**一种基于卷积神经网络的遥感图像融合算法**

**FusionCNN: a remote sensing image fusion algorithm based on deep convolutional neural networks**

**--Fajie Ye1,2 & Xiongfei Li1,2 & Xiaoli Zhang1,2（Springer Science+Business Media, LLC, part of Springer Nature 2018）**

**一、科学问题**

**1.1 本文所涉及科学问题**

卷积神经网络，图像融合

**1.2 同行专家如何解决**

在许多遥感图像融合算法中，融合规则都是基于单像素点设计的。

**1.3 本文所解决的问题**

用一种新的融合规则，即用卷积神经网络解决图像融合问题，输入一组源图片然后输出端到端的融合图像，通过大量训练数据使融合函数性能上有很强的适应性，避免了传统融合算法中人工设定规则的缺点。

**1.4 本文解决方案效果**

相较于5种经典的方法，实验表明本方法可以很好地处理不同频段的MS图像和不同类型的对象。在较强的鲁棒性上，它可以实现令人满意的融合质量：融合后的图像可以包含MS图像的频谱信息以及Pan图像的空间信息。

**二、研究内容**

**2.1 理论与方法介绍**

利用遥感图像融合的性能和卷积神经网络的优点，创建一个多尺度卷积神经网络模型，并将其用于泛色和多光谱图像。首先构造用于图像融合的卷积神经网络，这是一种典型的回归模型。通过仿真从自然图像集构建了一个新的PAN和MS图像数据集。将EPAN图像输入到训练好的FusionCNN中，使融合后的图像作为网络的输出。

**2.2 验证分析与实验效果**

Experimental results on Landsat images：通过与其他算法相比，该方法能较好地保存MS图像的光谱信息和Pan图像的空间信息。

Experimental results on Quickbird images：通过与其他算法相比，该方法可以最大程度的保留MS图像的频谱信息以及空间信息

**三、论文存在问题及后续研究重点**

**3.1 论文存在问题**

因为遥感领域没有数据集能被用于训练算法而从自然图像中构建新的数据集模拟Pan和MS图像，这个想法不错，但可能存在较大误差

**3.2 后续研究重点**

暂未发现

**四、该问题相关研究成果**

**4.1 相关论文一**

**（1）题目**：Remote sensing image fusion via wavelet transform and sparse

representation. ISPRS J Photogramm Remote Sens

**（2）作者介绍**：Cheng J, Liu H, Liu T et al

**（3）摘要**: It is very challenging to fuse remote sensing images with significant spatial resolution difference. In this paper, such kind of images are fusing with one posteriori probability model for the purpose of creating thematic maps. The outlined solution consists of parsing the low spatial resolution multispectral images into intensity-hue-saturation triples with spatial and spectral properties separated; the contribution of the old intensity and the panchromatic image is factorized into posteriori probabilities using their wavelet planes under the aid of support vector machine; the wavelet planes are probabilistically weighted to foresee one new pan-sharpened intensity, from which the fused images are produced by inverse IHS. The outlined method along with two representative methods are evaluated with two sets of imagery covering contrasting heterogeneous landscapes for different purposes of road extraction and wheat area map.

**4.2 相关论文二**

**（1）题目**：Fusion of IKONOS satellite imagery using IHS transform and local variation

**（2）作者介绍**：Chu H, Zhu W

**（3）摘要**：The intensity-hue-saturation (IHS) technique is a well-known merging approach for its computational efficiency and spatial definition holding. However, it results in color distortion particularly for the remote sensing images of IKONOS and QuickBird as some other fusion methods, such as principal component analysis, and Brovey transform. Although wavelet-based image fusion approaches can provide a better tradeoff between spatial and spectral quality, the fused images with these methods often have a spatial resolution that is less than that of the IHS-based algorithm. A remote sensing image fusion algorithm based on IHS transform and local variation and its modified approach with low computational complexity are proposed. Visual effect and quantity evaluation results show that the proposed simple algorithm outperforms the conventional image fusion methods in the spectral domain with the spatial quality similar to that of the undecimated wavelet transform-based scheme. The proposed modified method can obtain the similar spatial resolution of the merged image with the IHS-based fusion algorithm and the better spectral quality in the green vegetation areas.

