

基于离散余弦变换和深度网络的地貌图像分类

刘芳, 路丽霞, 黄光伟, 王洪娟, 王鑫

北京工业大学信息学部, 北京 100022

摘要 在未知环境中, 无人机(UAV)着陆地貌的自动识别和分类有着极其重要的研究意义, 传统的自然场景分类利用的是中层和底层特征信息, 但是无人机着陆地貌图像场景复杂、信息丰富, 需要较准确的高层语义特征表达。提出了一种基于离散余弦变换(DCT)和深度网络的地貌图像分类方法。首先将离散余弦变换能量集中的优势引入到卷积神经网络(CNN)的高效特征表达中, 以降低维度和计算复杂度; 然后根据地貌图像特点构建了 14 层的特征学习网络, 并改进了卷积神经网络结构; 最后将得到的深层特征输入到支持向量机(SVM)中, 快速准确地完成图像分类。实验结果表明, 该算法降低了数据冗余, 使训练时间大幅度减少, 可以自动地学习高层语义特征; 所提算法提取的特征具有较好的特征表达, 有效地提高了图像分类准确率。

关键词 光计算; 卷积神经网络; 离散余弦变换; 支持向量机; 无人机着陆地貌图像; 图像分类

中图分类号 TN911.73

文献标识码 A

doi: 10.3788/AOS201838.0620001

Landform Image Classification Based on Discrete Cosine Transformation and Deep Network

Liu Fang, Lu Lixia, Huang Guangwei, Wang Hongjuan, Wang Xin

Faculty of Information Technology, Beijing University of Technology, Beijing 100022, China

Abstract In the unknown environment, the automatic identification and classification of unmanned aerial vehicle (UAV) landing landforms are of great significance. The traditional natural scene classification uses the information of the middle- and the low-level features, but the UAV landing landform image has complex scene and rich information, which needs high-level semantic features to express more accurate information. A landform image classification algorithm based on discrete cosine transform (DCT) and deep network is proposed. First, the advantage of DCT energy concentration is introduced into the efficient feature representation of convolutional neural network (CNN) to reduce the dimensionality and computational complexity. Then a 14-layer feature learning network is constructed based on the characteristics of landform image, and the CNN structure is improved. Finally, the deep features are input into the support vector machine (SVM) to complete the image classification quickly and accurately. Experimental results show that the algorithm reduces data redundancy and training time greatly, and can automatically learn high-level semantic features. The features extracted by the proposed algorithm have better feature expressions and effectively improve the image classification accuracy.

Key words optics in computing; convolution neural network; discrete cosine transform; support vector machine; unmanned aerial vehicle landing landform image; image classification

OCIS codes 200.4260; 100.4996; 100.2000; 100.3008

1 引 言

随着技术的进步, 无人机(UAV)被广泛地应用于军事侦察、地质勘探、目标跟踪、交通监控等领域。无人机地貌图像有丰富的空间、纹理特征, 还包含了

大量场景语义信息, 由于场景的构成复杂且地物覆盖类别众多, 图像中包含不同尺度的物体, 大到大型建筑物, 小到路上行人; 此外地貌图像还具有密集性, 在一个小范围内会包含各种不同的物体。因此, 无人机地貌图像具有场景复杂、纹理信息丰富、大范

收稿日期: 2017-10-17; 收到修改稿日期: 2017-12-20

基金项目: 国家自然科学基金(61171119)、北京工业大学研究生科技基金(ykj-2015-12083)

作者简介: 刘芳(1971-), 女, 博士, 副教授, 硕士生导师, 主要从事图像处理、计算机智能和可视导航、机器视觉等方面的研究。E-mail: liufang@bjut.edu.cn

围、宽视角的特点。无人机飞行环境复杂多样,快速有效的地貌分类技术成为无人机执行飞行任务的安全保障。

近年来,深度学习被引入到图像分类中,Hinton等^[1]提出用深度学习思想对图像进行识别处理。其中卷积神经网络(CNN)在图像分类中具有较好的分类效果^[2]。文献[3]提出了基于深度学习的高分辨率影像分类方法,利用深度学习模型挖掘遥感影像的空间分布规律,将深度学习知识引入到高分辨率遥感影像分类问题中。方旭等^[4]获得了更为精细的高分类影像的边缘分类结果,将 Mean-shift 分割和全 CNN 相结合,减少了影像特征图自身特征丢失,提高了影像的分类精度。文献[5]提出了一种联合显著性采样和多层 CNN 的方法,利用显著性采样获取图像块样本集然后将其输入 CNN 中进行训练。

深度网络需要通过训练提高网络的分类性能,但是训练时间会受到各种因素的影响,例如图像维数越大,网络的训练时间越长^[6]。为了缩短训练时间,研究者采取缩小网络和减少训练图像的方法^[7],但是这会直接影响到网络的性能,不利于图像的特征学习。Agarwal等^[8]采用主成分分析(PCA)对数据进行除冗余降维处理,并提出采用离散余弦变换(DCT)通过选择 DCT 系数构造特征向量进行图像的识别,缩短了训练和识别的时间,取得了不错的识别效果^[9-11]。文献[12]中提出了基于 DCT 的稀疏自动编码器模型,在手写字体库上取得了不错的分类率。但是基于 DCT 的稀疏自动编码器模型比较简单,只在场景单一、内容简单的图像上具有不错的分类效果。

