

基于小波变换和深度网络的着陆地貌图像分类

刘 芳, 韩 笑

(北京工业大学信息学部, 北京 100124)

摘 要: 针对无人机着陆地貌图像场景复杂、纹理特征丰富等问题, 提出一种基于小波变换和深度网络的无人机着陆地貌图像分类算法. 利用非下采样小波变换(Non-Subsampled Wavelet Transform, NSWT)的快速压缩能力, 将小波变换后的前两层子图系数引入到卷积神经网络(CNN)中, 压缩数据量. 根据无人机着陆地貌图像的特点, 采用轻量化卷积模块设计了 15 层卷积神经网络. 通过支持向量机(SVM)实现复杂地貌场景的正确分类. 实验结果表明: 所提算法具有良好的特征表达能力, 提升了着陆地貌图像的分类准确率.

关键词: 非下采样小波变换; 卷积神经网络; 支持向量机; 图像分类

中图分类号: TN911.73; TP391

文献标识码: A

文章编号: 0372-2112(2021)11-2171-06

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20200870

Image Classification of Landing Landform Based on Wavelet Transform and Deep Network

LIU Fang, HAN Xiao

(Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China)

Abstract: Aiming at the problems of complex scenes and rich texture features of UAV landing landform image, a classification algorithm of UAV landing landform image based on wavelet transform and deep network is proposed. Using the fast compression capability of non-subsampled wavelet transform (NSWT), the first two layers of sub-image coefficients after wavelet transform are introduced into the convolutional neural network (CNN) to compress the amount of data. According to the characteristics of UAV landing landform image, a 15-layer convolutional neural network is designed using a lightweight convolution module. The correct classification of complex geomorphological scenes is realized by support vector machine (SVM). The experimental results show that the proposed algorithm has good feature expression ability and improves the classification accuracy of landing landform images.

Key words: non-subsampled wavelet transform; convolutional neural network; support vector machine; image classification

1 引言

随着无人机的广泛应用, 无人机着陆地貌图像分类成为无人机安全着陆和执行任务的关键. 由于无人机地貌图像中场景复杂, 类间相似度高, 导致传统地貌分类算法分类准确率低, 影响无人机安全着陆. 因此, 实现快速准确的无人机着陆地貌图像分类具有重要意义.

近年来, 卷积神经网络(CNN)在图像处理领域应用广泛. Krizhevsky 等^[1]首次将 CNN 用于图像分类, 提出了 AlexNet, 取得了当时最好的分类结果, 但 AlexNet 的特征提取能力有限. 方等^[2]采用全卷积网络代替机

器学习用于图像分类, 减少特征信息的丢失, 提升分类精度. 与传统方法相比, 基于 CNN 的方法具有更好的特征表达能力, 但是存在大量网络参数.

针对上述问题, 徐等^[3]采用小波变换选择中高频带后重构图像并计算共生矩阵特征值, 减少了数据量. 张等^[4]将 Haar 小波变换引入到 CNN 中, 提取图像的空间信息, 进行分层表达, 减少了参数量. 因此将小波变换引入卷积神经网络可增强特征表达能力, 减少参数量. 首先, 小波变换具有能量集中的特点, 能快速压缩图像信息, 降低数据量, 同时保留了最能表达原图像的重要特征. 其次, 小波变换具有多频域分解特点, 在不同频

域上表示不同的地貌特征信息. 此外,小波变换可分层表达彩色图像的颜色、轮廓、纹理等信息,依照某一明显特征进行分类能取得更好的效果.

综上所述,针对无人机着陆地貌图像场景复杂、纹理特征丰富的问题,提出一种基于小波变换和深度网络的无人机着陆地貌图像分类算法(WT-CNN). 首先将非下采样小波变换引入深度网络,在保留图像主要特征信息的前提下压缩数据信息,凸显信号的特征. 然后,根据无人机着陆地貌图像的特点,将轻量化卷积模块引入传统CNN网络模型,提取着陆地貌图像的深层次的特征表达,使其适用于无人机着陆地貌场景分类. 最后采用支持向量机(SVM)进行着陆地貌分类.

