



(12) 发明专利

(10) 授权公告号 CN 110443143 B

(45) 授权公告日 2020.12.18

(21) 申请号 201910614263.4

审查员 刘璇

(22) 申请日 2019.07.09

(65) 同一申请的已公布的文献号  
申请公布号 CN 110443143 A

(43) 申请公布日 2019.11.12

(73) 专利权人 武汉科技大学  
地址 430081 湖北省武汉市青山区和平大道947号

(72) 发明人 边小勇 陈春芳 张晓龙 盛玉霞

(74) 专利代理机构 武汉东喻专利代理事务所  
(普通合伙) 42224

代理人 王聪聪

(51) Int. Cl.

G06K 9/00 (2006.01)

G06K 9/62 (2006.01)

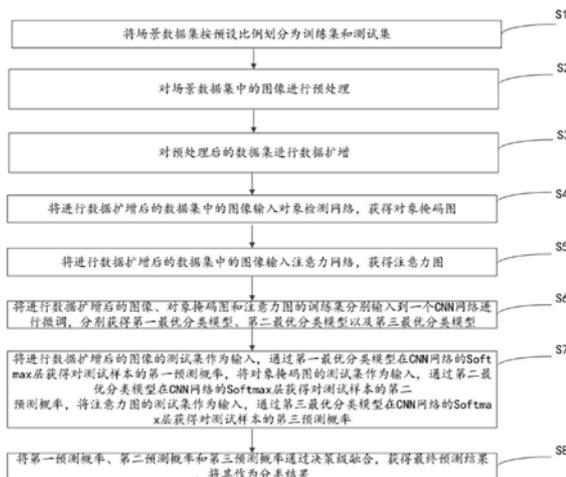
权利要求书3页 说明书11页 附图3页

(54) 发明名称

多分支卷积神经网络融合的遥感图像场景分类方法

(57) 摘要

本发明公开了一种多分支卷积神经网络融合的遥感图像场景分类方法,首先将场景数据集按比例随机划分成训练集和测试集;接着对数据集进行预处理和数据扩增;将处理后的数据分别通过对象检测网络和注意力网络获得对象掩码图和注意力图;然后将原图像、对象掩码图和注意力图训练集分别输入到一个CNN网络进行微调,分别获得最优分类模型,再将三组测试集作为输入分别通过最优分类模型获得Softmax层的输出,最后将三组Softmax层的输出通过决策级融合,获得最终的预测结果。本发明可以提高分类准确性和分类效果。



1. 一种多分支卷积神经网络融合的遥感图像场景分类方法,其特征在于,包括:

步骤S1:将场景数据集按预设比例划分为训练集和测试集;

步骤S2:对场景数据集中的图像进行预处理;

步骤S3:对预处理后的数据集进行数据扩增;

步骤S4:将进行数据扩增后的数据集中的图像输入对象检测网络,获得对象掩码图;

步骤S5:将进行数据扩增后的数据集中的图像输入注意力网络,获得注意力图;

步骤S6:将进行数据扩增后的图像、对象掩码图和注意力图的训练集分别输入到一个CNN网络进行微调,分别获得第一最优分类模型、第二最优分类模型以及第三最优分类模型;

步骤S7:将进行数据扩增后的图像的测试集作为输入,通过第一最优分类模型在CNN网络的Softmax层获得对测试样本的第一预测概率,将对象掩码图的测试集作为输入,通过第二最优分类模型在CNN网络的Softmax层获得对测试样本的第二预测概率,将注意力图的测试集作为输入,通过第三最优分类模型在CNN网络的Softmax层获得对测试样本的第三预测概率;

步骤S8:将第一预测概率、第二预测概率和第三预测概率通过决策级融合,获得最终预测结果,将其作为分类结果;

其中,步骤S4具体包括:

步骤S4.1:设置网络学习参数;

步骤S4.2:加载预训练模型,初始化网络权重,偏差参数;

步骤S4.3:将进行数据扩增后的数据集中的图像输入对象检测网络,以软最大化损失和辅助损失的组合方式进行微调,获得最优分割模型,组合的损失公式如下:

组合的损失:

$$L=L_1+L_2 \quad (1)$$

其中,软最大化损失 $L_1$ 为:

$$L_1 = -\sum_{i=1}^D y_i \log \frac{e^{z_i}}{\sum_{j=1}^D e^{z_j}} \quad (2)$$

辅助损失 $L_2$ 为:

$$L_2 = -\alpha \times \sum_{i=1}^D y_i \log \frac{e^{z'_i}}{\sum_{j=1}^D e^{z'_j}} \quad (3)$$

在式(2)中:

$z_i$ 表示输入软最大化损失层的第 $i$ 个类上的特征,

$y_i$ 表示第 $i$ 个类上的真实类别,

$D$ 表示类别数;

在式(3)中:

$\alpha$ 表示辅助损失权重,

$z'_i$ 表示输入辅助损失层的第 $i$ 个类上的特征;

步骤S4.4:将进行数据扩增后的图像的测试集作为输入,利用步骤S4.3获得的最优分割模型,得到对象分割图;

步骤S4.5:将对象分割图映射到进行数据扩增后的图像,产生对象掩码图;

步骤S8具体包括:

步骤S8.1:不同分支加权的预测,根据以上生成的三个分支的预测概率矩阵  $P_B = \{p_j\}_{j=1}^3$ ,采用加权系数进行聚合,并经过交叉验证找到权重系数  $\{w_j\}_{j=1}^3$ ,预测公式为:

$$l_p(s) = \operatorname{argmax}_{k \in \{1, \dots, D\}} \sum_{j=1}^3 w_j p_j \quad (4)$$

式(4)中:

$l_p(s)$ 表示具有最高概率的场景图像的预测类别,

$D$ 表示类别数;

步骤S8.2:基于大数表决的预测,找出三个预测分支结果中具有最大次数的类别  $l_m(s)$  作为场景图像  $s$  的预测类别,预测公式为:

$$l_m(s) = \operatorname{argmax}_{k \in \{1, \dots, D\}} (\operatorname{Count}([p_1 \ p_2 \ p_3]) = k) \quad (5)$$

式(5)中:

$\operatorname{Count}(\cdot)$ 表示计数器函数,统计三个分支上预测出的相同类别的次数,

$k$ 表示类别号;