**4.3 相关论文三**

**（1）题目**：Remote-sensing image fusion based on Curvelets and ICA

**（2）作者介绍**：Ghahremani M, Ghassemian H

**（3）摘要**：Improving the quality of pan-sharpened multispectral (MS) bands is the main aim of the recent research on pan-sharpening. In this article, we present a novel image fusion method based on combining the curvelet transform and independent component analysis (ICA). The idea is to map the MS bands onto a statistically independent domain to determine the intensity component, which contains the common information of the MS bands, and then to pan-sharpen it using curvelets and a modified adaptive fusion rule. The proposed method is evaluated by visual and statistical analyses and compared with the curvelet (CVT)-based method using a context-based decision model, the CVT-based method using the Dempster–Shafer evidence theory, the improved ICA method, and the combined adaptive principle component analysis (PCA)–Contourlet method. The experimental results using QuickBird and WorldView-2 data show that the proposed method effectively reduces the spectral distortion while injecting spatial details into the fused bands as much as possible.

**像素级投票网络的6自由度位姿估计**

**PVNet: Pixel-wise Voting Network for 6DoF Pose Estimation**

**--** **Sida Peng、Yuan Liu（cs.CV 2018）**

**一、科学问题**

**1.1 本文所涉及科学问题**

6D姿态估计，卷积神经网络，PVNet

**1.2 同行专家如何解决**

传统的方法已经表明位姿估计可以通过建立目标图像和模型之间的通信来实现，它们依赖于手工特点，这在图像变化或者背景杂乱时不具备鲁棒性。基于方法训练端到端神经网络的深度学习，输入一张图像然后输出它相应的位姿。然而，泛化仍是一个问题，因为目前还不清楚这种端到端方法是否能够学习足够的特征表示来进行姿态估计。

目前有一些方法使用卷积神经网络先退化2D关键字，然后用透视n点算法计算6D姿态参数。换句话说，被发现的关键点被用作姿态估计的中间表示。这样的两步由于关键点的鲁棒性检测实现了先进的性能。然而这些方法在处理阻塞和阻断图像时也有问题，因为看不到它们的部分关键点。尽管CNN可能通过记忆类似模型预测这些看不见的关键点，广化（退化）仍然是一个问题。

**1.3 本文所解决的问题**

图像不完整（阻塞或截断）时怎么选择关键点

**1.4 本文解决方案效果**

相对于直接从图像碎片中退化关键点位置，预测像素方向的任务迫使网络更加关注对象的局部特征，减轻了背景杂乱的影响。这种方法的另一个优点是能够表示被遮挡或图像外部的关键点。即使一个关键点是不可见的，它也可以根据从对象的其他可见部分估计的方向来正确定位。

**二、研究内容**

**2.1 理论与方法介绍**

提出一种像素级投票网络来广义化像素级单元向量指向关键点，即这些向量用RANSAC来选举出关键点位置。这为其中受到阻塞或者截断的关键点产生了灵活的代表性。这些有代表性的点的另一个重要特点是它提供了关键点位置的不确定性，根据PnP解释器可以让它进一步发挥作用。

**2.2 验证分析与实验效果**

为了防止过度拟合，将合成图像添加到训练集中。对于每个对象，我们渲染10000幅图像，这些图像的视点被均匀采样。我们使用[10]中提出的“剪切粘贴”策略进一步合成了另外10000张图像。每个合成图像的背景都是从SUN397中随机采样的。我们还应用在线数据增强，包括随机裁剪、调整大小、旋转和训练期间的颜色抖动。我们将初始学习率设置为0.001，每20次（epochs）减半一次。所有的模型都经过了200次的训练。

通过EPnP的方法确定R和T，然后通过 最小化 Mahalanobis distance最终确定6D姿态 （马氏距离也可以定义为两个服从同一分布并且其协方差矩阵为Σ的随机变量之间的差异程度）

实验证明上述方法在实时姿态估计时比LINEMOD，阻塞LINEMOD和YCBVideo数据集上的先进水平要好很多（a large margin），效率更高。方案还进一步创建了一个阻塞LINEMOD数据集来验证本文所述方法在解决处理截断问题。

**三、论文存在问题及后续研究重点**

**3.1 论文存在问题**

The images of the last column are the failure cases, where the visible parts are too ambiguous to provide enough information for the pose estimation.

