

基于残差伪孪生卷积互相关网络的异源遥感图像匹配算法

邹融平^{1,2}, 朱斌^{1,2*}, 王晨阳¹, 朱耀轩^{1,2}, 胡洋颀³

¹国防科技大学电子对抗学院, 安徽 合肥 230037;

²红外与低温等离子体安徽省重点实验室, 安徽 合肥 230037;

³中国人民解放军 32256 部队, 广西 桂林 541000

摘要 合成孔径雷达(SAR)成像和可见光成像是遥感卫星的常用成像方式。由于两者在成像信息上具有高度互补性, SAR和可见光的图像数据融合已成为了遥感的一个重要研究领域。异源数据匹配算法的性能直接影响获取地面控制点的精度, 匹配算法分为二阶段法和一阶段法, 现有的二阶段法难以适应地形复杂的遥感图像, 且在速度上无法满足实际的工程需求, 而速度满足要求的一阶段法在精度上仍有所欠缺。为解决这个问题, 提出了一个可端到端的高精度的基于残差伪孪生卷积互相关网络的异源遥感图像匹配算法。所提算法通过构建基于残差层的伪孪生网络, 对提取的SAR和可见光图像的特征进行卷积互相关操作从而实现异源遥感图像匹配。实验结果表明, 该算法在保持较高的速度下, 较大提升了SAR与可见光图像的匹配精度, 为深度学习方法在大规模异源遥感图像匹配任务中的工程应用奠定了基础。

关键词 遥感; 合成孔径雷达; 可见光; 异源匹配; 端到端

中图分类号 P407.8

文献标志码 A

DOI: 10.3788/LOP202259.1228002

Heterogeneous Remote Sensing Image Matching Algorithm Based on Residual Pseudo-Siamese Convolution Cross-Correlation Network

Zou Rongping^{1,2}, Zhu Bin^{1,2*}, Wang Chenyang¹, Zhu Yaoxuan^{1,2}, Hu Yangdi³

¹College of Electronic Engineering, National University of Defense Technology, Hefei 230037, Anhui, China;

²Key Laboratory of Infrared and Low Temperature Plasma of Anhui Province, Hefei 230037, Anhui, China;

³Army 32256 of PLA, Guilin 541000, Guangxi, China

Abstract Remote sensing satellites commonly use synthetic aperture radar (SAR) and visible light imaging. SAR and visible image data fusion have become an important research field of remote sensing owing to their high complementarity in imaging information. The accuracy of obtaining ground control points is directly influenced by the performance of a heterogeneous data matching algorithm. There are two methods of matching algorithms: two-stage and one-stage. The existing two-stage method is difficult to adapt to remote sensing images with complex terrain and it cannot meet the actual engineering needs in terms of speed, while the one-stage method meets the requirements in terms of speed but lacks in accuracy. To solve this problem, an end-to-end high-precision heterologous remote sensing image matching algorithm based on a residual pseudo-twin convolution cross-correlation network has been proposed. By constructing a pseudo twin network based on residual layer, the proposed algorithm

收稿日期: 2021-04-28; 修回日期: 2021-05-25; 录用日期: 2021-06-02

基金项目: 国家自然科学基金(61307025)

通信作者: *zhubin@nudt.edu.cn

performs convolution cross-correlation operation on the extracted features of SAR and visible images, so as to realize heterogeneous remote sensing image matching. The results show that this algorithm considerably improves the matching accuracy between SAR and visible images, maintaining a high speed and laying the foundation for the engineering applications of depth learning methods in large-scale heterogeneous remote sensing image matching tasks.

Key words remote sensing; synthetic aperture radar; visible light; heterogeneous match; end-to-end

1 引言

SAR 成像和可见光成像是空间遥感中常用的两种对地观测方式,两个波段差异极大,地物在这两个波段上的辐射反射特性有较大差异。SAR 成像具有全天时、全天候的独特优势^[1],在不良天气相较于可见光成像具有良好的抗干扰性,两者可以很好地在时空上进行信息的相互补充,所以 SAR-可见光的数据融合已经成为了遥感的一个相关研究领域^[2]。然而,SAR 独特的成像机制和复杂的电磁波散射过程,导致所采集图像含有大量相干斑噪声,且地物信息抽象、可读性差,而可见光图像正好与之相反,能够呈现色彩鲜明的地物且拥有高清晰的图像^[3],两者的匹配难度极大,数据融合的首要步骤就是匹配一定数量的同名像点来进行高精度的配准^[4],因此关于 SAR-可见光的高精度图像匹配的研究具有重大意义。

