

三 维 动 画 和 交 互 设 计 课 程 读 书 报 告



作者姓名 彭亮

作者学号 21851140

年级专业 2018级软件工程

所在学院 软件学院

完成日期 2018.12.17

Research on Computer Animation and Interaction Design

A Report Submitted to

Zhejiang University

Major Subject: Software Engineering

Advisor: Jianke Zhu

By

Liang Peng

College of Software Technology, Zhejiang University

2018

摘要

本文利用卷积神经网络训练得到了一种新的局部三维描述符，可直接应用于各种形状分析问题，如点对应，语义分割，交互区域预测和形状扫描匹配。此网络处理在三维形状上点及周围的表面邻域，通过从精心选择的摄像机位置上以多个尺度拍摄的视图捕获这些点。利用了两个非常大的数据集来训练此网络。此网络有效地编码了多尺度的本地上下文周围信息和细粒度的表面细节。可以通过训练以生成特定类别的描述符，也可以通过从多个形状类别中学习通用描述符。经过训练后，在测试时，网络就会提取形状的局部描述符，而不需要任何部分分割作为输入。即使对于类别未知或类型的形状，也可以生成有效的局部描述符。

**关键词 ：局部三维描述符，形状分析，卷积网络**

Abstract

In this paper, a new 3D local descriptor is obtained by using convolutional neural networ, which can be directly applied to various shape analysis problems, such as point correspondence, semantic segmentation, Interactive region prediction and shape scan matching. This network processes the point neighborhoods around the 3D shape and surrounding them, capturing these points by views taken at multiple scales from carefully selected camera positions. Two very large data sets were used to train this network. This network effectively encodes multi-scale local context information and fine-grained surface details. You can train to generate descriptors for a particular category, or you can learn general descriptors from multiple shape categories. After training, the network extracts the local descriptors of the shape during testing without any partial segmentation as input. Valid local descriptors can be generated even for shapes of unknown or type.

**Key Words : local 3D shape descriptors,shape analysis,convolutional networks**

目录

[一、论文题目及介绍 5](#_Toc533537658)

[二、论文背景及相关工作 6](#_Toc533537659)

[2.1 背景介绍 6](#_Toc533537660)

[2.2 论文创新点 7](#_Toc533537661)

[2.3 相关工作 7](#_Toc533537662)

[三、设计方案 8](#_Toc533537663)

[3.1 总览 8](#_Toc533537664)

[3.2 网络结构 8](#_Toc533537665)

[3.3 学习 10](#_Toc533537666)

[四、实验 11](#_Toc533537667)

[五、应用 14](#_Toc533537668)

[六、结论 16](#_Toc533537669)

[七、总结及思考 16](#_Toc533537670)

[参考文献 17](#_Toc533537671)

# 一、论文题目及介绍

论文题目：利用多视图卷积网络从部件对应中学习局部形状描述符(Learning local shape descriptors from part correspondences with multi-view convolutional networks)

