基于神经网络的遥感图像语义分割方法

王恩德[1,2,3]，齐凯[1,2,3,4*]，李学鹏[1,2,3]，彭良玉[1,2,3]

[1]中国科学院沈阳自动化研究所，辽宁 沈阳 110016；

[2]中国科学院机器人与智能制造创新研究院，辽宁 沈阳 110169；

[3]中国科学院光电信息处理重点实验室，辽宁 沈阳 110016；

[4]东北大学信息科学与工程学院，辽宁 沈阳 110819

摘要 为了提高遥感图像语义分割的效果和分类精度，设计了一种结合 ResNet18 网络预训练模型的双通道图像特征提取网络。将多重图像特征图进行拼接，融合后的特征图具有更强的特征表达能力。同时，采用批标准化层和带有位置索引的最大池化方法进一步优化网络结构，提升地表目标物的分类准确率。通过实验，将所提方法与多种神经网络方法进行准确率和 Kappa 系数比较。结果显示，所提的网络结构可以在小数据量样本下取得 90.68%的总体准确率，Kappa 系数达到了 0.8595。相比其他方法，所提算法取得了更好的语义分割效果，并且整体训练时间大幅缩短。

关键词 图像处理；全卷积神经网络；语义分割；双通道网络；多尺度特征；遥感图像

中图分类号 TP183 **文献标识码** A **doi**：10.3788/AOS201939.1210001

Semantic Segmentation of Remote Sensing Image Based on Neural Network

Wang Ende[1,2,3], Qi Kai[1,2,3,4*], Li Xuepeng[1,2,3], Peng Liangyu[1,2,3]

[1] *Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, Liaoning* 110016, *China*;

[2] *Institute for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang, Liaoning* 110169, *China*;

[3] *Key Laboratory of Opto-Electronic Information Processing, Chinese Academy of Sciences, Shenyang, Liaoning* 110016, *China*;

[4] *College of Information Science and Engineering, Northeastern University, Shenyang, Liaoning* 110819, *China*

Abstract To improve the effect and classification accuracy of semantic segmentation of remote sensing images, a two-channel image feature extraction network combining with ResNet18 pre-training model is designed. Images with multiple features are combined, and the combined feature map has stronger ability to express features. At the same time, batch normalization layer and maximum pooling with location index are adopted to optimize the network structure and improve the classification accuracy of surface object. The accuracy and Kappa coefficient of this method are compared with those of other neural network methods by experiments. The results show that the proposed network structure achieves an overall accuracy of 90.68% when the number of data samples is small, and the Kappa coefficient reaches 0.8595. Compared with other methods, the proposed algorithm achieves better semantic segmentation effect, and greatly reduces the overall training time.

Key words image processing; fully convolutional neural network; semantic segmentation; two-channel network; multiscale feature; remote sensing image

OCIS codes 100.4996; 100.2960; 100.3008

1 引 言

近年来，随着科学技术的发展进步，人类探索宇宙空间的能力得到了大幅提高，越来越多的卫星被送入到绕地球运行的轨道[1]。遥感卫星可以拍摄高分辨率的地球遥感图像。遥感图像中蕴含丰富的地

收稿日期：2019-07-09；**修回日期**：2019-07-18；**录用日期**：2019-08-19

* **E-mail**：qiqikai123456@163.com

理信息。如何对遥感图像进行目标识别和语义信息提取，逐渐成为图像分析领域的热点研究内容之一[2]。其中，语义分割研究被广泛应用于无人驾驶[3]、医疗影像分析等任务中。语义分割[4]是对图像中的目标进行像素级别的分割，为不同类别目标的所有组成的像素进行对应类别的颜色标注，本质上是对图像中的不同类别目标进行分类。在遥感图像中，语义分割指的是对图像的地表物体目标(包括河流、土地、建筑物等)进行分类和颜色标注。这对于很多目标识别任务来说，都是一项重要的基础工作。然而，像素级别的分类方法往往对噪声非常敏感，如果缺少目标的语义信息，很难获取目标的分类信息。地表包含的目标类型多样，容易受噪声、季节、光照等因素的影响，这给高分辨率遥感图像的目标分类任务带来了很大困难[5]。

学者们尝试了很多方法对高分辨率遥感图像进行目标分类。Blanzieri 等[6]采用支持向量机(SVM)的方法进行遥感图像目标物分类；Kluckner 等[7]采用非监督聚类算法对遥感图像中的房屋进行分割；Chen 等[8]改进传统的边缘检测方法，使遥感图像中的小物体也能被分割出来。但是，由于遥感图像包含丰富的光谱信息，传统的特征提取方法并不能取得很好的分割效果。从模式识别的角度来看，典型特征的选取是提高识别精度的瓶颈[9]。只使用一组特定的特征无法对所有类型的地面物体进行准确分类。所以，采用学习方法对相应数据集中的特征进行自动学习分类，相比人工设计特征，可以更有效地提高目标分类精度[10]。

神经网络研究方法由于 Hinton 等[11]提出深度学习理论而备受关注。深度学习的基本动机是建立一个深度神经网络来模拟人类大脑的学习和分析机制。与传统的机器学习算法相比，深度学习更加强调从庞大的数据中通过多层神经元组织自动学习特征。典型的深度学习结构包括递归神经网络(RNN)[12]、深度信念网络(DBN)[13]、卷积神经网络(CNN)[14]等。CNN 在图像分类、目标识别等计算机视觉任务上都取得了显著的效果，并且在 ImageNet、PASCAL VOC 等领域内权威数据集的竞赛中取得了优异成绩[15]。2012 年，Krizhevsky 等[16]研究设计了一种 7 层 CNN 的模型(被命名为 AlexNet)，赢得了 ILSVRC(ImageNet Large Scale Visual Recognition Challenge)竞赛的冠军。也有许多学者基于 CNN 的方法并针对遥感图像进行语义分析研究。Nguyen 等[17]提出了一种 5 层的网络结构，完成了遥感图像的目标分类工作；Hu 等[18]利用一种预训练的 CNN 模型对不同的遥感图像场景进行分类；Mnih[19]提出了一种基于 CNN 的航空图像大尺度上下文特征学习结构，但效果仍待提升。由于高分辨率图像的像素量巨大，因此要实现逐像素的分类十分困难，当前的像素级目标分类精度并不理想。

为了提升遥感图像的语义分割效果，本文设计了一种双通道不同尺度特征提取融合的全卷积神经网络(FCN)结构，并对高分辨率卫星遥感图像进行语义分割，以保证在提取整体特征的同时不丢失目标的细节特征。通过将图像目标的整体与细节特征相结合，提升了目标分类的准确率。同时，在池化层中采用最大池化索引解码的池化方式，来保留目标的边缘等细节信息，从而有效提高网络的识别效果。利用 ResNet18(Residual Neural Network)[20]预训练模型，结合 ResNet18 预训练得到的神经网络权重参数来获取遥感图像的特征图，并将其与文中双通道获得的特征图再次融合，增强了特征的表达能力，在保证整体网络结构训练效果的同时，大幅缩短了网络训练时间。