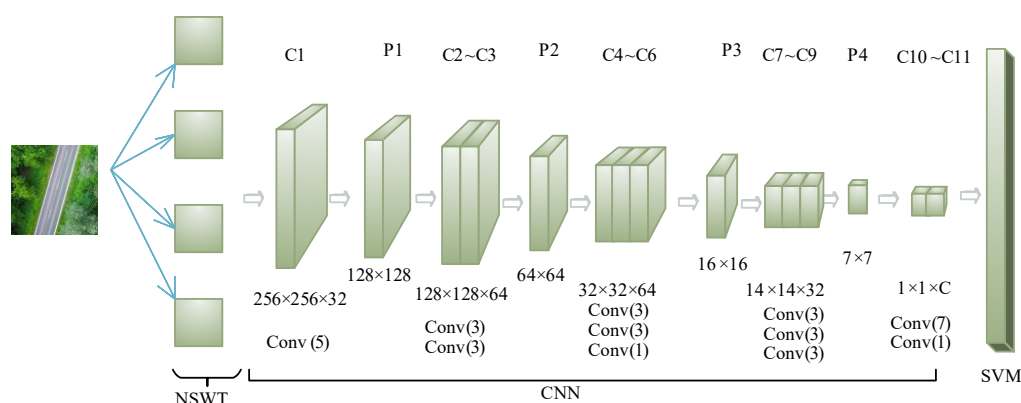


图1 算法网络结构图

2.1 非下采样小波变换初始化卷积核

非下采样小波变换(Non-Subsampled Wavelet Transform, NSWT)相比经典小波变换具有冗余性和平移不变性,可以较好的保存主要的边缘及细节信息,常应用在特征提取和边缘检测等领域.

图像作为二维数字信号,利用小波变换进行快速分解,分别得到低频子图和三个方向上的高频子图,如下式所示:

$$A_i(d, l) = \sum_{k_1} \sum_{k_2} g(k_1 - 2d) g(k_2 - 2l) f_{i-1}(k_1, k_2) \quad (1)$$

$$D_i^1(d, l) = \sum_{k_1} \sum_{k_2} h(k_1 - 2d) g(k_2 - 2l) f_{i-1}(k_1, k_2) \quad (2)$$

$$D_i^2(d, l) = \sum_{k_1} \sum_{k_2} g(k_1 - k_2) h(k_2 - 2l) f_{i-1}(k_1, k_2) \quad (3)$$

$$D_i^3(d, l) = \sum_{k_1} \sum_{k_2} h(k_1 - k_2) h(k_2 - 2l) f_{i-1}(k_1, k_2) \quad (4)$$

其中 $h()$ 和 $g()$ 分别表示高通和低通滤波器, k_1 表示图像数据水平方向平移相位差的倍数, k_2 表示图像数据垂直方向平移相位差的倍数, d 为位置变换参数, l 表示包含局部特征的子图数量.

NSWT可以将原始图像分解为低频子图以及水平、垂直、对角线三个方向上的高频子图,然后将低频子图以相同方式分解为第二层的四张不同频域的子图,以

2 基于小波变换和深度网络的着陆地貌图像分类

无人机地貌图像场景复杂、纹理信息丰富,具有较多的语义信息和较高的类间相似度,导致传统方法分类效果差. 因此,设计了适用于着陆地貌图像分类的15层卷积神经网络模型WT-CNN,网络结构如图1所示,主要分为三个部分:1)利用非下采样小波变换分解图像,使用前两层子图系数初始化卷积核,保留有用信息,压缩着陆地貌图像的数据量;2)采用轻量化模块改进网络模型,提取图像特征;3)用SVM对特征向量进行分类.

