CVPR2017 [paperlist](http://cvpr2017.thecvf.com/program/main_conference#program_summary)

CVPR2017 18个spotlight oral session

Machine learning 4个

3D computer vision 2+2

利用NN由单张图片进行3D场景重建

点云的配准，NN处理点云，进行重建，CAD

利用神经网络对单目图像的深度估计

物体识别场景理解 3+1

图像标题生成

Analyzing humans 2

低级中级的视觉 1

图像运动&tracking&视频分析 1

Computational Photography 1

Application 1

CVPR2017 3D vision 1

oral 1-1B

Semantic Scene Completion from a Single Depth Image

CVPR2017

文章标签：较新文章

1、Introduction

创新点

改进了TSDF

创建了一个合成数据集

2、Related work

RGBD-语义分割

形状补全

3D模型fitting

体素空间reasoning

从合成数据中学习

3、Methodology

语义场景补全网络

该网络的目标是将体素映射到一个label集合C中的某一个



体素占有是否用神经网络预测，这和一个个的桌子椅子凳子组成的库有什么关系？

类别概率是体素占有判别完了再做的吗，还是同时进行的吗？

体积数据编码

TSDF来编码3D空间

每个体素存储的

1、：该体素到离它最近的表面的距离

2、该体素是否为空

对标准TSDF做了以下改进

消除视角依赖

消除empty space的强梯度

网络结构

* 输入体积生成

3D场景编码成flipped TSDF，网格大小2cm，截断值0.24米？

4.8\*4.8\*2.88----240x240x144体素作为网络输入

* 抓取3D背景with大接收域
* 多尺度背景合成
* 数据平衡
* Loss：voxel-wisesoftmax
* 训练协议

合成训练数据

SUNCG一个大的从场景合成数据集

合成深度地图生成

体积groundtruth生成

4、Experiment

评价尺度

实验结果

与其他方法的比较

物体识别能帮助场景补充吗

场景补充能帮助物体识别码

合成数据有帮助吗

更大的接收域(receptive field)有帮助吗

多尺度合成有帮助吗

不同的编码方式有影响吗

数据平衡必要吗

限制

5、Conclusion

3DMatch: Learning Local Geometric Descriptors from RGB-D Reconstructions

3D匹配：从RGB-D重建学习局部几何描述子（CVPR2017）

文章标签：较新文章

1、Introduction

现今3D匹配的难点:noisy, lowresolution, and incomplete nature of 3D scan data

本文的方法是数据驱动的学习模型

2、Related work

Hand-crafted 3D 局部描述子

Learned 2D 局部描述子

Learned 3D 全局描述子

Learned 3D 局部描述子

自我监督深度学习

定义：自动获得标签，不需要人手工标注标签

例如：视频里的时间信息，可以作为监督的来源

本文获取数据的方法也是全自动的，从现存的RGBD重建数据中提取correspondence labels

3、Methodology

从重建中学习

目的：创建一个方程$fi$,将一个在3D表面上的点周围的体积区域映射为一个描述子向量。两个向量之间的距离l2越小，这两个点匹配上的可能性就越大。

方法：利用现存的高质量的RGB-D重建的数据，来学习这个方程$\fi$

该方法的三点优势：

* 重建数据集的点非常多，包含很多不同的视角
* 重建可以利用（leverage）时间信息以及全局优化，而用匹配对来训练一个强有力的描述子却不需要这些东西
* 鲁棒性,各种不同的训练集，不同的重建算法

生成训练对

1、在兴趣点周围不同的扫描角度提取局部3D patches

兴趣点：从重建中随机采样的

2、为了找到一个兴趣点对应点，我们把它在reconstruction中的3D位置投影到RGB-D frame中去

（这些3D点都在相机的视锥中，没有被遮挡）（不同视角的相机位置至少相隔1m，这样保证了观察对的大基线）

3、然后从两个RGBDframe中，在兴趣点周围提取两个局部3D patches，把他们两个当做匹配对。

（为了获得非匹配的点对，我们从随机抽取的关于兴趣点的两个深度帧抽取3D patches，兴趣点也是从重建的表面随机采样的）

我们把传感器噪声的扰动和重建的误差当做学习不变性使得其更鲁棒的契机

学习一个局部几何描述子

Volumetric 3D patch

512维描述子

3D ConvNet

3D数据表示

网络结构

网络训练

4、Experiment

Evaluation

特征点匹配

geometric registration

在场景中匹配局部几何

在重建过程中结合3D匹配

可以将3D匹配推广到新的领域吗

6自由度姿态估计通过模型对齐（model alignment）

Surface Correspondence on 3D meshes.