为了解决上述问题,相关领域的学者进行了大量研究。Suri 等^[5]提出了一种基于 SAR 图像和可见光图像之间互信息的方法来完成异源图像的匹配和配准的任务,该方法也被用于后来的二阶段匹配算法中。在此之后,为解决人工设计特征描述符的局限性,通过深度学习实现异源匹配的方法也逐渐发展。按照方法的步骤,可以将匹配的方法分为二阶段法和一阶段法,而最先发展成熟的是二阶段法。2015 年,Simo-Serra 等^[6]使用孪生网络的 128 维向量描述符通过相似度度量的方法进行图像匹配。2017 年,Mou 等^[7]将匹配作为二进制分类问题,通过训练孪生神经网络来得到相同尺度下 SAR 图像与可见光图像的相似度关系,从而进行匹配。2019 年,Bürgmann 等^[8]提出了一种通过衡量神经网络输出的特征向量信息来获取相似度测度进行匹配的方法。除使用相似度信息进行匹配的方法外,基于条件生成对抗网络生成人造同源图像后进行匹配的方法也被 Merkle 等^[9]应用于匹配任务中。随后一阶段法也开始发展,2020 年,Hughes 等^[10]提出了一种 SAR 与可见光图像匹配的端到端的框架,并在城镇

区域上实现了较高的准确率。上述方法虽然极大推进了异源遥感图像匹配领域的发展,并在城镇区域等部分场景中具有较高的精度和准确率,对数据融合领域有着较大的贡献,但是在面对部分特征稀疏或较为复杂的场景时,难以完成匹配任务。

本文为解决现有方法在部分场景下表现不佳的问题,提出了一个可端到端的高精度的基于残差伪孪生卷积互相关网络的异源遥感图像匹配算法。所提算法通过特征提取网络提取网络的深度特征,然后使用卷积互相关的方法进行端到端匹配,从而得到异源遥感图像之间同名点的对应关系,并通过难样本学习的方法来提升模型在复杂场景下的精度和稳健性。实验结果表明,所提算法可实现高精度的异源遥感图像匹配。

2 残差伪孪生卷积互相关网络设计

残差伪孪生卷积互相关网络(RPSCCNet)总体结构如图 1 所示,神经网络的数据流有两个输入分支以及一个输出总支。首先,SAR 图像(大小为 128×128 ,图 1 中记为 SAR)和可见光图像(大小为 256×256 ,图 1 中为 Optical)分别输入特征提取网络(第 1 层卷积层,图 1 中的 Conv 7×7 的卷积核在通道上的大小分别与两幅图像的通道数相等,其余的网络结构相等)。而通过特征提取网络提取出来的高维特征,经过降维(图 1 中的 Conv 1×1)后输出只有一个通道的特征图,为了使网络能够学习到异源图像之间的互相关性,将两个伪孪生神经网络分支的输出进行卷积互相关操作,进行有关相关性特征的提取。在卷积互相关网络中,孪生神经网络其中的 SAR 图像数据流分支的特征图经过变形(图 1 中的 Reshape 操作)得到一个卷积核,并在另外一个分支的特征图上进行卷积,最后得到两个特征图的卷积互相关图,得到最终的 SAR 图像关于可见光图像的相似度测度分布图。