此类推. 选取小波变换后第一、二层子图系数输入卷积神经网络中初始化卷积核,压缩数据量,保持压缩后的特征不变性. 首先,低频子图对应原图的背景整体信息,高频子图对应原图的纹理细节、边缘、轮廓信息,有利于突出局部纹理特征. 随着小波变换不断分解,得到的高频子图中所含细节信息越来越少,前两层高频子图包含的纹理特征细节相对丰富,第三层及更高层子图包含的细节特征过少不利于特征的提取,因此采用第一、二层子图系数,有利于提取完整的特征. 其次,NSWT在信号的分解过程中未进行下采样操作,而是在高通和低通滤波器中每两个系数之间插入零来实现对滤波器的扩展,在图像信号分解过程中不采用下采样操作,能完整保留原图像的纹理结构信息,不会造成特征的缺失. 同时卷积神经网络也是对整幅图像进行处理,可以增强特征表达能力.

2.2 CNN模型网络结构

针对传统卷积神经网络参数量大、准确率低等问题,提出了用于无人机着陆地貌图像的15层的深度网络结构. 如表1所示, C1~C11表示卷积层, P1~P4表示池化层,每个卷积层后添加批量归一化(BN)和PReLU激活函数,加快收敛速度.

卷积核的设计是图像进行卷积操作有效提取特征

表1 基于NSWT和CNN的网络结构

网络层	类型	卷积核	零填充	输出
x	input			256×256
C1	Convolution	5×5	2	256×256×32
	PReLU			
P1	Max Pooling	3×3		128×128
C2	Convolution	3×3	1	128×128×64
C3	Convolution	3×3	1	128×128×64
	PReLU			
P2	Max Pooling	3×3		64×64
C4	Convolution	3×3	1	32×32×64
C5	Convolution	3×3	1	32×32×64
C6	Convolution	1×1		32×32×32
	PReLU			
P3	Max Pooling	3×3		16×16
C7	Convolution	3×3		14×14×32
C8	Convolution	3×3		14×14×32
C9	Convolution	3×3		14×14×32
	PReLU			
P4	Global Mean Pooling	3×3		7×7
C10	Convolution	7×7		1×1×32
C11	Convolution	1×1		1×1×C
	SVM			v

的重要前提,卷积核尺寸过大会增加网络计算复杂度,过小会导致特征不完整.为了提取完整的高层语义特征,本文设计了11层卷积的结构.如表1所示,由于前两层输入图像分辨率相对较大,第一层(C1)卷积核大小为5×5,步长为1,能有效的保持地貌特征的完整性,减小计算量.然后设计了三个由3×3卷积构成的轻量化卷积模块(C2~C3, C4~C6, C7~C9),分别采用大小为3×3和1×1两种卷积方式,针对较小的图像分辨率和地貌特征,该卷积核能保持较大的感受野,同时能控制网络的参数量,提取到地貌图像深层特征,其中1×1卷积主要用于降低数据维度.最后设计一个包含两层卷积(C10~C11)的网络模块,分别采用大小为7×7和1×1卷积层,步长为1,虽然增加了尺寸,但是提取到了更加完整、准确的特征表达.这种设计既能有效的提取完整的全局特征,又能减少计算开销,其中1×1卷积层的主要计算发生在通道维上,在减少模型参数的同时通过调整网络之间的通道数来控制模型的复杂度,用于整合特征,以便分类器进行分类.此外,WT-CNN在网络模型的不同卷积层中增加了零填充(Zero-Padding)操作,有利于网络提取到更完整的地貌图像特征.

在卷积层与池化层之间,采用了批量归一化(BN)^[5]操作和激活函数PReLU^[6].选择PReLU函数代替ReLU作为线性修正单元,PReLU函数具有非饱和性,在负半轴输出不为零,避免了梯度值为零的问题,

其既能克服输入为正时的梯度饱和问题,又能缓解输入为负时ReLU函数的均值漂移和神经元死亡问题^[7].如式(5)所示:

$$\text{PReLU}(x) = \begin{cases} x, & x > 0 \\ ax, & x \leq 0 \end{cases} \quad (5)$$

其中 x 为特征输入, a 为可学习参数.

