引文格式：高玉慧,朱武,张锐煊,等. D-Net：一种动态模拟道路形态的路网提取模型[J]. 地理空间信息, 2024, 22(6): 25-28. doi:10. 3969/j. issn. 1672-4623.2024.06.006

D-Net：一种动态模拟道路形态的路网提取模型

高玉慧[1]，朱　武[2*]，张锐煊[2]，王　越[2]
（1. 云南省交通投资建设集团有限公司，云南 昆明 650103；2. 长安大学 地质工程与测绘学院，陕西 西安 710054）

摘　要：对于当前利用深度学习技术实现遥感影像中道路提取时，由于网络中的矩形卷积而导致无法捕捉道路形态特征、提取精度低等问题。基于此，提出一种基于形变卷积的卷积神经网络（D-Net，Deformation Network），该网络旨在从遥感影像中准确提取道路区域。为了验证模型的性能，在马萨诸塞州道路数据集和DeepGlobe道路数据集上进行了测试。此外，还选择了一景GF-2影像对模型的泛化能力进行评估。实验结果显示，D-Net显著提升了道路分割的精度，为通过形变卷积优化提取结果提供了有力支持。该研究的开展对于进一步探索道路智能提取具有重要的理论和实践指导意义。

关键词：形变卷积；深度学习；遥感技术；道路提取
中图分类号：P237　　**文献标志码：**B　　**文章编号：**1672-4623（2024）06-0025-04

D-Net: A Road Network Extraction Model for Dynamical Simulation of Road Morphology

GAO Yuhui[1], ZHU Wu[2], ZHANG Ruixuan[2], WANG Yue[2]
(1. Yunnan Communications Investment & Construction Group Co., Ltd., Kunming 650103, China;
2. School of Geological Engineering and Geomatics, Chang'an University, Xi'an 710054, China)

Abstract: In the current road extraction from remote sensing images using deep learning techniques, the issue arises due to the use of rectangular convolutions in the network, leading to difficulties in capturing road morphology features and low extraction accuracy. We proposed a convolutional neural network based on deformable convolution (D-Net), aiming to accurately extract road areas from remote sensing images. To validate the model's performance, we conducted experiments on the road datasets of Massachusetts and DeepGlobe. Additionally, we chose a GF-2 satellite image to evaluate the model's generalization ability. Experimental results demonstrate that D-Net significantly improves road segmentation accuracy, providing strong support for optimizing extraction results through deformable convolutions. The progress of this research holds significant theoretical and practical implications for further exploration in intelligent road extraction.

Key words: deformable convolution, deep learning, remote sensing technology, road extraction

随着我国社会进程的飞速发展，基础设施建设达到了空前的水平。其中，道路作为一种连续的线性地物，在社会发展和城市建设中具有重要的地位[1]。准确获取当前路网覆盖情况一直是一大热门研究领域[2]。在过去，对于道路的制图工作通常采用实地勘察的方式，这些受到工作量大、人工成本高的限制。近年来，随着遥感技术的不断发展，以非接触方式对道路进行解译逐渐取代传统外业[3]。相关研究以像素或是对象作为最小单元，进行道路识别[4-5]。然而，这些方式的成功与否很大程度取决于先验知识，且较难设计一种具备普适性的模型[6]。深度学习技术的快速发展为该领域带来了新的活力，一种以数据驱动为手段的研究方式逐渐成为主流[7]。在此过程产生许多优异的模型，但是，仍会存在受阴影遮挡而混淆地物的误判，以及缺乏对上下文信息捕捉而漏判的现象。具体表现为，现有研究中模型的感受野会受到规则卷积方式的影响，模型获取的特征仅来自于中心像素的周边邻域范围，并不能较好地关联全局信息，将语义上下文特征更好表达。基于此，本文结合形变卷积[8]，通过在卷积中引入偏移项，一次动态调整模型在编码过程中对影像不同地物的关注度，设计在训练过程可动态变化的卷积神经网络（D-Net，Deformation Network），并将其应用在道路提取任务中。D-Net结构上采用编解码器对称的方式，在编码模块中添加形变卷积，并以跳跃

