受野与计算量 的明智之选,在增大感受野的同时又不增加计算量 与参数量。空洞卷积就是在一般的卷积核中添加空 洞,空洞的值为0,膨胀率越大添加的0越多,感受野 也越大,如图5所示。Deeplabv3+采用 ASSP 模块进 一步提取多尺度的特征信息,ASSP使用不同膨胀率 的空洞卷积来实现多尺度特征提取,其中包括一个 1×1卷积、膨胀率为6的3×3空洞卷积、膨胀率为 12 的 3 × 3 空洞卷积、膨胀率为 18 的 3 × 3 空洞卷 积和一个特征图池化层,最后将这5个部分的特征 图讲行堆叠。



1.3.3 类别不平衡修正

在神经网络的学习中,寻找最优参数时,要寻找 使损失函数的值尽可能小的参数。为了找到使损失 函数的值尽可能小的参数,需要计算损失函数关于 参数的导数(梯度),然后以这个梯度为指引,逐步 更新参数的值,在这个过程中损失函数的确定尤为 重要,它可以影响分类的精度。目前,大多数语义分 割模型采用交叉熵损失函数,其表达式 L1 为:

$$L1 = -\frac{1}{N} \sum_{i=1}^{N} \left[y_i \ln \dot{y}_i^{\wedge} + (1 - y_i) \ln(1 - \dot{y}_i^{\wedge}) \right], (1)$$

式中: N 为样本总数; i 为第 i 个样本; y_i 为样本 i的实际标签; $\hat{y_i}$ 为样本 i 的预测标签。在遥感影像 道路语义分割中它有一个明显缺点,道路像素数量 远小于背景像素数量,就会使得模型严重偏向于背 景,导致分割效果不好。针对这种情况,本文采用 Dice Loss 与交叉熵损失函数相加作为损失函数的 方法来解决样本极度不均衡的情况,其中 Dice Loss 表达式 L2 为:

$$L2 = 1 - \frac{2\sum_{i=1}^{N} y_i \hat{y}_i}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \hat{y}_i} \circ (2)$$

可以从式(2)中看到 Dice Loss 是一种区域性代价函数,即网络的所有输出都会影响样本 *i* 的梯度,因此遥感影像中道路与背景样本极度不均衡问题得以缓解。

2 实验数据与评价指标

2.1 实验数据集

本文的实验数据来自 Massachusetts Roads^[16]遥 感道路语义分割数据集, Massachusetts Roads 遥感道 路语义分割数据集覆盖了美国马萨诸塞州超过 2 600 km²的面积,包含城市、城镇、农村和山区等多 种地区的道路信息,图像大小为1500像素×1500 像素,空间分辨率约为1m。由于该数据集的样本 存在严重的图像与标签数据丢失现象,经过人工筛 选去除不符合要求的图像与标签,对于存在漏标与 错标的标签使用 Labelme 软件重新标注。由于计算 机性能的限制原始大小图像无法输入网络训练,所 以将原始遥感图像和标签图像以256 步长进行裁 剪,裁剪过后的图像大小为256 像素×256 像素。 同时对训练集应用图像增广,图像增广在对训练图 像进行一系列的随机变化之后,生成相似但不同的 训练样本,从而扩大了训练集的规模,并且随机改变 训练样本可以减少模型对某些属性的依赖,从而提 高模型的泛化能力,本文采取旋转、水平和垂直翻 转、改变颜色、高斯模糊等多种图像增广技术。最终 得到8575张道路图像以及对应标签图像,并按照 9:1 分配训练集与测试集。

2.2 评价指标

为了从多个维度对提取结果进行评价,实验选 取整体精度、精确率、召回率、F1分数、平均帧数及 参数量作为评价指标。其中前4个评价指标计算公 式分别为:

$$OA = \frac{TP + TN}{TP + TN + FP + FN} , \qquad (3)$$

$$P = \frac{TP}{TP + FP} \quad , \tag{4}$$

$$R = \frac{TP}{TP + FN} \quad , \tag{5}$$

$$F1 = \frac{2TP}{2TP + FN + FP} \quad , \tag{6}$$

式中: TP 为正确预测的道路像素数; FP 为将非道

路预测为道路的像素数; FN 为将道路预测为非道路的像素数; TN 为将非道路预测为非道路的像素数; TN 为将非道路预测为非道路的像素数; P 为精确率,即预测正确的道路像素占总真实道路像素的比例,反映道路提取的完整程度; OA 为整体精度,即被预测正确的像素数占总像素的比例,反映模型的整体准确率; F1 为精确率和召回率的调和平均值。另外,平均帧数为模型每秒能处理图像的数量,其越高代表模型的速度越快。参数量为考察模型体积的重要指标,其值越低代表模型的轻量化程度越高。