5、Conclusion

6、附录

RGB-D重建数据集

实施细节

运行时间信息

TDF转换

3D匹配描述子

Geometric Registration.

PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation

(CVPR2017) PointNet：用于3D分类和分割的基于点集的深度学习

之前处理点云的方式都是编码成体素网格，会导致冗余，我们这里直接将点云喂给网络，有很多用途，物体分类，部分分割，场景语义解析

1、Introduction

2、Related work

**点云feature**

**Deep Learning on 3D Data**

**Deep Learning on Unordered Sets**

3、Methodology

点集性质

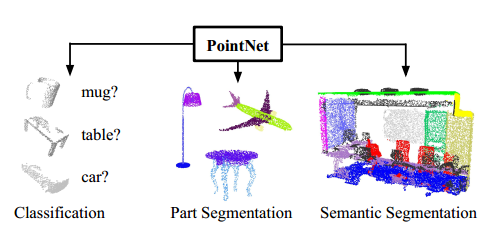
PointNet结构

理论分析

Visualizing PointNet

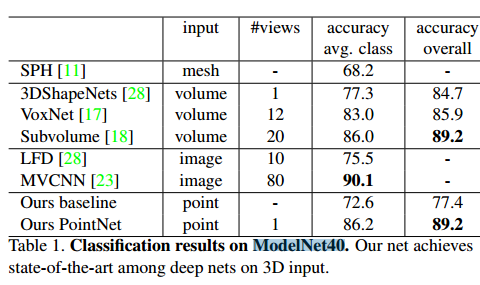
4、Experiment

应用



3D物体分类

测试数据集**ModelNet40**

评估指标：每一类的准确率，总体的准确率，MVCNN最好，但是输入的是多个面，最主要的提升是本文方法只用一个view就能取得与其他相当的方法， 

问题：与3Dshapenet是持平的呀，baseline是什么？

评估：准确率已经做到90%，overall有5个百分点的增长

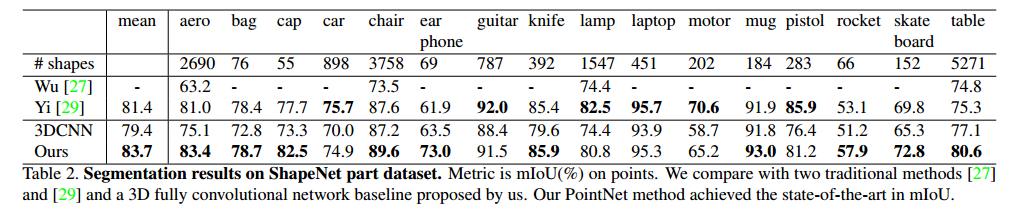
3D Object Part Segmentation

**测试数据集：**ShapeNet part data set from [29]

16,881 shapes from 16 categories, annotated with 50 parts in total

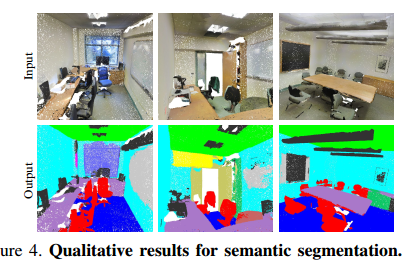
**评价尺度：**mIoU on points

方法只有之前的两种，3DCNN也是他们做的baseline

****

评估：平均下来只有两个百分点的增长

场景语义分割

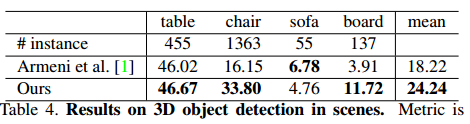


**测试数据集：**Stanford 3D semantic parsing dataset [1]

The dataset contains 3D scans from Matterport scanners in 6 areas including 271 rooms

**Metric** ： average IoU over 13 classes

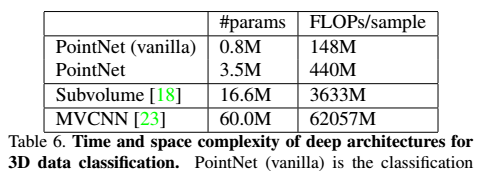
Metric is average precision with threshold IoU 0.5 computed in 3D volumes



