图像分割

2022年4月9日　17:23

前期资料调查

网课：

区别 语义分割：每个像素打上标签（只区分类别不区分具体单位）；

实例分割：不仅要区分类别，还要区分个体

python：Deep-learning-of-machine-vision-main

matlab：[十六、数字图像处理之图像分割_Liaojiajia-2020的博客-CSDN博客_图像分割]

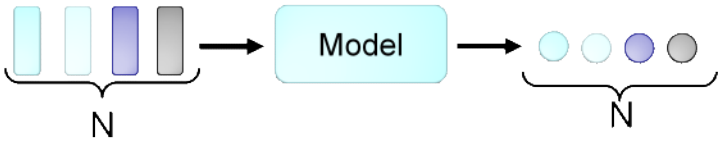
基于卷积神经网络的图像分割——CNN

理论

常用方法原理：
transformer运用于语义分割
后边很多模型是基于transformer进行模改，因此注意一些transformer的细节部分：

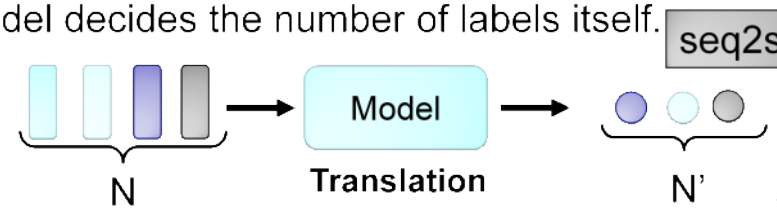
Sequence to Sequence：NLP machine translation
3 models：

- Each vector has a label.

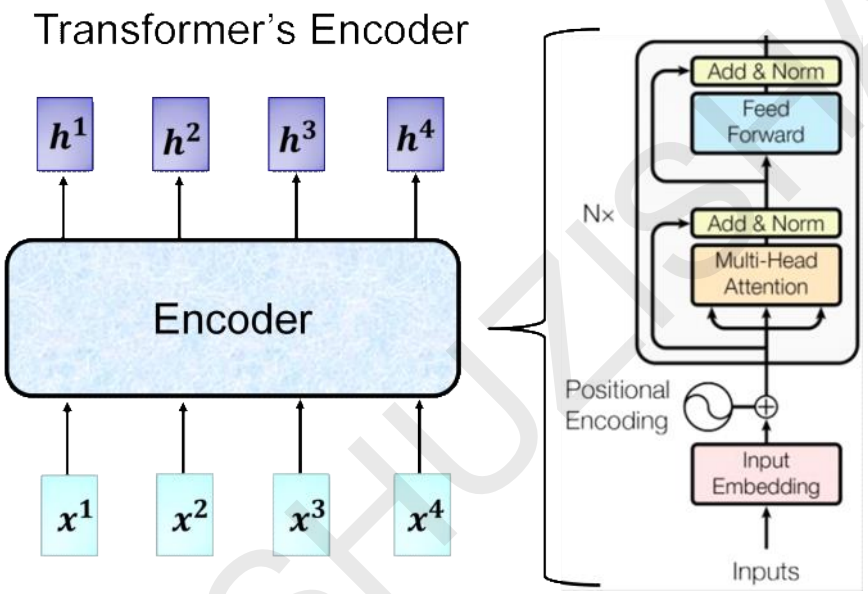


- The whole sequence has a label.