RPSCCNet 的核心部分为特征提取网络 ResNet34,由 ResNet^[11]中的卷积层所组成,能够采用残差层来解决随着网络的加深而导致的网络退

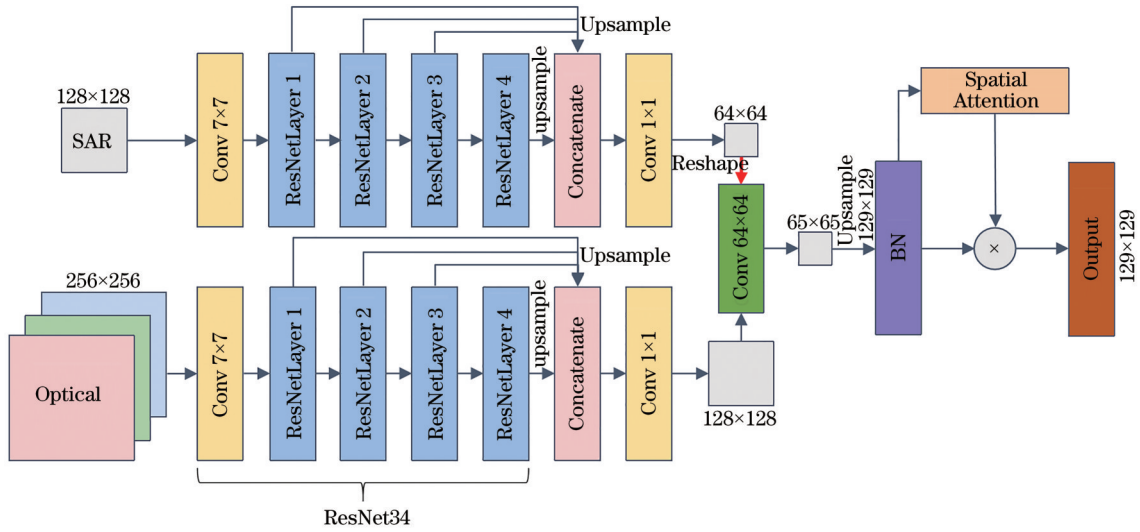


图 1 残差伪孪生卷积互相关网络结构

Fig. 1 Structure of residual pseudo-siamese convolution cross-correlation network

化问题(即网络的精度出现饱和甚至出现下降的现象),使得神经网络的层数能够尽可能的加深。为了使得网络能够学习到各层的特征,每个残差层(图 1 中的 ResNetLayer,右边数字表层数)输出的特征经上采样(图 1 中的 Upsample)至同一输入图像的一半大小并在通道上进行拼接(图一中的 Concatenate)。通过神经网络自动学习 SAR 图像和可见光图像之间相同的特征,为后面的操作奠定基础,是整个网络的基础组成部分,也是这个网络参数量最大的部分。为了保证网络的推理速度,也为了使得网络能够学习到对于每个通道所对应像素的重要性,即通过权重相乘的方式来表示每个像素的显著性,以类似于全连接层的方式^[12]通过 1×1 的卷积核学习在对应坐标位置上各个通道的权重,目的是为了能够先将特征提取网络得到的高维特征进行跨通道信息交互,并在保证特征图尺寸不变的同时不增大参数量。

然后使用 SAR 图像以及可见光两个分支数据流进行卷积互相关操作,其中构成卷积层的卷积核的是 SAR 图像经特征提取网络所得到的特征图,该操作主要通过类似于模板匹配的方式提取两个图像之间的互相关性^[13]。卷积的具体过程可描述为

$$\begin{cases} \mathbf{M}_S \rightarrow \mathbf{M}_S' \\ \mathbf{M}_T = \mathbf{M}_O * \mathbf{M}_S' \end{cases} \quad (1)$$

式中: \mathbf{M}_T 为输出图像的矩阵; \mathbf{M}_O 为可见光图像数据流分支的特征图矩阵; \mathbf{M}_S' 为 SAR 图像数据流分支的特征图 \mathbf{M}_S 变形后的矩阵; \mathbf{M}_T 为两者的卷积结果。 \mathbf{M}_T 矩阵各位置像素与 \mathbf{M}_O 和 \mathbf{M}_S 矩阵之间的对

应关系为

$$\mathbf{M}_T(x, y) = \sum_{i=x}^{x+w_s} \sum_{j=y}^{y+h_s} \mathbf{M}_O^{(i+w_s, j+h_s)} \mathbf{M}_S^{(w_s, h_s)},$$

$$x \in [0, w_o - w_s + 1], y \in [0, h_o - h_s + 1], \quad (2)$$

式中: $\mathbf{M}_{(x_0, y_0)}^{(x_1, y_1)}$ 是矩阵 \mathbf{M} 中点 (x_0, y_0) 到点 (x_1, y_1) 两个点之间所对应的所有矩阵元素所构成的矩阵; w_s 和 h_s 分别为 SAR 图像数据流分支特征图的宽度和高度; w_o 和 h_o 分别为可见光图像数据流分支特征图的宽度和高度。经过卷积后,可以得到宽为 $w_o - w_s + 1$ 、高为 $h_o - h_s + 1$ 的图像。

在卷积互相关操作结束后,为了提高网络的精度,需要对得到的特征图进行上采样,从而达到足够分辨率。最后为了加速网络的训练,在图像完成上采样后,将数据送入批归一化层(Batch normalization layer,图 1 简称为 BN)^[14]。随后使用空间注意力机制(图 1 中为 Spatial Attention)^[15]对特征图每个位置像素的显著性进行学习,该层的最终输出限定在 $[0, 1]$ 之间,通过与输入的乘积完成显著性特征提取,得到最终结果。

3 实验与数据分析

3.1 实验条件与评价方法

为了能够兼具场景的多样性和复杂性且使实验具有代表性的同时减少网络训练的成本,本实验组选择了 SEN1-2 数据集^[16]的夏季图片来制作网络所需要的训练集和测试集,训练集和测试集的比例为 0.95:0.05,并将 RPSCCNet 与其他 3 种算法进行了对比实验,分别为 Adapting HardNet^[8]、