综上所述,与传统图像分类方法相比,CNN 对图像具有较好的特征表达能力,在图像分类方面具有较高的准确率,因此选用 CNN 进行特征提取。但是在训练时,运用 CNN 模型存在输入高维图像会大大增加训练时间、影响深层次特征的学习以及参数过多导致调参困难等问题。

针对以上问题,提出一种基于 DCT 和深度网络的地貌图像分类方法。为了在不损失原始输入信息的前提下降低输入数据的冗余信息,本文根据 DCT 能量集中的特点提出了系数判别法,保留最能表达图像信息的系数。并根据 DCT 系数的特点以及无人机地貌图像的特点对传统的 CNN 模型进行改进,构建了 DCT-CNN 模型,最后将提取的特征输入到支持向量机(SVM)分类器中完成图像分类。

2 基本原理

2.1 DCT

DCT 是一种实数域变换,其变换核是实数余弦函数。一幅图像经过 DCT 后,图像的可视信息主要集中在 DCT 的一小部分系数中,即图像的低频部分。对于一幅 $M \times N$ 的图像 $f(x, y)$ 矩阵,其 DCT 定义为:

$$F(m, n) = \alpha(m)\alpha(n) \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \times \cos\left[\frac{(2x+1)\pi m}{2M}\right] \cos\left[\frac{(2y+1)\pi n}{2N}\right], \quad (1)$$

其中:

$$\alpha(m) = \begin{cases} \frac{1}{\sqrt{M}}, & m = 0 \\ \sqrt{\frac{2}{M}}, & 1 \leq m \leq M-1 \end{cases}, \quad \alpha(n) = \begin{cases} \frac{1}{\sqrt{N}}, & n = 0 \\ \sqrt{\frac{2}{N}}, & 1 \leq n \leq N-1 \end{cases}. \quad (2)$$

图像经过 DCT 处理是从时域的表达变换到频域。变换的前后能量不发生变化,但是能量的分布发生了改变,图像的大部分能量聚集在低频系数中,如图 1 所示。图像的左上角颜色较亮,因此图像经过 DCT 后,能量主要集中在左上角的低频分量中,左上角对应的 DCT 系数较大,而右下角区域的高频细节分量对应的高频 DCT 系数较小^[13]。

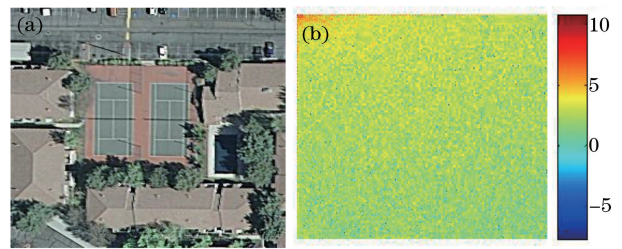


图 1 (a)原图像与(b) DCT 后的能量分布

Fig. 1 (a) Original image and (b) energy distribution after DCT

2.2 CNN

CNN 是一种特殊的深层前馈神经网络,其结构一般由输入层、多个卷积层、多个池化层、全连接层以及输出层组成,如图 2 所示。作为一种多层感知器,CNN 对于图像的比例缩放、平移以及形变等具有一定的稳健性。最后将图像特征输入到全连接层进行输出,计算样本与输出结果之间的误差,通过反向传播(BP)算法对网络进行微调。

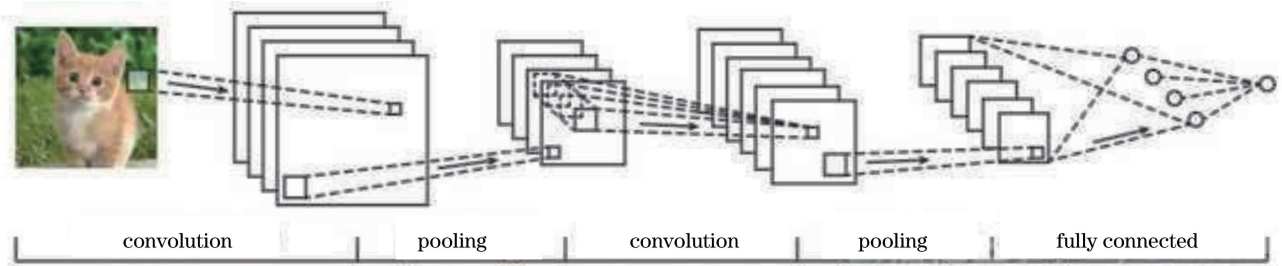


图 2 CNN 结构图

Fig. 2 Structure of CNN

CNN 通过卷积操作对图像进行逐层的特征提取,并且利用权值共享的思想在卷积操作时大大减少了网络的训练参数,降低了网络的复杂度。卷积过程就是用一个可学习的卷积核或滤波器去卷积一个输入的图像,加上偏置然后通过激活函数得到卷积特征图。公式表示为:

$$x_j^l = f\left(\sum_{i \in M_j} y_i^{l-1} \otimes k_{ij}^l + b_j^l\right), \quad (3)$$

式中 x_j^l 为第 l 层第 j 个特征图的输入, y_j^{l-1} 为第 $l-1$ 层第 j 个特征图的输出, k_{ij}^l 为前一层第 i 个特征图与当前层第 j 个特征图之间的卷积核, b_j^l 为第 l 层第 j 个特征图的偏置, $i \in M_j$ 为前一层与当前层第 j 个特征图连接的所有特征图, $f(\ast)$ 表示激活函数。