2　相关工作

得益于计算机硬件的发展，在图像多目标分类相关研究中，神经网络方法得到了蓬勃发展，其间出现了 CNN、FCN、SegNet、ResNet、空洞卷积[21]等一系列优秀的网络结构和方法。

2.1　CNN 基本方法

CNN 是当前目标分类和目标检测领域采用最多的方法。传统的 CNN 模型在网络层的最后部分都会加上若干个全连接层，用于目标类别的划分。FCN 最早由 Shelhamer 等[22]在 2017 年提出，如图 1 所示。FCN 将传统 CNN 中的全连接层替换为卷积层，以减少网络结构中的参数，加速网络训练进程。同时，采用全卷积的结构，使得输入图像的尺寸不再受约束。

2.2　ResNet 网络结构

ResNet(Residual Network)网络结构由 He 等[20]在 2016 年提出，此论文获评 2016 年 CVPR 会议(Conference on Computer Vision and Pattern Recognition)最佳论文，对应的模型在 ILSVRC2015 比赛中夺得冠军。ResNet 网络结构的参数比 VGGNet[23]更少，进一步简化了网络的结构。

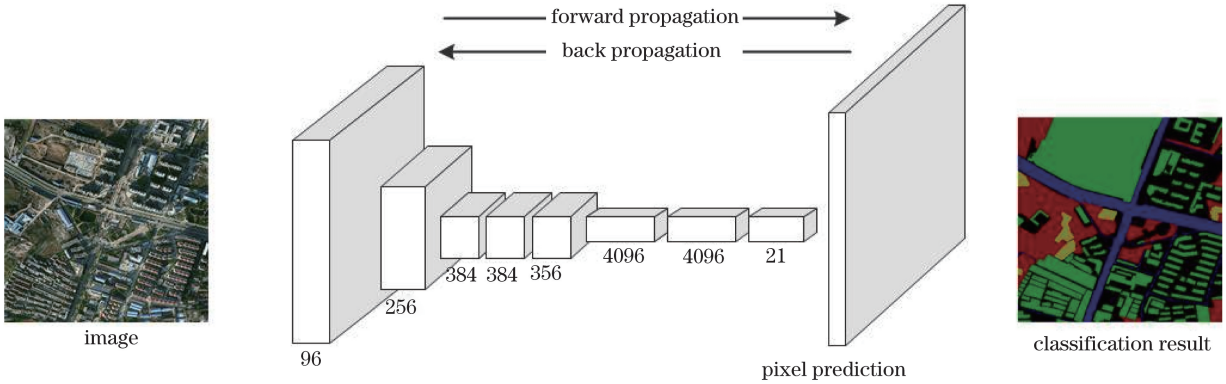


图 1　FCN 结构示意图

Fig. 1　FCN structural diagram

从传统意义上来讲，网络的深度对图像最后的分类和识别效果有很大影响[24]，所以常规的想法是网络设计得越深越好；但是当网络堆叠很深时，效果却会变差。通过分析效果变差的原因，发现梯度消失是产生此问题的主要因素。ResNet 可以很好地解决网络层深度增加时梯度消失的问题。

图 2 展示了 ResNet 的基本结构单元。ResNet 提出了两种映射（mapping）：一种是恒等映射（identity mapping），指的是图 2 中弯曲的线；另一种残差映射（residual mapping），指的就是除弯曲的线外的部分。最后的输出为

$$y=F(x)+x, \tag{1}$$

式中：y 为残差映射与恒等映射相加的神经元输出；x 为输入的神经元参数；$F(x)$ 为残差映射的神经元参数。

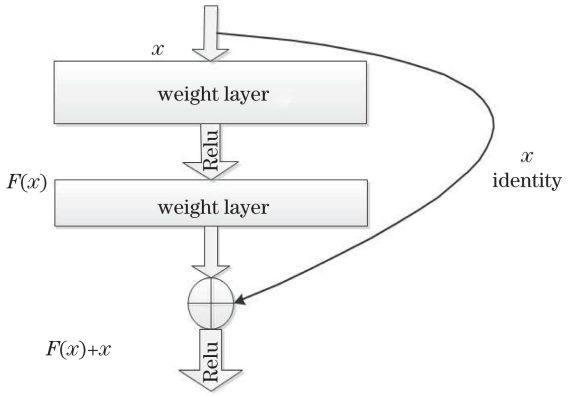
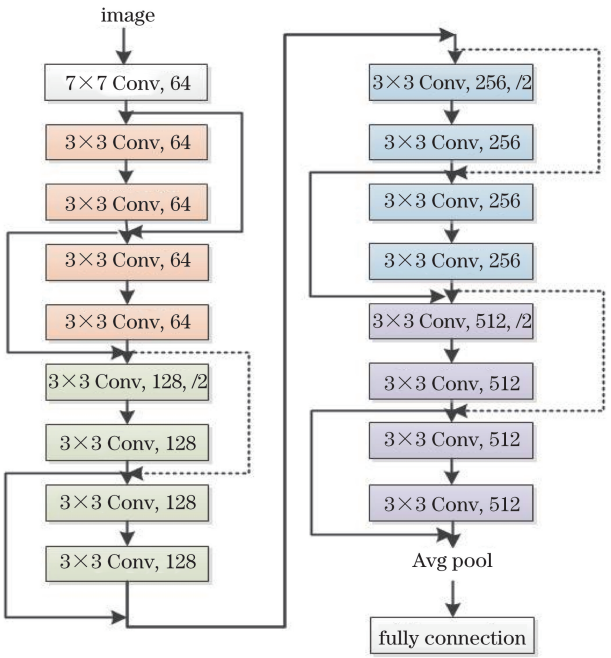


图 2　ResNet 基本结构单元

Fig. 2　ResNet basic structural unit

恒等映射从字面意思理解就是指其本身，也就是(1)式中的 x（即神经元中的参数）；而残差映射指的是差，即 $y-x$，所以残差指的就是 $F(x)$。理论上，对于“随着网络加深，准确率下降”的问题，ResNet 提供了上述两种选择方式。如果网络已经到达最优，继续加深网络，残差映射将逐渐变为 0，只剩下恒等映射，这样理论上网络一直处于最优状态，网络的性能也就不会随着深度增加而降低了。

图 3 所示为 ResNet18 的网络结构。采用 ResNet18 所训练的权重参数，将其输出得到的图像特征图与由设计的双通道 CNN 得到的特征图相结合，使网络可以得到更加准确的图像目标分类结果。图 3 中 Conv 代表卷积，Avg 代表平均，最后进行平均池化，/2 代表对卷积通道数的减半。



Note：“/2” represents halving the number of convolution channels.