3 模型训练实验及其结果分析

3.1 模型训练实验

3.1.1 模型训练实验环境及参数设置

本文实验环境为:操作系统为64 位 Windows10, CPU 型号为 Intel i5 - 10400f, 内存为 16 GB, GPU 型 号为 GeForce1060 显存为 6 GB 及 CUDA 并行计算 平台实现 GPU 编程加速。模型基于 Pytorch 深度学 习框架搭建网络结构,损失函数加入 Dice Loss 以解 决正负样本不均衡带来的训练问题,并采用 Adam (adaptive moment estimation)优化器算法优化权重 更新。实验还引入迁移学习的策略,将在 ImageNet 上训练好的 MobileNetv2 网络权重作为实验的初始 权重,使用迁移参数初始化网络能够提升泛化性能。 所以训练分为2个阶段:阶段一为冻结主干特征提 取网络阶段,冻结主干特征提取网络意味着权重不 进行更新,是为了防止权重在一开始就被破坏;阶 段二是解冻主干特征提取网络阶段,也就是需要更 新主干网络的权重来达到更高的精度。阶段一的初 始学习率为 0.000 5, 批处理大小为 8, 迭代 30 轮, 每迭代5轮学习率变为原来的0.02倍。阶段二初 始学习率为0.000 05,批处理大小为4,迭代70轮, 每迭代5轮学习率变为原来的0.02倍。

3.1.2 实验结果

使用 Massachusetts Roads 数据集对改进的 Deeplabv3 + 网络模型进行训练,将训练好的最优模型 权重提取并保存用于测试数据集的道路提取,得到 被提取道路的混淆矩阵,并依据混淆矩阵计算整体 精度、精确率、召回率和 F1 分数4项评价指标,然后 计算模型的平均帧数及参数量。

本文所改进的 Deeplabv3 + 模型对 Massachusetts Roads 测试集提取道路的整体精度达到 98.71%,精 确率、召回率和 F1 分数都在 87% 以上,证明了模型 在高分辨率遥感影像中道路提取的有效性。模型参数量为5.81 M,相比一般深度神经网络模型大幅减少,说明了模型的轻量化。平均帧数达到120.05,即每秒平均能提取120张256 像素×256 像素图像的道路,证明了模型在高分辨率遥感影像中道路提取的时效性。部分道路提取结果如表2所示,可以看到提取结果与真实道路基本吻合,也不存在传统提取方法所产生的椒盐噪声,提取道路边缘较为平滑。表2中绿色框表示标签图错误标注的道路,蓝色框表示道路存在被遮挡现象,橙色框表示标签图中没有振动的道路。可以看到对于标错的道路模型并没有提取,对于漏标的道路模型提取出来了,表明模型具有强大的学习能力,能较好地对道路特征进行提取,对于有遮挡现象的道路模型还具有屏蔽遮挡的能力。





3.2 对比实验及其分析

3.2.1 实验结果分析

为了进一步验证本文深度学习模型在高分辨率 影像道路提取中性能和轻量化的优越性。在 Massachusetts Roads 数据集上,将本文模型与现有 3 种经 典语义分割模型进行道路提取效果比较,分别是 U – Net,PSPNet 和 Deeplabv3 +,且为了比较的公平性和 客观性,所有模型都在统一的环境配置中训练,测试 所用的数据集也一致。U – Net 是基于编码 – 解码 结构的深度学习语义分割网络,其主干特征提取网 络使用 VGG – 16。U – Net 经常被用于医学影像的 分割,道路类似于医学影像中的血管,所以 U – Net 可以被用来提取道路。PSPNet 是基于全局金字塔 池化的空洞全卷积语义分割网络,其主干网络使用 与本文一致的 MobileNetv2。Deeplabv3 + 使用 Xception 作为主干网络。对比结果如表 3 所示。

表3 不同模型对比结果

Tab. 3 Comparing the results for the different models

模型	OA/%	P/%	R/%	F1/%	参数 量/M	平均 姉数
PSPNet	96.59	63.59	77.16	69.72	2.38	129.56
U – Net	98.26	82.46	82.33	83.39	24.89	49.71
Deeplabv3 +	98.45	87.00	81.92	84.38	54.71	30.77
改进 Deeplabv3 +	98.71	87.49	87.06	87.27	5.81	120.05

由表3可知,改进 Deeplabv3+模型在整体精度、精确率、召回率、F1分数、参数量及平均帧数较其他3个模型都有不同程度的提高。PSPNet 作为语义分割网络并不适用于高分辨率遥感影像的道路提取,其精度在4种网络中最低,但因其使用 Mo-bileNetv2 轻量化网络使得模型参数量很小,只有