PSiamNet^[7]、CorASLNet^[10]。在同一数据集上使用深度学习框架 Pytorch^[17],并采用原文中默认的参数对这 3 算法种进行了重新训练。

RPSCCNet 采用 Stochastic Gradient Descent (SGD) 优化方法进行训练,批大小为 25,学习率为 0.005,权重衰减 0.001,最大迭代次数 200。为了使得网络能够主动学习困难样本,在使用交叉熵作为损失函数的情况下,采取了 TopK Loss^[18]的方式进行梯度的反向传播,对 Top Loss 设定的阈值为 0.5。因为批大小为 25,所以采用前 13 个交叉熵损失最大的样本进行反向传播,这样可以保证网络能够主动学习困难样本。采用 L2 距离(L2 distance)、准确率(ACC)、热图质量(Heatmap quality)作为评价指标,L2 距离是网络推理所得中心点与实际中心点之间的欧几里得距离,单位为 pixel;准确率是 L2 距离误差小于设定阈值的样本数量与总样本量的比值;热图质量是网络推理结果与标签之间的交叉熵损失,数值越大,热图的质量越差,数值越小两者的分布差异越小。

$$H(O, L) = - \sum_i^x \sum_j^y O(i, j) \ln[L(i, j)],$$

$$i \in [0, h], y \in [0, w], \quad (3)$$

式中: O 为网络的推理结果; L 为理想分布图; $H(O, L)$ 为两者之间的交叉熵; $O(i, j)$ 为网络推理结果上 (i, j) 坐标对应的像素值; h 和 w 为图片的宽和高。

3.2 实验结果

3.2.1 四种方法结果的可视化与假设检验

在第 2 节中谈及了网络的相关设计理念和一些基于部分先验知识的假设,将其归于以下几点:

1) 伪孪生结构能够更好地提取异源数据之间的特征,能够较高提升网络的精度和稳健性。

2) 端到端的训练方式会使得网络的表现能够优于非端到端训练的网络。

3) 更好的特征提取网络能够让网络在精度上有更加优秀的表现。

4) 将空间注意力机制用于卷积互相关操作之后能够更好地抑制局部极大值。

5) 因为网络采用了难样本学习的方式进行训练,所以网络在对多种复杂地形有更高的精度与稳健性的同时,在简单样本上表现可能会稍微低于没有进行难样本学习的方法。

图 2 和表 1 罗列了部分实验数据,实验选取了

几类地形类别差异的图片,覆盖地形有森林、高原、河流、城市、水面、戈壁、农田、山川等,以此来对比检验模型的稳健性和精度并验证以上内容。

从图 2 可以看出,对比同为二阶段法的 PSiamNet 和 Adapted HardNet,采用孪生结构的 Adapted HardNet 的表现不太理想,采用伪孪生结构的 PSiamNet 的效果明显更好。PSiamNet 的地形适应能力明显更强,在热图质量和 L2 距离误差上都优于 Adapted HardNet。同时,其他两个一阶段法的网络 RPSCCNet 和 CoaASLNet 也有更好的表现,这也证明了采用伪孪生结构的网络能够更好地适应异源的数据,说明了伪孪生结构在异源数据上的优越性。

此外,对比一阶段法和二阶段法网络的热图质量和 L2 距离误差可以发现,CorASLNet 虽然在 L2 距离误差的均值上略低于 PSiamNet,但是在中值上高于 PSiamNet,这说明了 CorASLNet 推理失败的情况下误差较大,会极大地影响总体的评价指标。但是在采用中值进行评价的情况下,CorASLNet 又优越于 PSiamNet,这说明了大部分情况下 CorASLNet 的推理结果是较为不错的。此外,RPSCCNet 无论是在热图质量还是 L2 距离误差上都有更为优秀的表现,两个一阶段法的网络的总体表现是较为优秀的。这是因为两个一阶段法的网络都采用了端到端的方式进行训练,使得网络能够学习到两张异源图像的总特征信息。同时为了保证网络训练的收敛性对生成的结果进行了 Softmax 化,使得所有像素点的值的总和为 1,从而较大地提升了热图质量。这些正是一阶段法的网络能够优于二阶段法的重要原因。

但是不可否认的是,虽然 PSiamNet 推理的速度较慢,但是在[图 2(10)~(15)]的复杂地形下也拥有较低的 L2 距离误差,这得益于 PSiamNet 拥有网络参数量为 CorASLNet 的 8 倍的较强的特征提取网络。同时在相同种类的方法中进行对比可以发现,拥有更多参数量的方法表现更好,即二阶段法的 PSiamNet 及一阶段法的 RPSCCNet,这也证明了特征提取能力能够极大地影响网络的表现,拥有较强特征提取网络能够拥有更高的精度。