步骤S8.3:将上述S8.1和S8.2两种预测方法中具有最高准确度的预测作为最终的预测结果。

2.如权利要求1所述的方法,其特征在于,步骤S2具体包括:

步骤S2.1:将场景图像名与对应的类别编号保存为txt文件;

步骤S2.2:将输入的常规图像格式转换为CNN网络要求的lmdb格式;

步骤S2.3:将转换为lmdb格式的图像的尺寸缩放到CNN网络要求的输入大小。

3.如权利要求1所述的方法,其特征在于,步骤S3具体包括:

对进行预处理后的数据集中包含的图像的中心区域和四个角落进行裁剪得到5个位置的新的图像数据,同时水平翻转裁剪后的图像。

4.如权利要求1所述的方法,其特征在于,步骤S5具体包括:

步骤S5.1:将由空间变换网络、网格生成器和采样器组成的空间变换模块嵌入到一个标准的CNN网络中,构成注意力网络;

步骤S5.2:将数据扩增后的图像的训练集作为输入,对注意力网络进行微调,获得最优注意力模型;

步骤S5.3:将数据扩增后的图像的测试集作为输入,利用步骤S5.2获得的最优注意力模型,产生注意力图。

5.如权利要求1所述的方法,其特征在于,步骤S6具体包括:

步骤S6.1:将进行数据扩增后的图像、对象掩码图和注意力图的训练集作为不同的分支输入图像;

步骤S6.2:加载在ImageNet数据集上预训练好的CNN模型;

步骤S6.3:采用随机梯度下降算法和软最大化损失分别对CNN进行微调,达到最大迭代次数或最高验证精度时结束训练,分别获得第一最优分类模型、第二最优分类模型以及第三最优分类模型。

6.如权利要求1所述的方法,其特征在于,步骤S7具体包括:

步骤S7.1:将进行数据扩增后的图像测试集作为输入,利用第一最优分类模型,获得在CNN网络的Softmax层获得对测试样本的第一预测概率,其中,第一预测概率为大小为num\_class $\times$ 1的概率分布特征向量 $p_1$ ;

步骤S7.2:将对象掩码图测试集作为输入,利用第二最优分类模型,获得在CNN网络的Softmax层获得对测试样本的第二预测概率,其中,第二预测概率为大小为num\_class $\times$ 1的概率分布特征向量 $p_2$ ;

步骤S7.3:将注意力图测试集作为输入,利用最优分类模型,获得在CNN网络的Softmax层获得对测试样本的第三预测概率,其中,第三预测概率为大小为num\_class $\times$ 1的概率分布特征向量 $p_3$ 。

## 多分支卷积神经网络融合的遥感图像场景分类方法

### 技术领域

[0001] 本发明涉及遥感图像场景分类技术领域,具体涉及多分支卷积神经网络融合的遥感图像场景分类方法。

### 背景技术

[0002] 遥感图像场景分类任务作为遥感图像处理技术的重要分支,在军用和民用领域都具有重要的意义。场景分类旨在通过学习的分类器自动地为每个场景图像预测一个语义类别。然而,遥感图像场景具有不同颜色,视点,姿态,空间分辨率等丰富的变化和多种混合对象,且不同类别的若干图像场景在许多方面可能彼此相似。确切地说,由于类内多样性和类间相似性的问题,使得遥感图像场景分类依然面临挑战。

[0003] 传统上的方法主要集中在手工特征提取和编码上。这些方法包括尺度不变特征变换,稀疏表示和其他空间特征提取。此外,还有基于视觉单词编码,混合高斯模型,Fisher向量编码和改进的Fisher核编码等方法,用于编码颜色、纹理、大小和结构信息的低级特征,这些方法均可以概括为中级特征方法,需要较为繁琐的参数调整,而且不是端到端的学习过程。

[0004] 最近,深度学习方法越来越多地用于场景图像分类,该类方法可以进行端到端的学习,比手工特征具有更好的判别能力。目前深度学习的遥感图像场景分类方法主要有:面向网络不同学习策略的方法和判别增强的特征学习方法。K.Nogueira,O.Penatti and J.Santos等人发表的《用于遥感场景分类的更好卷积神经网络的探索》,模式识别,vol.61,pp.539-556,2016(《Towards better exploiting convolutional neural networks for remote sensing scene classification》(Pattern Recognition,vol.61,pp.539-556,2016.))中分别采用预训练网络,微调网络和完全训练网络模式对几个常见遥感图像场景进行分类,由于采用结构简单的卷积神经网络,因此特征提取的不充分从而导致较低的分类准确率。在遥感图像上的分类缺少针对性。Q.Wang,S.Liu,J.Chanussot and X.Li等人发表的《重复注意力非常高分辨率遥感图像的场景分类》,IEEE国际地理科学与遥感学报,vol.57,no.2,pp.1155-1167,2019(《Scene classification with recurrent attention of VHR remote sensing images》(IEEE Trans.Geoscience Remote Sensing,vol.57,no.2,pp.1155-1167,Feb.2019.))中利用多个长短期记忆模型对3个典型遥感图像场景进行多注意力区域特征提取及分类,由于注意力模型主要关注对象类型的场景并不能充分学习到自然场景的特征信息,因此模型的判别能力仍显不足。

[0005] 由此可知,现有技术中的方法存在分类效果不佳的技术问题。

### 发明内容

[0006] 有鉴于此,本发明提供了一种多分支卷积神经网络融合的遥感图像场景分类方法,用以解决或者至少部分解决现有技术中存在的分类效果不佳的技术问题。

[0007] 为了解决上述技术问题,本发明提供了一种多分支卷积神经网络融合的遥感图像

场景分类方法,包括:

[0008] 步骤S1:将场景数据集按预设比例划分为训练集和测试集;

[0009] 步骤S2:对场景数据集中的图像进行预处理;

[0010] 步骤S3:对预处理后的数据集进行数据扩增;

[0011] 步骤S4:将进行数据扩增后的数据集中的图像输入对象检测网络,获得对象掩码图;

[0012] 步骤S5:将进行数据扩增后的数据集中的图像输入注意力网络,获得注意力图;

[0013] 步骤S6:将进行数据扩增后的图像、对象掩码图和注意力图的训练集分别输入到一个CNN网络进行微调,分别获得第一最优分类模型、第二最优分类模型以及第三最优分类模型;