图 3　ResNet18 结构图

Fig. 3　Diagram of ResNet18 structure

2.3　SegNet 网络结构

SegNet 网络最初由 Badrinarayanan 等[25]在 2017 年的 CVPR 会议上提出。这个网络展示了一

种编解码结构的深度全卷积神经网络结构，可用于图像中逐个像素的语义分割。网络整体可概括为一个编码网络和一个对应的解码网络，并跟随一个像素级别的分类层。此网络结构提出了带位置索引的最大池化方法，即在进行最大池化操作时，实现了对原来特征图中关键信息的位置保留，而保留的位置信息在上采样过程中加以利用。这样做可以使得网络在对原图进行特征提取时能够提取到更加有用的特征信息，最终实现更好的语义分割效果。该网络结构同样只有卷积层和池化层，比全连接网络的参数少很多，加速了网络训练的过程，节省了计算资源，但是分割效果仍有待提高。

3　本文方法

在相关研究与测试的基础上，针对高分辨率卫星遥感图像，提出一种双通道不同尺度特征提取融合的 FCN 结构，即在双通道中实现不同尺度的图像特征提取，并分别实现各自通道的训练，在实现最终分类目的之前，将双通道提取的图像特征加以融合，同时结合预训练模型提取的特征图，保证整个网络结构的精确性和稳健性，最后利用分类函数实现图像中各类目标的准确分类。

所提的网络主体结构如图 4 所示，该网络主要包含两个大的模块。模块一为两个独立通道，分别实现不同尺度图像特征的提取，并将分别提取的特征图进行融合。将初始大小为 $N\times N$ 和 $N/2\times N/2$的遥感图像送入两个全卷积网络通道，其中，$N/2\times N/2$ 的图像是从 $N\times N$ 图像中随机分割得到的图像，以保证训练的随机性，从而使图像不同位置的细节特征得以补充。模块二为 ResNet18 预训练过程，通过预训练模型与前面两个独立通道获得的特征图再次融合，实现了高层特征的联合增强表达，达到提高精确度和稳健性的目的，同时缩短了网络训练时间。

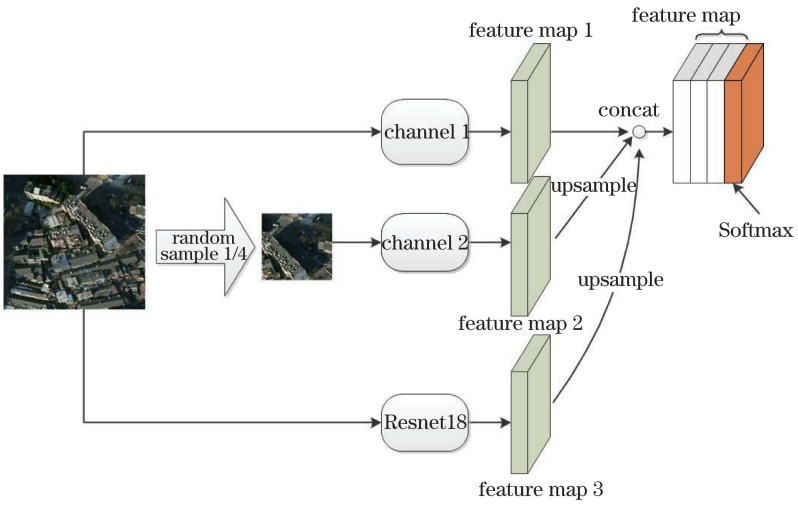


图 4　网络主体结构

Fig. 4　Main structure of network

在这两个模块的基础上，网络中同时采取了 SegNet 中带有位置索引的最大池化方法和批标准化加速层方法。通过加入这些优化策略，整体模型在小样本训练集上即可实现较高精度的分类。最后，采用 Softmax 分类函数实现对图像中不同像素对应类别的划分。

3.1　双通道不同尺度特征提取

如图 4 所示，网络整体结构主要由两个单通道卷积网络组成。Upsample 为上采样操作，Softmax 为采用的函数，Concat 代表对特征图进行维度上的叠加。图 5 展示了单通道的具体设计。每个通道都包含 4 个下采样和 4 个上采样部分，每个部分都由批标准化层(BN)、卷积层、Relu 激活函数和带位置索引的最大池化层组成。图 5 中，IndexMaxpool 表示带位置索引的最大池化操作，ConvOutlayer 为卷积输出层，downsample 表示整体的下采样过程，upsample 表示整体的上采样过程。

两个通道输入的遥感图像大小分别为 $N\times N$ 和 $N/2\times N/2$(实际训练中，图像输入大小 N 取值为 224)。在网络训练过程中，第一个通道对输入大小为 $N\times N$ 的图像进行处理，此通道主要对整张图像的全局特征进行提取；第二个通道为整张图像随机取样所得到的大小为 $N/2\times N/2$ 的图像，加入这一通道的目的主要是补充第一通道在特征提取过程

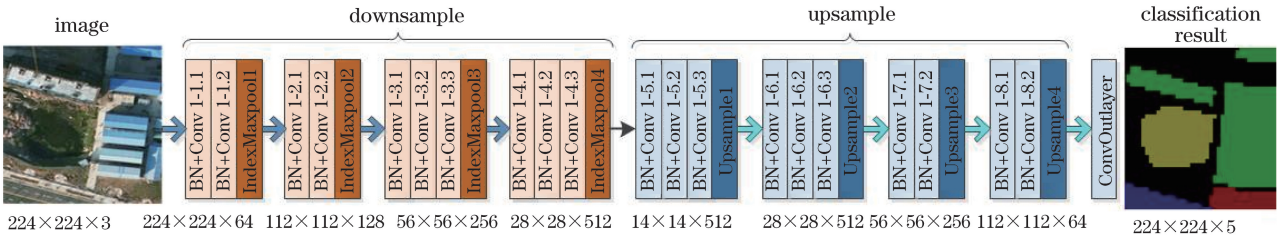


图 5　$N\times N$ 通道结构图

Fig. 5　Diagram of $N\times N$ channel structure

中所遗漏的一些细节信息，使整体网络结构在提取图像特征时更加准确、全面。

图 5 为 $N\times N$ 通道的结构图。其中，下采样部分中，卷积层从 Conv1.1 到 Conv4.3，共包含 10 个卷积层和 4 个池化层，每个卷积层的卷积核大小为 3×3，步长大小为 1。输入大小为 $N\times N$ 的遥感图像，第一个卷积块使用通道数为 64 的卷积核对图像进行操作，得到 224×224×64 的特征图，经过 4 个卷积块操作之后得到 14×14×512 的特征图，整个过程通过网络参数自动调整优化的方式获得遥感图像不同尺度的特征。

$N/2\times N/2$ 的通道主要作用是对提取特征中的细节作补充，整个过程与 $N\times N$ 通道类似，下采样最终得到 7×7×512 的特征图。不同之处在于，此通道是随机提取的 $N\times N$ 特征图的一部分。随机提取保证了在图像特征提取网络的反复训练中能够充分提取图像的细节特征，以提高整体训练的准确性。