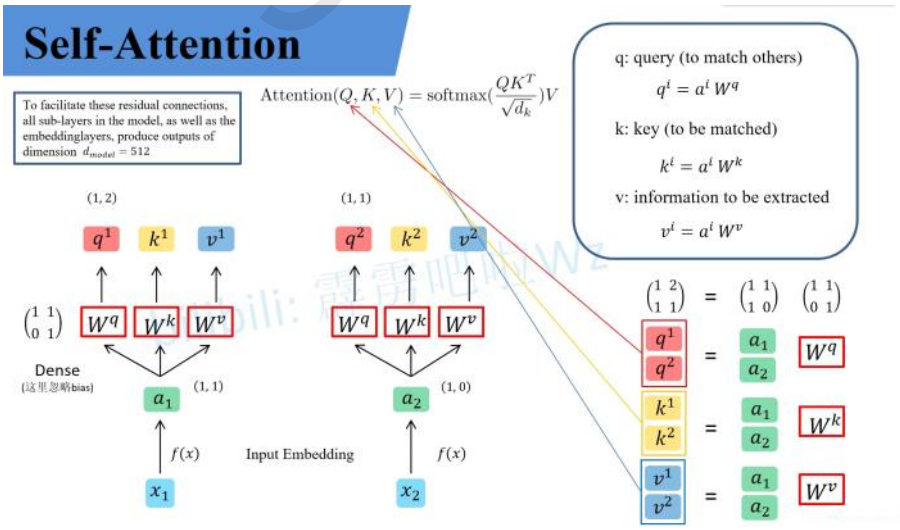
- Model decides the number of labels itself.



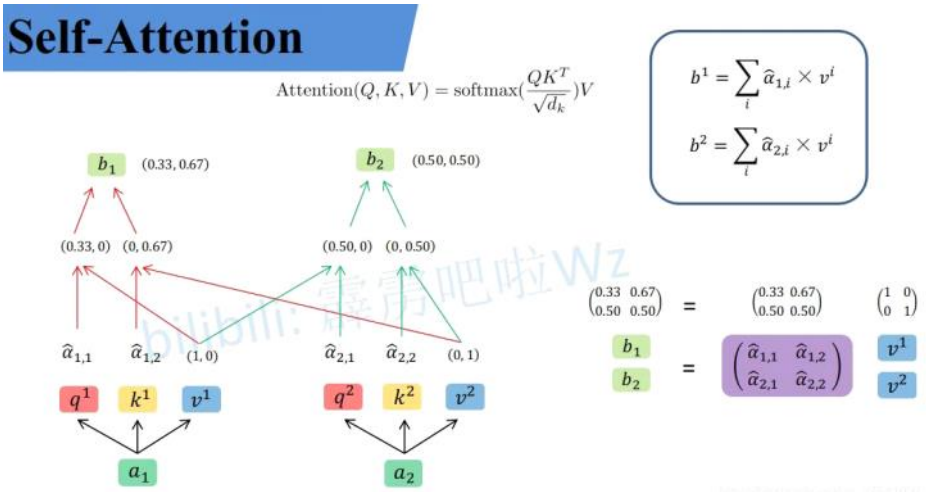
I like learning->我喜欢学习（不定长，所以需要sequence to sequence）