然后对比一阶段法的网络 CorASLNet 和 RPSCCNet,两者均采用了伪孪生的结构以及卷积互相关的方法,但是可以明显看出,RPSCCNet 的表现优于 CorASLNet。RPSCCNet 不仅在[图 2(1)~(3)]的常规地形中表现较好,也能够适应[图 2(8)~

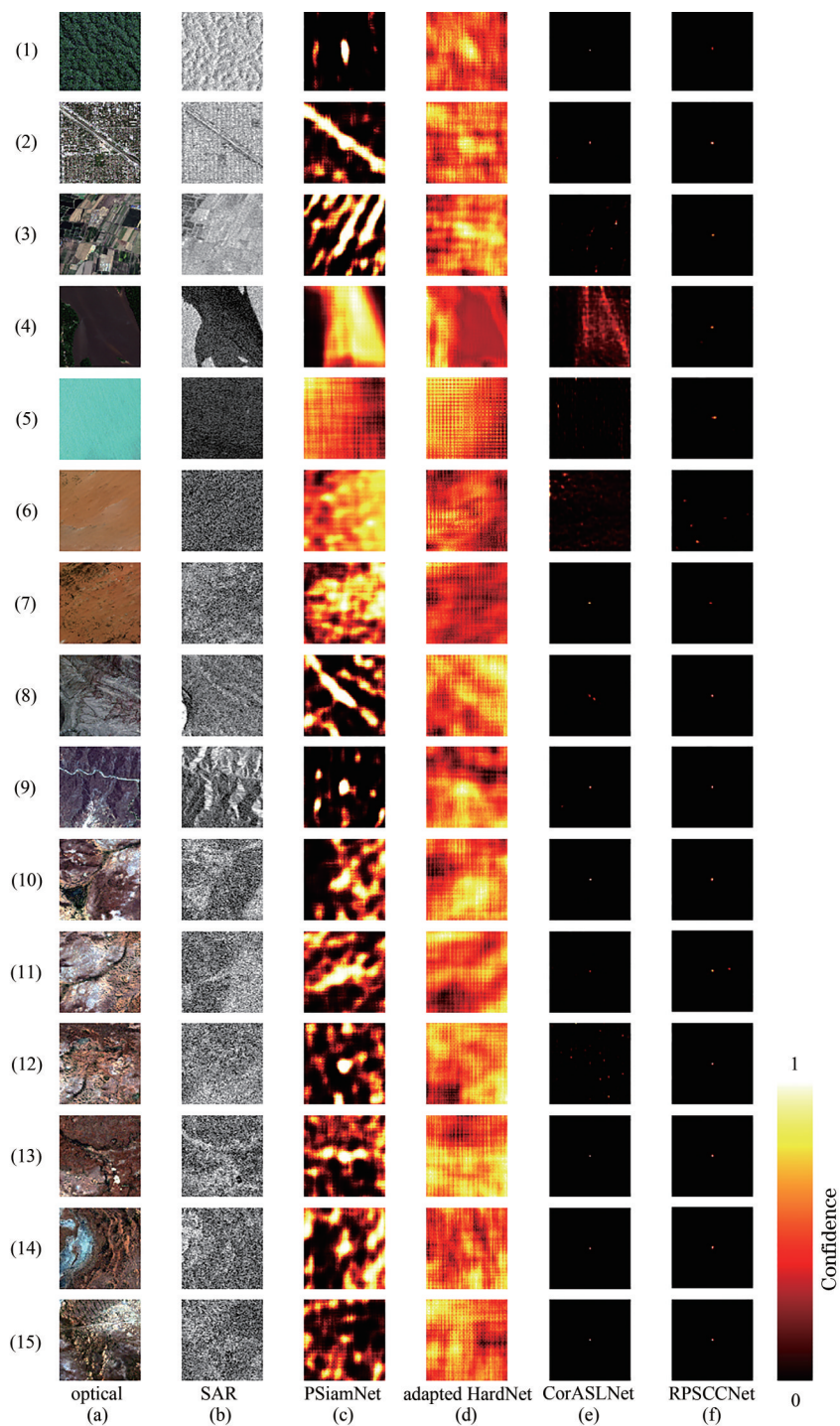


图 2 不同地形下不同网络的推理结果热图

Fig. 2 Heat map of inference results for different networks in different terrains

(15)]的复杂地形,甚至在[图 2(4)~(5)]的水面这类肉眼难以识别的地形下也有不错的表现。这是因为 RPSCCNet 拥有更好的设计理念,即采用了较强的特征提取网络使得网络拥有较强的特征提取能力及更为恰当的空间注意力机制的使用方式。图 3 是网络中间层特征的可视化。从图 3 可以看出,CorASLNet 的结构没有办法对局部极大值进行