采用的激活函数为 Relu 激活函数，即修正线性单元，表达式为

$$f(x)=\max(0,x), \tag{2}$$

式中：$f(x)$为隐层神经元的输出。

Relu 激活函数使神经网络中的神经元具有稀疏激活性，可以起到单侧抑制的作用，即把所有负值变为 0，而正值不变。对于线性函数而言，ReLu 的表达能力更强；而对于非线性函数而言，Relu 由于非负区间的梯度为常数，因此不存在梯度消失问题，可使得模型的收敛速度维持在一个稳定状态。

下采样过程中，每个通道同时包含 4 个池化层，如图 5 中的 IndexMaxpool1 ～ IndexMaxpool4 所示。其中，池化窗口的大小为 2×2，步长为 2。池化操作中要同时记录每个池化窗口最大值的位置索引，以便在上采样中加以利用，从而使得图像特征保留得更充分。池化对输入的特征图进行压缩，一方面使特征图变小，简化网络复杂度；另一方面在压缩过程中提取主要特征。

3.2　双通道不同尺度特征联合

在分别得到不同尺度的图像特征之后，进行特征联合。特征联合部分如图 6 所示。将 $N/2\times N/2$通道经过上采样得到的 112×112×64 的特征图再进行一次上采样操作，得到 224×224×64 的特征图，并与 $N\times N$ 通道得到的 224×224×64 特征图进行拼接，得到 224×224×128 的联合特征图。根据联合特征图得到了增强之后的高层特征表达。

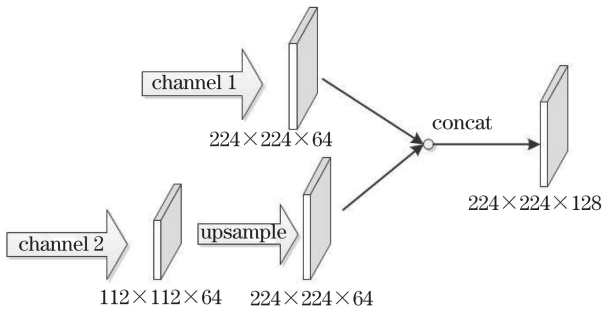


图 6　双通道不同尺度特征融合图

Fig. 6　Two-channel feature fusion with different scales

3.3　位置索引最大池化

所设计的双通道卷积网络，每个独立通道采用的都是下采样加上采样的结构设计，整体网络可以实现图像的像素级分类。下采样操作用于实现对遥感图像的特征提取，上采样操作时将下采样提取的特征图还原至原输入图像大小并实现像素分类。

在设计池化层时，采用带有位置索引的最大池化方法。该池化方法可以在下采样过程中保存图像特征关键信息的位置，并将保存下来的位置信息在上采样时加以利用，从而使得下采样得到的特征图按位置还原。相比于普通大反卷积上采样，该方法可以更好地保留图像目标区域的细节特征，实现更好的语义分割效果。经过最大池化索引上采样后，两个通道分别获得 224×224×64 和 112×112×64 的特征图。图 7 为位置索引最大池化下采样到上采样的过程示意图，x_1，x_2，…，x_{16}均

表示像素值。可以看出，该池化方法实现了关键信息的位置保留，而在不重要的位置上进行采样时则以 0 填充。图 7 中 index max pool 表示带有位置索引的最大池化操作，upsample 为对应的上采样过程。

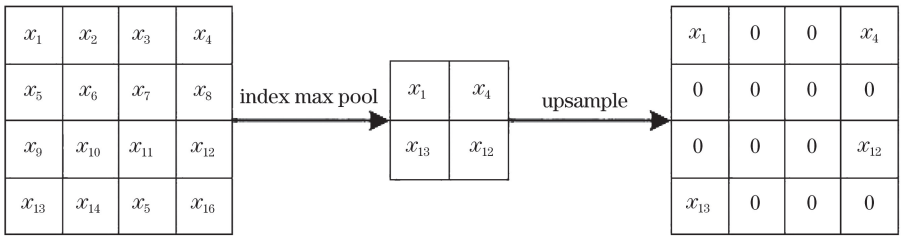


图 7　位置索引最大池化下采样和上采样

Fig. 7　Downsampling and upsampling of location index max pool

3.4　批标准化层

设计两个独立通道的 FCN 结构，且每个通道的网络结构较为复杂，导致网络训练速度较慢。为此，特地加入 BN 的网络优化策略。在训练深度神经网络时，前一层参数的变化会导致每层的输入分布在训练过程中发生变化。通常情况下，要求较低的学习率和详细的参数初始化会导致训练过程减慢。批标准化利用数据标准化的方式，将神经网络每层的输入标准化为方差为 1、均值为 0 的正态分布，从而加速神经网络的收敛。

在网络训练过程中处理多个特征图，求取每批次特征图所有神经元的均值和方差。在实验中，针对多个图像进行小批次标准化操作以加速网络训练过程。引入批标准化处理后，网络训练速度明显加快，有效缩短了训练时间。学习率也不再受限于很小的范围，对提高网络的分类性能也十分有益。

3.5　引入 ResNet18 预训练参数

从 CNN 的发展历史来看，神经网络模型的层数和复杂度发生了巨大变化。随着模型层数和复杂度的增加，模型在相应数据集上的错误率也随之降低。为了提高遥感图像目标分类结果的准确率、解决数据量不充分所带来的过拟合问题，采取迁移学习方式，加入 ResNet18 的预训练权重参数，即将处理后的数据图像送入预训练好的 ResNet18 权重参数中进行特征向量的提取，然后再将提取的图像特征与所设计的双通道 FCN 提取的特征图进行拼接，最终得到增强后的联合表达的特征图。

图 3 已展示了 ResNet18 的网络结构，这种网络结构可以在加深网络深度的同时防止梯度消失或梯度爆炸，是一种十分优异的网络结构。实验中，将 ResNet18 最后一层的全连接层除去，舍弃最后两个卷积层，最终获取 $14\times14\times256$ 的特征图，将此特征图上采样至 $224\times224\times64$，然后与 $N\times N$ 通道和 $N/2\times N/2$ 通道拼接后的通道再次拼接，最终得到 $224\times224\times192$ 的特征图，再经过 Softmax 函数将其每个像素类别区分出来。

3.6　网络训练过程

CNN 训练过程主要包含正向传播和反向传播两个过程：正向传播中，给定输入，输入经过所设计的 FCN 结构进行加权求和与激活，响应值在最后的网络层中输出；反向传播时，根据正向传播得到的输出值求得与真实值的误差，之后将此误差值通过反向传递来调节神经元之间的连接权值，通过反复调整最终获得误差最小条件下的网络模型。