池化层能缓解卷积层对空间的过度敏感性,给特征引入平移与旋转不变性^[6].为了保持特征的完整性和有效性,如表1所示,WT-CNN模型中采用了两种池化方式,用于降低数据维度,保留全局特征,最后两层卷积层输出维度小,不需要池化.针对分辨率较高的图像,前三层(P1、P2、P3)采用最大池化,选取图像的最大值作为该区域的池化后的值,有效保留了图像的纹理特征,有利于提取图像的主要特征.最后一层(P4)采用全局平均池化,选取区域图像的平均值作为该区域池化后的值,有效保留了图像全局特征和突出背景信息.

在CNN结构中,全连接层作为分类器整合卷积层中具有类别区分性的局部信息^[8].CNN结构是基于经验风险最小化原则,最小化训练集的误差.当反向传播算法找到第一个分类超平面时,不管是否最优,训练过程都会终止.而支持向量机(SVM)是基于结构化风险最小化原则,能最小化经验误差,快速找到满足具有“最大间隔”的分类超平面对特征向量进行正确分类.因此,通过使用SVM替换CNN的全连接层将会加强分类的准确性^[9].此外,采用SVM代替全连接层进行分类可以减少全连接方式产生的参数量,加快分类速度.

2.3 算法流程

所提算法流程如图2所示,主要步骤为:

(1)对训练样本进行非下采样小波变换,对原图进行不同频域的分解,分解为低频子图和水平、垂直、对角线三个方向的高频子图,快速压缩图像信息并保持压缩后的特征不变性,突出主要特征.

(2)将训练样本输入到网络模型中训练,初始化卷积核,逐层提取特征,得到特征向量.采用交叉熵损失函数对网络进行训练.

(3)将得到的特征向量输入到SVM中训练,调整更新网络参数,直至损失函数收敛至最优值,得到最终结果.其中SVM采用Hinge Loss进行训练.

(4)将测试图片输入到训练好的网络模型中进行测试,提取特征,得到特征映射,输入训练好的SVM中进行分类,得到分类结果.

3 实验结果与分析

根据实际应用需求,实验数据选取了UC Merced LU数据库和实验室构建的无人机着陆地貌图像数据集UAV Landform Dataset2和UAV Landform Dataset1验

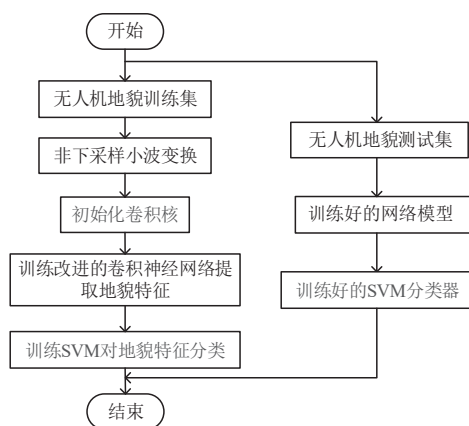


图2 算法流程图

证所提出的网络的性能。其中,UC Merced LU 数据集包含 21 类,共 2100 幅图像,选取 70% 作为训练集,30% 作为测试集。UAV Landform Dataset2 和 UAV Landform Dataset1 是实验室自建数据库,所有图像和视频全部来自实际无人机拍摄。其中,UAV Landform Dataset2 包含 10 类共 7200 张,其中 5000 张作为训练集,2200 张作为测试集,UAV Landform Dataset1 包含 6 类共 4800 张图像,其中 3300 张是训练集,1500 张是测试集。网络模型采用随机梯度下降算法训练,其中,训练参数 batch 大小设为 64,权重衰减设为 0.0005,动量参数为 0.9,学习率设为 0.001,每个数据集分别迭代 60 个 epoch。实验所用操作系统为 Windows 10,处理器为 i7-9800X,内存为 32GB 的计算机,GPU 为 NVIDIA GeForce GTX 1080Ti。