卷积后的特征图个数增加,根据图像的局部相关性原理,通过池化层对卷积后得到的特征图进行下采样,降低了图像的维度,避免了维数灾难。公式表示为:

$$x_j^l = f[\beta_j^l d(x_j^{l-1}) + b_j^l], \quad (4)$$

式中 β_j^l 为乘性偏置, $d(\ast)$ 表示下采样函数。

3 基于 DCT 和深度网络的无人机地貌图像分类

无人机地貌图像场景复杂、纹理信息丰富,在图像分类时只需要利用图像中的主要目标特征信息。DCT 具有能量集中的特点,能去除数据冗余信息,保留最能表达图像特征的信息。图像经过 DCT 后将其能量的大部分集中在频率域的一个小范围,包含图像中主要的视觉信息,可以根据实际需要选择和提取图像特征。DCT 原理简单、计算复杂度低,可以对地貌图像进行初级的处理,获得原有图像在变换域上具有良好不变性和区分性的描述,避免了对多层特征提取方法中滤波器的学习。CNN 是通过卷积运算由浅层到深层提取图像不同层次的特征,对图像具有较好的特征表达能力。针对无人机

地貌图像场景复杂、信息丰富的特点,以及 CNN 模型层数较多、结构复杂和训练花费时间较长的问题,将基于 DCT 能量集中的优势引入到基于 CNN 的特征学习中,结合两者优点构建了 DCT-CNN 模型。

3.1 DCT 系数的选择

一幅图像经过 DCT 后,图像的大部分能量聚集在低频 DCT 系数中,所以低频系数比高频系数更重要,但是完全把高频系数丢弃就会损失较多的图像细节信息。在 DCT 系数选择时为了不丢失太多的细节信息,并保留最能表达图像特征的系数,研究采用保留矩阵左上角 10 个低频系数以及保留具有较高能量信息的中高频系数,提出了系数判别法,并利用其对中高频系数进行选择。该方法具体步骤为:

1) 对图像进行 8×8 分块,然后对每个子块分别进行 DCT;

2) DCT 后的系数矩阵,按照 ZigZag 扫描的方式保留矩阵左上角的 10 个低频系数;

3) 对剩余的 54 个系数采用系数判别法进行系数选择,首先求每一个子块的平均值,并设定平均值为每块的系数阈值,然后对每个子块中剩余的 54 个系数进行选择,如果系数小于阈值则该系数置 0,大于阈值则保留该系数;

4) 最后把每幅图像筛选的 DCT 系数进行整合。

图像经过 DCT 后能量主要分布在左上角,右下角的能量分布比较低。图 3 为三维频谱,图 3(a) 是原图像经过 DCT 后的系数频谱,频谱集中于中心一点向上,只是会聚了能量较高的左上角系数;图 3(b) 是经过系数判别法后得到的系数频谱图,会聚了能量较高的左上角系数和能量较高的中低频系数。经过系数判别法得到的 DCT 系数矩阵具有较集中的能量信息,包含了图像中主要的视觉信息,在不损失图像原始数据的前提下去除了冗余信息。

3.2 DCT-CNN 模型

由于 CNN 模型层数较多、结构复杂,将高维的图像直接输入 CNN 模型中进行训练,会使模型的训练时间较长、训练参数增加,另外还会导致深层的特征不容易被学习,从而丢失特征信息。在不损失输入数据信息的前提下,降低输入数据的冗余信息,可以减少参数的数量并缩短网络的训练时间。将 DCT 作为深度网络的第一层,图像经过 DCT 后得到和原图像相同维度的 DCT 系数,为了减少数据冗余,选择少量的 DCT 系数输入到深度网络中进行网络训练。为了提高无人机地貌图像的分类准确率,对传统 CNN 结构进行改进,最后将得到的深层特征输入到 SVM 分类器中完成图像分类。DCT-CNN 模型结构如图 4 所示。

为了减少网络训练时间,并且在降低网络计算复杂度的前提下提高无人机地貌图像分类的准确率,研究构建了一个 14 层用于无人机地貌图像分类的 DCT-CNN 网络结构,包括一个输入层、5 个卷积层、3 个池化层、4 个线性修正单元层(ReLU 层)和一个输出层。为了提取高层语义特征并减少网络的计算量,将采集到的无人机地貌图像大小统一缩放到 $128 \text{ pixel} \times 128 \text{ pixel}$,构建的 DCT-CNN 网络结构参数如表 1 所示。

表 1 DCT-CNN 网络结构的各层参数

Table 1 Layer parameters of DCT-CNN network structure

Layer	Type	Patch size	Stride	Zero padding	Output size
x	Input				128×128
h_1	Convolution	5×5	1	2	$128 \times 128 \times 32$
h_2	ReLU				
h_3	Mean pooling	3×3	2		64×64
h_4	Convolution	3×3	2	0	$32 \times 32 \times 32$
h_5	ReLU				
h_6	Max pooling	3×3	2		16×16
h_7	Convolution	7×7	1	2	$14 \times 14 \times 64$
h_8	ReLU				
h_9	Max pooling	3×3	2		7×7
h_{10}	Convolution	7×7	1	0	$1 \times 1 \times 64$
h_{11}	ReLU				
h_{12}	Convolution	1×1	1	0	$1 \times 1 \times 10$
o	SVM				$n(\text{class})$