作者：HAIBIN HUANG, University of Massachusetts Amherst

EVANGELOS KALOGERAKIS, University of Massachusetts Amherst

SIDDHARTHA CHAUDHURI, IIT Bombay

DUYGU CEYLAN, Adobe Research

VLADIMIR G. KIM, Adobe Research

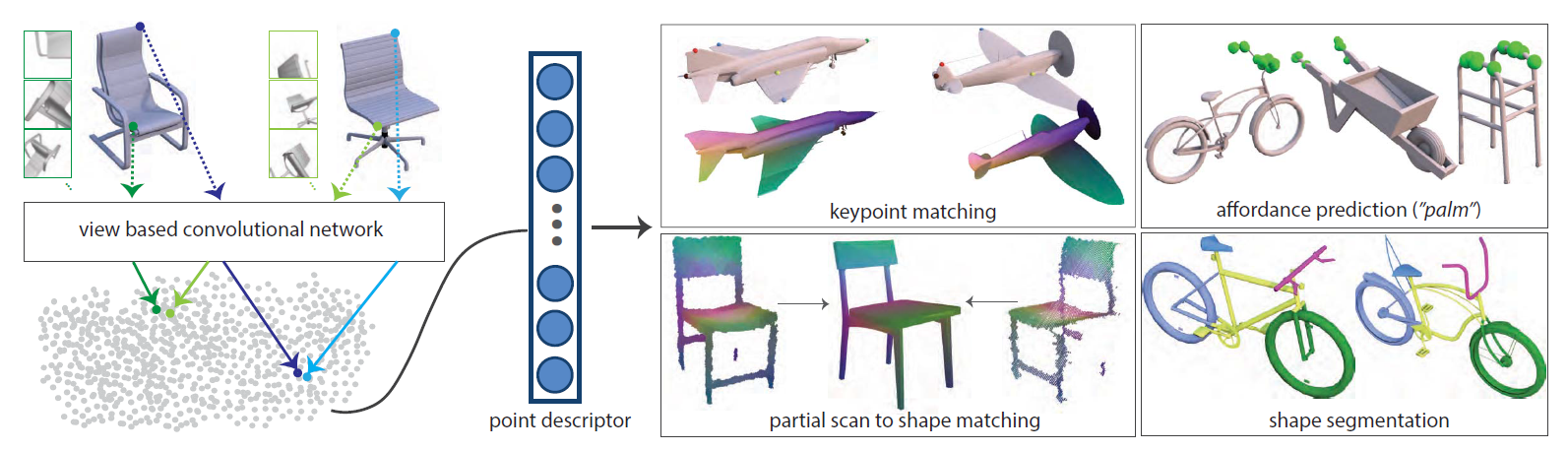
ERSIN YUMER, Adobe Research

发表于：ACM Transactions on Graphics，2018年1月

# 二、论文背景及相关工作

2.1 背景介绍

在各种3D形状分析问题中，表面点的局部描述符是很核心的部分，可以应用在关键点检测，形状对应，语义分割，区域标记和三维重建等场景中。目前根据不同的应用场景构造了相应的描述符。这些描述符可以表示出局部的分析属性，比如曲率以及形状中的点与周围结构相关的一些概念。



图一

图一展示了局部描述符的一些应用，除了在几何上可以应用形状分析之外，在功能性或语义、点和局部形状区域的相似性层面上均有用途。大多数现有的描述符依赖于两个主要假设：（a）可以直接指定哪些几何特征与语义相似性相关；（b）强大的正则化后处理。然而后面的条件通常计算的成本高，难以开发，应用在特定任务中，很难实际应用。本论文的工作主要关注第一个假设。

在这样的已知情况下，由于很难先验地确定形状几何的哪些方面与点相似性或多或少的相关，本论文采用深度学习的方法来学习局部描述符，从而隐含地捕获点之间更高级别相似性​​的概念，同时保持结构变化，噪声以及数据源和表示的差异。因此，本论文不需要手动猜测哪些几何特征可能与它对应或相关，而是通过深度学习的方式学习得到。  
     目前，学习得到的3D描述符相较于2D任务而言，有一定的缺点。（1）依赖于源自小规模形状数据库的有限训练数据，（2）在低空间分辨率下计算导致细节灵敏度的损失，（3）被设计为操作特定形状类，例如可变形的形状。

为了克服这些挑战，本论文引入了多尺度，基于视图的投影表示的局部描述符，并且在3D形状上进行学习。给定网格或点云，本论文的方法为形状上的任何点生成局部描述符。通过围绕它的一组渲染视图来表示查询点。为了更好地捕捉局部和全局背景，作者在多个尺度上渲染视图，并提出一种新的视点选择方法，以避免不希望的遮挡。 这种表示适用于在视图上操作的2D卷积神经网络（CNN）。基础网络的最后一层为每个视图生成一个特征向量，然后通过池化层跨视图组合，以产生该点的单个描述性向量。该网络以孪生风格在成对的对应点上进行训练。

2.2 论文创新点

总体来说，本论文的创新及贡献点体现在如下几点：

1、一种新的基于点的局部描述符，适用于一般3D形状，可直接应用于各种形状分析任务，对细粒度的局部和上下文信息都很敏感。

2、利用卷积网络，用于将多个比例的表面点周围的渲染视图组合成单个紧凑的局部描述符。

3、一种选择每点视图的方法，避免自我遮挡并提供有益渲染投影的集合。

4、一个用于训练目的的相应点对的大量合成数据集。

此描述符可以直接用于许多应用程序，包括关键点检测，人工交互的区域标记以及完整3D形状上的形状分割任务。此外，它们可用于非结构化扫描数据上的部分形状匹配任务，而无需任何微调。本论文的基于点的描述符明显优于传统的手工制作和基于体素的形状描述符。