3.1 非下采样小波变换对分类性能的影响

为了验证小波变换分解层数对图像分类准确率的影响,分别在上述三个数据集上进行对比试验,如表 2、3 所示。其中 Method 1 是指图像不进行非下采样小波分解,直接输入卷积神经网络进行网络训练。Method 2 是指图像经过非下采样小波变换后,取所有分解层作为输入初始化卷积核,进行模型训练。WT-CNN 是所提方法取图像进行小波变换后前两层子图系数初始化卷积核,进行网络训练。表 2~表 6 中的 Dataset1、Dataset2 和 Database 分别代表数据集 UAV Landform Dataset1、UAV Landform Dataset2 和 UC Merced LU Database。

表 2 每个数据集中不同方法的分类准确率

准确率	Dataset1	Dataset2	Database
Method 1	89.73%	88.61%	88.75%
Method 2	97.49%	97.02%	97.38%
WT-CNN	97.37%	96.85%	97.25%

从表 2、3 中可以看出,本文算法在三个数据集中分别取得了 97.25%、96.85% 和 97.37% 的分类准确率,

表 3 每个数据集中不同方法的推理时间

推理时间(s)	Dataset1	Dataset2	Database
Method 1	0.033	0.032	0.031
Method 2	0.068	0.061	0.063
WT-CNN	0.036	0.035	0.034

相比于 Method1 分别提升了 8.5%、8.24%、7.64%。相比于 Method2,虽然没有准确率的明显提升,但降低了推理速度,推理时间缩短了 40% 左右。主要原因是非下采样小波变换能有效降低数据维度,同时具有良好的特征表达能力,在信号的分解过程中未进行下采样操作,保留完整的特征信息,对图像进行 NSWT 后作为卷积神经网络的第一层可以提取到更完整、辨识度更高的特征,有利于提升分类性能。经过非下采样小波分解后,低频子带描述地貌图像的整体背景信息,高频子带描述不同地貌的纹理、边缘等特征,随着子图分解层数增多,所包含的细节信息就会减少。如果取小波分解后前三层或者更高层子图系数作为改进的卷积神经网络模型的输入,虽然对地貌特征的表达能力略有提升,但会造成更大的计算代价,难以满足实时性的需求。如果分解层数较少,容易丢失部分重要的特征信息,导致提取到的地貌特征不完整,降低准确率。因此,经过非下采样小波变换后图像的主要特征信息主要集中在前两层子带中,选择这两层作为卷积神经网络的输入初始化卷积核更有利于地貌特征的提取。

3.2 支持向量机对分类性能的影响

为了验证支持向量机和全连接层对图像分类算法的影响,设计了对比实验,对图像进行特征提取之后,分别采用全连接层和支持向量机的分类效果。结果如表 4 和表 5 所示,可以看出,所提网络在特征提取之后,采用支持向量机分类得到的分类效果优于全连接层的分类结果,在不同数据集中准确度分别上涨 1.5% 左右,同时,算法的推理时间大幅减少,推理速度提升了约 45%。主要原因是 CNN 能够从输入图像有效地学习不变特征,但不一定能得到最优的分类结果,支持向量机虽然不能学习复杂的不变特征,但具有快速全局寻优能力,快速找到满足具有“最大间隔”的分类超平面,在学习到的特征空间中得到分类问题的最优解。而且 SVM 替换全连接层进行特征分类舍弃了原有的全连接方式,提升了分类速度。因此,本文算法结合了 CNN 和 SVM 的优点,补偿了它们的局限性,在无人机地貌图像上取得了更好的分类效果。

表 4 每个数据集中不同方法的分类准确率

准确率	Dataset1	Dataset2	Database
WT-CNN(FC)	95.79%	95.28%	95.83%
WT-CNN(SVM)	97.37%	96.85%	97.25%

表 5 每个数据集中不同方法的推理时间

推理时间(s)	Dataset1	Dataset2	Database
WT-CNN(FC)	0.079	0.073	0.075
WT-CNN(SVM)	0.043	0.039	0.041