卷积过程要针对具体图像设计卷积核大小,如果卷积核的尺寸过大,会增加网络运算量;如果尺寸过小,则不能提取到完整的特征。此外,卷积核个数过少会使特征提取不充分从而导致分类准确率低;数量过多会增加计算量,使卷积时间较长,从而大大增加运行时间。

基于以上分析,为了提取高层语义特征并减少

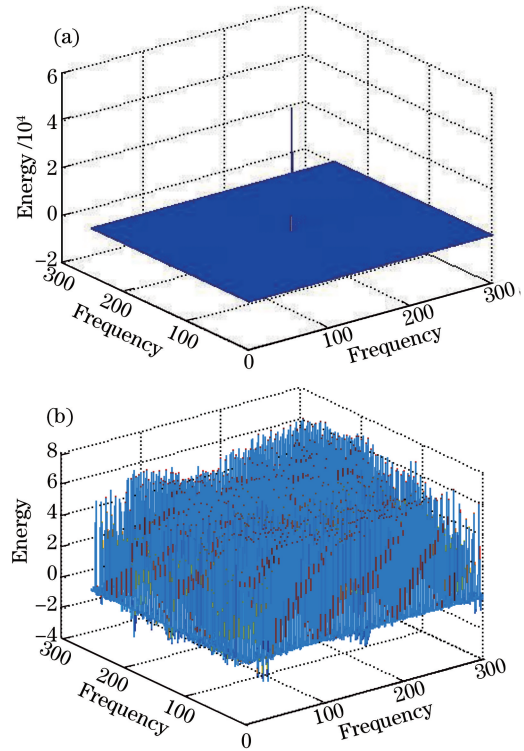


图 3 三维频谱图。(a)原图像 DCT 系数频谱图;
(b)系数选择后的频谱图

Fig. 3 Three-dimensional spectrum diagram.

(a) DCT coefficient spectrum of original image;

(b) spectrum after the coefficient selection

网络的计算量,DCT-CNN 模型中的 5 层卷积层为 5-3-7-7-1 结构,采用 5 层卷积层可以提取到较好的图像特征,获得较完整的特征表达。第一层卷积核的大小是 5×5 ,控制卷积核尺寸,有效地减少参数的数目,既能提取到深层特征表达,又能保证运算量不会大幅度增加;第二层的卷积核大小是 3×3 ,用以控制参数数目,降低网络运算的复杂度;第三层和

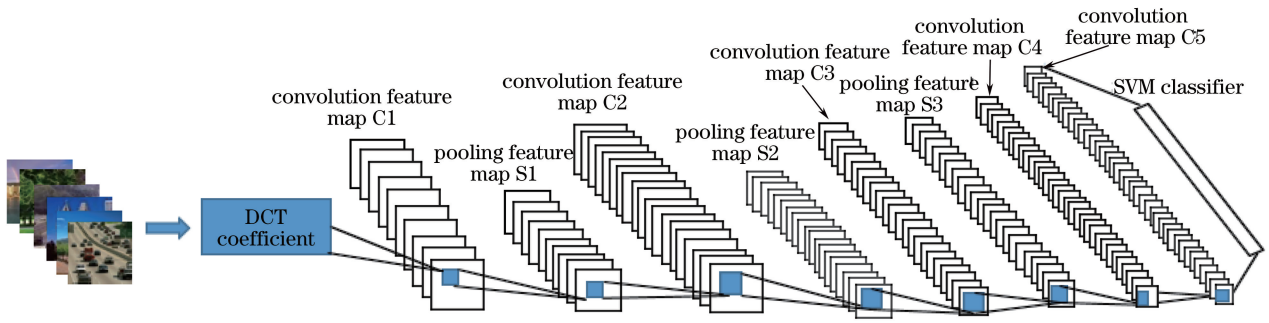


图 4 DCT-CNN 模型结构

Fig. 4 Structure of DCT-CNN model

第 4 层的卷积核大小是 7×7 , 虽然卷积核尺寸较大, 会增加网络运算量, 但是得到的特征比较完整、准确; 最后一层卷积核大小是 1×1 , 可以与任何层直接相连, 相当于一层全连接层, 直接输入到分类器中得到分类结果。

DCT-CNN 模型中的池化层采用了平均池化和最大值池化两种方法, 总共有三层池化层, 前三层卷积层后面都有一层池化层, 用于数据降维、去冗余信息, 最后两层卷积层输出的特征维度为 1, 不需要经过池化层处理。第一层池化层采用平均池化, 因为平均池化能保留整体数据的特征和突出背景的信息; 后两层池化采用最大值池化, 因为最大值池化能更好地保留纹理上的特征, 能提取重要、突出的特征, 舍弃其他弱的某类特征。