[0014] 步骤S7:将进行数据扩增后的图像的测试集作为输入,通过第一最优分类模型在CNN网络的Softmax层获得对测试样本的第一预测概率,将对象掩码图的测试集作为输入,通过第二最优分类模型在CNN网络的Softmax层获得对测试样本的第二预测概率,将注意力图的测试集作为输入,通过第三最优分类模型在CNN网络的Softmax层获得对测试样本的第三预测概率;

[0015] 步骤S8:将第一预测概率、第二预测概率和第三预测概率通过决策级融合,获得最终预测结果,将其作为分类结果。

[0016] 在一种实施方式中,步骤S2具体包括:

[0017] 步骤S2.1:将场景图像名与对应的类别编号保存为txt文件;

[0018] 步骤S2.2:将输入的常规图像格式转换为CNN网络要求的lmdb格式;

[0019] 步骤S2.3:将转换为lmdb格式的图像的尺寸缩放到CNN网络要求的输入大小。

[0020] 在一种实施方式中,步骤S3具体包括:

[0021] 对进行预处理后的数据集中包含的图像的中心区域和四个角落进行裁剪得到5个位置的新的图像数据,同时水平翻转裁剪后的图像。

[0022] 在一种实施方式中,步骤S4具体包括:

[0023] 步骤S4.1:设置网络学习参数;

[0024] 步骤S4.2:加载预训练模型,初始化网络权重,偏差参数;

[0025] 步骤S4.3:将进行数据扩增后的数据集中的图像输入对象检测网络,以软最大化损失和辅助损失的组合方式进行微调,获得最优分割模型,组合的损失公式如下:

[0026] 组合的损失:

$$[0027] \quad L = L_1 + L_2 \quad (1)$$

[0028] 其中,软最大化损失 $L_1$ 为:

$$[0029] \quad L_1 = -\sum_{i=1}^D y_i \log \frac{e^{z_i}}{\sum_{j=1}^D e^{z_j}} \quad (2)$$

[0030] 辅助损失 $L_2$ 为:

$$[0031] \quad L_2 = -\alpha \times \sum_{i=1}^D y_i \log \frac{e^{z'_i}}{\sum_{j=1}^D e^{z'_j}} \quad (3)$$

[0032] 在式(2)中:

[0033]  $z_i$ 表示输入软最大化损失层的第 $i$ 个类上的特征,

[0034]  $y_i$ 表示第*i*个类上的真实类别，

[0035]  $D$ 表示类别数；

[0036] 在式(3)中：

[0037]  $\alpha$ 表示辅助损失权重，

[0038]  $z'_i$ 表示输入辅助损失层的第*i*个类上的特征。

[0039] 步骤S4.4:将进行数据扩增后的图像的测试集作为输入，利用步骤S4.3获得的最优分割模型，得到对象分割图；

[0040] 步骤S4.5:将对象分割图映射到进行数据扩增后的图像，产生对象掩码图。

[0041] 在一种实施方式中，步骤S5具体包括：

[0042] 步骤S5.1:将由空间变换网络、网格生成器和采样器组成的空间变换模块嵌入到一个标准的CNN网络中，构成注意力网络；

[0043] 步骤S5.2:将数据扩增后的图像的测试集作为输入，对注意力网络进行微调，获得最优注意力模型；

[0044] 步骤S5.3:将数据扩增后的图像的测试集作为输入，利用步骤S5.2获得的最优注意力模型，产生注意力图。

[0045] 在一种实施方式中，步骤S6具体包括：

[0046] 步骤S6.1:将进行数据扩增后的图像、对象掩码图和注意力图的训练集作为不同的分支输入图像；

[0047] 步骤S6.2:加载在ImageNet数据集上预训练好的CNN模型；

[0048] 步骤S6.3:采用随机梯度下降算法和软最大化损失分别对CNN进行微调，达到最大迭代次数或最高验证精度时结束训练，分别获得第一最优分类模型、第二最优分类模型以及第三最优分类模型。

[0049] 在一种实施方式中，步骤S7具体包括：

[0050] 步骤S7.1:将进行数据扩增后的图像测试集作为输入，利用第一最优分类模型，在CNN网络的Softmax层获得对测试样本的第一预测概率，其中，第一预测概率为大小为num\_class $\times$ 1的概率分布特征向量 $p_1$ ；

[0051] 步骤S7.2:将对象掩码图测试集作为输入，利用第二最优分类模型，在CNN网络的Softmax层获得对测试样本的第二预测概率，其中，第二预测概率为大小为num\_class $\times$ 1的概率分布特征向量 $p_2$ ；

[0052] 步骤S7.3:将注意力图测试集作为输入，利用最优分类模型，在CNN网络的Softmax层获得对测试样本的第三预测概率，其中，第三预测概率的为大小为num\_class $\times$ 1的概率分布特征向量 $p_3$ 。

[0053] 在一种实施方式中，步骤S8具体包括：

[0054] 步骤S8.1:不同分支加权的预测，根据以上生成的三个分支的预测概率矩阵 $P_B = \{p_j\}_{j=1}^3$ ，采用加权系数进行聚合，并经过交叉验证找到权重系数 $\{w_j\}_{j=1}^3$ ，预测公式为：

$$[0055] \quad l_p(s) = \operatorname{argmax}_{k \in \{1, \dots, D\}} \sum_{j=1}^3 w_j p_j \quad (4)$$

[0056] 式(4)中：

[0057]  $l_p(s)$ 表示具有最高概率的场景图像的预测类别，

[0058] D表示类别数；

[0059] 步骤S8.2:基于大数表决的预测,找出三个预测分支结果中具有最大次数的类别 $l_m(s)$ 作为场景图像s的预测类别,预测公式为:

$$[0060] \quad l_m(s) = \operatorname{argmax}_{k \in \{1, \dots, D\}} (\operatorname{Count}([p_1 \ p_2 \ p_3]) = k) \quad (5)$$

[0061] 式(5)中:

[0062]  $\operatorname{Count}(\cdot)$ 表示计数器函数,统计三个分支上预测出的相同类别的次数,

[0063] k表示类别号;