收稿日期：2024-03-19。
项目来源：国家重点研发计划项目（2020YFC1512000）；陕西省杰出青年基金项目（2023-JC-JQ-24）。
第一作者简介：高玉慧（1969—），本科，高级工程师，从事高原公路养护智慧化等工作，E-mail：1059216837@qq.com。
通信作者：朱武（1982—），博士，教授，研究方向为影像大地测量学，E-mail： zhuwu@chd.edu.cn。

连接的方式沟通不同层级的特征信息。实验部分，以公开数据集和实际案例证明模型的可行性。

1 研究方法

1.1 形变卷积

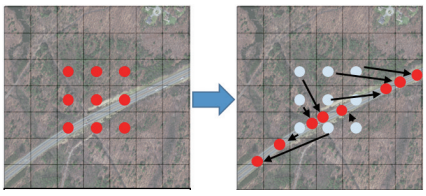
卷积是卷积神经网络的基础，常规卷积形式可被公式表示为[9]：

$$O_{(,,z)}=\sum_{k=1}^{C}(M_{(,,k)}*F_{(,,k,z)})+\boldsymbol{B}_{(,,z)} \tag{1}$$

式中，$O_{(,,z)}$ 为输出的第 z 通道特征图；$M_{(,,k)}$ 为输入的第 k 通道特征图；$*$ 为卷积运算；$F_{(,,k,z)}$ 为卷积核；$\boldsymbol{B}_{(,,z)}$ 为偏置矩阵；C 为输入的总通道数。

因此，若卷积核尺寸为2，即卷积掩膜构成了一个2×2网格。可将网格范围表示为 $\text{Range}\in\{(i,i),(0,i+1),(i+1,0),(i+1,i+1)\}$（$i$ 表示掩膜左上起始点）。采用此类方式进行卷积，会存在网络探测受限的情况，即模型在对道路形态进行学习时只能学习到中心像元周边四邻域（或八邻域）的内容。因此对于道路这类具备连续特征的地物，并不能较好满足[10]。

在卷积中引入可学习的偏置项，以此使网络在训练过程可以将更多的注意力聚焦在道路区域。具体而言，当引入偏置后，卷积掩膜由原本范围变为 $\text{Range}'\in\text{Range}+\Delta b$（$\Delta b$ 为可学习的偏置）。图1展示了2种卷积的差异。可以看出形变卷积具备更高的学习能力，因此以该方式为基础构建网络模型具有一定的理论基础和依据。



a 常规卷积　　b 形变卷积

图1 卷积方式差异的可视化

1.2 D-Net网络结构

在网络模型设计上参照UNet模型结构[11]，结合形变卷积，提出一种编解码器对称的动态分割网络，形变网络（D-Net）。模型结构如图2所示。D-Net整体由编码器和解码器组成。编码器在网络前向传播过程中执行特征提取操作，即不断压缩数据的二维尺寸，同时拓展其特征维度。在数据特征上表现为将浅层的表征信息聚焦为深层的语义信息。解码其则是将网络在编码过程中产生的高密度的特征图逐级解包，恢复至原影像大小，实现端到端、像素到像素的训练过程[12]。