评估：测试训练数据集太小，可参考资料太少

**总体评价**

参数少，在3D分类项目中，该网络参数少，时间快,且一个网络可以做三个用途



5、Conclusion

Multi-view Supervision for Single-view Reconstruction via Differentiable Ray Consistency

单视角重建的多视角监督，通过 ray一致性

1、Introduction

研究了2D和3D形状之间的consistency这个概念，提出的这个方法能计算3D形状的梯度（给出任意视角的2D observation）

用DRC重新定义了view consistency

这个formulation能将不同类型的信息（forground mask，深度，彩色图像，语义）融合到learning framework中去，作为学习单视角预测的监督。

这个方法可以提高单视角重建，在PASCAL VOC数据集上

2、Related work

**基于图像注解的物体重建**

**基于3D监督的物体重建**

**Multi-view Instance Reconstruction**

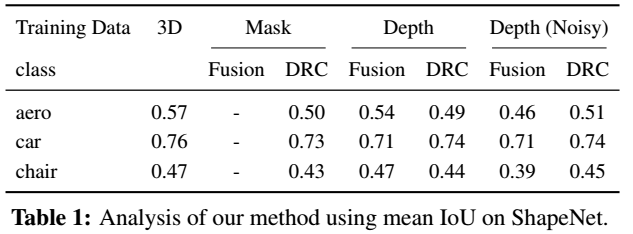
**Multi-view Supervision for Single-view Depth Prediction.**

3、Methodology

4、Experiment

5、Conclusion

**5.1、ShapeNet的经验分析**



**Setup**

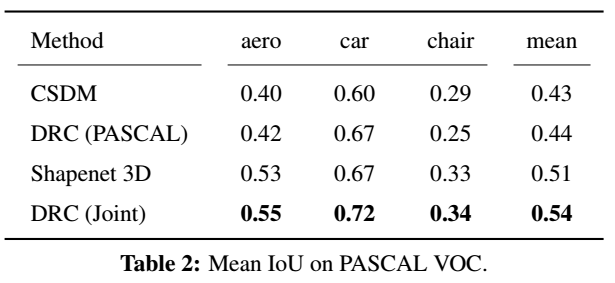
**Evaluation metric**

Ground-truth occupencies和predicted 3D occupencies之间的交集（IoU）

**Results**

**Ablation切除**

**5.2、Object Reconstruction on PASCAL VOC**



评估：都是对自己的结果分析，并没有对比？

**5.3. 3D Scene Reconstruction from Ego-motion**

**5.4. Object Reconstruction from RGB Supervision**

Shapenets 数据集 ShapeNet: An Information-Rich 3D Model Repository

2015

3D CAD数据库，根据WordNet分类法组织它们

它是为每个3D模型提供许多语义注释的数据集的集合，例如一致的刚性对齐，零件和双边对称平面，物理尺寸，关键字以及其他计划注释。

ShapeNet已经索引了超过3百万个型号，其中22万个型号分为3,135个类别（WordNet synsets）

1、Introduction

通过部分扫描来识别形状是计算机视觉以及图形学共同的难题。

机器人的自主导航，以及抓取物体操作是两大受益于3D形状理解的方向。

这些问题的根源是给3D形状贴上语义，

2、Related work

3.1、数据采集

3.2、标签类型

3.3、标注方法

4、注释获取和验证

5、当前统计

ShapeNets是目前最大的3D独立物体的数据集，人造物品居多，不像imageNet那样自然物品比如说植物和动物。（自然物体的CAD建图不容易）它包含以下子集：

**5.1. ShapeNetCore**

单一干净的3D模型的子集，55个类别，5万个独立的3D模型，包括常用的PASCAL 3D+（12个物体类别）也在里面

**5.2. ShapeNetSem**

这是一个小的密集标注的子数据集

270个类别的12000个类

有在真实世界的尺寸，材料的估计，以及总体积和重量的估计

On-the-Fly Adaptation of Regression Forests for Online Camera Relocalisation

1、Introduction

相机的重定位问题

常用技术可以将当前图像与来自跟踪器的已知姿势的关键帧匹配，或者在当前图像中的关键点和场景中的点之间建立2D至3D对应关系，以估计相机姿态。

现在比较流行的是回归森林，但是需要对特定的场景进行离线训练，但是当遇到新场景时，无法表现出适应性，本文提出了一中适应森林adapted forest，可以绕过这个障碍。