2.3 相关工作

主要在传统的形状描述符，通过学习得到的全局描述符，通过学习得到的表面描述符三方面进行介绍。

传统的形状描述符。传统上，研究人员和工程师依靠他们的先验知识人为构造形状描述符。这些描述符捕获低级几何信息，这些信息通常无法可靠地映射到功能或语义形状属性。此外，这些描述符通常缺乏对噪声，数据缺失或大结构形状变化的鲁棒性。它们经常依赖于一小组手动调整参数，这些参数是针对特定数据集或形状处理场景而定制的。

通过学习得到的全局描述符。目前通过学习的全局3D特征已经得到在3D对象检测，分类和检索任务中有希望的结果。本文中的重点是学习基于点的形状描述符，可用于密集特征匹配，关键点检测，可供性标记和其他局部形状分析应用。

通过学习得到的表面描述符。在诸如人体之类的可变形模型的背景下，已经设计了深度学习网络以在内在表面表示上操作。由这些结构网络产生的学习的内在描述符表现出等长或近等距变形的不变性。然而，在几种形状类别中，特别是人造物体，即使是部件的刚性旋转也可以改变它们的基本功能和与其他部分的语义对应。

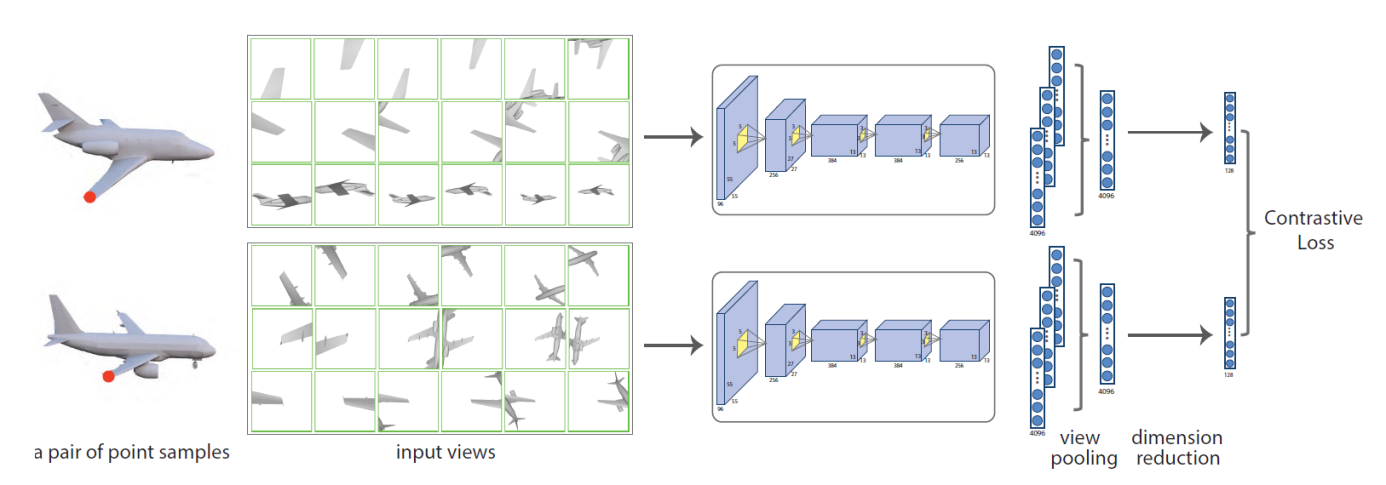
本论文的方法受到用于形状分类的多视图卷积网络的启发。采用的是一种在训练过程中比较点几何相似性的方法，而不是优化形状分类或部件标注。本论文使用自适应选择的局部视图来取代固定的视图与选择最大化表面覆盖率的视图，来捕捉表面点周围的多尺度环境信息。本论文还提出了一种自动方法来创建大量的逐点对应数据，以基于部件引导的非刚性配准来训练本论文的结构。

# 三、设计方案

3.1 总览

在本论文提出的方法中，输入为3D形状的任何表面点p，经过处理之后输出该点的描述符。将具有不同结构形状，但是几何和语义相似的表面点得到尽可能接近的描述符（欧几里德度量）。此外，不限制输入形状具有一致的对齐或任何特定的方向。本论文的主要假设是输入形状表示为多边形网格或点云，对其连接或拓扑没有任何限制。本论文遵循机器学习方法的基本路线。该功能可以作为特定类别的（例如针对椅子上的匹配点）或作为跨类别（例如针对匹配跨越椅子，自行车，推车等的人类接触区域）来学习。