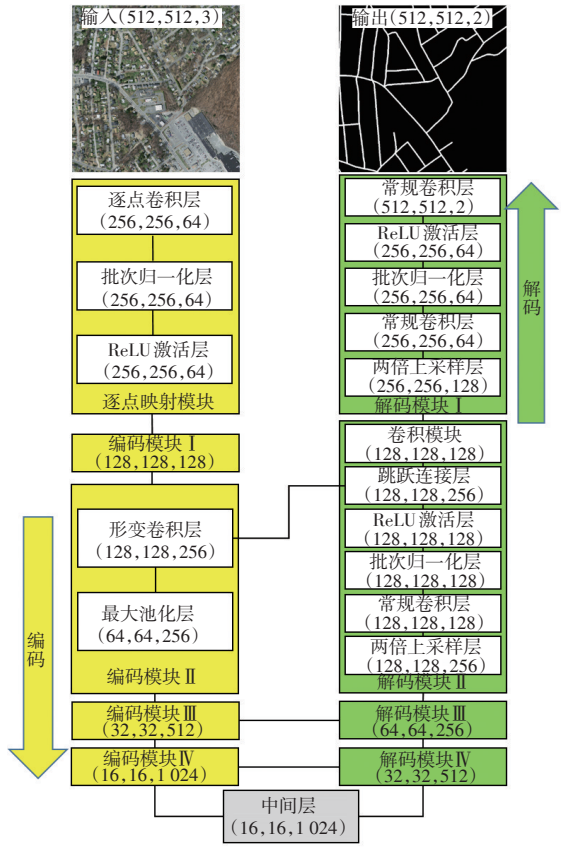


图2 D-Net结构示意图

2 实验设计

2.1 评价指标

基于二分类混淆矩阵计算，采用交并比（IoU）和 F_1-score 作为评价指标，对模型的精度进行定量化评估。IoU衡量了预测分割掩码与真实分割掩码之间的重叠程度，是语义分割任务中常用的评价方式。取值范围在[0,1]，数值越大说明预测结果越贴近真值。具体来说，IoU是通过计算预测分割区域和真实分割区域的交集与它们的并集之比来定义。准确率是指预测为正类别的样本中实际为正类别的比例，召回率是指实际为正类别的样本中被正确预测为正类别的比例。F_1-score是准确率和召回率的调和平均，它在平衡了两者之间的关系后给出了一个单一的指标，是一种常用的二分类评价方式。2种指标对应公式如下：

$$\text{precision}=\frac{TP}{TP+FP} \tag{2}$$

$$\text{recall}=\frac{TP}{TP+FN} \tag{3}$$

$$F_1\text{-score}=\frac{2\times\text{precision}\times\text{recall}}{\text{precision}+\text{recall}} \tag{4}$$

$$\text{IoU}=\frac{TP}{TP+FP+FN} \tag{5}$$

式中，TP 为正确分割为道路的像元数；FP 为识别

为道路但标记为背景的像元数；*FN* 为提取作为背景但标记为道路的像元数。

2.2　公开数据集

采用马萨诸塞州道路数据集[15]和DeepGlobe道路数据集测试模型的精度。马萨诸塞州道路数据集为航拍数据，覆盖了美国马萨诸塞州超过2 600 km²的地理区域，包括城市、城镇、农村和山区等多种地形类型的道路信息。每张图像的尺寸1 500，其空间分辨率约为1 m。该数据集包含1 108张训练样本、49张测试样本以及14张验证样本。DeepGlobe道路数据集中图像尺寸为1 024，分辨率0.5 m。标签中白色代表道路，黑色代表背景。该数据集共有8 579张图像，其中训练用图像6 226张。由于DeepGlobe道路数据集中测试和验证集未提供相应标签，因此，的模型测试是在马萨诸塞州道路数据集全集和DeepGlobe道路数据集训练集完成的。对于马萨诸塞州道路数据集，首先将其尺寸采样至1 024，保证与DeepGlobe道路数据集一致。随后，考虑实验设备性能，将其裁剪为512尺寸的图像块。最终，将马萨诸塞州道路数据集构建为图像尺寸512，训练集、验证集、测试集比例4 432∶56∶196。对于DeepGlobe道路数据集，将其中6 226个训练样本以同样方式裁剪为512大小。按照7∶2∶1的比例，最终生成了一个训练集、验证集、测试集比例14 267∶4 076∶2 039的数据集。