Table 6 shows quantitative results in terms of the 2D projection and ADD(-S) metrics. We also test the released model from [39], but it does not obtain reasonable results as it is not designed for this case.

**3.2 后续研究重点**

进一步完善算法

**四、该问题相关研究成果**

**4.1 相关论文一**

**（1）题目**：Making deep heatmaps robust to partial occlusions for 3D object pose estimation

**（2）作者介绍**：Markus Oberweger,Mahdi Rad,Vincent Lepetit

**（3）摘要**: We introduce a novel method for robust and accurate 3D

object pose estimation from a single color image under large occlusions.Following recent approaches, we first predict the 2D projections of 3D points related to the target object and then compute the 3D pose from these correspondences using a geometric method. Unfortunately, as the results of our experiments show, predicting these 2D projections using a regular CNN or a Convolutional Pose Machine is highly sensitive to partial occlusions, even when these methods are trained with partially occluded examples. Our solution is to predict heatmaps from multiple small patches independently and to accumulate the results to obtain accurate and robust predictions. Training subsequently becomes challenging because patches with similar appearances but different positions on the object correspond to different heatmaps. However, we provide a simple yet effective solution to deal with such ambiguities. We show that our approach outperforms existing methods on two challenging datasets:The Occluded LineMOD dataset and the YCB-Video dataset, both exhibiting cluttered scenes with highly occluded objects.

**4.2 相关论文二**

**（1）题目**：Uncertainty-driven 6d pose estimation of objects and scenes from a single rgb image.

**（2）作者介绍**：E. Brachmann, F. Michel

**（3）摘要**：In recent years, the task of estimating the 6D pose of object instances and complete scenes, i.e. camera localization, from a single input image has received considerable attention. Consumer RGB-D cameras have made this feasible, even for difficult, texture-less objects and scenes. In this work, we show that a single RGB image is sufficient to achieve visually convincing results. Our key concept is to model and exploit the uncertainty of the system at all stages of the processing pipeline. The uncertainty comes in the form of continuous distributions over 3D object coordinates and discrete distributions over object labels. We give three technical contributions. Firstly, we develop a regularized, auto-context regression framework which iteratively reduces uncertainty in object coordinate and object label predictions. Secondly, we introduce an efficient way to marginalize object coordinate distributions over depth.This is necessary to deal with missing depth information.Thirdly, we utilize the distributions over object labels to detect multiple objects simultaneously with a fixed budget of RANSAC hypotheses. We tested our system for object pose estimation and camera localization on commonly used data sets. We see a major improvement over competing systems.

**4.3 相关论文三**

**（1）题目**：When regression meets manifold learning for object recognition and pose estimation.

**（2）作者介绍**：M. Bui, S. Zakharov

**（3）摘要：**In this work, we propose a method for object recognition and pose estimation from depth images using convolutional neural networks. Previous methods addressing this problem rely on manifold learning to learn low dimensional viewpoint descriptors and employ them in a nearest neighbor search on an estimated descriptor space. In comparison we create an efficient multi-task learning framework combining manifold descriptor learning and pose regression. By combining the strengths of manifold learning using triplet loss and pose regression, we could either estimate the pose directly reducing the complexity compared to NN search, or use learned descriptor for the NN descriptor matching. By in depth experimental evaluation of the novel loss function we observed that the view descriptors learned by the network are much more discriminative resulting in almost 30% increase regarding relative pose accuracy compared to related works. On the other hand, regarding directly regressed poses we obtained important improvement compared to simple pose regression. By leveraging

the advantages of both manifold learning and regression tasks, we are able to improve the current state-of-the-art for object recognition and pose retrieval that we demonstrate through in depth experimental evaluation.