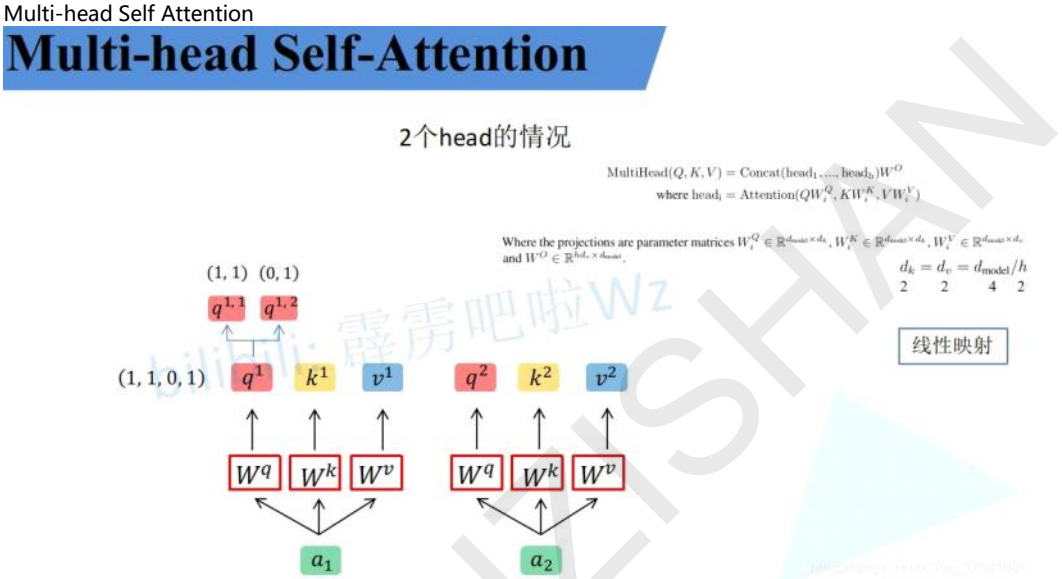
Self-attention：RNN，LSTM

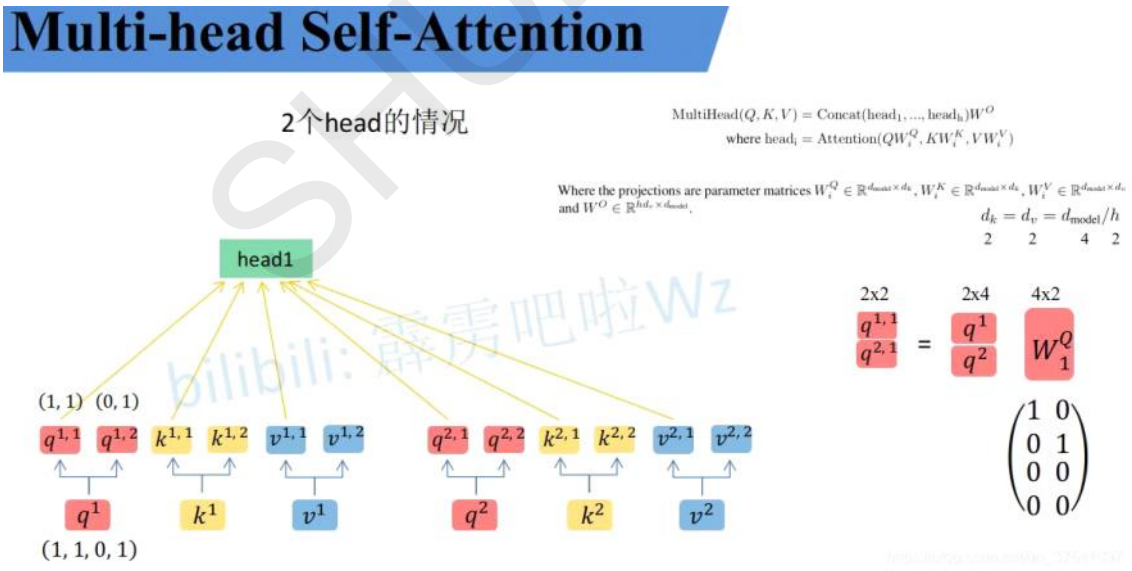


点乘——矩阵乘法



Multi-head Self Attention

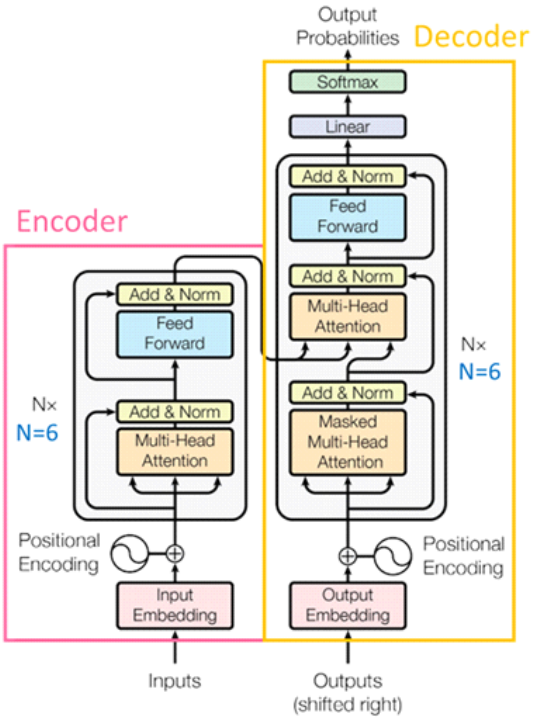




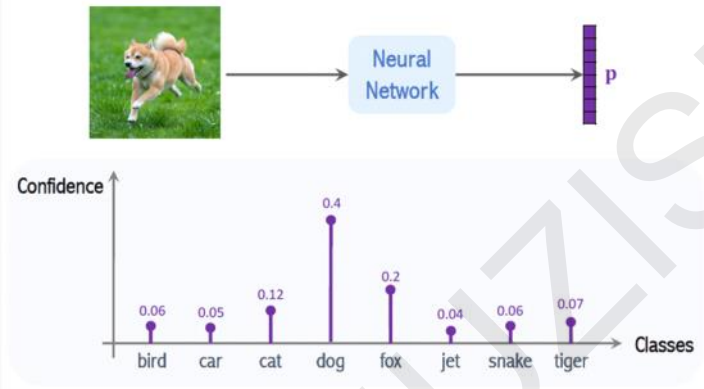
Query 由一个 Query 变成两个 Query，或者说更多的 Query，有不同类型的关联度

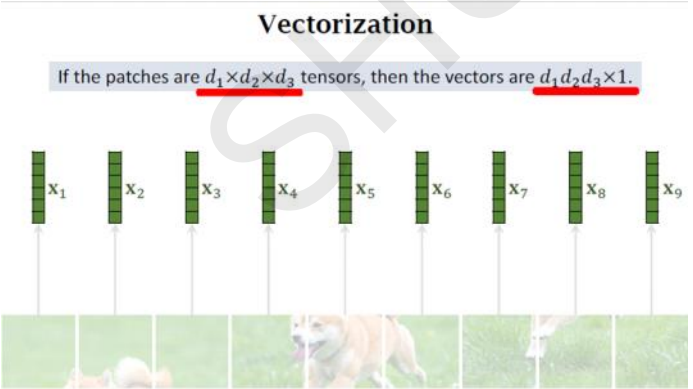