[0064] 步骤S8.3:将上述S8.1和S8.2两种预测方法中具有最高准确度的预测作为最终的预测结果。

[0065] 本申请实施例中的上述一个或多个技术方案,至少具有如下一种或多种技术效果:

[0066] 本发明采用多分支卷积神经网络融合的遥感图像场景分类方法,首先利用对象检测网络训练结合图像上下文全局信息检测场景语义对象,以及利用注意力网络训练不断注意关键区域,最后学习局部重要区域。这两个分支网络的运用有助于提高模型对场景语义对象的检测能力。训练过程同时利用原图像、对象掩码图和注意力图组成多分支卷积神经网络来抽取具有判别且互补的多个特征,建立判别性强的分类模型。同时多分支卷积神经网络通过数据扩增技术从大量训练数据中学习到有益的特征信息。最后,由多分支卷积神经网络预测的概率矩阵进行组合表决,也就是使用基于预测概率的加权或大数表决中具有最高准确率的预测方法来赋予场景类别,具有提高遥感图像场景分类效果的特点。

[0067] 因此,本发明通过多分支卷积神经网络模型训练及其融合,优化损失函数,提取遥感图像场景的深度特征并分类,具有提高遥感图像场景分类效果的特点,解决了现有技术中的分类效果不佳的技术问题。

## 附图说明

[0068] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0069] 图1为本发明实施例中多分支卷积神经网络融合的遥感图像场景分类方法的流程图;

[0070] 图2为本发明实施例中遥感图像场景分类训练和测试的流程图;

[0071] 图3为本发明中对象检测网络结构流程图;

[0072] 图4为本发明中对象检测网络损失层结构图;

[0073] 图5为本发明中注意力网络结构流程图;

[0074] 图6为本发明中注意力模块嵌入到一个标准CNN网络的结构图;

[0075] 图7为本发明的实施例中遥感图像场景的图像示例;

[0076] 图8为图7中采用本发明的对象检测网络获得的对象分割图;

[0077] 图9为图7中采用本发明的注意力网络获得的注意力图。

## 具体实施方式

[0078] 本发明的目的在于针对现有技术中由于对类内多样性和类间相似性的遥感图像场景特征表示的不足,而导致的分类效果不佳的技术问题,提供一种基于多分支的卷积神经网络融合的遥感图像场景分类方法,从而达到提高分类准确性和分类效果的目的。

[0079] 为了实现上述技术效果,本发明的主要构思如下:

[0080] 提供一种多分支卷积神经网络融合的遥感图像场景分类方法,首先将场景数据集按比例随机划分成训练集和测试集;接着对数据集进行预处理和数据扩增;将处理后的数据分别通过对象检测网络和注意力网络获得对象掩码图和注意力图;然后将原图像、对象掩码图和注意力图训练集分别输入到一个CNN网络进行微调,分别获得最优分类模型,再将三组测试集作为输入分别通过最优分类模型获得Softmax层的输出,最后将三组Softmax层的输出通过决策级融合,获得最终的预测结果。

[0081] 本发明提供的多分支卷积神经网络融合的遥感图像场景分类方法,由于通过对象检测网络结合图像全局信息检测场景语义对象,又通过注意力网络学习局部重要区域,并且通过决策级融合,在遥感图像场景分类上取得了更好的效果。

[0082] 为使本发明实施例的目的、技术方案和优点更加清楚,下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0083] 为实现上述目的,先将本发明采用的技术方案中所涉及到的标记统一说明如下:

[0084] CNN表示卷积神经网络,

[0085] Softmax表示卷积神经网络分类层,

[0086] Caffe表示深度学习底层框架,

[0087] lmdb表示在Caffe中卷积神经网络训练需要的一种数据格式,

[0088] num\_class表示场景数据集的类别数。

[0089] 实施例一

[0090] 本实施例提供了一种多分支卷积神经网络融合的遥感图像场景分类方法,请参见图1,该方法包括:

[0091] 步骤S1:将场景数据集按预设比例划分为训练集和测试集。

[0092] 具体来说,场景数据集是指是某一个开源的图像场景数据集,该场景数据集中包含多个类别,每个类别包括多个图像。预设比例可以根据需要进行设置,例如1:9、2:8、3:7等等。

[0093] 具体示例中以NWPU-RESISC45场景图像数据集为例子,该数据集包括45个场景类,每个类包含700个图像,像素为 $256 \times 256$ 。

[0094] 步骤S2:对场景数据集中的图像进行预处理。

[0095] 具体来说,对场景数据集中的图像进行预处理,就是将场景数据集中的图像的格式和尺寸调整成CNN网络可以处理的形式。

[0096] 在一种实施方式中,步骤S2具体包括:

[0097] 步骤S2.1:将场景图像名与对应的类别编号保存为txt文件;

[0098] 步骤S2.2:将输入的常规图像格式转换为CNN网络要求的lmdb格式;