简要翻译

摘要

本篇论文从严重阻塞或截断下的单张RGB图片着手，列出6~~维~~自由度姿态估计所面临的挑战。许多作品都指出两步走，首先发现关键字然后解决姿态估计的透视n点问题，达到明显的效果。然而这些方法中大多数都只能根据退化它们的图像坐标系或者热点图来集中一系列稀少的关键点，这对阻塞或者截断太敏感（容易受到阻塞或截断的影响）。（为了避免这样，）本文介绍一种像素级投票网络来广义化像素级单元向量指向关键点，即这些向量用RANSAC来选举出关键点位置。这为集中受到阻塞或者截断的关键点产生了灵活的代表性。这些有代表性的点的另一个重要特点是它提供了关键点位置的不确定性，根据PnP解释器可以让它进一步发挥作用。实验证明上述方法在实时姿态估计时比LINEMOD，阻塞LINEMOD和YCBVideo数据集上的先进水平要好很多（a large margin），效率更高。我们进一步创建了一个阻塞LINEMOD数据集来验证本文所述方法在解决处理截断问题。

介绍

目标姿态估计旨在发现目标和根据相关典型框架估计它们的orientations和translations。精确姿态估计对诸如增强真实性，自动驾驶和机器人操作的各种应用非常重要。例如，快速健壮的姿态估计对亚马逊采摘挑战非常重要，机器人需要从仓库架上选择目标。本文集中在对一张RGB图像在3维模式下旋转翻译的特定设置下**复原6自由度姿态的目标**（此句不是很理解）。这个问题从很多构面上都很有难度，包括在严重阻塞，光照和外观变化，让背景目标变得杂乱等条件下检测目标。

传统的方法已经表明位姿估计可以通过建立目标图像和模型之间的通信来实现，它们依赖于手工特点，这在图像变化或者背景杂乱时不具备鲁棒性。基于方法训练端到端神经网络的深度学习，输入一张图像然后输出它相应的位姿。然而，泛化仍是一个问题，因为目前还不清楚这种端到端方法是否能够学习足够的特征表示来进行姿态估计。

目前有一些方法使用卷积神经网络先退化2D关键字，然后用透视n点算法计算6D姿态参数。换句话说，被发现的关键点被用作姿态估计的中间表示。这样的两步由于关键点的鲁棒性检测实现了先进的性能。然而这些方法在处理阻塞和阻断图像时也有问题，因为看不到它们的部分关键点。尽管CNN可能通过记忆类似模型预测这些看不见的关键点，广化仍然是一个问题。

我们认为列出（adressing？）阻塞和截断需要密集的预测，也就是为最后的输出或中间表示进行像素级或者碎片集估计。最后，我们用像素级投票网络（PVNet）为6D姿态估计提出新颖的框架。PVNet预测表示从图像各个像素点到关键点方向的单元矢量，取代了直接退化图像的坐标系或热点图。然后这些方向根据RANSAC为关键点位置投票。这种投票机制的好处是一旦我们看到某些局部就可以推断出其他部分的相关方向。

**我们的方法本质上创建了代表关键字位置的矢量区域。相对于基于坐标系或热点图的表示，学习这种方法迫使网络集中于图像的局部特征和图像部分间的空间联系。作为结果，看不见的部分可以根据看的见的部分推断出。而且这种矢量区域表示能够代表输入图像外的图像关键点。所有这些优点让这种表示方法成为被阻塞或截断图像的理想方法。一些组员提出相似的想法来表示对象，下面来用它局部化关键点。**

**另一个优点是密集输出为PnP解释器解决不确定关键点的预测提供了丰富信息。基于SANSAC的投票结果剔除了异常预测，并给出了每个关键点的空间概率分布。这种关键点位置的不确定性给予PnP解释器为预测最终的姿态辨别一致的对应关系更大的自由。实验表明不确定性驱使PnP算法提高姿态估计的精确度。**

我们在LINEMOD，阻塞LINEMOD和YCB-Video数据集上评估了（所提出的）方法，这些是被广泛用于6自由度姿态估计的数据集。在所有数据集上，PVNet展示了最先进的性能。我们也在截断LINEMOD数据集上验证了我们的方法在解决截断图像上的能力，这个数据集是通过随机裁剪LINEMOD上的图像生成的。此外，我们的方法是非高效的，在GTX 1080ti GPU上运行25帧帧频，用于实时姿态估计。

**总而言之，这项工作有以下贡献**：1>提出了一种新的基于像素级投票网络(PVNet)的6D位姿估计框架，该框架学习了一种向量场表示方法，用于健壮的2D关键点定位，并能自然地处理阻塞和截断问题。2>基于PVNet的密集预测，提出了一种基于不确定性驱动的PnP算法来解决2D关键点定位中的不确定性问题。3>与基准数据集上的最新技术相比，我们展示了我们的方法的显著性能改进(ADD: 86.3% vs. 79% on LINEMOD and 40.8% vs. 30.4% on OCCLUSION)。我们还创建了一个新的数据集，用于对截断的对象进行评估。