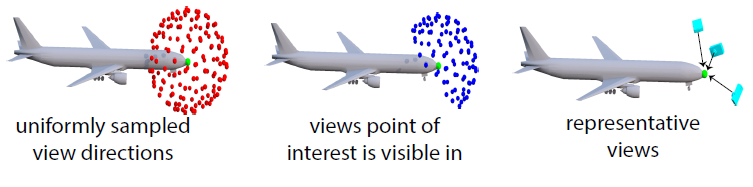
3.2 网络结构



图二

图二展示了本网络的结构，作者使用1024个表面成对点输入形状表面进行均匀采样，并为每个点计算描述符。在测试时，可以对形状上的任意点进行采样并计算其描述符。

预处理。在预处理阶段，首先均匀地采样由球面坐标参数化的形状周围的观察方向。从每个观察方向渲染形状，使得每个像素将索引存储到通过透视投影映射到该像素上的表面点。因此，对于每个表面采样点，可以检索点可见的查看方向。由于相邻的观察方向产生非常相似的表面邻域的渲染视图，作者进一步修剪每个表面采样点的观察方向集合，减少冗余图像数量。通过在所选择的观察方向上执行K-mediods聚类算法来完成修剪（使用它们的球面坐标来测量它们之间的球面距离）。为了捕获每个表面点及其选定的观察方向的多尺度周围环境，本论文创建了3个视点，放置在形状的边界球半径的不同位置。图三展示了代表性的视图方向。



图三

输入。本论文的网络的输入是一组渲染图像，其基于上述视点配置所描绘的表面点p周围的局部表面邻域。由于旋转3D形状将均匀采样视图方向视图，在代表性视图中可以看到关注点。为了促进旋转不变性，以90度间隔旋转输入图像 4次。图像以227×227分辨率渲染。

视图池化。本论文的基础架构为AlexNet，并且在原网络中去除了全连接层。通过上述架构传递每个渲染图像产生4096维描述符。由于每个点总共有36个渲染视图，本论文需要将这36个基于图像的描述符聚合成一个紧凑的点描述符。原因是评估点的每个单个基于图像的描述符与另一个点的所有其他36个基于图像的描述符的距离的代价将是非常昂贵的。为了产生单点描述符，本论文通过使用选择跨视图的最有辨别力的描述符（最大特征条目）的元素方式最大操作来聚合输入36渲染视图中的描述符。类似的“最大视图池化”策略也被有效地用于形状识别。在本论文的例子中，池化应用于本地（而不是全局）基于图像的描述符。

降维。给定由视图池化聚合产生的点描述符Yp，本论文进一步降低其维度以使最近邻查询更有效并且还权衡任何不包含有用信息的维度（例如，阴影信息）。通过在执行线性变换的视图池化之后，再添加一个层来执行降维。

3.3 学习

为了训练本论文中的网络，需要大量的对应点的数据集。现有的数据集对应基准只有有限数量的形状，或专注于特定情况，例如可变形的形状。最大的此类数据集是分段的ShapeNetCore数据集，包括分为16个类别的17K人造形状。通过观察发现，虽然这些人造形状在其零件的数量和排列方面存在显着差异，但具有相同标签的单个零件通常通过简单的变形来关联。通过在具有相同标签的成对部件上执行的非刚性配准来计算这些变形，可以获得训练点对应的大型数据集。虽然可能这种方法产生的对应关系并不如人工标注的精确，但是巨大的数量可以弥补这个缺点。

基于部件的配准。给定一对一致分割的形状和，本论文的配准算法旨在非对称地对齐两个形状中具有相同标签的所有区段对。首先，本论文从它们的网格表示中采样10K个点和。这些点用相应面的零件标签标记。令，表示分别来自和中具有相同标签的一对相应部分的点集。对于每个部分，本论文计算的初始仿射变换，使其具有与相同的定向边界框。然后对于每个点，本论文寻找一个平移（偏移），它尽可能地将它移动到由表示的表面，而中相邻点的偏移尽可能相似。确保顺利变形。以相同的方式，对于每个点，本论文计算一个偏移，它使部分平滑地变形为。为了计算偏移，本论文最小化了一个变形能量，它会惩罚点集之间的距离。这两部分，以及相邻点之间的不一致偏移：