- [0099] 步骤S2.3:将转换为lmbd格式的图像的尺寸缩放到CNN网络要求的输入大小。
- [0100] 步骤S3:对预处理后的数据集进行数据扩增。
- [0101] 在一种实施方式中,步骤S3具体包括:
- [0102] 对进行预处理后的数据集中包含的图像的中心区域和四个角落进行裁剪得到5个位置的新的图像数据,同时水平翻转裁剪后的图像。
- [0103] 具体来说,通过上述数据扩增操作,可以得到10倍于原来图像规模的数据集。
- [0104] 步骤S4:将进行数据扩增后的数据集中的图像输入对象检测网络,获得对象掩码图。
- [0105] 其中,步骤S4是利用对象检测网络产生对象掩码图。
- [0106] 具体地,步骤S4具体可以通过下述步骤来实现:
- [0107] 步骤S4.1:设置网络学习参数;
- [0108] 步骤S4.2:加载预训练模型,初始化网络权重,偏差参数;
- [0109] 步骤S4.3:将进行数据扩增后的数据集中的图像输入对象检测网络,以软最大化损失和辅助损失的组合方式进行微调,获得最优分割模型,组合的损失公式如下:
- [0110] 组合的损失:
- [0111]  $L=L_1+L_2$  (1)
- [0112] 其中,软最大化损失 $L_1$ 为:
- [0113] 
$$L_1 = -\sum_{i=1}^D y_i \log \frac{e^{z_i}}{\sum_{j=1}^D e^{z_j}} (2)$$
- [0114] 辅助损失 $L_2$ 为:
- [0115] 
$$L_2 = -\alpha \times \sum_{i=1}^D y_i \log \frac{e^{z'_i}}{\sum_{j=1}^D e^{z'_j}} (3)$$
- [0116] 在式(2)中:
- [0117]  $z_i$ 表示输入软最大化损失层的第 $i$ 个类上的特征,
- [0118]  $y_i$ 表示第 $i$ 个类上的真实类别,
- [0119]  $D$ 表示类别数;
- [0120] 在式(3)中:
- [0121]  $\alpha$ 表示辅助损失权重,
- [0122]  $z'_i$ 表示输入辅助损失层的第 $i$ 个类上的特征。
- [0123] 步骤S4.4:将进行数据扩增后的图像的测试集作为输入,利用步骤S4.3获得的最优分割模型,得到对象分割图;
- [0124] 步骤S4.5:将对象分割图映射到进行数据扩增后的图像,产生对象掩码图。
- [0125] 具体来说,步骤S4.1中的网络学习参数包括:批处理大小,学习率,动量,最大迭代次数,权重衰减率等;
- [0126] 在具体实施时,可以将批处理大小设置为8,学习率设置为0.001,动量设置为0.9,最大迭代次数设置为50000,权重衰减率0.0001等,预训练模型为加载在ADE20K数据集上预训练好的对象检测网络模型,辅助损失权重 $\alpha$ 可以设置为0.4。
- [0127] 上述步骤S4.3中,所述对象检测网络结构如图3所示,具体包括:
- [0128] (1) 给定输入图像,使用CNN获取最后一个卷积层的特征,本实施例选择ResNet50

网络；

[0129] (2) 应用金字塔池化模块来收集不同子区域表示, 然后进行上采样和连接以形成最终的特征。其中, 金字塔池化模块包括: 不同步长的池化层以提取k个不同尺度的子区域特征和k个对应的卷积层以降低特征维数;

[0130] (3) 将连接后的特征通过最后一个卷积层, 获得对象分割图。

[0131] 其中, 图4为本发明中对象检测网络损失层结构图, 图7为本发明的实施例中遥感图像场景的图像示例, 最终生成的对象分割图如图8所示。

[0132] 步骤S5: 将进行数据扩增后的数据集中的图像输入注意力网络, 获得注意力图。

[0133] 其中, 步骤S5是利用注意力网络获得注意力图。

[0134] 具体地, 步骤S5具体可以通过下述步骤来实现:

[0135] 步骤S5.1: 将由空间变换网络、网格生成器和采样器组成的空间变换模块嵌入到一个标准的CNN网络中, 构成注意力网络;

[0136] 步骤S5.2: 将数据扩增后的图像的训集作为输入, 对注意力网络进行微调, 获得最优注意力模型;

[0137] 步骤S5.3: 将数据扩增后的图像的测试集作为输入, 利用步骤S5.2获得的最优注意力模型, 产生注意力图。

[0138] 具体来说, 步骤S5.1中, 将由空间变换网络, 网格生成器和采样器组成的空间变换模块嵌入到一个标准的CNN网络中, 如图6所示, 可以嵌入到多个卷积层前, 本实施方式中选择嵌入到第一个卷积层前, 最后构成如图5所示的注意力网络。

[0139] 上述步骤S5.1中, 所述的空间变换模块中具体步骤如下:

[0140] (1) 将输入图像U作为空间变换网络输入, 获得变换参数 $\theta$ ;

[0141] (2) 网格生成器通过矩阵运算计算出目标图像V中的每个位置对应到原图像U中的位置, 即生成T(G)。变换公式如下所示:

$$[0142] \quad \begin{pmatrix} x_i^s \\ y_i^s \\ 1 \end{pmatrix} = T_\theta(G_i) = A_\theta \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix} \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix} \quad (6)$$

[0143] 式(6)中:

[0144] s表示原始图的坐标,

[0145] t表示目标图的坐标,

[0146] A表示定位网络拟合出的6个角度值, 即变换参数 $\theta$ 。

[0147] (3) 将T(G)中的坐标信息以及原始图像经过采样器处理, 得到变换后的图像V。变换公式如下:

$$[0148] \quad V_i^c = \sum_n^H \sum_m^W U_{nm}^c k(x_i^s - m; \Phi_x) k(y_i^s - n; \Phi_y) \quad \forall i \in [1 \dots H'W'] \quad \forall c \in [1 \dots C] \quad (7)$$

[0149] 式(7)中:

[0150]  $\Phi_x$ 和 $\Phi_y$ 表示定义图像插值的采样核 $k(\cdot)$ 的参数;

[0151] C表示通道数;

[0152]  $U_{nm}^c$ 表示输入通道c中位置(n,m)的值;

[0153]  $V_i^c$ 表示输出值像素i在通道c中的位置 $(x_i^t, y_i^t)$ 。

[0154] 其中,产生的注意力图如图9所示。

[0155] 上述步骤S5.1中,所述的一个标准的CNN网络具体可以是:CaffeNet,GoogleNet,VGG-VD16和ResNet50,本实施例选择VGG-VD16网络。

[0156] 步骤S6:将进行数据扩增后的图像、对象掩码图和注意力图的训练集分别输入到一个CNN网络进行微调,分别获得第一最优分类模型、第二最优分类模型以及第三最优分类模型。

[0157] 具体来说,步骤S6是进行不同分支的CNN模型训练。

[0158] 在一种实施方式中,步骤S6具体包括:

[0159] 步骤S6.1:将进行数据扩增后的图像、对象掩码图和注意力图的训练集作为不同的分支输入图像;

[0160] 步骤S6.2:加载在ImageNet数据集上预训练好的CNN模型;

[0161] 步骤S6.3:采用随机梯度下降算法和软最大化损失分别对CNN进行微调,达到最大迭代次数或最高验证精度时结束训练,分别获得第一最优分类模型、第二最优分类模型以及第三最优分类模型。

[0162] 具体来说,ImageNet数据集可以为CaffeNet,GoogleNet,VGG-VD16和ResNet50等。其中,步骤S6.3中,软最大化损失仍使用上述步骤S4.3中公式(3)进行。

[0163] 上述步骤S4.3、步骤S5.2和步骤S6.3中,所述利用给定输入图像微调卷积神经网络的具体步骤如下:

[0164] (1) 正向传递学习,计算在当前网络参数下,网络每一层的输出特征。正向传播过程如下:

$$[0165] \quad a^l = \tau(z^l) = \tau(a^{l-1}w^l + b^l) \quad (8)$$

[0166] 式(8)中:

[0167]  $l$ 表示第 $l$ 层,

[0168]  $w^l$ 表示第 $l$ 层的权重,

[0169]  $b^l$ 表示第 $l$ 层的偏移值,

[0170]  $\tau(*)$ 表示激活函数,一般是ReLU,

[0171]  $z^l$ 表示卷积操作之后的输出,

[0172]  $a^l$ 表示第 $l$ 层的输出。

[0173] (2) 反向传播更新,采用随机梯度下降方法(SGD)算法,每迭代一次从训练样本中随机抽取一组样本,通过网络获得Softmax层的输出,计算梯度,更新网络参数,网络权重和偏差参数更新公式如下:

$$[0174] \quad w^l = w^l - \mu \frac{\partial L(w,b)}{\partial w^l}, \quad b^l = b^l - \mu \frac{\partial L(w,b)}{\partial b^l} \quad (9)$$

[0175] 式(9): $L(w,b)$ 表示损失函数, $\mu$ 表示学习率。

[0176] (3) 重复上面正向传递学习和反向传播更新两个步骤,直到达到最大迭代次数结束微调。

[0177] 步骤S7:将进行数据扩增后的图像的测试集作为输入,通过第一最优分类模型在CNN网络的Softmax层获得对测试样本的第一预测概率,将对象掩码图的测试集作为输入,通过第二最优分类模型在CNN网络的Softmax层获得对测试样本的第二预测概率,将注意力

图的测试集作为输入,通过第三最优分类模型在CNN网络的Softmax层获得对测试样本的第三预测概率。

[0178] 其中,步骤S7提取不同分支的Softmax层输出。

[0179] 在一种实施方式中,步骤S7具体包括:

[0180] 步骤S7.1:将进行数据扩增后的图像测试集作为输入,利用第一最优分类模型,在CNN网络的Softmax层获得对测试样本的第一预测概率,其中,第一预测概率为大小为num\_class $\times$ 1的概率分布特征向量 $p_1$ ;

[0181] 步骤S7.2:将对象掩码图测试集作为输入,利用第二最优分类模型,在CNN网络的Softmax层获得对测试样本的第二预测概率,其中,第二预测概率为大小为num\_class $\times$ 1的概率分布特征向量 $p_2$ ;

[0182] 步骤S7.3:将注意力图测试集作为输入,利用最优分类模型,在CNN网络的Softmax层获得对测试样本的第三预测概率,其中,第三预测概率为大小为num\_class $\times$ 1的概率分布特征向量 $p_3$ 。

[0183] 步骤S8:将Softmax层的第一预测概率、第二预测概率和第三预测概率通过决策级融合,获得最终预测结果,将其作为分类结果。

[0184] 其中,步骤S8是进行决策级融合,由多分支网络预测的概率矩阵来获得预测类别。

[0185] 具体地,步骤S8具体包括:

[0186] 步骤S8.1:不同分支加权的预测,根据以上生成的三个分支的预测概率矩阵 $P_B = \{p_j\}_{j=1}^3$ ,采用加权系数进行聚合,并经过交叉验证找到权重系数 $\{w_j\}_{j=1}^3$ ,预测公式为:

$$[0187] \quad l_p(s) = \operatorname{argmax}_{k \in \{1, \dots, D\}} \sum_{j=1}^3 w_j p_j \quad (4)$$

[0188] 式(4)中:

[0189]  $l_p(s)$ 表示具有最高概率的场景图像的预测类别,

[0190]  $D$ 表示类别数;

[0191] 步骤S8.2:基于大数表决的预测,找出三个预测分支结果中具有最大次数的类别 $l_m(s)$ 作为场景图像 $s$ 的预测类别,预测公式为:

$$[0192] \quad l_m(s) = \operatorname{argmax}_{k \in \{1, \dots, D\}} (\operatorname{Count}([p_1 \ p_2 \ p_3]) = k) \quad (5)$$

[0193] 式(5)中:

[0194]  $\operatorname{Count}(\cdot)$ 表示计数器函数,统计三个分支上预测出的相同类别的次数,

[0195]  $k$ 表示类别号;

[0196] 步骤S8.3:将上述8.1)和8.2)两种预测方法中具有最高准确度的预测作为最终的预测结果。

[0197] 总体来说,如图2所示,为本发明实施例中遥感图像场景分类训练和测试的流程图。现有技术中还有基于多尺度特征卷积神经网络的图像场景分类方法,构建一种多尺度特征卷积神经网络,将多个不同尺度特征提取模块嵌入到卷积神经网络中,提取不同尺度特征进行概率预测,但是这种方法在卷积神经网络中嵌入了多个模块增加了网络训练的难度,并且多尺度的融合特征对于高度相似的图像还不具有足够的判别性。此外,还有“基于卷积神经网络的遥感图像场景自动提取和分类方法”,提出首先对遥感图像场景进行逐像素地物类别分类,再将分类后的结果专题图与原遥感场景图像拼接输入卷积神经网络进行

训练和分类。在基于像元的分类方法中利用像元的光谱信息、纹理信息、空间关联在内的信息对地物类别属性或者传统聚类方法进行分类,这种方法只能简单地获取一些细节信息,并不能足以获得地物类别语义信息,对于简单的图像场景反而增加了其复杂性,不能选择性地学习好的判别特征。

[0198] 本发明采用多分支卷积神经网络融合的遥感图像场景分类方法,首先利用对象检测网络训练结合图像上下文全局信息检测场景语义对象,以及利用注意力网络训练不断注意关键区域,最后学习局部重要区域。这两个分支网络的运用有助于提高模型对场景语义对象的检测能力。训练过程同时利用原图像、对象掩码图和注意力图组成多分支卷积网络来抽取具有判别且互补的多个特征,建立判别性强的分类模型。同时多分支网络通过数据扩增技术从大量训练数据中学习到有益的特征信息。最后,由多分支网络预测的概率矩阵进行组合表决,也就是使用基于预测概率的加权或大数表决中具有最高准确率的预测方法来赋予场景类别,具有提高遥感图像场景分类效果的特点。