有效的抑制,将空间注意力机制用于卷积互相关操作后的 RPSCCNet 很好地抑制了局部极大值,这也证明了该操作的合理性和有效性。

最后观察可以发现,如[图 2(1)、(2)、(7)、(11)、(13)、(14)、(15)]所示,在 CorASLNet 能够预测成功的大部分结果中,热图质量都高于 RPSCCNet,这里很大的一部分原因是因为

表 1 图 2 中对应归一化热图的质量与推理结果的 L2 距离误差

Table 1 Quality of corresponding normalized heat map and L2 distance error of inference result in Fig. 2

Terrain	Evaluating indicator	PSiamNet (c)	Adapted HardNet (d)	CorASLNet (e)	RPSCCNet (f)
Forest	Heatmap quality	2862.88	11939.88	0.27	28.14
	(1) L2 distance /pixel	1.00	6.08	0.00	2.00
Urban Area	Heatmap quality	11379.40	13830.62	7.08	11.78
	(2) L2 distance /pixel	3.00	4.24	0.00	0.00
Farmland	Heatmap quality	17970.94	15000.8073	60.99	27.42
	(3) L2 distance /pixel	4.47	5.00	45.69	2.00
River	Heatmap quality	18175.43	9731.15	1813.56	32.71
	(4) L2 distance /pixel	51.24	78.81	64.28	2.00
Water Surface	Heatmap quality	11373.06	14464.50	136.06	41.58
	(5) L2 distance /pixel	80.60	66.24	67.67	4.47
Gobi(flat)	Heatmap quality	17252.40	9498.53	474.13	54.71
	(6) L2 distance /pixel	45.18	32.57	69.07	53.66
Gobi	Heatmap quality	16796.47	7643.43	27.01	27.87
	(7) L2 distance /pixel	7.07	87.69	2.00	2.00
Hilly Area	Heatmap quality	13250.74	16251.06	35.37	2.66
	(8) L2 distance /pixel	3.16	66.67	10.00	0.00
River valley	Heatmap quality	4118.53	12901.60	4.73	1.53
	(9) L2 distance /pixel	3.00	8.60	0.0	0.00
Highland	Heatmap quality	9106.88	16033.42	8.69	4.92
	(10) L2 distance /pixel	1.00	29.83	0.00	0.00
Highland	Heatmap quality	11021.97	12959.89	30.17	33.03
	(11) L2 distance /pixel	9.21	63.97	2.00	2.00
Highland	Heatmap quality	7603.97	16355.28	68.42	1.44
	(12) L2 distance /pixel	5.00	86.37	84.85	0.00
Highland	Heatmap quality	8796.72	18123.40	0.39	2.86
	(13) L2 distance /pixel	2.23	69.77	0.00	0.00
Highland	Heatmap quality	13597.99	12910.65	0.93	4.24
	(14) L2 distance /pixel	3.60	83.63	0.00	0.00
Highland	Heatmap quality	5318.50	15361.87	0.48	2.00
	(15) L2 distance /pixel	70.17	74.81	0.00	0.00
Mean Value of Heatmap quality		11247.00	13532.41	177.88	18.46
Mean Value of L2 distance /pixel		19.33	50.95	23.03	4.54
Median Value of Heatmap quality		11373.49	13830.62	27.01	11.78
Median Value of L2 distance /pixel		4.47	66.24	2.00	0.00

RPSCCNet 采用了难样本学习的方法主动学习预测结果较差的样本,主动减少了对预测成功样本的学习,这使得其对于在对于简单样本的热图质量上有所下降。但是这些能够增强其对于难样本的学习,使其在综合表现上优于 CorASLNet,这也验证了难样本学习在难样本上表现较好,在简单样本上表现稍逊的特性。

综上所述,实验结果较好地验证了之前所提出

的 5 个论点,有效地论证了 RPSCCNet 结构设计的合理性。

3.2.2 四种方法的匹配精度

表 2 是不同网络对于测试集的总体推理结果,其中 $ACC(L2 \leq n \text{ pixel})$ 表示网络预测的中心点位置与标签的欧几里得距离小于 $n \text{ pixel}$ 的结果所占的比例。