2.38 M。U-Net与 Deeplabv3+有着不错的分割精度,但这两者的网络参数繁多、计算量庞大导致平均帧数相对较低,不能够达到时效性的要求。反观改进 Deeplabv3+模型,相比原 Deeplabv3+模型,因其采用更加轻量化的主干特征提取网络 MobileNetv2,使得参数量减少了 89.38%,平均帧数提高了 3.90 倍;因改进损失函数来平衡样本不均衡问题,使整体精度、精确率、召回率和 F1 分数分别提升了 0.26,0.49,5.14 和 2.89 百分点。改进后的模型在不损失精度的同时大大减少了参数量,满足了时效性的要求。

3.2.2 不同情况下道路提取分析

为了更加直观地对比4种模型的道路提取效 果,选取测试集中不同情况的道路对其细节进行对 比,如表4所示。

表4 4种道路提取模型结果对比

Tab. 4 Comparison of the results of four road extraction models



细节表现方面,在表4 区域 a,b 和 d 中,因 PSP-Net 网络模型结构不适合道路提取,导致其道路细 节提取能力较弱,不能有效地分离双车道和多车道 等复杂的道路情况,U - Net 与 Deeplabv3 + 模型可 以分离上述道路情况,但其分离效果不佳,会产生不 同道路局部粘连问题,而本文改进的 Deeplabv3 + 模 型可以有效地提取双车道和多车道信息,不同道路 之间少有粘连现象,与真实道路细节基本一致;在 表4 区域 c 中, PSPNet, U - Net 与 Deeplabv3 + 模型

均不能完整地提取左下方的圆形道路,其提取结果 与真实道路差别过大,结构上存在缺陷,而本文改进 的 Deeplabv3 + 模型能够完整地提取该目标区域的 道路。可以看出本文方法适应于道路细节特征的提 取,可以有效地提取出一些局部细节特征,在细节表 现方面优于其他3种模型。在模型的适应性方面, 在表4 区域 c 中,遥感影像中道路存在被树木、建筑 物及阴影遮挡或阻断现象,除了本文改进模型,其他 3 种模型均有不同程度不能提取被遮挡或阻断道路 的情况,也就是说改进 Deeplabv3 + 模型的屏蔽遮挡 的能力较强,对遮挡道路提取效果良好。说明了本 文改进模型对于复杂环境下的道路提取具有较好的 适应能力。在提取准确性方面,在表4中,PSPNet 模型道路提取分辨率较低,道路边缘粗糙并且存在 缺失问题,道路结构损失最为严重; U-Net 与 Deeplabv3+提取结果接近,也都存在道路边缘粗糙和 结构不完整等问题。相比于这3种方法,本文改进 方法提取结果噪声较少,并且道路结构最完整。由 于本文针对遥感影像中道路像素数量远小于背景像 素数量,模型严重偏向于背景,导致分割效果不好的 现象,提出了 Dice Loss 与交叉熵损失函数相加作为 损失函数的方法来解决样本极度不均衡的情况,使 得本文方法对各种情况下道路的提取均优于 Deeplabv3+模型,改善了 Deeplabv3+错提和漏提的现 象,提取出的道路结构完整度较高,总体上提取效果 较好。

4 结论

本文以深度学习语义分割中的 Deeplabv3 + 网 络模型为基础,提出结合 MobileNetv2 主干特征提取 网络与 Dice Loss 函数的改进 Deeplabv3 + 高分辨率 遥感影像道路提取模型。该方法针对遥感影像中背 景与道路像素比例极度失衡的特点引入 Dice Loss 函数,并结合轻量化特征提取网络大大减少了模型 的参数量。研究表明:本文提出的道路提取模型在 高分辨率遥感影像上具有可行性,提取道路的整体 精度达到98.71%,具有较高的提取精度;与3种经 典的语义分割网络相比,在道路提取的完整性和精 确度,以及树木、建筑物和阴影遮挡方面,本文方法 均有较好的提取效果;在提取道路的速度方面本文 方法平均帧数达到120.05,模型参数量仅为5.81 M, 表明本文方法满足了时效性的要求。本文方法在大 幅减少参数量满足时效性的同时保证了提取的精确 度,为提高基于高分辨率影像的道路提取精度和速 度提供了一种新的改进思路和方法。

在下一步的研究中,可以针对道路的结构信息 及特点,将在本文模型的基础上加入注意力机制,从 而提高模型对道路的提取能力;在数据集制作方 面,标签的获取采用 Labelme 软件对道路进行人工 标注,在这个过程中存在漏标和错标的问题,会产生 一定的误差,所以可以从如何快速获取高质量的标 签出发,进一步提高道路提取精度。

参考文献(References):

[1] Herold M, Roberts D. Spectral characteristics of asphalt road aging

and deterioration: Implications for remote - sensing applications [J]. Applied Optics,2005,44(20):4327-4334.