其中，分别是每个点的邻域，dist计算转换后的点到其余点集的最近兼容点的距离。使用基于ICP的处理过程可以最小化能量：在最初两个部分上给出最接近的兼容点对，通过最小化上述能量来计算偏移，然后更新最近的对。最终偏移提供和的最近兼容点之间的密集对应。

训练网络。本论文构造了一个损失函数来惩罚对应点对的描述符的差异，以及非对应点对的描述符小差异。其中包含一个正则化项，以防止参数值变得任意大。损失函数如下：

其中是一组对应点，为一对输入描述符之间的欧几里德距离。正则化参数被设置为0.0005。数量（称为边距）设置为1(仅缩放距离，使得非对应点对倾向于具有至少一个单位距离的余量)。本论文的卷积层使用训练好的AlexNet进行初始化。优化方法采用Adam法。通过批量梯度下降最小化成本函数。在每次迭代中，随机选择32对对应点。 这些对来自随机的形状对，本论文已经预先执行了基于部件的配准。此外，选择了32对非对应点，使得本论文的总批量大小等于64。

# 四、实验

本节主要描述了使用本论文提出的方法在目前的公开数据集上的实验结果及分析。

数据集。在BHCP基准上评估本论文的产生的描述符。该基准包括404个人造形状，包括自行车，直升机，椅子和来自Trimble库的飞机。这些形状具有显着的结构和几何多样性。每个形状具有6-12个一致选择的具有语义对应的特征点（例如翼尖）。好的描述符值可以将这些特征点与其余特征点区分开来，并在描述符空间距离相近。除此之外还需要评估旋转不变性。BHCP中的大多数形状始终是直立的，没有足够泛化性，偏向某些方面。因此对每个BHCP形状应用随机3D旋转。

方法。本论文对比了多个目前流行并且先进的方法，包括由3DMatch的体积CNN产生的学习描述符，和几个人为构造的描述符。3DMatch方法将深度图像投影回3D空间以在体积网格中获得截断距离函数值，此方法使用了体积CNN。在本论文的比较中，本论文为了和体积CNN使用了相同类型的输入（在3D表面点周围提取体素TDF贴片）。为了确保本论文的方法与3DMatch之间的公平比较，本论文在与CNN相同的训练数据集上训练了3DMatch的体积CNN。本论文尝试了两种针对3DMatch的训练策略：（a）在本论文的数据集上从头开始训练他们的体积CNN，以及（b）用他们公开的模型初始化体积CNN，然后在本论文的数据集上微调它。微调策略比从头开始训练他们的CNN更好，因此本论文在此策略下报告结果。

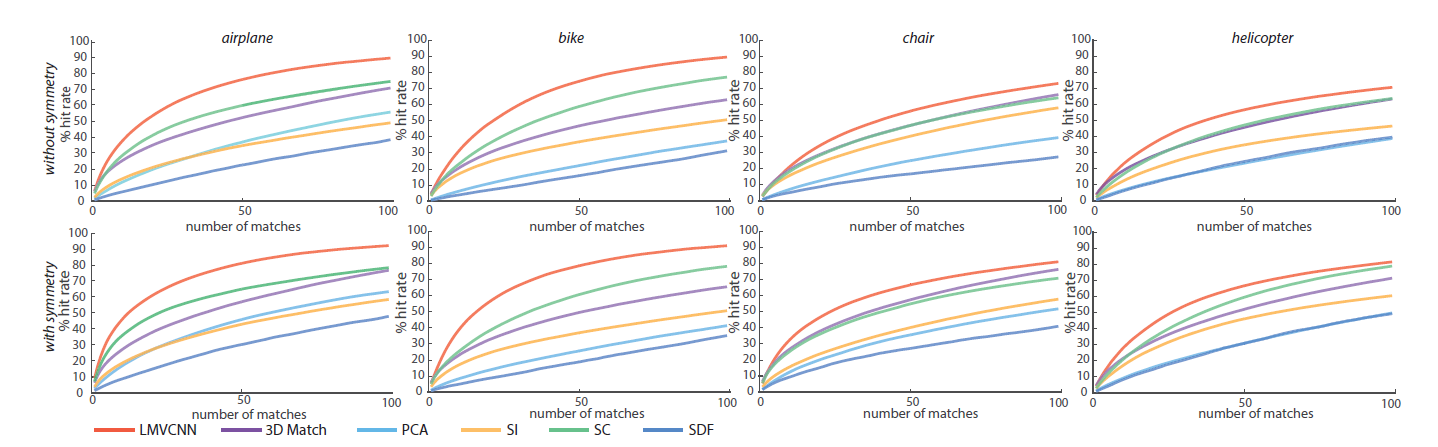
训练设置。本论文在两种训练设置下进行评估。在本论文称为“单一类别”设置的第一个训练设置中，本论文从单个类别训练本论文的方法对点对应数据（在第5节中描述）并测试相同或另一个类别的形状。为了构建一个更通用的描述符，本论文还在“跨类别”设置中训练本论文的方法，为此本论文训练本论文的方法来训练跨分段的ShapeNetCore数据集的几个类别的数据，并测试相同的形状或其他类别。