3.3 所提方法与已有方法对比实验

在相同实验条件下,将本文算法与SAE-SVM^[10]、CRF^[11]、VGG-16^[12]、ResNet^[13]、DenseNet^[14]和DCT-CNN^[15]等多种算法进行实验对比.表6是不同方法在三个数据集上的分类结果,从表中可以看出,所提算法的分类准确率分别为96.85%、97.37%和97.25%,高于其他分类算法.主要原因是非下采样小波变换具有多分辨率特性和压缩不变性,地貌图像经过非下采样小波变换后分解为不同频率的子图,使其能量集中在包含图像视觉信息的某一频率范围内,在保留图像主要特征信息的前提下压缩数据、去除图像冗余信息,压缩比高且压缩后特征具有不变性,凸显信号的特征,既能有效提取地貌图像的主要特征,又能减少数据量.另一方面,传统卷积神经网络模型复杂,训练过程中产生大量参数,训练过程耗时长,网络中的轻量化卷积模块采用了小尺寸卷积核对地貌图像进行卷积,多个3×3卷积核可以得到与大尺寸卷积核相同的感受野,既保证相同的输出大小,又能减少参数量.此外采用1×1卷积固定通道维度,减少卷积核的通道维数,计算主要发生在通道维上,能极大地减少计算量.

表 6 现有算法在不同数据集的分类性能比较

算法	Dataset1	Dataset2	Database
SAE-SVM	86.26%	85.45%	88.97%
CRF	89.07%	89.16%	92.46%
VGG-16	91.47%	89.78%	89.98%
ResNet	94.43%	93.63%	93.64%
DenseNet	96.03%	95.46%	95.97%
DCT-CNN	94.38%	94.65%	95.76%
WT-CNN	97.37%	96.85%	97.25%

为了验证所提算法在实际场景应用中的效果,给出了本文算法在无人机着陆地貌图像数据集UAV Landform Dataset2上的分类混淆矩阵.如图3所示,可以看出,本文算法的地貌分类结果均在95%以上,分类准确率较高.此外,该算法对实验数据集中Slope、Road等边缘、纹理特征差异不明显的无人机着陆地貌图像具有较好的分类准确率.

4 结论

针对无人机着陆地貌图像场景复杂,纹理特征丰富,类间相似度高问题,提出一种基于小波变换和深度网络的无人机地貌场景分类算法(WT-CNN).采用非

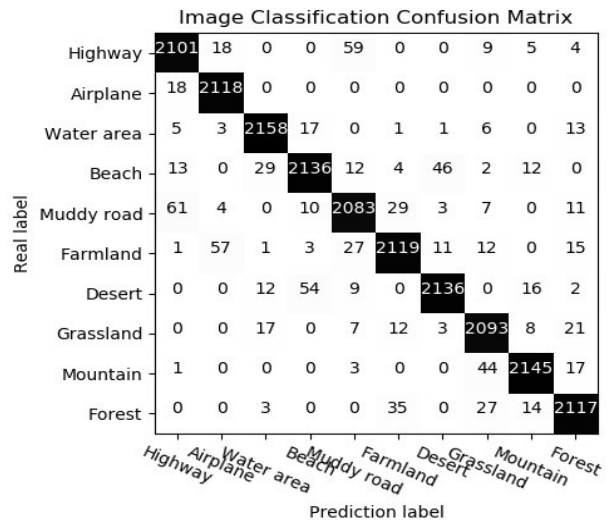


图 3 UAV Landform Dataset2图像分类混淆矩阵

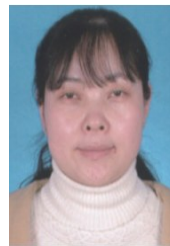
下采样小波变换将着陆地貌图像分解为不同频域的子图系数初始化卷积核,在保持原图像有效信息的前提下压缩数据量,突出图像的局部特征;为了提高对着陆地貌图像的分类准确率,设计了15层卷积神经网络模型,提升网络的特征表达能力,降低计算量.实验结果表明,本文算法在着陆地貌图像中达到了较高分类准确率,实现了复杂环境中无人机着陆地貌场景的准确分类.