所提模型采用 Softmax 分类器，损失函数设定为交叉熵损失，并采用正则化项进行校正，以进一步防止过拟合现象的发生。设 n 为样本类别数，经过 Softmax 函数后输出的向量为 $\boldsymbol{Y}\in\boldsymbol{R}^{1\times n}$。其中，$\boldsymbol{R}^{1\times n}$ 为 $1\times n$ 的向量空间；$\boldsymbol{Y}=(y_1,y_2,y_3,\cdots,y_n)$，为 $1\times n$ 的向量，y_i 代表向量中第 i 个元素的预测值。则损失函数表达式为

$$L=-\sum_i y_i'\ln(y_i)+\frac{1}{2}\lambda\|\boldsymbol{W}_i\|^2,\tag{3}$$

式中：第一项为交叉熵损失表达式，其中，y_i' 为相应的真实值；第二项为权值的 L2 正则项，$\boldsymbol{W}_i$ 为正则化项卷积核参数，λ 为正则项的系数，由各权值的衰减系数之积决定。

4　实验过程及结果分析

4.1　数据集及预处理

数据集选用佳格天地科技公司开源的遥感卫星高分辨率影像数据集，对数据集进行人工目视标记。数据集由高分辨遥感图像组成，每张图像包含的地表目标像素点数量为 10^7。图像中的地表目标分为 5 类：1）植被（vegetation），包括草地、林地和农用耕地等绿植，在标签图中记为 1；2）建筑物（building），

各类建筑物均归为此类，标签为 2；2）水体（water），包含湖泊、河流等，标签为 3；4）道路（road），即常见的各种通行道路，标签为 4；5）其他类（others），除上述 4 类地表目标外，均归为其他类，并标记为 0。数据集中地表物类别的多样性和差异性均得到了保证，能够确保网络学习到的模型具有较好的泛化性能。图 8 展示了数据集中的部分图像及其可视化标记。

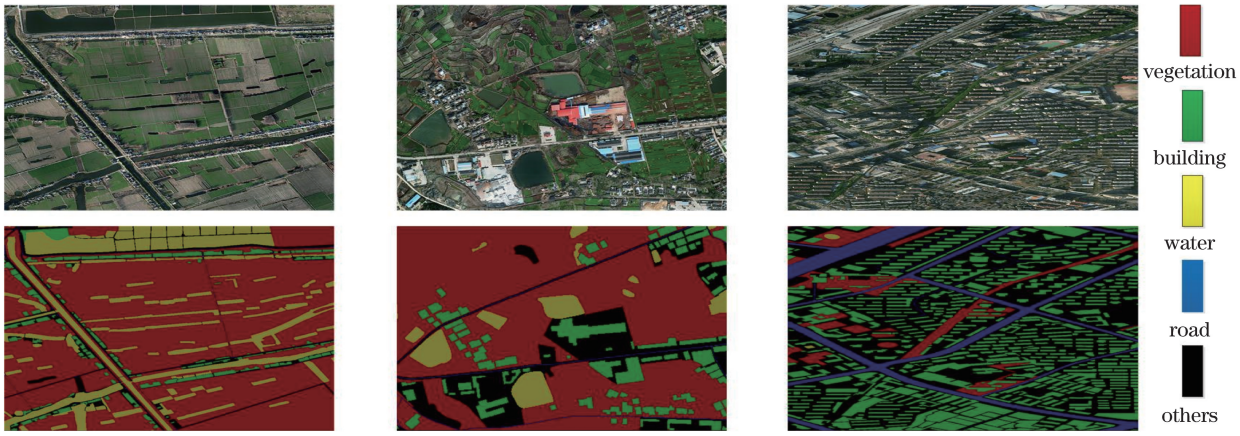


图 8　数据集部分图像及其对应的可视化标签

Fig. 8　Partial images of data sets and corresponding visual labels

该数据集所包含的图像均为大尺寸高分辨率卫星遥感图像，而深度 CNN 的参数量巨大，对计算机的性能要求很高，像该数据集这样的图像并不能对其进行直接处理，需要先将其切割成小尺寸图像。将此数据集中的高分辨率图像切割成 224 pixel×224 pixel 的小尺寸图像，从图像左上角开始，进行步长为 224 pixel×224 pixel 的滑动切割。切割后的图像及其对应的可视化标签如图 9 所示。

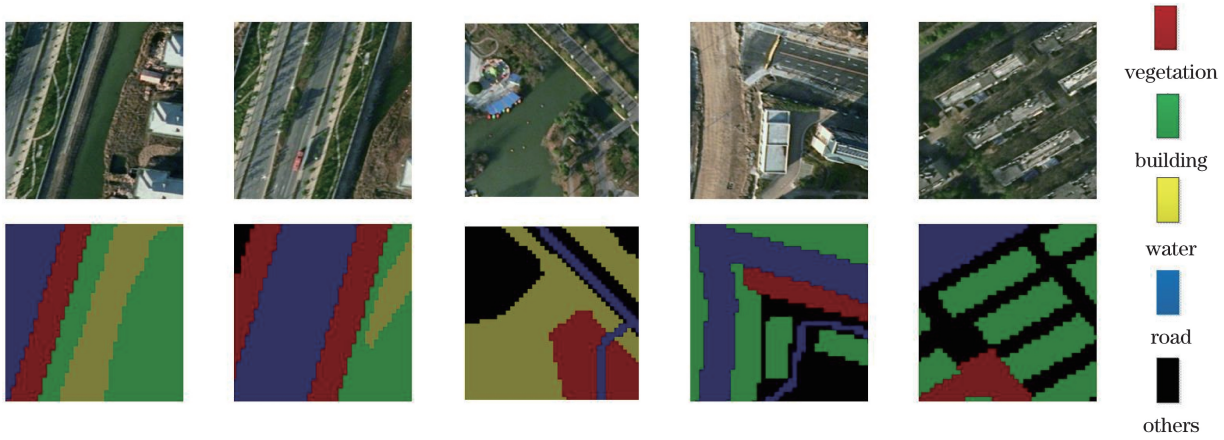


图 9　切割后的部分数据集图像及其可视化标记展示

Fig. 9　Partial data set images and their visual mark display after cutting

深度 CNN 神经元之间的权重参数数量巨大，对训练样本的数量具有较高要求。为此，采用一些方法对现有样本进行扩充：对前面切割得到的小尺寸图像，分别进行翻转变换，顺时针旋转 90、180°、270°。这些方法可以使样本的数据量得到很好的扩充，从而有效防止过拟合现象的发生。部分变换后的图像如图 10 所示。

4.2　实验设置

实验环境配置：CPU 为 Intel（R）Core（TM）i7-9700K 处理器，显卡为两块 NVIDA GeForce GTX1080Ti 显卡，内存总容量为 32 GB。运行的所有 CNN 模型均在 Pytorch 框架下进行。