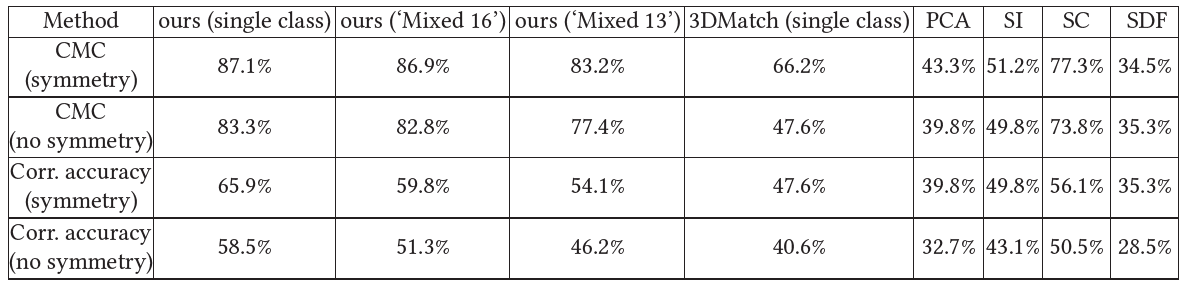
图三

图三表示了学习到的描述符对代表点采样后对应点的可视化，相应的点RGB值相同。

性能指标。本论文使用两种指标来评估方法生成的特征描述符。首先，本论文使用累积匹配特征（CMC）测量，该测量旨在捕获描述符空间中对应点的接近度。特别地，给定一对形状和其中一个形状上的输入特征点，本论文检索另一个形状上的点的排序列表。根据这些检索点与描述符空间中的输入点之间的欧几里德距离对列表进行排序。通过记录所有形状对的所有特征点的等级，本论文创建一个图，其Y轴是真实值对应点的部分，其等级等于或低于在X轴上标记的等级。应该将最高等级分配给真实对应点。