DCT-CNN 模型中增加了零填充 (zero padding), 为了更好地对图像进行特征学习, 根据不同情况对特征图进行 0 填充, 之后再行卷积、池化过程。在该网络结构中, 每一层卷积层后接一层激活函数层, 给网络加入非线性的因素, 使得 CNN 更好地解决复杂的问题, 网络中采用的是线性修正单元 (ReLU) 激活函数, 因为 ReLU 函数原理近似于人脑神经元在接触外界信号时的工作状态, 可以拟合更多的非线性过程, 其公式为:

$$f(x) = \max(0, x) \quad (5)$$

设每张输入特征图的尺寸为 $i \times i$, 卷积核尺寸为 $k \times k$, 步长为 s , 0 填充范围为 p , 则每张输出特征图的尺寸 $l \times l$ 满足下式 (各层中特征图尺寸变化如表 1 所示):

$$l = \begin{cases} (i - k) + 1, & s = 1, p = 0 \\ (i - k) + 2p + 1, & s = 1, p > 0 \\ \lfloor (i - k) / s \rfloor + 1, & s > 1, p = 0 \\ \lfloor (i + 2p - k) / s \rfloor + 1, & s > 1, p > 0 \end{cases} \quad (6)$$

在训练过程中, 使用梯度下降法对网络参数进

行更新, 参数更新规则如下:

$$W_{ij}^{(l)} = W_{ij}^{(l)} - \alpha \frac{\partial}{\partial W_{ij}^{(l)}} J(W, b) \quad (7)$$

$$b_i^{(l)} = b_i^{(l)} - \alpha \frac{\partial}{\partial b_i^{(l)}} J(W, b) \quad (8)$$

式中 α 为学习率, $\frac{\partial}{\partial W_{ij}^{(l)}} J(W, b)$ 和 $\frac{\partial}{\partial b_i^{(l)}} J(W, b)$ 分别是 CNN 的损失函数 $J(W, b)$ 对权重参数 W 和 b 的偏导数。

为了提高无人机地貌图像的分类准确率, DCT-CNN 模型中最后一层采用 SVM 分类器。由于 SVM 在解决非线性及高维分类问题时表现出特有的优势^[14], 具有较好的分类性能, 通过特征学习得到图像深层次的特征向量, 将其输入到 SVM 分类器中, 完成图像分类。

3.3 DCT-CNN 模型的算法流程

图 5 是基于 DCT 和深度网络的地貌图像分类算法流程图。

算法步骤如下:

- 1) 首先对图像进行 8×8 分块, 然后对每个子块分别进行 DCT;
- 2) DCT 后的系数矩阵, 按照 ZigZag 扫描的方式保留矩阵左上角的 10 个低频系数;
- 3) 对剩余的 54 个系数采用系数判别法进行系数选择, 首先求每一个子块的平均值, 并设定平均值为每块的系数阈值, 然后对每个子块中剩余的 54 个系数进行选择, 如果系数小于阈值则该系数置 0, 大于阈值则保留该系数;
- 4) 把每幅图像筛选的 DCT 系数进行整合;
- 5) 将训练集的 DCT 系数输入到改进的 DCT-CNN 模型中训练, 采用 (7)、(8) 式进行网络参数更新, 直到损失函数收敛于一个较小的值, 训练结束;
- 6) 输入测试集, 利用训练好的模型对测试图像进行逐层学习, 通过 (3) 式计算得到卷积特征图, 通

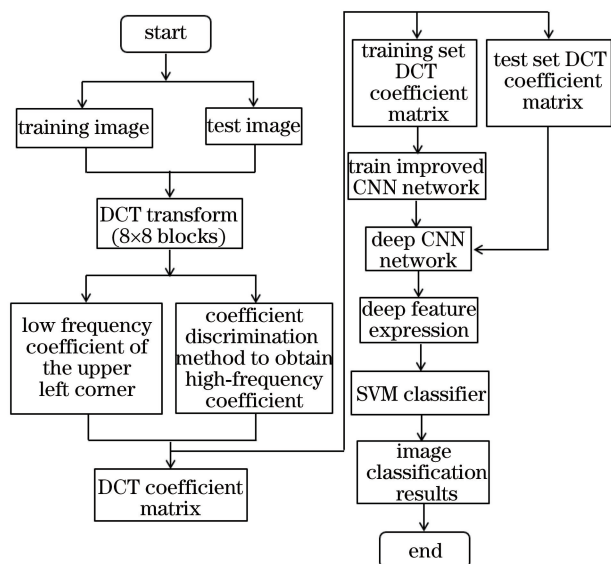


图 5 基于 DCT 和深度网络的地貌图像分类算法流程图

Fig. 5 Flow chart of landform image classification algorithm based on DCT and deep network

过(4)式得到下采样后的特征图,最后将得到的一维特征向量输入到 SVM 分类器中进行分类,得到图像分类结果。

4 仿真实验

4.1 实验数据

实验数据采用 UC Merced LU 数据库和采集到的无人机着陆地貌图像。UC Merced LU 共有 21 类,每类 100 幅图像,从该数据库中每类随机选取 80 张图像作为训练集,其余 20 张图像作为测试集,实验设置与文献[5]、[15]相同。无人机着陆地貌数据库总共有 8 类 7800 幅图像,每类图像的数量为 500~800,从该数据库中每类随机选取 400 张图像作为训练集,100 张图像作为测试集。两个数据库的地貌图像都具有场景复杂、内容丰富的特点,所以图像类别的判断取决于整幅图像的特征,两个数据库的部分示例图像如图 6 所示。

4.2 实验结果与分析

实验采用图 4 构建的基于 DCT 的 CNN 结构,在实验中采用 10 折交叉验证方法,交叉验证重复 10 次,每个子样本验证一次,将 10 次结果取平均,最终得到一个单一估测值。实验所用计算机操作系统为 Windows 7,中央处理器为 Intel(R) Core(TM) 2 Duo 3.0 GHz,图形处理器(GPU)为 NVIDIA GeForce GTX 860 M,内存为 16 GB,仿真平台为 64 位的 MATLAB R2014a。