实验中，采用随机梯度下降优化算法对设计的神经网络结构进行迭代求解，epoch 设置为 60，初始学习率设置为 0.001。在训练过程中，逐步调整学习率随损失值的变化，既能保证学习速度不会太慢，又能保证在训练的后面可以找到全局最优解。通过与损失函数比较，反复进行正向传播和反向传播，当满足误差阈值或者迭代次数时停止训练，即可得参数最优的网络模型。

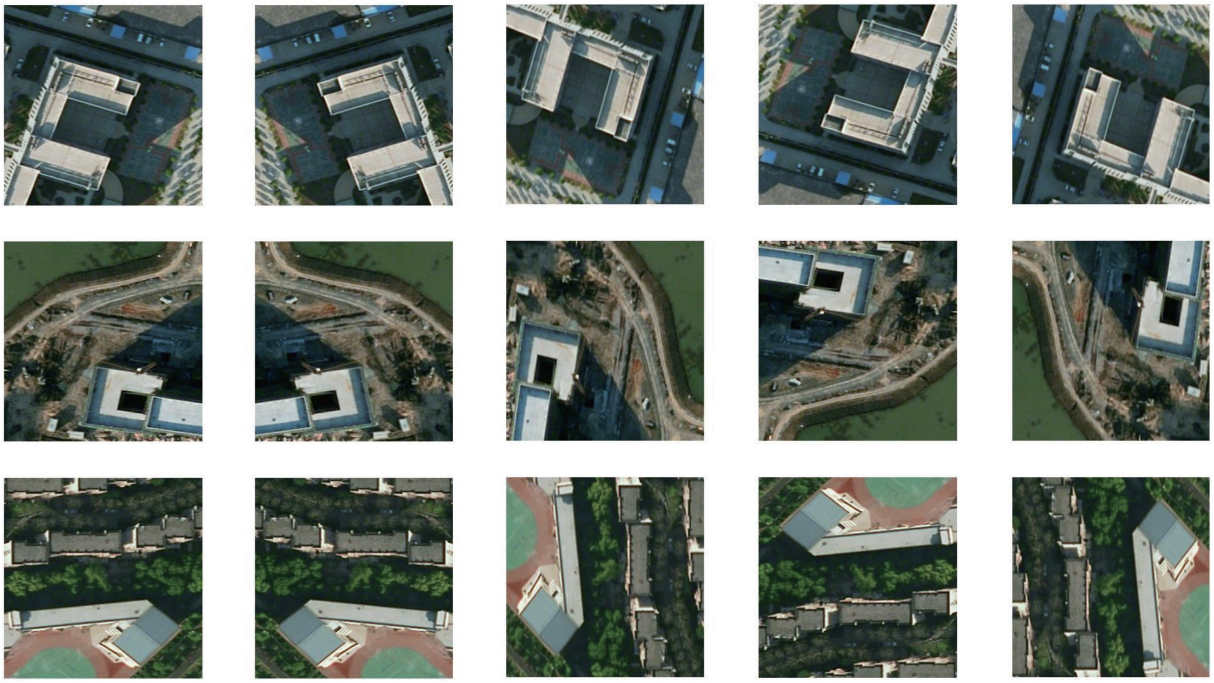


图 10　部分原图像及经过翻转、旋转后的图像展示

Fig. 10　Partial original images and display of flipped and rotated images

4.3　实验结果与分析

为了更好地表明所提的 FCN 结构的性能优势，分别利用 FCN 发展历史中几个优异的网络结构对高分辨率遥感图像进行训练与测试，并与所提算法进行比较。对比结构主要包括 FCN-8s 模型、SegNet 模型、Unet[26]模型，这些模型由不同专家在 CVPR 会议上提出，是现在语义分割领域中性能较为突出的算法结构，被广泛地应用在无人驾驶、医疗影像分析等语义分割任务中。

将前述模型获得的实验结果与所提模型得到的实验结果进行比对，部分可视化实验结果如图 11 所示。其中：第 1 列表示原始彩色遥感图像；第 2 列是人工标记的地物目标真实类别，包括植被（vegetation）、建筑（building）、水体（water）、道路（road）和其他（others）5 种；第 3～5 列分别表示使用 FCN-8s 网络、Unet 网络、SegNet 网络得到的实验结果；第 6 列是所提算法的遥感图像分类结果。

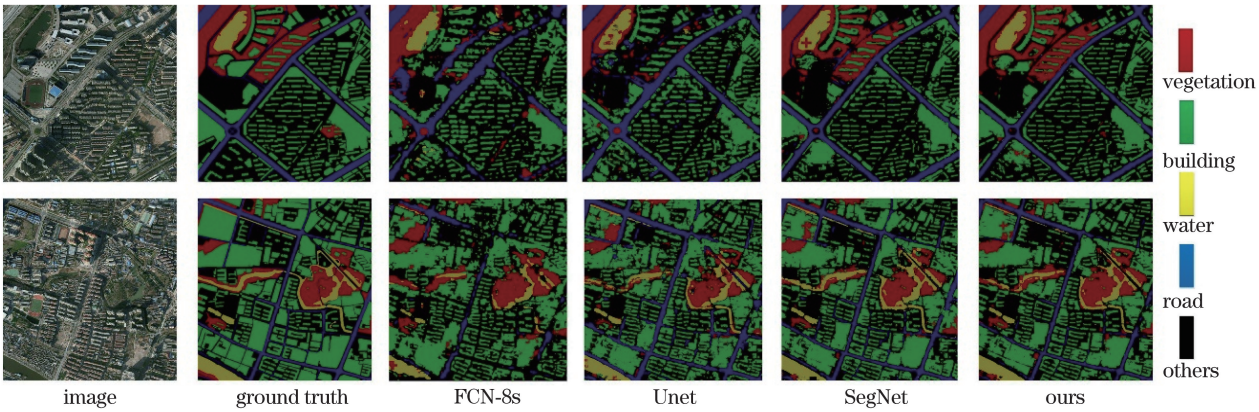


图 11　不同分类算法对遥感图像的可视化分类结果

Fig. 11　Visual classification results of remote sensing images by different classification algorithms

从图 11 中能够看出，FCN-8s 由于网络结构比较简单，最后输出的分类结果误差较大，有些区域出现了大面积误判为其他类（others）的预测结果；Unet 网络的表现相较于 FCN-8s 略好一些，对于一

些大面积为某一类别的区域几乎未出现错误，但是在细节性的区域表现很差，对于细小的道路类，很多都没能预测出来；SegNet 网络和 Unet 的表现相似，大面积同类别分类效果不错，但细节表现很差；所提算法表现得最好，不仅将河流、植被这些复杂区域分割得很好，而且在一些十分细小的道路类区分中也有较好的表现。这表明设计的双通道不同尺度特征提取融合算法，可以更好地保留目标的细节信息，将目标的边缘轮廓细致地划分出来，从而获得更好的语义分割效果。从实验效果中也可以直观反映出，即便是针对样本类别较少的道路，所提算法也有很好的效果。但是，对于一些极其细小的道路和其他类区域，所提算法的表现还不够理想。

