**Transferring Deep Convolutional Neural Networks for the Scene Classification of High-Resolution Remote Sensing Imagery**

**摘要：**将CNN框架应用到高分辨率遥感影像场景识别。本文提出两种从CNN不同的层中生成影像特征的方案，第一种，从全连接层中提取激活向量，作为最终的影像特征；第二中方案，从多尺度对最后一个卷积层提取密集特征，并通过常用的特征编码方法将此特征编码到全局影像特征。

In the first scenario, the activation vectors extracted from fully-connected layers are regarded as the final image features; in the second scenario, we extract dense features from the last convolutional layer at multiple scales and then encode the dense features into global image features through commonly used feature coding approaches.

近几年，高分辨率遥感影像场景分类进展缓慢，主要是因为一般方法无法提供足够的特征表达，而词袋模型BOW与一些无监督特征提取方法只能够建立“中层次”特征表达。而更高级的特征，是较低层次特征的抽象，能有效进行场景分类。

很难通过一个小的样本数据集来训练出高性能的卷积神经网络。

However, it is difficult to train a high-powered deep CNN with small datasets in practice.

很多研究表明，从使用大型数据集预先训练好的CNNs中学习到的中间激励层可以转移到其他有限数据集的识别任务中。

At present, many recent works have demonstrated that the intermediate activations learned with deep CNNs pre-trained on large datasets such as ImageNet can be transferable to many other recognition tasks with limited training data.

**提出问题：**是否可以将与预训练的CNNs成功转移到高分辨率遥感影像的场景分类中，而高分辨率遥感影像场景分类同样具有有限的数据集。

**方案：**

本文通过移除CNNs的最后一层，将剩下的部分作为一个特征提取器。提出两种方案来从不同层提取CNNs的特征：

1. 在整个图像场景中计算CNN的激活向量，并将全连接层中的激活向量视作场景的全局特征表达。

we simply compute the CNN activations over the entire image scene and regard the activation vectors of the fully-connected layer as the global feature representations for scenes;

1. 最后一个卷积层具有原始影像场景的多个尺度，首先从其中生成密集的CNN激活向量，然后使用传统的特征编码方式将密集的卷积特征聚合到一个全局特征表达中，如BOW和Fisher编码。最终，密集的CNN激活向量就可以描述多尺度的空间信息。

we first generate dense CNN activations from the last convolutional layer with multiple scales of the original input image scenes, and then we aggregate the dense convolutional features into a global representation via the conventional feature coding scheme, e.g., the BOW and Fisher encoding. These dense CNN activations describe multi-scale spatial information.

**深度卷积神经网络介绍：**

典型的卷积神经网络有多个级联层次组成。前几个阶段是由卷积层与池化层组成（下图）。卷积层输出特征映射，其每个元素都是通过计算输入特征映射层的某个局部区域与权重（滤子或卷积核）的点积获得的。通常，还要在这些特征映射层上增加非线性的激活函数。池化层进行了一步降采样操作。全连接层出现在多个堆叠的卷积层与池化层之后，最后一个全连接层是一个Softmax层，计算每个类的得分。CNNs将输入的原始像素值转换成最终的类别得分