实验中两个数据库内的图像像素不相同,首先

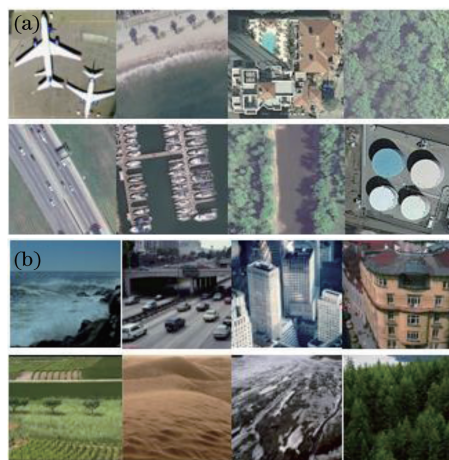
图 6 数据库示例图像。(a) UC Merced LU 数据库;
(b) 无人机着陆地貌数据库

Fig. 6 Example images in database. (a) UC Merced LU database; (b) UAV landing landform database

将所有数据库图像统一缩放到 $128 \text{ pixel} \times 128 \text{ pixel}$,然后采用所构建的 DCT-CNN 网络结构进行图像训练和分类。图像经过 DCT 后,大部分能量聚集在低频 DCT 系数中,所以低频系数比高频系数更重要,但是完全把高频系数丢弃就会损失太多的图像细节信息。在 DCT 系数选择时为了不丢失太多的细节信息,保留最能表达图像特征的系数,研究采用保留左上角 10 个低频系数以及具有较高能量信息的中高频系数,提出了系数判别法对中高频系数进行选择。为了验证系数判别法的有效性以及 DCT-CNN 方法是否能提取丰富的特征信息并减少训练时间,在 UC Merced LU、无人机着陆地貌数据库上进行实验,其中,SVM 分类器中惩罚系数 C 和核函数参数 g 的值分别为 3.3786、0.1168。

DCT 在分块时,如果子块的尺寸越大,算法的复杂度会增加,相反如果子块的尺寸过小,分成的子块较多,在系数选择时会丢失部分重要的特征信息,因此选择分块大小为 8×8 。DCT 预处理对图像分类准确率的影响结果如表 2、表 3 所示,表中 Method 1 是指对图像进行 8×8 分块,每一子块采用 ZigZag 扫描的方式保留矩阵左上角的 10 个低频系数,其他数置 0,然后将该矩阵作为输入数据输入到改进的 DCT-CNN 模型中进行网络训练和分类;Method 2 是采用所提系数判别法进行 DCT 系数的筛选,将得到的矩阵作为输入数据输入到改进的 DCT-CNN 模型中进行网络训练和分类;Method 3 是指图像不经过 DCT 预处理,将原图像直接输入到改进的 CNN 模型(无 DCT 处理,CNN 网络参数与表 1 相同)中进行训练和分类。

表 2 不同方法对 UC Merced LU 数据库的分类影响
Table 2 Effect of different methods on classification of UC Merced LU database

Method	Accuracy /%	SD	Training time /h
Method 1	84.25	0.78	0.8
Method 2	95.76	0.28	1.0
Method 3	92.83	0.52	3.3

表 3 不同方法对无人机着陆地貌数据库的分类影响
Table 3 Effect of different methods on classification of UVA landing landform database

Method	Accuracy /%	SD	Training time /h
Method 1	83.73	0.85	1.0
Method 2	94.38	0.34	1.3
Method 3	92.10	0.61	3.9

从表 2、表 3 中可以看出,图像经过 DCT 处理后的分类准确率明显高于未经过 DCT 预处理的分类准确率。如果只保留每个子块中左上角的 10 个低频系数,训练时间最少,但是分类准确率却不是很高,标准差(SD)系数最大,因为只保留低频系数会损失一部分重要的中高频信息,有用的特征信息较少,导致样本错分的数量较多。对于 UC Merced LU 数据库,Method 2 的 SD 系数为 0.28,得到的分类准确率为 95.76%,比 Method 3 提高了 2.93%,训练时间也减少了 2.3 h;对于无人机着陆地貌数据库,Method 2 的 SD 系数最小,得到的分类准确率为 94.38%,比 Method 3 提高了 2.28%,而且训练时间减少了 2.6 h。由于 DCT 具有紧凑表达能力,能有效地降低数据冗余信息,因此在大大减少了训练时间的同时提高了分类准确率,减少了离散错分样本数。

4.3 所提方法与其他已有方法的比较

为了验证所提方法的有效性,分别采用联合显著性卷积神经网络(CS-CNN)、离散余弦变换-稀疏自动编码器(DCT-SAE)、空间关系金字塔(PSR)、多尺度卷积神经网络(MS-DCNN)以及经典方法随机森林(RF)法、线性判别分析-随机森林(LDA-RF)法与所提 DCT-CNN 方法对标准 UC Merced LU 数据库和采集的无人机着陆地貌数据库进行分类,分类效果见表 4、表 5。实验是在同一设备、相同实验条件下进行的,文献[5]、[15-16]中的算法是针对无人机图像分类算法中具有较好分类效果的算法,文献[12]中的算法同样经过 DCT,但是系数选择和网络结构与所提方法不一样。