为了更好地体现所设计的双通道加 ResNet18 预训练通道的结构优势，同时进行单独的通道 1、通道 1+2 和整体结构的实验对比，结果如图 12 所示。

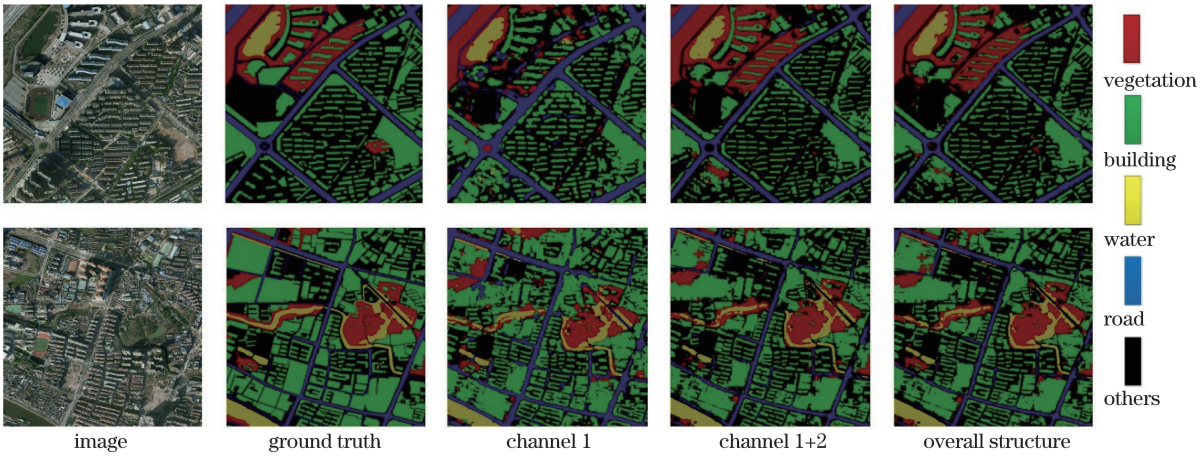


图 12　通道 1、通道 1+2、整体结构三种情况的可视化分类结果

Fig. 12　Visual classification results of channel 1, channel 1+2, and overall structure

从图 12 中不难看出，三种情况下的语义分割效果逐渐变好。单独通道（通道 1）的效果不够好，例如第一张图的大面积其他类中出现了误分类为植被和建筑的情况，从其他一些细节区域也能看出，只有通道 1 时，分割效果较为粗糙；在采用双通道（通道 1+2）之后，语义分割的细节效果得到了大幅提升，如第二张图中左上角实际为建筑却被通道 1 误分类为植被的区域，在这里得到了很大程度的纠正，整体细节上，分类结果也与真实标签图更加接近；进一步加入 ResNet18 预训练通道提取的特征之后，不同类别的边缘分割效果再次得到了微小提升，加入这一通道后，因为利用了 ResNet18 的预训练模型参数，整个训练过程的耗时减少了许多，即获得稳定模型的时间更短了，对于节约计算资源有很大帮助。为了定量分析，表 1 列出了所提算法的分类结果混淆矩阵（表中为像素数）。

分类结果的混淆矩阵是一种直观地评价分类算法性能的指标，混淆矩阵中列出所有目标类别及每个类别目标分类正确及错误的数量，以此为基础，可以得到样本中各类别、总体的准确率。总体的准确率由所有被正确分类的样本数除以样本总数得到，准确率越高，说明分类算法的效果越好。

表 1　所提算法的分类结果混淆矩阵

Table 1　Obfuscation matrix of classification results of proposed algorithm　　pixel

Category	Vegetation	Building	Water	Road	Others
Vegetation	9225390	3939997	187818	75047	281427
Building	783726	12505800	50563	126407	290737
Water	182936	9015	2069612	6761	29298
Road	20196	50490	5049	3256576	176714
Others	83475	144690	8348	41738	2672437

表 1 中，植被、建筑、水体、道路、其他类的预测准确率分别为 90.77%、90.82%、90.18%、90.3% 和 90.57%。通过设计的分类算法，训练得到了参数最优的分类算法模型。用此模型对测试集中的大尺寸遥感图像进行像素级分类，将所有测试图像统计到的每个类别的像素数加到一起进行计算。从表 1 中不难看出，采用所提算法时，每个类别得到的分类准确率虽稍有不同，但相差不大，整体的分类效果也很好，达到了 90%左右的准确率。但从实验过程中可以发现，对建筑和周围的植被进行分类时易混淆，而对于处在其他类中的极细小道路同样也容易出现错误分类。

为了进一步分析不同神经网络结构的算法性

能，统计了以上测试模型对测试集的像素分类结果，并计算了每个模型中各类别的分类准确率和总体准确率(OA)，结果如表 2 所示。各类别的分类准确率，即分类正确像素数除以该类别像素总数；总体准确率，即所有正确分类的像素数除以总像素数。同时，计算各算法的 Kappa 系数(R_{Kappa})。R_{Kappa}系数的计算是基于混淆矩阵的，也是常用的衡量分类精度的指标之一，其数值代表预测结果和真实结果的一致性，数值越高说明算法越好。R_{Kappa}可表示为

$$R_{\mathrm{Kappa}}=\frac{T\sum_{i=1}^{t}a_{ii}-\sum_{i=1}^{t}(g_i\cdot p_i)}{T^2-\sum_{i=1}^{t}(g_i\cdot p_i)},\tag{4}$$

式中：T 代表样本总数；a_{ii}代表混淆矩阵对角线上的数目，即被正确分类的样本数；g_i 代表第 i 个类别样本的实际样本总数；p_i 代表被分类为第 i 类的样本总数。

表 2　不同算法的分类准确率和 R_{Kappa}

Table 2　Classification accuracy and R_{Kappa} of different algorithms

Algorithm	Classification accuracy /%						R_{Kappa}
	Vegetation	Building	Water	Road	Others	OA	
FCN-8s	71.88	71.43	70.09	69.97	71.23	71.30	0.6603
Unet	80.92	80.27	80.10	79.89	80.52	80.44	0.7522
SegNet	83.66	83.18	82.64	82.52	83.07	83.21	0.7810
Channel 1	81.16	81.23	81.47	80.97	81.32	81.20	0.7623
Channel 1+2	89.56	89.78	89.58	89.16	89.76	89.63	0.8406
Ours	90.77	90.82	90.18	90.30	90.57	90.68	0.8595

从表 2 可以看出，所提算法经测试得到的总体准确率达到了 90.68%，与其他算法相比，准确率有较大幅度的提升，并且在和单独通道(通道 1)、双通道(通道 1+2)的对比中也能体现所提算法在结构完整后性能更好。这说明在所提出的双通道卷积结构网络的基础上加入其他优化策略的设计是一种优良的遥感图像语义分割的方法。所提算法相比其他算法在细小道路的分割效果突出，这得益于其双通道结构，及加入的带位置索引的最大池化方法，使得网络模型在细节处也能取得很好的分割效果。从表 2 中的 R_{Kappa}也能够看出，相比其他算法，所得算法的预测结果和遥感图像的真实情况一致性更好，得到的学习模型更加可靠。