从表 2 可以看到,无论是在哪个指标上,RPSCCNet 均取得了最优秀的结果,结果在误差

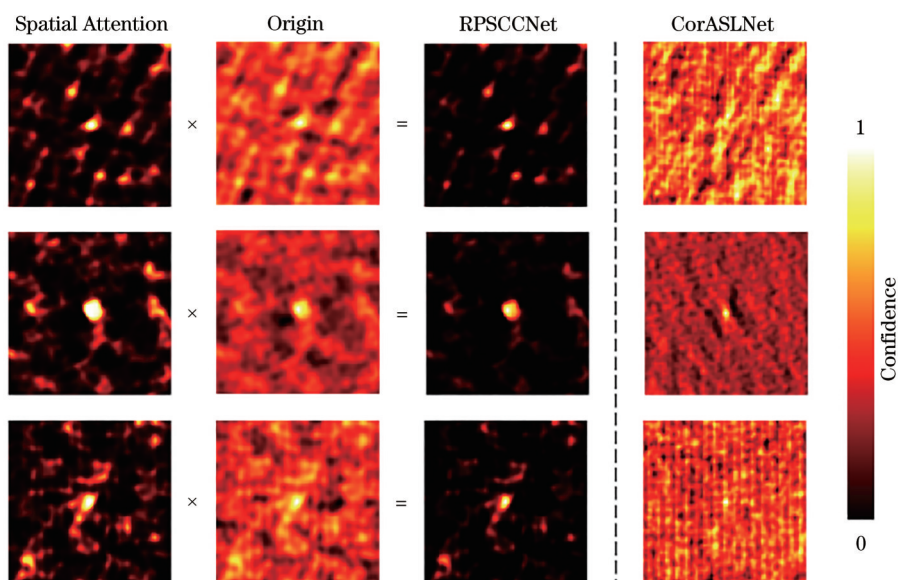


图3 RPSCCNet中间层特征热图以及两个网络进行Softmax操作之前的结果热图

Fig. 3 Heat map of middle layer features in RPSCCNet and heat map of result before Softmax operation for two networks

表2 四种方法在测试集的匹配精度和平均L2误差

Table 2 Matching accuracy and mean value of L2 distance of four methods in test set

Evaluating indicator	RPSCCNet	CorASLNet	PSiamNet	Adapted HardNet
Mean value of L2 distance /pixel	5.347	12.27	17.86	35.03
ACC(L2≤3 pixel) /%	91.03	79.65	55.32	16.72
ACC(L2≤2 pixel) /%	82.49	74.53	23.04	6.01
ACC(L2≤1 pixel) /%	55.49	52.47	6.50	1.75
ACC(L2=0 pixel) /%	17.37	17.02	0.00	0.00

3 pixel 内的比例更是达到了 91.03%，超过了 CorASLNet 的 79.65%，远高于 PSiamNet 和 Adapted HardNet 的 55.32% 和 16.72%。总而言之，RPSCCNet 具有最好的稳健性和最高的精度。

3.2.3 四种方法的性能对比

除了精度之外，网络的推理速度也是一个非常重要的指标，因为任何方法，如果无法应用到工程中就没有办法产生实际的价值。为了能够衡量网络推理的速度，采用 frame per second(FPS)比较可见光图片大小为 256×256、SAR 图片大小为 128×128 时的推理速度，其中，FPS(Network)表示网络自身推理速度，FPS(Single match)表示网络每秒完成图像匹配的次数，结果如表 3 所示，RPSCCNet 和

CoaASLNet 的完成单张图像匹配的速度远超过 PSiamNet 和 Adapted HardNet，这是两个网络同为端到端完成任务的性质所决定的。这也是两阶段法的局限性，即完成任务所需要的时间远远超过一阶段法，而这也是两阶段法无法适应大规模异源遥感匹配任务的一个重要原因。

RPSCCNet 在拥有最大的参数量(CorASLNet 的 17 倍)下，速度只降低为 CorASLNet 的 45.19%，这说明了特征提取能力强的网络可以保证网络拥有较高的精度，同时合理的设计结构能够保证速度不会下降太多。这说明了 RPSCCNet 具有优良的工程适应性，可以用较高的精度和不低的速度进行异源遥感图像的匹配。

表3 四种方法在 Nvidia RTX 2080ti 上的性能对比

Table 3 Performance comparison of four methods in Nvidia RTX 2080Ti

Parameter	RPSCCNet	CorASLNet	PSiamNet	Adapted HardNet
Size of Network /MB	85.40	4.96	41.00	5.16
FPS(Network)	82.88	183.37	165.46	364.82
FPS(Single match)	82.88	183.37	0.01	0.02