2个公开数据集的训练集、验证集、测试集将分别执行训练、验证和评估模型的性能的任务。具体而言，训练集用于训练模型的参数和权重。模型通过与训练集中的样本交互，学习特征之间的关系，并调整自身参数以最小化损失函数。验证集用于监控模型在训练过程中的性能变化并决定本次迭代权重是否保留。通过在验证集上评估模型的表现，可以及时发现模型是否过拟合或欠拟合。测试集用于最终评估模型的性能。

2.3　实验细节

采用Pytorch框架完成对整个实验的代码实现，并采用NVIDIA GeForce RTX 3090（24 G）完成实验。在预设参数设计上，将总迭代次数设置为150，批次数设置为4，学习率设置为0.000 1，采用Adam优化器和学习率余弦退火算法优化训练过程。Adam优化器根据每个参数的梯度的一阶矩估计和二阶矩估计来调整学习率，是一种常用于训练网络的优化算法，它结合了动量法和自适应学习率方法。学习率余弦退火算法是一种用于优化神经网络训练过程中学习率的调度方法。它通过在训练过程中逐渐降低学习率的大小来帮助模型更好地收敛到最优解。该算法的思想是模拟余弦函数的曲线，即学习率在训练过程中以余弦形式逐渐减小。同时，使用交叉熵作为损失函数帮助网络学习道路特征。交叉熵对概率分布敏感，是分类（分割）任务中常见损失函数。可以帮助模型更好地区分不同类别，并且能够促使模型输出的概率分布接近真实分布，从而提高分类准确性。

3　分析与讨论

3.1　公开数据集表现

D-Net在公开数据集中的测试结果如表1所示。观察表1可看出，在2种公开数据集中D-Net具有一个相对稳定的表现能力，可以实现对道路区域的稳定识别。

表1　D-Net实验结果评价

数据集	IoU/%	F_1-score/%
马萨诸塞州道路数据集	62.63	77.02
DeepGlobe道路数据集	62.66	77.05

从马萨诸塞州道路数据集和DeepGlobe道路数据集中各选取2个典型案例，并进一步展示了D-Net对道路识别中的一些细节。从结果中案例Ⅰ和Ⅱ来自马萨诸塞州道路数据集，其余2个来自DeepGlobe道路数据集。每个案例依次展示网络输入图像、用于展示网络聚焦状态的梯度-类激活图（Grad-CAM）、D-Net提取效果、以及将提取结果叠置到影像中的结果。结果叠置图中黄色表示被正确分类的目标像元，蓝色表示误判像元、红色表示漏判像元。通过观察结果可知，D-Net在从遥感影像提取道路这项任务中的良好表现，验证了其一定的生产价值。

具体而言，观察案例Ⅰ可知，D-Net可以准确关注路网的整体形态，实现对图像中道路区域的整体提取。这一现象在案例Ⅱ中也得到了印证。然而，案例Ⅰ和Ⅱ中也同样反映了D-Net的2点局限性。首先，通过观察案例Ⅰ中的叠加结果可以看出存在一定的误判现象（蓝色标注）。该类误判在提取结果上表现为D-Net将未在标签（真值）中标记的区域识别为道路。可以看出该现象主要出现在建筑物的内部，例如建筑物的间隙。这可能是由于D-Net缺少对全局信息的学习，因此，当地物在小范围内表现出道路特征时，模型就会将其识别并分割出。其次，通过观察案例Ⅱ中的叠加结果可以看出存在少部分错判现象（红色标注）。该类错判在提取结果表现为D-Net未能将标签中标记的道路区域识别完全。该类问题主要集中在道路尽头或连接处，即道路

特征发生一定改变的时候。该问题与案例Ⅰ所呈现问题在原理上可能一致，因此，在后续的研究中，可以将模型搭建的重点放置在扩展网络的感受野上，采用空洞卷积等离散的卷积方式来改善学习效果。上述情况在DeepGlobe道路数据集上也有所表现。

3.2 现实场景测试

为测试模型的泛化能力，选取一景国产高分辨率遥感影像（GF-2），并对其全图幅进行道路提取应用。GF-2卫星是国内自主研制的高空间分辨率卫星。实验所选取的GF-2影像对应具体参数如表2所示。