相关工作（概括）

1. Holistic methods.

Traditional methods mainly rely on template matching techniques（sensitive to cluttered environments and appearance changes） →→PoseNet introduces a CNN architecture to directly regress a 6D camera pose from a single RGB image, a task similar to object pose estimation.（directly localizing objects in 3D is difficult due to a lack of depth information and the large searchspace）→→PoseCNN localizes objects in the 2D image and predicts their depths to obtain the 3D location（directly estimating the 3D rotation is also difficult, since the non-linearity of the rotation space makes CNNs less generalizable.）→→discretize the rotation space and cast the 3D rotation estimation into a classification task. Such discretization produces a coarse result and a post-refinement is essential to get an accurate 6DoF pose（将旋转空间离散化，将三自由度旋转估计转换为分类任务。这样的离散化产生了一个粗糙的结果，后细化（post-refinement）是得到一个精确的6自由度位姿的必要条件。）

2> Keypoint-based methods.

First predict 2D keypoints of the object and then compute the pose through 2D-3D correspondences with a PnP algorithm（have difficulty in handling texture-less(结构不清晰的)objects and processing low-resolution images）→→define a set of semantic keypoints and use CNNsas keypoint detectors.

3> Dense methods.

every pixel or patch produces a prediction for the desired output, and then casts a vote for the final result in a generalized Hough voting scheme use a random forest to predict 3D object coordinates for each pixel and produce 2D-3D correspondence hypotheses using geometric constraints.

Proposed approach

给定一幅图像，(姿态估计的任务是)检测目标并估计其三维方向和平移量(orientations and translations)。6D姿态从对象的坐标系统到相机的坐标系统由严格的转换（R；t）代表，R代表3D旋转，t表示3D偏移量。

。。。用两步评估图像位姿：首先用CNN检测2D图像关键点，然后用PnP算法计算6D姿态参数。我们的创新点在于2D目标关键点的新表示，以及被改进过的PnP算法用于姿态估计。具体来说，本方法用像素级投票网络以一种类似RANSAC的方法检测2D关键点，这种方法能有效的处理阻塞和截断的目标。基于RANSAC的投票也给出了每个关键点的空间概率分布，允许我们用不确定驱使的PnP估计6自由度位姿。

Votingbased keypoint localization（基于投票的关键点定位）

相对于直接从图像碎片中退化关键点位置，预测像素方向的任务迫使网络更加关注对象的局部特征，减轻了背景杂乱的影响。这种方法的另一个优点是能够表示被遮挡或图像外部的关键点。即使一个关键点是不可见的，它也可以根据从对象的其他可见部分估计的方向来正确定位。

。。。直观地说，较高的投票分数意味着假设更有信心，因为它符合更多的预测方向。。。所得到的假设代表了图像上关键点的空间概率分布。

关键点的选择：在本方法中，关键点应该从图像的表面选取，而且这些关键点应该分散到对象上以使PnP算法更加稳定 两点。，考虑到上述两点要求，我们用最远点样本算法（FPS）选择了K个关键点。

训练机制

为了防止过度拟合，我们将合成图像添加到训练集中。对于每个对象，我们渲染10000幅图像，这些图像的视点被均匀采样。我们使用[10]中提出的“剪切粘贴”策略进一步合成了另外10000张图像。每个合成图像的背景都是从SUN397中随机采样的。我们还应用在线数据增强，包括随机裁剪、调整大小、旋转和训练期间的颜色抖动。我们将初始学习率设置为0.001，每20次（epochs）减半一次。所有的模型都经过了200次的训练。

结论

提出了一种基于像素级投票网络的6自由度目标姿态估计框架(PVNet)用于关键点定位和不确定驱动最后姿态估计的PnP。结果表明，对向量场进行预测，然后基于RANSAC对关键点位置进行投票，其性能优于关键点坐标的直接回归，特别是对被遮挡或截断的对象。在求解PnP问题时，考虑了预测关键点位置的不确定性，进一步改进了位姿估计。