可以看出,基于深度网络的算法(CS-CNN、MS-DCNN、DCT-SAE 和 DCT-CNN)可以学习到高层

次的语义特征,具有较好的特征表达,分类准确度高 于中层语义和底层特征的 PSR、RF 和 LDA-RF 方法。

表 4 不同方法对 UC Merced LU 数据库的分类准确率比较
Table 4 Comparison of the classification accuracy of different methods for UC Merced LU database

Method	Accuracy /%	SD
RF	79.25	0.82
LDA-RF	82.92	0.69
CS-CNN ^[5]	92.86	0.59
PSR ^[15]	89.10	0.69
MS-DCNN ^[16]	91.34	0.63
DCT-CNN	95.76	0.28

表 5 不同方法对无人机着陆地貌数据库的分类准确率比较
Table 5 Comparison of the classification accuracy of different methods for UAV landing landform database

Method	Accuracy /%	SD
RF	77.10	0.70
LDA-RF	80.23	0.74
CS-CNN ^[5]	91.78	0.62
DCT-SAE ^[12]	86.49	0.96
MS-DCNN ^[16]	90.16	0.71
DCT-CNN	94.38	0.34

所提方法与传统方法相比准确率有大幅度的提高,在 UC Merced LU 数据库上,实验方法与文献[5]、[15-16]中的无人机地貌分类方法相比,分类准确率有不同程度的提高。在无人机着陆地貌数据库上 DCT-CNN 方法的分类准确率最高达到 94.38%,与文献[12]中的 DCT-SAE 模型相比,DCT-CNN 模型层数较多,结构复杂,且分类准确率有很大提升,比 DCT-SAE 方法的准确率提高了 7.89%。因此,DCT-CNN 模型在解决场景复杂、内容丰富的无人机地貌图像方面具有较好的分类效果,能减少离散错分样本数量。

DCT-CNN 网络有 5 层卷积层,复杂度为 $O(cf_1^2n_1+n_1f_2^2n_2+n_2f_3^2n_3+n_3f_4^2n_4+n_4f_5^2n_5)$,其中 f 为单个卷积核的尺寸, n 为卷积核的个数, c 为输入图像的通道数(只对 Y 通道进行处理,故 $c=1$),那么该网络一次迭代的计算复杂度为 311672,文献[5]中 CS-CNN 网络一次迭代的计算复杂度为 780000,是 DCT-CNN 网络的两倍多,文献[16]中 MS-DCNN 网络一次迭代的计算复杂度为 2840160,每次迭代训练时参数计算量约是 DCT-CNN 网络的 9 倍。LDA 方法在特征提取过程中的计算复杂度为 $O(n^3)$, n 为特征的维数,所以 LDA 方法的计算复杂度为 2097152。因此,DCT-CNN

网络计算复杂度比其他网络计算复杂度小,训练耗时最少。

图 7 为不同分类方法中训练样本的数量对分类准确率的影响,可以看出,随着每类训练样本图像数量的增加,各方法的分类准确率不断增加。对于两种数据集来说,所提 DCT-CNN 方法在不同训练样本数量中的分类准确率都是最高的。

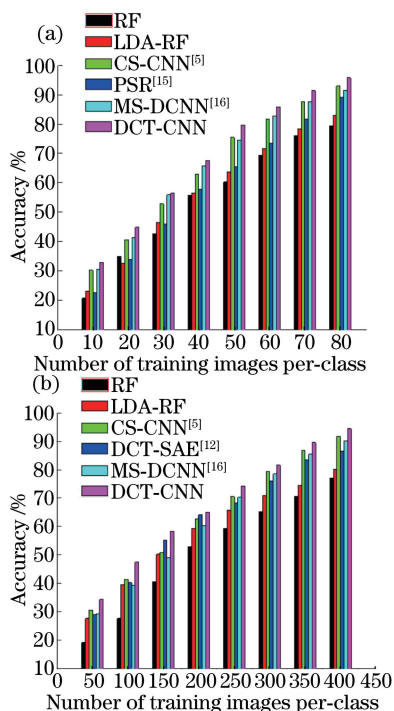


图 7 训练样本数量不同时各分类方法的分类性能。

(a) UC Merced LU 数据库; (b) 无人机着陆地貌数据库

Fig. 7 Classification performance of each method when the number of training samples is different. (a) UC Merced LU database; (b) UAV landing landform database

为了验证所提方法在地貌分类中的实际应用效果,给出了无人机着陆地貌数据库的分类混淆矩阵结果,如图 8 所示。从混淆矩阵图中可以看出,该数据集 8 类地貌类型分类准确率都在 90% 以上,主要是内容较为相似的图像容易错分。综上,与其他各方法比较,所提方法的分类性能更优异,实验结果表明了所提 DCT-CNN 方法的有效性,该方法提取到的特征能有效地反映出图像信息,具有较好的图像分类准确率,同时该方法使网络训练时间大幅度减少。

5 结 论

提出了一种基于 DCT 和深度网络的无人机地貌图像分类方法,针对无人机地貌图像场景复杂、信息丰富且需要较准确的高层语义特征表达等特点,以及 CNN 模型层数较多、训练花费时间较长等问

coast	0.92	0.08	0.00	0.00	0.00	0.00	0.00	0.00
forest	0.00	0.95	0.00	0.00	0.00	0.05	0.00	0.00
highway	0.00	0.01	0.93	0.00	0.00	0.00	0.04	0.00
desert	0.02	0.00	0.00	0.95	0.00	0.01	0.01	0.00
mountain	0.00	0.00	0.00	0.00	0.98	0.00	0.00	0.02
open country	0.00	0.03	0.01	0.00	0.00	0.96	0.00	0.00
street	0.01	0.01	0.07	0.01	0.00	0.90	0.90	0.00
tall building	0.00	0.01	0.00	0.00	0.02	0.01	0.01	0.96