Splotlight 1-1B

Context-aware Captions from Context-agnostic Supervision

感知背景的图像标题生成

Global Hypothesis Generation for 6D Object Pose Estimation

从RGB-D图片估计3D物体对象的6D姿态，用一种全局的方法替代了ransac

A Practical Method for Fully Automatic Intrinsic Camera Calibration Using Directionally Encoded Light

全自动相机内参标定方法，使用定向编码光

CATS: A Color and Thermal Stereo Benchmark

色彩和温度双目基准

Elastic Shape-From-Template With Spatially Sparse Deforming Forces

完全不知道在说什么

Distinguishing the Indistinguishable:Exploring Structural Ambiguities via Geodesic Context

完全不知道在说什么

Multi-Scale Continuous CRFs as Sequential Deep Networks for Monocular Depth Estimation

用于单目深度估计的一个网络

Dynamic Time-Of-Flight

动态ToF

3D Vision 2 Oral 2-1C

**A Point Set Generation Network for 3D Object Reconstruction from a Single Image**

(CVPR2017)

1、Introduction

用神经网络生成3D数据非常热门，大部分方法用体积格或者图像集合来表示。但是这个表示方法模糊了3D形状在集合变换下天然的不变性？

本文：由单张图片（RGB或者RGBD）进行3D重建，输出一个点云坐标系

条件形状采样器*conditional shape sampler* ：由输入图片预测多个可能的3D点云

2、Related work

由单张图像进行3D重建

深度学习对几何物体合成

3、Methodology

问题及注解

定义groundtruth为基于输入图像I的概率分布P（·|I）

神经网络G作为P的条件采样器

方法

方法概览

点集估计网络

点集间的距离尺度

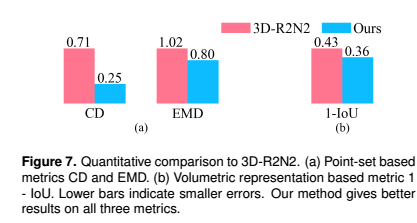
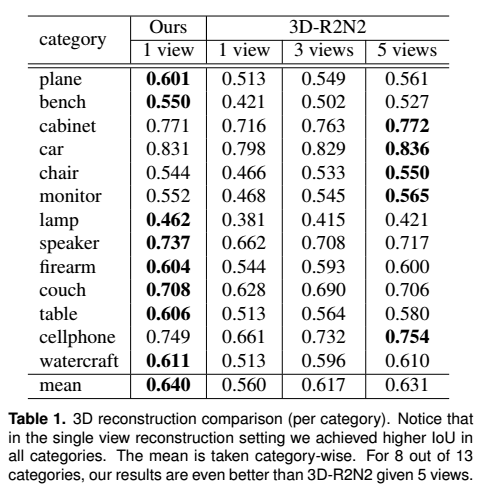
4、Experiment

5.1、合成训练数据

5.2、从RGB-D重建物体外形

对13个类别进行3D重建比较，对比数据有他们之前的work 3D-R2N2

The results are compared under three different metrics CD, EMD and IoU (intersection over union).



5.3、从RGB-D补全外形

通过加入一部分先验形状，来预测看不见的部分

5.4、预测多种可能的外形

我们网络的随机性使得对同一个输入，有多种不同输出的可能性

5.5、网络设计分析

5.6、真实世界的更多结果和应用

5.7、人类单视角3D重建能力的分析

错误结果分析

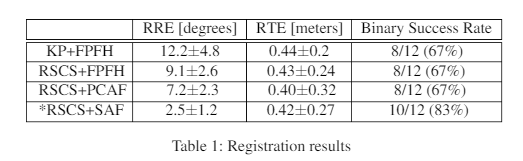
实施细节

5、Conclusion

3D Point Cloud Registration for Localization Using a Deep Neural Network Auto-Encoder

用于定位的3D点云注册，基于深度神经网络自动编码

什么是点云注册？



Flight Dynamics-Based Recovery of a UAV Trajectory Using Ground Cameras

使用地面摄像机 基于飞行动力学的无人机轨迹恢复

DSAC - Differentiable RANSAC for Camera Localization

用于相机定位的差分RANSAC算法

Spotlight 2-1C

Efficient Solvers for Minimal Problems by Syzygy-Based Reduction

基于XXX的最小问题求解器

HSfM: Hybrid Structure-from-Motion

混合SFM算法

Efficient Global Point Cloud Alignment Using Bayesian Nonparametric Mixture

全局点云对齐算法

A New Rank Constraint on Multi-View Fundamental Matrices, and Its Application to Camera Location Recovery