[0199] 因此,本发明通过多分支卷积神经网络模型训练及其融合,优化损失函数,提取遥感图像场景的深度特征并分类,具有提高遥感图像场景分类效果的特点。

[0200] 为了验证本发明提出的遥感图像场景分类方法的优势,使用45类场景图像数据集(NWPU-RESISC45),将由原图像网络,对象检测网络和注意力网络组成的本实例算法与其他算法做分析和对比,如表1所示:

[0201] 表1:不同算法的分类正确率比较

分类算法	NWPU-RESISC45 (45 类划分比例)	
	10%	20%
[0202] 本实例算法	89.39 ± 0.41	92.35 ± 0.08
微调 VGG-VD16	88.02 ± 0.14	91.07 ± 0.12
D-CNN + VGG-VD16	89.22 ± 0.50	91.89 ± 0.22
IOR4 + VGG-VD16	87.83 ± 0.16	91.30 ± 0.17
[0203] LLC + SIFT	38.81	40.03
BoVW + SIFT	41.72	44.97

[0204] 其中,D-CNN+VGG-VD16算法参见Cheng等人发表的论文《当深度学习结合度量学习:通过学习判别性CNN进行遥感图像场景分类》(IEEE国际地理科学与遥感学报,vol.56,no.5,pp.2811-2821,2018.);IOR4+VGG-VD16算法参见Wang等发表的论文《IORN:一种有效的遥感图像场景分类框架》(IEEE地理科学与遥感快报,vol.15,no.11,pp.1695-1699,2018.);LLC+SIFT算法参见Wang等人发表的论文《用于图像分类的局部约束线性编码》(2010年第23届IEEE计算机视觉与模式识别会议,旧金山,2010);BoVW+SIFT算法参见Zhu等人发表的论文《用于高空间分辨率遥感图像的具有局部和全局特征的视觉词袋场景分类器》(IEEE地理科学与遥感快报,vol.13,no.6,pp.747-751,2016.)。由表1可知,基于卷积神经网络的遥感图像场景分类算法效果优于使用中低级特征分类的算法,而在基于卷积神经网络算法中,本发明提出的对分支深度卷积神经网络融合的遥感图像场景分类方法具有更好地效果,进一步说明通过多分支卷积神经网络模型训练及其融合,优化损失函数,提取遥感图像场景的深度特征并分类,能够提高遥感图像场景分类效果。

[0205] 为了进一步说明本实例算法各个分支的不同贡献,不同分支网络的分类结果如表2所示:

[0206] 表2:不同分支网络上的分类结果

	各分支算法	NWPU-RESISC45 (45 类划分比例)	
		10%	20%
[0207]	本实例算法 + VGG-VD16	89.39 ± 0.41	92.35 ± 0.08
	对象检测网络 + VGG-VD16	88.53 ± 0.22	91.58 ± 0.30
	注意力网络 + VGG-VD16	88.45 ± 0.28	91.47 ± 0.26
	VGG-VD16	88.02 ± 0.14	91.07 ± 0.12

[0208] 由表2可知,在45类图像场景数据集上,三个单独分支的网络都达到了一定的效果,由此可见,本算法在每个分支上的有效性,并且通过三支融合后分类结果有了进一步提升,说明了本算法中三个分支是具有互补性的,融合后的特征具有更加丰富的判别性。

[0209] 尽管已描述了本发明的优选实施例,但本领域内的技术人员一旦得知了基本创造性概念,则可对这些实施例做出另外的变更和修改。所以,所附权利要求意欲解释为包括优选实施例以及落入本发明范围的所有变更和修改。

[0210] 显然,本领域的技术人员可以对本发明实施例进行各种改动和变型而不脱离本发明实施例的精神和范围。这样,倘若本发明实施例的这些修改和变型属于本发明权利要求及其等同技术的范围之内,则本发明也意图包含这些改动和变型在内。