4 结 论

提出了一个端到端的实现 SAR 图像和可见光图像匹配的一阶段深度学习网络,使用 SEN1-2 数据集的数据对该网络进行训练,经过实验检验,对比 PSiamNet 和 Adapted HardNet 两阶段网络,该网络具有快速高精度的特性,与同为一阶段法的 CorASLNet 相比在各种类型的地形上网络具有稳健性,实现了对异源遥感图像的高精度端到端匹配,在匹配准确率上大大超过了以往的网络,该网络在水面等特征极少的地形甚至拥有超越人类的性能,并且在提高参数量的情况同时保证了较高的速度,这也使得 RPSCCNet 在异源遥感图像的匹配能够较高的工程应用价值,为今后大规模异源遥感图像匹配任务奠定了基础。同时该方法虽然较大提升了网络对于各类样本的处理精度,但是仍有部分预测失败的情况,相信在未来的一段时间里,能够通过工程上的手段例如条件约束等,将网络的稳健性实现进一步的提升。

参 考 文 献

- [1] 李莹莹, 吴昊, 常学立, 等. 基于星载光学和 SAR 影像的多视立体定位误差分析[J]. 光学学报, 2018, 38(5): 0528003.
Li Y Y, Wu H, Chang X L, et al. Multi-view stereo positioning error analysis based on spaceborne optics and SAR images[J]. Acta Optica Sinica, 2018, 38(5): 0528003.
- [2] Schmitt M, Tupin F, Zhu X X. Fusion of SAR and optical remote sensing data: challenges and recent trends[C]//2017 IEEE International Geoscience and Remote Sensing Symposium, July 23-28, 2017, Fort Worth, TX, USA. New York: IEEE Press, 2017: 5458-5461.
- [3] 韩子硕, 王春平, 付强, 等. 基于空间分离表征 GAN 的遥感图像模式互转[J]. 光学学报, 2021, 41(7): 0728003.
Han Z S, Wang C P, Fu Q, et al. Remote sensing image mode translation by spatial disentangled representation based GAN[J]. Acta Optica Sinica, 2021, 41(7): 0728003.
- [4] 谢志华, 刘晶红, 孙辉, 等. 可见光图像与合成孔径雷达图像的快速配准[J]. 激光与光电子学进展, 2020, 57(6): 062803.
Xie Z H, Liu J H, Sun H, et al. Fast registration of visible light and synthetic aperture radar images[J]. Laser & Optoelectronics Progress, 2020, 57(6): 062803.
- [5] Suri S, Reinartz P. Mutual-information-based registration of TerraSAR-X and ikonos imagery in urban areas[J]. IEEE Transactions on Geoscience and Remote Sensing, 2010, 48(2): 939-949.
- [6] Simo-Serra E, Trulls E, Ferraz L, et al. Discriminative learning of deep convolutional feature point descriptors[C]//2015 IEEE International Conference on Computer Vision, December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 118-126.
- [7] Mou L C, Schmitt M, Wang Y Y, et al. A CNN for the identification of corresponding patches in SAR and optical imagery of urban scenes[C]//2017 Joint Urban Remote Sensing Event (JURSE), March 6-8, 2017, Dubai. New York: IEEE Press, 2017: 16868046.
- [8] Bürgmann T, Koppe W, Schmitt M. Matching of TerraSAR-X derived ground control points to optical image patches using deep learning[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2019, 158: 241-248.
- [9] Merkle N, Auer S, Müller R, et al. Exploring the potential of conditional adversarial networks for optical and SAR image matching[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2018, 11(6): 1811-1820.
- [10] Hughes L H, Marcos D, Lobry S, et al. A deep learning framework for matching of SAR and optical imagery[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 169: 166-179.
- [11] 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, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [12] Lin M, Chen Q, Yan S C. Network in network[EB/OL]. (2013-12-16) [2021-01-02]. <https://arxiv.org/abs/1312.4400>.
- [13] Briechle K, Hanebeck U D. Template matching using fast normalized cross correlation[J]. Proceedings of SPIE, 2001, 4387: 95-102.
- [14] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]//Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, July 6-11, 2015, Lille, France. Cambridge: JMLR, 2015: 448-456.
- [15] Laskar Z, Kannala J. Context aware query image

- representation for particular object retrieval[M]// Sharma P, Bianchi F M. Image analysis. Lecture notes in computer science. Cham: Springer, 2017, 10270: 88-99.
- [16] Schmitt M, Hughes L H, Zhu X X. The SEN1-2 dataset for deep learning in SAR-optical data fusion [EB/OL]. (2018-07-04) [2021-01-03]. <https://arxiv.org/abs/1807.01569>.
- [17] Paszke A, Gross S, Massa F, et al. PyTorch: an imperative style, high-performance deep learning library[EB/OL]. (2019-12-03) [2021-02-01]. <https://arxiv.org/abs/1912.01703>.
- [18] Zhang K P, Zhang Z P, Li Z F, et al. Joint face detection and alignment using multitask cascaded convolutional networks[J]. IEEE Signal Processing Letters, 2016, 23(10): 1499-1503.