多视图fundamental matrix的一种新的秩约束，以及它在相机定位恢复中的应用

IM2CAD

给出了一张房间的单张照片和一个大型家具CAD模型数据库，我们的目标是重建与照片中描绘的场景尽可能相似的场景，并从数据库中绘制的对象组成。

ScanNet: Richly-Annotated 3D Reconstructions of Indoor Scenes

一个网络：丰富标注的室内3D场景重建

Noise Robust Depth From Focus Using a Ring Difference Filter

一个滤波器，噪声深度

Group-Wise Point-Set Registration Based on Rényi's Second Order Entropy

点集配准算法

Analyzing humans with 3D vision

oral 4-1B

Recurrent 3D Pose Sequence Machines

递归3D姿态顺序机，完全不知道在讲什么

Learning Detailed Face Reconstruction From a Single Image

从单一图像学习的面部细节重建

Thin-Slicing Network: A Deep Structured Model for Pose Estimation in Videos

一个网络，用于视频中的姿态估计

Dynamic FAUST: Registering Human Bodies in Motion

运动人体配准

Spotlight 4-1B

Machine Learning for 3D Vision

Oral 4-2B

Unsupervised Monocular Depth Estimation With Left-Right Consistency

无监督单目深度估计，根据左右一致性

Unsupervised Learning of Depth and Ego-Motion From Video

从视频中进行无监督学习深度和自由运动

OctNet: Learning Deep 3D Representations at High Resolutions

OctNet：在高分辨率下学习深度3D表示

3D Shape Segmentation With Projective Convolutional Network

用投影卷积网络进行 3D shape分割

Spotlight 4-2B

Multi-View 3D Object Detection Network for Autonomous Driving

利用多视角进行3D物体检测的网络，自动驾驶

UltraStereo: Efficient Learning-Based Matching for Active Stereo Systems

双目匹配的一种新方法 active？

Shape Completion Using 3D-Encoder-Predictor CNNs and Shape Synthesis

形状补全，利用3D编码预测CNNs以及形状合成

Geometric Loss Functions for Camera Pose Regression With Deep Learning

深度学习的几何代价函数，用于相机姿态回归

CNN-SLAM: Real-Time Dense Monocular SLAM With Learned Depth Prediction

CNN-SLAM:一种实时稠密单目SLAM，深度预测是通过学习得来

Learning From Noisy Large-Scale Datasets With Minimal Supervision

利用最少的监督，从大规模有噪声的数据集中学习

SyncSpecCNN: Synchronized Spectral CNN for 3D Shape Segmentation

一个网络，用于3D形状分割的同步光谱CNN

Non-Local Deep Features for salient Object Detection

一种非局部的深度特征，用于突出物体检测

Poster： 3D Computer Vision

Poster1-1

Previous Paper

Structured Prediction of Unobserved Voxels From a Single Depth Image

使用单张深度图对遮挡体素的结构预测（CVPR 2016）

标签：较新文章

1、介绍

创新点：

用随机森林来训练预测

2、相关工作

3D primitives（图元，基元）

具体的形状模型

表面补全

体素空间resoning

数据集

3、Methodology

体素算法回顾

Support regions

R就是表面体素附近的一个长方体区域，用我们的模型能够进行TSDF预测

R的上下面与地面平行

方向：没看懂

Voxlets

对R遮挡的部分进行的几何预测

方法：在测试时，用观察到s对R做一个feature descriptor x(s)，再用训练出来的判别模型对遮挡部分进行预测，随机森林预测

学习从feaures到voxlets的映射（mapping）



.输入：feature vector

输出：编码了TSDF的R（w\*d\*h）

从低维映射到高维，用结构化的随机森林来学习f

训练

Feature

从s提取特征描述子

1、将深度图像D用相机内参投影到三维空间，重投影

2、用kinect fusion里的方法生成TSDF体素网格*VD*

3、特征向量x是直接从*VD*提取的，在s周围的三维空间内

3D volume features

从以s为中心的一个球体采样的

半径:?

