# 基于深卷积神经网络的室内RGB-D语义分割研究综述

## 1. 语义分割

### 1.1. 公式化：

输入RGB图像：

产生语义分割图：**，**其中**C**是语义类的数目

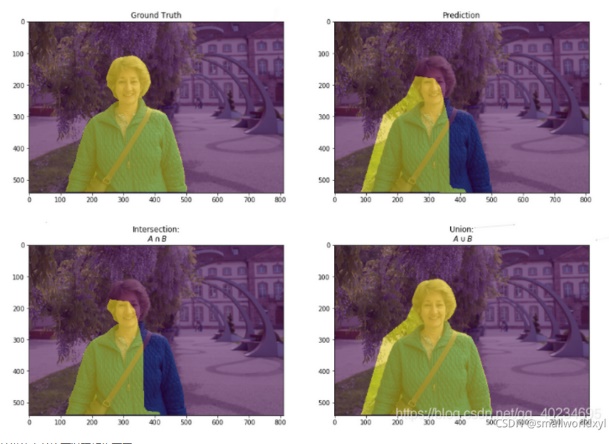
在RGB-D背景下，深度图：也可以作为输入，以增强预测分割图的准确性。

### 1.2. 度量：

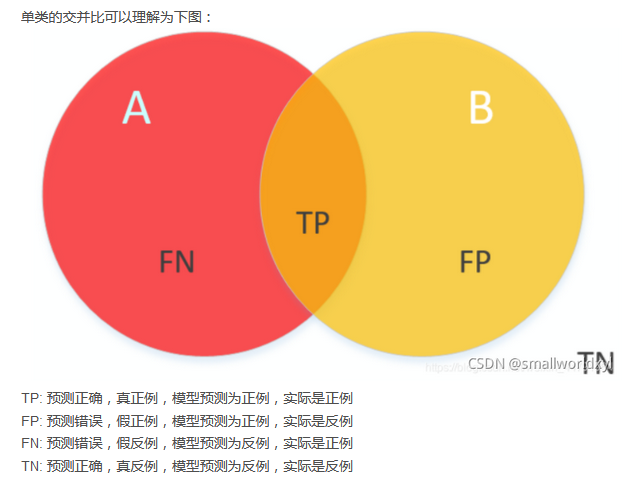
Pixel Accuracy (PA)：像素准确度为预测结果S中被正确预测的像素的比率

mean Intersection over Union (mIoU)：预测结果S和真实值之间的所有交集的平均值

交并比：该类的真实标签和预测值的交和并的比值



单类的交并比：



TP: 预测正确，真正例，模型预测为正例，实际是正例

FP: 预测错误，假正例，模型预测为正例，实际是反例

FN: 预测错误，假反例，模型预测为反例，实际是正例

TN: 预测正确，真反例，模型预测为反例，实际是反例

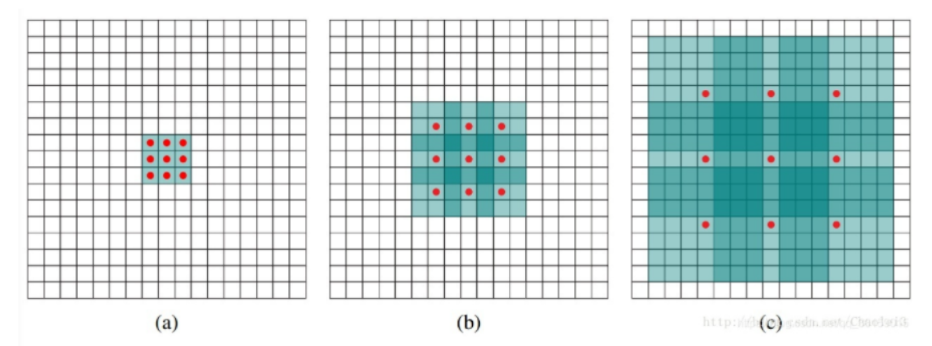
Intersection over Union（IoU）定义为：

MIOU就是该数据集中的每一个类的交并比的平均，定义为：

### 1.3. 分割网络分类：

① encoder-decoder结构：由编码器、解码器两个模块组成。编码器通常是一个标准的主干网络，其目的是提取将被馈送到解码器部分的特征。解码器恢复编码器深层部分丢失的空间信息，重建语义分割图。

② 基于atrous卷积：空洞卷积是针对图像语义分割问题中下采样会降低图像分辨率、丢失信息而提出的一个卷积网络。利用添加空洞扩大感受野，让原本3×3的卷积核，在相同参数计算量下拥有5×5（dilated rate=2）或者更大的感受野，从而无需下采样。空洞卷积引入了一个成为“扩张率（dilated rate）”的新参数，该参数定义了卷积核处理数据时各值的间距，换句话说，相比原来的标准卷积，空洞卷积多了一个 hyper-parameter（超参数）称之为 dilation rate（扩张率），指的是 kernel各点之间的间隔数量。

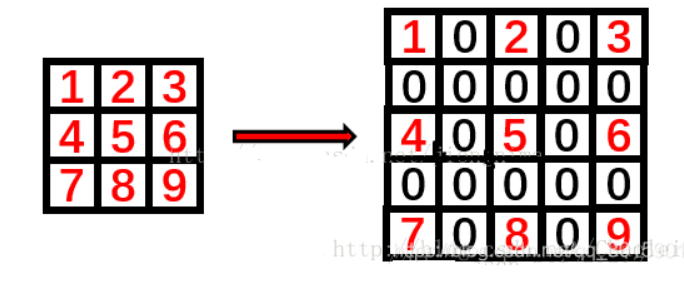


图（a）dilated rate=1，和普通卷积操作相同

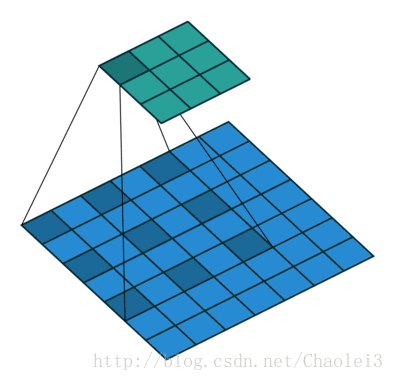
图（b）dilated rate=2，实际的卷积核大小还是3×3，空洞为1，所以卷积核大小变成了5×5