- [2] Kass M, Witkin A, Terzopoulos D. Snakes: Active contour models
 [J]. International Journal of Computer Vision, 1988, 1(4):321 331.
- [3] 罗庆洲,尹 球,匡定波.光谱与形状特征相结合的道路提取方法研究[J].遥感技术与应用,2007,22(3):339-344.
 Luo Q Z, Yin Q, Kuang D B. Research on extracting road based on its spectral feature and shape feature[J]. Remote Sensing Technology and Application,2007,22(3):339-344.
- [4] Ghaziani M, Mohamadi Y, Koku A B, et al. Extraction of unstructured roads from satellite images using binary image segmentation [C]//2013 21st Signal Processing and Communications Applications Conference (SIU). IEEE, 2013:1-4.
- [5] Sirmaçek B, Ünsalan C. Road network extraction using edge detection and spatial voting [C]//2010 20th International Conference on Pattern Recognition. IEEE, 2010;3113 – 3116.
- [6] 贾建鑫,孙海彬,蒋长辉,等. 多源遥感数据的道路提取技术研究现状及展望[J]. 光学精密工程,2021,29(2):430-442.
 Jia J X, Sun H B, Jiang C H, et al. Road extraction technology based on multi source remote sensing data: Review and prospects
 [J]. Optics and Precision Engineering,2021,29(2):430-442.
- [7] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017,60(6):84 – 90.
- [8] 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. IEEE, 2016:770 – 778.
- [9] Mnih V, Hinton G E. Learning to detect roads in high resolution aerial images [C]//European Conference on Computer Vision. Springer, Berlin, Heidelberg, 2010:210 – 223.
- [10] 叶雪娜. 基于卷积神经网络的遥感图像道路提取研究[D]. 西安:陕西师范大学,2017.
 Ye X N. Research on remote sensing image road extraction based on convolutional neural network[D]. Xi'an: Shaanxi Normal University,2017.
- [11] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation [C]//Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC:IEEE Computer Society, 2015:3431 – 3440.
- [12] Chen L C, Papandreou G, Kokkinos I, et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs [J]. Computer Science, 2014(4):357-361.
- [13] Chen L C, Zhu Y, Papandreou G, et al. Encoder decoder with atrous separable convolution for semantic image segmentation [C]// Proceedings of the European Conference on Computer Vision, 2018:801 – 818.
- [14] 魏云超,赵 耀.基于 DCNN 的图像语义分割综述[J].北京交 通大学学报,2016,40(4):82-91.
 Wei Y C,Zhao Y. A review on image semantic segmentation based on DCNN[J]. Journal of Beijing Jiaotong University, 2016,40 (4):82-91.
- [15] Sandler M, Howard A, Zhu M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018:4510 -

4520.

[16] Mnih V. Machine learning for aerial image labeling [D]. Toronto:

University of Toronto, 2013.

An information extraction model of roads from high – resolution remote sensing images based on improved Deeplabv3 +

ZHAO Linghu¹, YUAN Xiping^{2,3}, GAN Shu^{1,2}, HU Lin¹, QIU Mingyu¹

(1. School of Land and Resources Engineering, Kunming University of Science and Technology, Kunming 650093, China;

2. Surveying and Mapping Geo - Informatics Technology Research Center on Plateau Mountains of Yunnan Higher

Education, Kunming 650093, China; 3. School of Earth Science and Engineering, West

Yunnan University of Applied Sciences, Dali 671000, China)

Abstract: Aiming at the problems of poor extraction effect and slow extraction speed of traditional road extraction methods in the information extraction of roads from high – resolution remote sensing images, this study proposed a new information extraction model based on improved Deeplabv3 +. In the new model, the combination of the MobileNetv2 backbone feature extraction network with the Dice Loss function effectively balanced the contradiction between the precision and speed of road information extraction from high – resolution remote sensing images. As a result, high extraction precision was achieved while meeting timeliness requirements by reducing model parameters. The experimental results based on the open – source road information extraction remote sensing images, with overall accuracy of up to 98.71%; (2) In terms of the information extraction speed, the new model had an average frame number of 120.05 and parameter amount of only 5.81 M. Therefore, the new model was more lightweight lighter than original models, meeting the timeliness requirements. Therefore, the model proposed in this study meets the timeliness requirements by greatly reducing the parameter amount while ensuring high extraction accuracy. This study provides a new philosophy and method for improving the accuracy and speed of road information extraction from high – resolution while ensuring high extraction accuracy information extraction from high – resolution images.

Keywords: remote sensing image; road information extraction; deep learning; semantic segmentation; Deeplabv3 + model

(责任编辑:陈理)