球体的方向：aligned to the normal and world up direction at **s**

在球内采集260个offsets作为候选特征

在测试时间预测占有

选择最好的预测

实施细节

数据集

4、实验

基线

结果

5、结论和未来工作

我们的监督算法有效地结合了局部形状的选择和姿态估计，使用简单的特征测试评估来预测局部几何占有率。

1、物理证明未解决

2、将这个方法应用到slam

随着新数据的到来，下一个最佳视图算法[42,31]可以利用我们的预测来将相机引导到捕捉到最多几何信息的位置来验证我们的提案。

6、相关需了解知识

Surface reconstruction

surface construction的核心是TSDF及其更新。

TSDF全程Truncated Signed Distance Function。它将空间场景模型化为一个立方体栅格，每个栅格中都存有两个值，一个是距离值F，另一个是权重值W。

网格生成算法的输入一般为三维点云，输出为三维网格。输入的点云一般为若干个点云的集合。输出的网格一般为单一的网格。

3D volume features

Kinect fusion里有，还有[7]里面也有

随机森林

A Volumetric Method for Building Complex Models from Range Images (1996)

Truncated Signed Distance Function (TSDF) --

标签：基础理解

1、Introduction

2、Related work

3、Methodology

如图4所示，TSDF模型中存储的距离值，surface处值为零，传感器一侧的值为正，距离表面越远值越大，另一侧值为负，距离表面越远值越小。而权重值则与表面测量的不确定度有关。

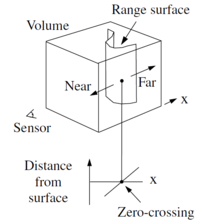


图4.TSDF模型

而在实际的测量中，接近零的位置的值才是有效的，因此需要设定一个阈值u，将与surface距离大于u置为无效，不予考虑，而小于u的值进行归一化截断。  
图5给出了当前帧TSDF距离值的计算方法。

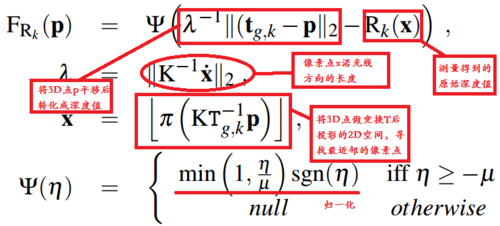


图5.TSDF距离值F的计算公式

全局的TSDF是由每一帧单个的TSDF加权平均得来的，如图6所示。

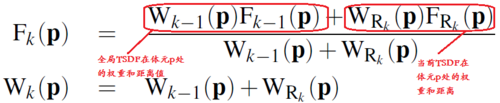


图6.全局TSDF

虽然理论上权重值和表面测量的不确定度有关，即与顶点光线方向和法线方向夹角的余弦值成正比，与相机中心和顶点的距离值成反比。但是在实际中，权重恒等于1可以得到很好的结果，并且权重的累积不是一直增大，而是设置了最大值阈值，如图7所示。

http://upload-images.jianshu.io/upload_images/5134720-17466219d09505be.png?imageMogr2/auto-orient/strip%7CimageView2/2/w/300

图7.设置最大阈值

参考论文：

A Volumetric Method for Building Complex Models from Range Images VRIP 1996用深度图重建网格

kinect fusion Real-Time Dense Surface Mapping and Tracking

博客

从点云到[网格](http://www.cnblogs.com/luyb/p/5658784.html)

[**KinectFusion**](http://www.jianshu.com/p/7a447c633ae9) **论文精析**

4、Experiment

5、Conclusion

kinect fusion Real-Time Dense Surface Mapping and Tracking

(ISMAR 2011)

IM2CAD

Challenge:将图片中的和CAD模型匹配起来，尤其是模型和图片里的物体不完全匹配

本文的贡献是写了一个系统，可以自动的将图片变成全景CAD模型（房间+家具）

全景优化算法：迭代的调整了物体的拜访和尺寸

有语义，用shapenet里的物体ID，和方向，位置，尺度耦合起来

Related work

房间形状估计 地板，墙

用cuboids逼近家具

3D物体的姿态识别

单张图片中的物体识别

Hand-crafted features，本文使用了CNN feature

Methodology

**Room Geometry Estimation**

ranking the room 3D box hypothesis using deep convolutional features.