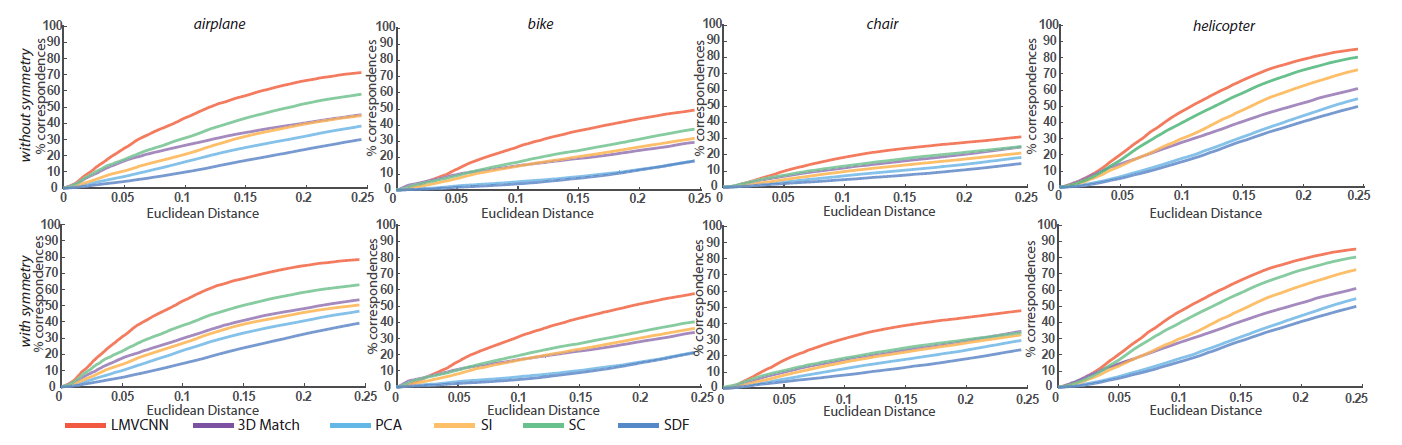


图四



表一

另一种衡量标准是对应准确性。该度量旨在捕获预测的对应点与3D空间中的真实对应点的接近度。具体来说，给定一对形状和其中一个形状上的输入特征点，本论文在另一个形状的描述符空间中找到最近的特征点，然后测量其3D位置与真实位置对应点之间的欧几里德距离。通过在所有形状对上收集欧几里德距离，本论文创建了一个图，其Y轴显示了在X轴上显示的给定欧几里得误差阈值以下正确预测的对应部分。根据应用，匹配对称点是可以接受的。因此，对于这两个指标，本论文讨论下面的结果，其中本论文接受对称（例如从左到右翼尖）匹配，或不接受它们。



图五

图四，表一，图五分别表明了在CMC指标，对应准确性指标下评估的结果。可以看出，本论文提出的方法相较于其余的方法占有较大优势。

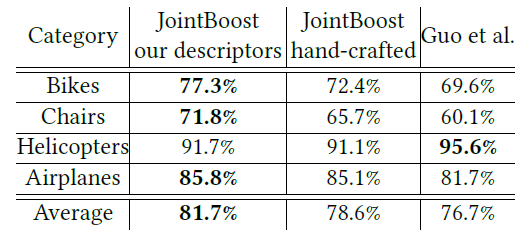
# 五、应用

主要有形状分割、使用3D扫描匹配形状、预测交互区域。



图六

形状分割。给定一个输入形状，使用本论文的描述符根据一组零件标签标记表面点。图六是分割的一个可视化例子。按照图切割能量公式，图切割能量依赖于评估网格面与零件标签的一致性的一元术语，以及提供相邻面是否应具有相同标签的线索的成对项。为了评估一元项，原始实现依赖于每个网格面计算的本地手工描述符。用本论文的方法提取的描述符替换了所有这些人为构造的描述符，以检查分割结果是否得到改进。表二总结了结果。使用本论文学到的描述符平均提高标签准确性。



表二

使用3D扫描匹配形状。本论文的描述符的另一个应用是扫描和3D模型之间的密集匹配，这反过来可以有益于形状和场景理解技术。为了渲染点云，使用以每个点为中心的小球。即使扫描有噪声，包含异常值，缺少整个部分，或者有嘈杂的法线和随后的阴影瑕疵，也可以生成健壮的描述符，以便将它们与完整的形状密切匹配。



图七

预测交互区域。如图七所示，本论文使用可供性基准与人工选择的接触区域，以便人们与各种物体相互作用（例如骨盆和手掌的接触点）。具体来说，给定参考形状上这些区域的标记点，本论文根据它们与描述符空间中标记点的距离来检索其他形状上的点。从这些结果中本论文可以看出，本论文的方法还可以推广到在不同类别的形状之间匹配局部区域，并且具有很大不同的全局结构。

# 六、结论

本论文提出了一种通过在多个尺度中获取形状区域的多个渲染视图并通过学习的深度卷积网络处理它们来计算局部形状描述符。通过视图池化和降维，生成紧凑的局部描述符，可用于各种形状分析应用程序。本论文的结果证实了使用这种基于视图的架构的好处。本论文还提出了一种生成综合训练数据的策略，以自动化学习过程。迭代训练数据生成，学习和非刚性对齐的流程可以进一步提高性能。

# 七、总结及思考

不仅仅局限于在三维空间中，本论文提到的创新点比如主动探寻更大数据集的方法等等，在二维的图像处理中也值得借鉴。但是存在很多局限性。目前，本论文提出的方法依赖于基于启发式的配置和渲染过程。我们可以探索相关的优化策略来获得更好性能。并且 依靠透视投影来捕获当地表面信息，也许还有其他更好的方案来获得表面信息。本论文的自动非刚性对齐方法往往会对几何和拓扑变化很大的训练形状部分产生不太准确的训练对应关系。太多错误的训练对应将反过来影响描述符的判别性能。

# 参考文献

[1] Huang, H., Kalogerakis, E., Chaudhuri, S., Ceylan, D., Kim, V.G., Yumer, E.: Learning local shape descriptors from part correspondences with multiview convolutional networks. ACM Transactions on Graphics 37(1) (2018)