表2 影像参数

时间	分辨率	传感器	中心坐标
2021-01-19	3.24 m	PMS2	117.6°E，34.7°N

实验结果表明，D-Net可以完成对于研究区域中的路网的整体提取。模型所产生的识别结果可以用作路网初级产品，为后续进一步加工提供基础。

具体而言，D-Net成功的获取了影像全景范围的道路区域。并且，在提取结果上极大地保证了道路区域的连贯性。这得益于D-Net模型内部所采用的可在训练过程中依据可训练的形变卷积偏置项，而产生的对道路区域的高度关注。

然而，进一步分析实验结果。D-Net仍然有一些可以进一步优化的方面。首先，模型对于群落密集的城镇区域识别能力有限。当建筑物密度增加时，由于建筑物与道路具备同质纹理，当建筑物在影像中也展现一定的线性特征时，D-Net极大程度会将其余道路区域视为整体，不进行识别。在提取结果上表现为，在城区中解译出的道路区域会有所下降，相较于建筑低密度区域。其次，对于一些其他的非道路且具备线性特征的地物缺乏区分能力，表现为误判现象。例如，在影像中的山区，D-Net会将部分山地边界识别为道路。

4 结 论

本文以形变卷积为模型改进切入点，设计并应用一种编解码器对称的卷积神经网络（D-Net），实现遥感影像中道路区域的全面提取。同时，本研究也揭示了D-Net在广域感受野探测上的局限性。后续，将继续针对动态学习过程对模型作进一步改进，以生产更高精度的路网产品。

参考文献

[1] 刘洋,康健,管海燕,等. 基于双注意力残差网络的高分遥感影像道路提取模型[J]. 地球信息科学学报,2023,25(2):396-408

[2] 何哲,陶于祥,罗小波,等. 基于改进U-Net的遥感图像道路提取[J]. 激光与光电子学进展,2023,60(16):385-392

[3] 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

[4] 阎逸飞. 基于缓冲区的道路中心线提取方法与应用[J]. 地理空间信息,2023,21(11):36-38

[5] 徐翰文,艾波,史绍雨. 基于纹理特征融合与核模糊聚类的道路提取[J]. 地理空间信息,2019,17(11):92-95

[6] 孔令然,刘立志,吴优,等. 一种基于特征点过程的遥感影像道路中心线提取算法[J]. 无线电工程,2023,53(2):340-351

[7] 龙伊娜,谷玉海,吴文昊,等. 基于改进D-Linknet的高分遥感影像道路提取方法[J]. 激光杂志,2023,44(5):162-168

[8] Dai J,Qi H,Xiong Y,et al. Deformable Convolutional Networks[C]//Proceedings of the IEEE International Conference on Computer Vision,2017:764-773

[9] Zhang R, Zhu W, Li Z, et al. Re-Net:Multi-branch Network with Structural Reparameterization for Landslide Detection in Optical Imagery[J].IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2024,17:2 828-2 837

[10] Zhang R,Zhu W,Li Y,et al. D-fusion Net:Road Extraction from Remote Sensing Images Using Dilated Convolutional Block[J]. GIScience & Remote Sensing, 2023, 60(1):557-559

[11] Ronneberger O, Fischer P, Brox T. U-net:Convolutional Networks for Biomedical Image Segmentation[C]//Medical Image Computing and Computer-assisted Intervention-MICCAI 2015:18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, Springer International Publishing,2015:234-241

[12] Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2015:3 431-3 440

[13] Ioffe S, Szegedy C. Batch Normalization: Accelerating deep Network Training by Reducing Internal Covariate shift[C]//International Conference on Machine Learning. Pmlr, 2015: 448-456

[14] Nair V, Hinton G E. Rectified linear Units Improve Restricted boltzmann machines[C]//Proceedings of the 27th International conference on machine learning (ICML-10). 2010: 807-814

[15] Mnih V. Machine learning for Aerial Image Labeling[M]. University of Toronto (Canada),2013