训练一个网络来估计每个像素表面的标签，天花板，底板，左中右墙

**Object Detection**

该网络有两层

先检测出bounding box，再计算likelyhood进行归类

**CAD Model Alignment**

用卷积网络，计算生成图片和检测出来的bounding box之间cosine similarity

检测出了是椅子，我们要决定是什么样的椅子，外形，近似的3D姿态

估计3D姿态，通过将他的外形和很多个不同角度去对比，用一个深度全集的距离尺度。只取8类，椅子、桌子、沙发、书架，床头柜，箱子、窗子。每个3D模型有32个视角

**Object Placement in the Scene**

首先用三个正交vanish pointtts来估计出相机内参K和旋转矩阵R

选择一个课件的房间角作为世界坐标系的原点，如果没有角落可见，选择可见墙边和底板的交点作为原点

Shpenet里的3D模型用一个bounding box 归一化到一个单位立方体，根据alignment里的结果，我们可以决定输入图片中对应立方体八个角的pixel location

可以找到物体的位置和x，y上的尺度，通过地面与相机中心投射的光线相交，光线穿过对应CADcube四个底角的对应像素位置。

为了计算Z上的尺度，计算投影cube的四个垂直边和平面行这些边的比例…。 消失线的高度等于相机高度.

把窗子当做一个特殊的例子， 因为它在墙上额而不是在地板上。为了放置窗子，我们找到窗子的bounding box和每个墙的交点，把窗子放到交叠最大的墙上。

窗子的bounding box反投影到墙上的一个四边形，姿态和位置，最大的…

**Scene Optimization via Render and Match**

全局优化通过最小化代价函数，为生成图片和输入图片的cosine distance

照片里的N个物体，每个物体都有7个变量，xyz坐标，xyz尺度和旋转，给出一个参数向量，我们就能生成对应的图片。

具体网络优化如下：

**Coloring CAD models**

scanNet 3D dataset

RGB-D scan

丰富标注 ，larger

1513个scann 707 个地方

用于RGB-D 场景理解

1，怎么收集，2 怎么用

用处以下三个方面：

**3D Object Classification**

3D分类网络

训练集 shapeNet合成数据集训练 ，独立物体

输入是什么 3D合成物体的什么，图片、RGB-D信息？

输出是什么 可能性最高的类

问题：训练时用合成数据集，测试时real-world数据，虽然有一定的依赖性，但是还是有矛盾，比如噪声和遮挡问题

ScanNet ：对bounding box里的scanned points进行分类，训练也是用扫描数据训练，色是也是用扫描数据测试

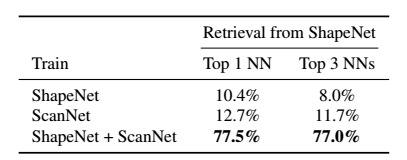
**3d retrieval**

两个网络结合，如何结合

The core idea is to train a network on a shape classification task where a shape embedding can be learned as *byproduct* of the classification task

Training on both ShapeNet and ScanNet together is able to find an embedding of shape similarities between both data modalities, resulting in much higher retrieval accuracy

In Table 6, object retrieval results using objects from ScanNet to query for nearest neighbor models from ShapeNetCore



**Semantic Voxel Labeling**

对所有可见的表面体素预测语义标签in a given 3D scene;

SceneNet在一个大型合成的数据集上训练 fine-tunes on NYU2

voxel classification results on ScanNet in Table 7

semantic voxel labeling results on the NYU2 dataset Table 5

## Recovering the Spatial Layout of Cluttered Rooms（iccv2009）

[Code](http://vision.cs.illinois.edu/~vhedau2/Research/research_spatialLayout.html)