图（a）dilated rate=4

空洞的位置全填进去0，填入0之后再卷积即可，例如下图：



空洞卷积动态过程：



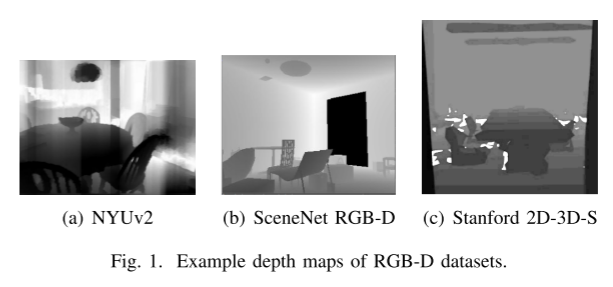
## 2. 数据集

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 数据集（年份） | 分辨率 | 样本数 | Object Classes | Scene Classes | Data Type | 备注 |
| NYU-V1  (2011) | 640 × 480 | 2347 | 13 | 7 | RGB, D | 2D |
| NYU-V2  (2012) | 640 × 480 | 1449 | 894, 40, 4 | 26 | RGB, D | 2D |
| SUNRGB-D  (2015) | Variable | 10335 | 800, 37 | 47 | RGB, D, Pc | 2D, 3D |
| SceneNet  (2016) | 320 × 240 | 5M | 255 | 5 | RGB, D | 2D |
| Stanford2D-3D-S  (2017) | 1080 × 1080 | 70496 | 13 | 11 | RGB, D, Pc | 2D, 3D, P |
| Matterport3D  (2017) | 1280 × 1024 | 194400 | 50811, 40 | 61 | RGB, D, Pc | 2D, 3D, P |

用于语义分割的室内RGB-D数据集综述，[Data Type: **RGB**, **D**epth and **P**oint **c**loud. Annotation: 2D, 3D, and **P**anorama annotation mask.]

当前数据集存在以下问题：

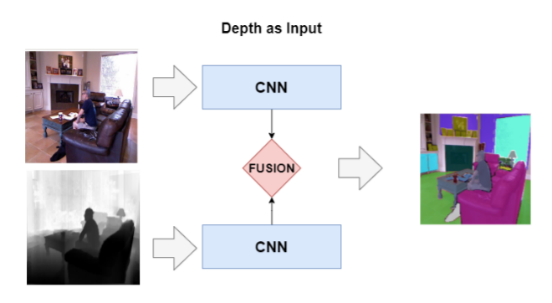
1. 类别不平衡：例如，“墙”“地板”的标签几乎覆盖了整个数据集，导致一些较少的对象的表现性能较差（例如，“电视”）；
2. 收集到的图片深度特征不准确，提供了具有许多伪影的非平滑深度图；
3. 一些数据集中可用图像的数量有限，不适合深度学习等数据密集型机器学习算法。



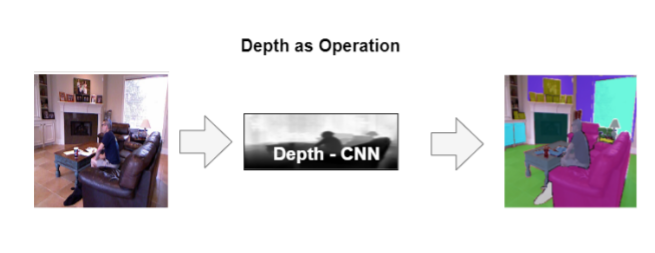
## 3. 深度信息

深度提供了额外的几何信息，可以使RGB语义分割模型受益。根据将深度特征融入DCNN的方式，分为三种：

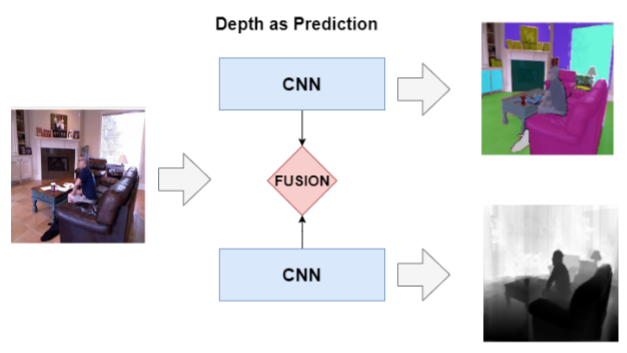
1. 深度作为输入：使用深度图作为RGB图像的附加输入，以便提取更多特征。深度图和RGB图像分别输入到DCNN的分支，然后融合所提取的特征以产生分割掩模。



1. 深度作为操作：修改DCNN中的一些操作，例如卷积和池化，以考虑深度信息。取代使用深度图作为输入，直接相对于深度修改DCNN的操作。



1. 深度作为预测：设计一个DCNN，从RGB图像中预测分割图像和深度图，以这种方式，模型利用辅助深度信息提取互补几何信息，可以将两个任务相关特征合并在一起以改进两个预测，包括目标分割任务。



### 4． 性能分析

语义分割的室内应用通常运行在低功耗器件上，因此需要轻量化和快速的模型，为了解决这一问题， “Depth as Operation”是一个很好的解决方案，因为其在DCNN中合理地利用了深度信息。