图 8 无人机着陆地貌数据库图像分类混淆矩阵

Fig. 8 Image classification confusion matrix for UAV landing landform database

题,结合 CNN 和 DCT 优点构建了 DCT-CNN 学习网络。根据 DCT 能量集中的特点提出了 DCT 系数判别法,在不损失原始输入信息的前提下降低输入数据的冗余信息。为了提高分类准确率,对传统 CNN 结构进行改进,构建了 14 层的学习网络。针对改进方法进行实验,将结果和其他基于 CNN 的无人机地貌图像分类方法进行对比。改进方法提取到的特征能有效地反映出图像信息,具有较高的图像分类准确率,且在训练时间上耗时最少,改进方法可以利用视觉导航系统实现无人机着陆地貌的识别和分类,降低无人机对于外界信息和其他机载设备的依赖性,提高无人机着陆的自主性和准确性。

参 考 文 献

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

[2] Szegedy C, Liu W, Jia Y Q, *et al.* Going deeper with convolutions[C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 1-9.

[3] Liu D W, Han L, Han X Y. High spatial resolution remote sensing image classification based on deep learning[J]. *Acta Optica Sinica*, 2016, 36 (4): 0428001.

刘大伟, 韩玲, 韩晓勇. 基于深度学习的高分辨率遥感影像分类研究[J]. *光学学报*, 2016, 36 (4): 0428001.

[4] Fang X, Wang G H, Yang H C, *et al.* High resolution remote sensing image classification combing with mean-shift segmentation and full convolution neural network [J]. *Laser & Optoelectronics Progress*, 2018, 55(2): 022802.

方旭, 王光辉, 杨化超, 等. 结合均值漂移分割与全卷积神经网络的高分遥感影像分类[J]. *激光与光电*

- 子学进展, 2018, 55(2): 022802.
- [5] He X F, Zou Z R, Tao C, *et al.* Combined saliency with multi-convolutional neural network for high resolution remote sensing scene classification [J]. *Acta Geodaetica et Cartographica Sinica*, 2016, 45(9): 1073-1080.
- 何小飞, 邹峥嵘, 陶超, 等. 联合显著性和多层卷积神经网络的高分影像场景分类[J]. *测绘学报*, 2016, 45(9): 1073-1080.
- [6] Le Q V, Ranzato M, Monga R, *et al.* Building high-level features using large scale unsupervised learning [C]. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013: 8595-8598.
- [7] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks [J]. *Science*, 2006, 313(5786): 504-507.
- [8] Agarwal A, El-Ghazawi T, El-Askary H, *et al.* Efficient hierarchical-PCA dimension reduction for hyperspectral imagery[C]. *IEEE International Symposium on Signal Processing and Information Technology*, 2008: 353-356.
- [9] Pan Z J, Rust A G, Bolouri H. Image redundancy reduction for neural network classification using discrete cosine transforms [C]. *IEEE-INNS-ENNS International Joint Conference on Neural Networks*, 2000, 3: 3149-3155.
- [10] Dabbaghchian S, Ghaemmaghami M P, Aghagolzadeh A. Feature extraction using discrete cosine transform and discrimination power analysis with a face recognition technology [J]. *Pattern Recognition*, 2010, 43(4): 1431-1440.
- [11] Sun J P, Liu J Q. Coal and rock recognition method based on low frequency component characteristics of discrete cosine transform and learning vector quantization [J]. *Industry & Mine Automation*, 2015, 41(11): 1-6.
- 孙继平, 刘剑桥. 基于离散余弦变换低频分量特征及学习向量量化的煤岩识别方法[J]. *工矿自动化*, 2015, 41(11): 1-6.
- [12] Zou X Y, Xu X M, Qing C M, *et al.* High speed deep networks based on Discrete Cosine Transformation[C]. *IEEE International Conference on Image Processing*, 2014: 5921-5925.
- [13] Yan Z Z, Liu J J. Improved image compression coding method based on discrete cosine transform[J]. *Computer Technology and Development*, 2016(1): 147-149.
- 严珍珍, 刘建军. 基于离散余弦变换的图像压缩编码方法及改进[J]. *计算机技术与发展*, 2016(1): 147-149.
- [14] Dai C N. SVM visual classification based on weighted feature of genetic algorithm[C]. *IEEE Sixth International Conference on Intelligent Systems Design and Engineering Applications*, 2015: 786-789.
- [15] Chen S Z, Tian Y L. Pyramid of spatial relations for scene-level land use classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2014, 53(4): 1947-1957.
- [16] Xu S H, Mu X D, Zhao P, *et al.* Scene classification of remote sensing image based on multi-scale feature and deep neural network[J]. *Acta Geodaetica et Cartographica Sinica*, 2016, 45(7): 834-840.
- 许凤晖, 慕晓冬, 赵鹏, 等. 利用多尺度特征与深度网络对遥感影像进行场景分类[J]. *测绘学报*, 2016, 45(7): 834-840.