Introduction

**Recovering Spatial Layout.**

1.如何参数化场景空间

2.how do we estimate the parameters using image cues

**Clutter**

**Summary of Contributions.**

3. Estimating the Box Layout

**3.1. Estimating the Box Orientation**

利用消失点估计方向

**3.2. Getting the Box Translation**

**3.3. Learning to Rank Box Layouts with Structured Outputs**

4. Estimating Surface Labels

5.experiment

Pano2CAD: Room Layout From A Single Panorama Image

用网络拟合出一个模型

对齐

深度关系的融合

输入：多张图片 输出：立体图

输入：深度图 输出：立体图

输入：多张CAD图 输出：全景CAD

只需要墙面的融合 ，多视角空间的融合

1.墙面的分割，墙面的配准,生成一个相对的立体空间

五面的墙面就可以重新用texture去配准

2.物体的位置，怎么计算object相对空间的立体位置

那个是怎么做到的？

位置摆放：利用3D cube对应图片中像素的位置

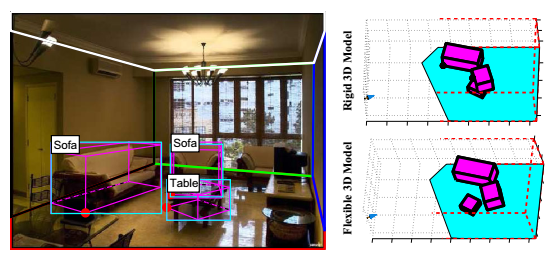
输入：全景图：输出立体图

畸变object的检测

NN是分类，能不能把一个长方体分成很多个网格，再判断这个object属于哪个网格里的

最简单的，仅考虑两个极端，房间两个朝向各拍一张

Indoor Scene Understanding with Geometric and Semantic Contexts （IJCV2015）



想法

只研究box layout那个网络

网络

输入特征

输出结果

很多的特征不知道和结果是什么关系

不知道它是什么模型

想法：看看scene understanding 相关文章

Semantic segmentation

计算IoU二维照片计算IoU

准确率并不高

想法2

假设相机的旋转角度已知

可以完全根据坐标系之间的转换来做

可以结合low level feature

多角度数据集有没有

没有，能不能根据两张照片之间的对应关系，取出这个变换矩阵

评价标准：

IoU：

再看看pano2CAD

问题

1尺度是怎么确定的

2上周看的两篇 3d groundtruth是怎么来的

3两张图片能不能恢复旋转矩阵

接下来

1.看pano2CAD

看他的评价尺度

看他的统计方法是什么

2.看MVG彻底弄清楚由2D图片进行三维重建的原理

3D bounding box estimation Using Deep Learning and Geometry

之前：只回归出3D物体的orientation

2Dbounding box+估计出的几何约束，创造出3D bounding box

第一个网络：估计出3D物体的orientaiton

第二个网络:

a[**rXiv:1511.02300**](https://arxiv.org/abs/1511.02300)**[cs.CV]**

2D Object Detector in RGB-D Images

3D Object Detector

3D Feature Learning

3D Deep Learning

Several recent works [17, 5, 31, 13] also extract deep learning features for retrieval and classification of CAD models. While these works are inspiring, none of them focuses on 3D object detection in RGB-D images

Region Proposal

3D Object Detection in RGB-D Images

Complex Event Recognition from Images with Few Training Examples

事件概念作为识别social events的中间的feature，识别social events

从flicker tags生成的分割，

**3.1. Event Concept Discovery**

Discovered Concept包含一系列物体，场景，动作，事件下属类型

为图片生成一个简短的表示

1. 从维基上挑选150 generic social events

每个events从flicker上挑选200张图片 收集这些图片的tag tag的分割我不懂

1. 通过谷歌的数据集，为每个event选择20个最近的neighbor
2. 将谷歌和flicker的segments结合

最终有856个event concepts ，Concept不经包含物体，场景和动作，也包含下属的events和他们的类型

**3.2. Training Concept Classifiers**

比起直接用深度CNN feature把图片分到对应的events类中用有限的训练数据，我们的方法要好。

什么方法？

**3.3. Predicting Concept Scores for Classification**

总体的准确率要高

Unseen event catogories ？

主要贡献：

1.从网络数据发现事件相关的概念，删除某些杂乱的概念

2.一个concept-based表示方法

3.SocEID，包含8个社会种类的事件，37000张图片

RED 包含稀少的事件，21个种类，7000张图片，更有挑战性。

获取图片

对图片进行表示，主要是表示

对图片事件进行分类

One shot learning？

word2vec space？

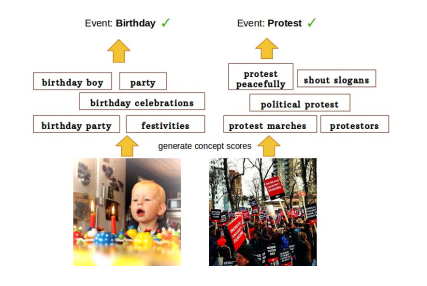
Action and Event Recognition in Videos by Learning From Heterogeneous Web Sources

主要强调从网上扒数据学习？

计划

目的：

由单张图片进行事件检测



应用：

场景的监控，比如监控视频，该算法实现后，可根据具体的场景建立自己的数据集。

数据集

SocEID，包含8个社会种类的事件，37000张图片

RED 包含稀少的事件，21个种类，7000张图片

UIUC SPORTS

WIDER

RED

创新点

1.在事件识别中加入人表情的识别

2.将识别出来的所有东西转化成向量

人，物体，标语 目前已经有转化成向量的了

计划