参考文献

- [1] Krizhevsky A, Sutskever II, Hinton G. ImageNet classification with deep convolutional neural networks[A]. Proceedings of the Advances in Neural Information Processing Systems[C]. USA, 2012. 1097 – 1105.
- [2] 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辨遥感影像分类[J].激光与光电子学进展, 2018, 55(2):446 – 454.
Fang X, Wang G H, Yang H C, et al. High resolution remote sensing image classification combining with mean-shift segmentation and fully convolution neural network[J]. Laser & Optoelectronics Progress, 2018, 55 (2):446 – 454. (in Chinese)
- [3] 徐婷婷,吉晓东,李文华,等.基于颜色和纹理特征的胶囊内镜图像分类[J].现代电子技术, 2018, 41(19):66 – 70.
Xu T T, Ji X D, Li W H, et al. Capsule endoscope image classification based on color and texture features [J]. Modern Electronics Technique, 2018, 41 (19): 66 – 70. (in Chinese)
- [4] 张慧娜,李裕梅,傅莺莺.基于Haar-CNN模型的自然场景图像分类的研究[J].四川师范大学学报(自然科学版), 2017,40(01):119 – 126.

- Zhang H N, Li Y M, Fu Y Y. Research on natural scene image classification based on Haar-CNN model [J]. Journal of Sichuan Normal University (Natural Science), 2017, 40 (01): 119 – 126. (in Chinese)
- [5] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift[J]. Machine Learning, arXiv:1502.03167, 2015.
- [6] He K, Zhang X, Ren S, et al. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification[A]. IEEE International Conference on Computer Vision (ICCV) [C]. Chile: IEEE, 2015. 1026 – 1034.
- [7] Clevert D E, Unterthiner T, Hochreiter S. Fast and accurate deep network learning by exponential linear units (elus) [J]. arXiv:2015, 1511.07289.
- [8] Sainath TN, Mohamed A, Kingsbury B, et al. Deep convolutional neural networks for LVCSR[A]. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing[C]. Vancouver: IEEE, 2013. 8614 – 8618.
- [9] 李明威. 图像分类中的卷积神经网络方法研究[D]. 南京邮电大学, 2016.
Li M W. Research of Convolution Neural Network in Image Classification [D]. Nanjing University of Posts and Telecommunications, 2016. (in Chinese)
- [10] Ju Y, Guo J, Liu S. A deep learning method combined sparse autoencoder with SVM[A]. 2015 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery[C]. USA: IEEE, 2015. 257 – 260.
- [11] Zeggada A, Benbraika S, Melgani F, et al. Multilabel conditional random field classification for UAV images[J]. IEEE Geoscience and Remote Sensing Letters, 2018, 15 (99): 1 – 5.
- [12] Karen S, Andrew Z, et al. Very deep convolutional networks for large-scale image recognition[J]. Computer Science, arXiv 1409.1556, 2014.
- [13] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[A]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [C]. USA: IEEE, 2016. 770 – 778.
- [14] Huang G, Liu Z, Laurens V D M, et al. Densely connected convolutional networks[A]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [C]. USA: IEEE, 2017. 2261 – 2269.
- [15] 刘芳, 路丽霞, 黄光伟, 等. 基于离散余弦变换和深度网络的地貌图像分类[J]. 光学学报, 2018, 38(6): 266 – 274.
Liu F, Lu L X, Huang G W, et al. Landform image classification based on discrete Cosine transformation and deep network [J]. Acta Optica Sinica, 2018, 38 (6): 266 – 274. (in Chinese)

作者简介



刘 芳 女, 1972 年生, 陕西西安人. 现为北京工业大学信息学部副教授、硕士生导师, 主要研究领域为图像处理、深度学习、视频跟踪.
E-mail: liufang@bjut.edu.cn



韩 笑 (通信作者) 男, 1996 年生, 河北沧州人. 现为北京工业大学信息学部硕士研究生, 主要研究方向为数字图像处理、深度学习.
E-mail: hxdeemail@163.com