同时，计算上述算法的卷积核参数总量、单次前向传播时间、模型达到稳定的训练时间，并进行进一步对比分析(表 3)。

从表 3 可以直观地看出，FCN-8s 的卷积核参数最多，达到了 1.343×10^8，单次前向传播时间和训练时间也最长。综合对比可以发现，所提算法的卷积核参数虽然也比较多、单次前向传播时间也较长，但是由于加入了 ResNet18 预训练的模型参数，在卷积核参数调整中起到了很大作用，可以使模型快速稳定，相比于其他算法，训练效率大大提升，极大地缩短了训练时间，只需要2.9 h就可以获得稳定的训练模型，有利于节省计算资源。从最终的语义分割效果、整体训练时间来看，所提算法的性能较以往算法有很大程度的提升。

表 3　不同算法的卷积核参数、单次前向传播时间和训练时间

Table 3 Convolution kernel parameters, single forward propagation time, and training time of different algorithms

Algorithm	Total number of parameters /10^6	Forward time /ms	Train time /h
FCN-8s	134.3	221	5.56
Unet	23.6	56	4.72
SegNet	29.4	78	4.88
Channel 1	15.3	50	4.55
Channel 1+2	30.5	108	5.06
Ours	30.5	116	2.90

5　结　论

为了能够取得更好的高分辨率遥感图像语义分割效果，设计了双通道不同尺度特征提取融合的 CNN 结构，同时加入了许多小的优化策略，使得算法能够在小规模数据集上取得良好的语义分割效果，总体准确率达到了 90.68%，R_{Kappa}为 0.8595。与 FCN-8s、Unet、SegNet 等网络模型，及单通道、双通道的对比实验显示，所提方法具有优异的语义分割

性能，训练速度也更快。但是从实验结果也可看出，所提方法对于图像中某些极其细微的区域，分割结果不佳，在今后的研究中，还应对此进行进一步探索。同时，也要研究更加快速稳定的网络模型，以切实投入到科学应用之中。

参 考 文 献

[1] Wang Y, Xing L N. Remote sensing satellite networking technology and remote sensing system: a survey[C]//2015 12th IEEE International Conference on Electronic Measurement & Instruments (ICEMI), July 16-18, 2015, Qingdao, China. New York: IEEE, 2015: 1251-1256.

[2] Zhao Y C, He X, Feng W T, *et al*. Design of common aperture coaxial field-bias optical system used in area array imaging sensor[J]. Infrared and Laser Engineering, 2018, 47(7): 0718004.
赵宇宸, 何欣, 冯文田, 等. 同轴偏视场共孔径面阵成像光学系统设计[J]. 红外与激光工程, 2018, 47(7): 0718004.

[3] An Z, Xu X P, Yang J H, *et al*. Design of augmented reality head-up display system based on image semantic segmentation[J]. Acta Optica Sinica, 2018, 38(7): 0710004.
安喆, 徐熙平, 杨进华, 等. 结合图像语义分割的增强现实型平视显示系统设计与研究[J]. 光学学报, 2018, 38(7): 0710004.

[4] Meletis P, Dubbelman G. Training of convolutional networks on multiple heterogeneous datasets for street scene semantic segmentation[C]//2018 IEEE Intelligent Vehicles Symposium (IV), June 26-30, 2018, Changshu, China. New York: IEEE, 2018: 1045-1050.

[5] Paisitkriangkrai S, Shen C H, van den Hengel A. Pedestrian detection with spatially pooled features and structured ensemble learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(6): 1243-1257.

[6] Blanzieri E, Melgani F. An adaptive SVM nearest neighbor classifier for remotely sensed imagery[C]// 2006 IEEE International Symposium on Geoscience and Remote Sensing, July 31-August 4, 2006, Denver, CO, USA. New York: IEEE, 2006: 3914-3917.

[7] Kluckner S, Bisch H. Image-based building classifiation and 3D modeling with super-pixel[C]// ISPRS Technical Commission Ⅲ Symposium on Photogrammetry Computer Vision and Image Analysis, September 1-3, 2010, Paris, France. [S.l.: s.n.], 2010: 233-238.

[8] Chen B, Qiu F, Wu B F, *et al*. Image segmentation based on constrained spectral variance difference and edge penalty[J]. Remote Sensing, 2015, 7(5): 5980-6004.

[9] Benchaou S, Nasri M, Melhaoui O E. Feature selection based on evolution strategy for character recognition[J]. International Journal of Image and Graphics, 2018, 18(3): 1850014.

[10] Song X F, Duan Z, Jiang X G. Comparison of artificial neural networks and support vector machine classifiers for land cover classification in Northern China using a SPOT-5 HRG image[J]. International Journal of Remote Sensing, 2012, 33(10): 3301-3320.

[11] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507.

[12] Graves A, Liwicki M, Fernandez S, *et al*. A novel connectionist system for unconstrained handwriting recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(5): 855-868.

[13] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.

[14] LeCun Y, Bottou L, Bengio Y, *et al*. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[15] Xie Y T, Richmond D. Pre-training on grayscale ImageNet improves medical image classification[M]// Leal-Taixé L, Roth S. Computer vision-ECCV 2018 Workshops. Lecture notes in computer science. Cham: Springer, 2019, 11134: 476-484.

[16] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C]//Neural Information Processing Systems, December 3-6, 2012, Lake Tahoe, Nevada, United States. Canada: NIPS, 2012.

[17] Nguyen T, Han J, Park D C. Satellite image classification using convolutional learning[C]. AIP Conference Proceedings, 2013, 1558(1): 2237-2240.

[18] Hu F, Xia G S, Hu J W, *et al*. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery[J]. Remote Sensing, 2015, 7(11): 14680-14707.

[19] Mnih V. Machine learning for aerial image labeling [M]. Canada: University of Toronto, 2013: 12-16.

[20] 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 (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 770-778.

[21] Yu F, Koltun V, Funkhouser T. Dilated residual networks[C]//2017 IEEE Conference on Computer

Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 636-644.

[22] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.

[23] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J/OL]. (2015-04-10)[2019-07-08]. https://arxiv.org/abs/1409.1556.

[24] He L, Wang G H, Hu Z Y. Learning depth from single images with deep neural network embedding focal length[J]. IEEE Transactions on Image Processing, 2018, 27(9): 4676-4689.

[25] Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.

[26] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[M]// Navab N, Hornegger J, Wells W, *et al*. Medical image computing and computer-assisted intervention- MICCAI 2015. Lecture notes in computer science. Cham: Springer, 2015, 9351: 234-241.