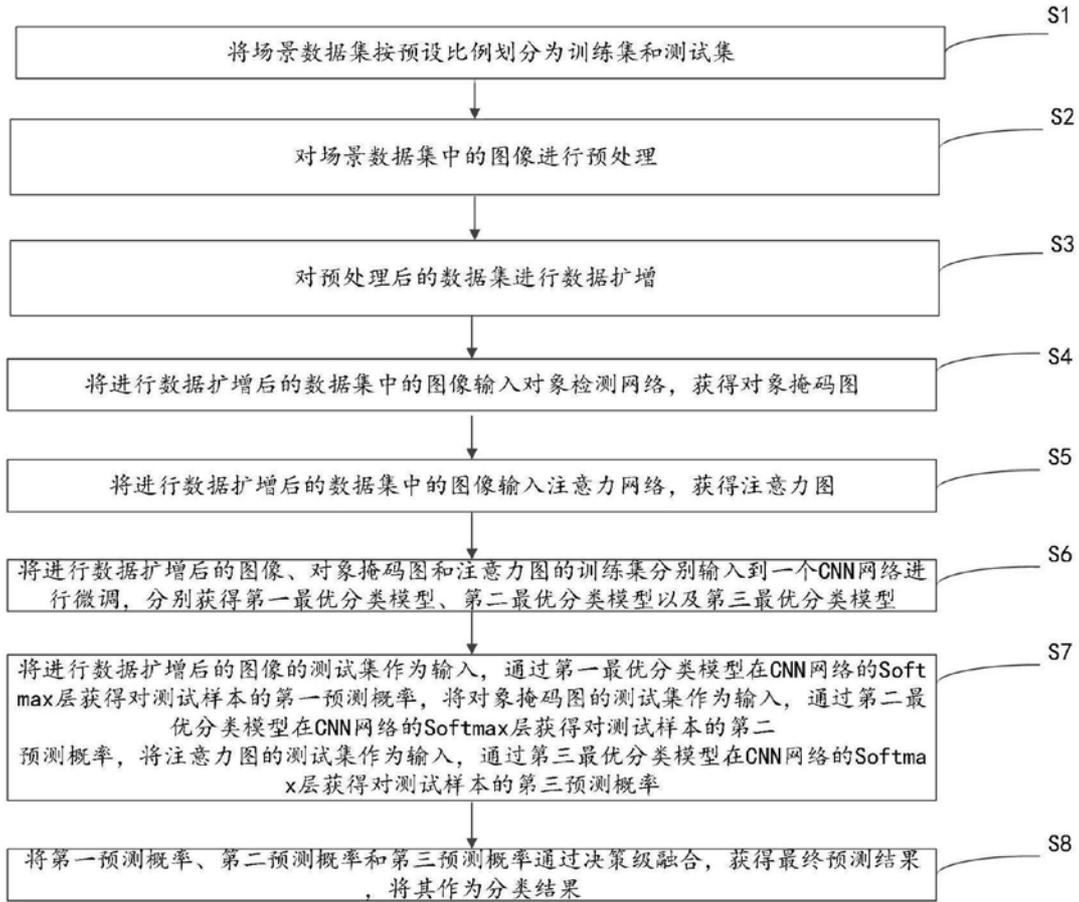


图1

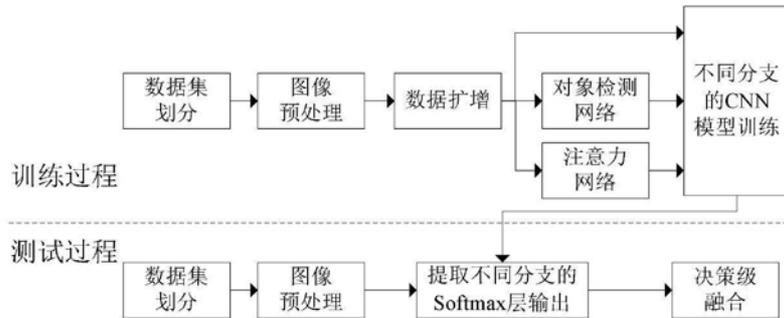


图2

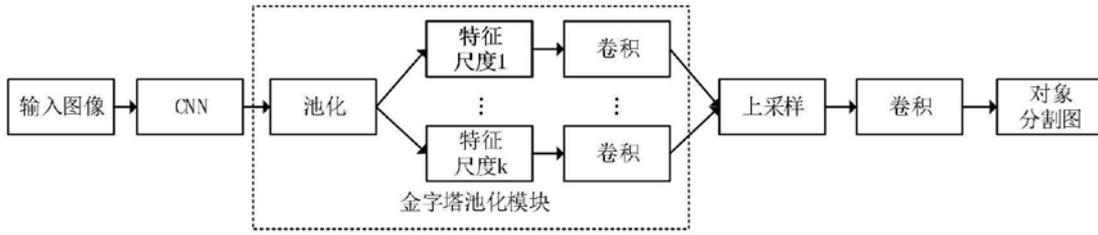


图3

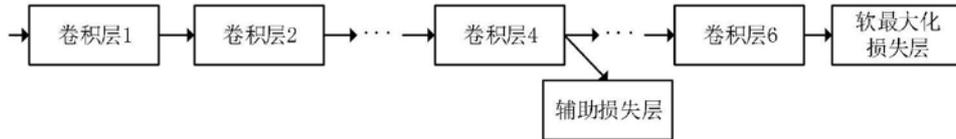


图4

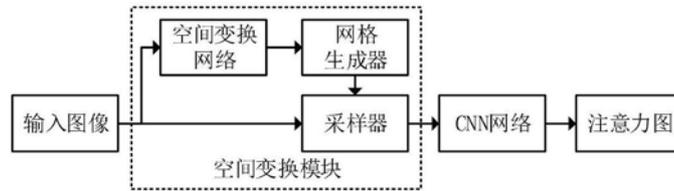


图5

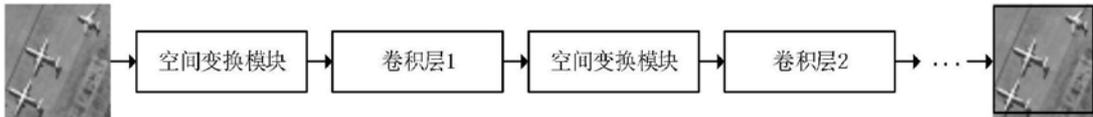


图6

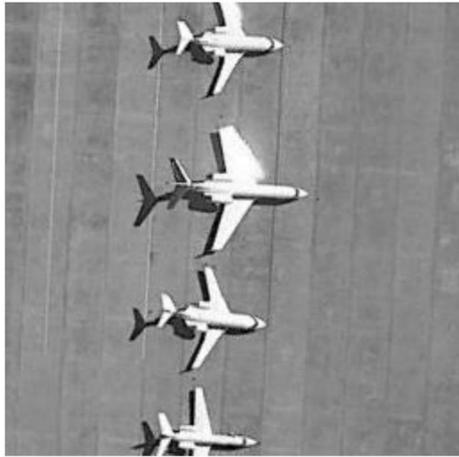


图7

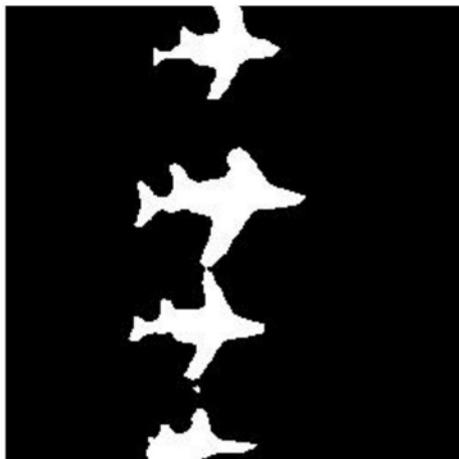


图8

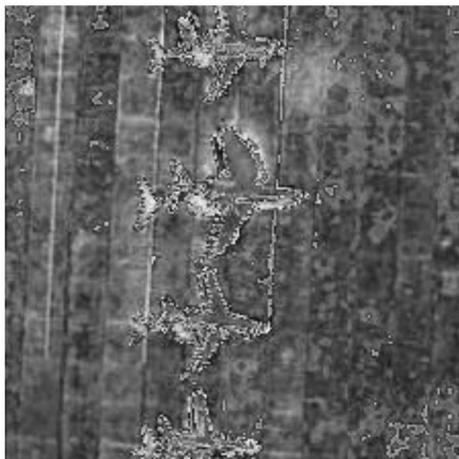


图9