Multi-head attention allows the model to jointly attend to information from different representation subspaces at different positions.

Encoder Decoder

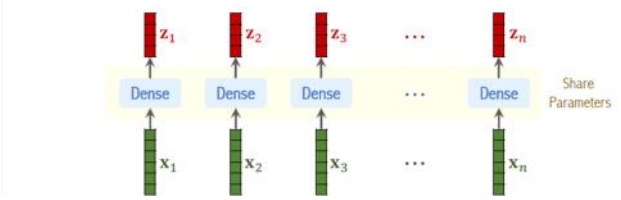


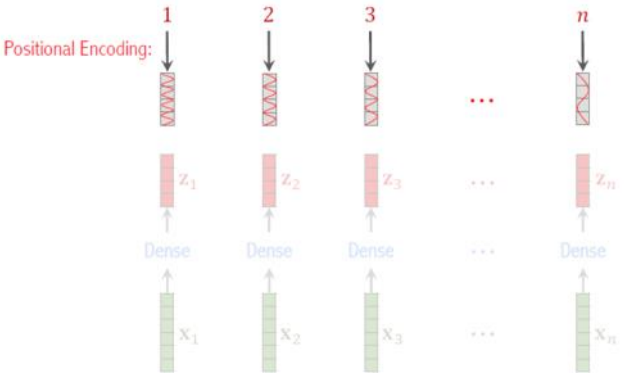
Transformer architecture适用于NLP，在CV领域则有了vision transformer；
Vision transformer主要用来做 image classification的



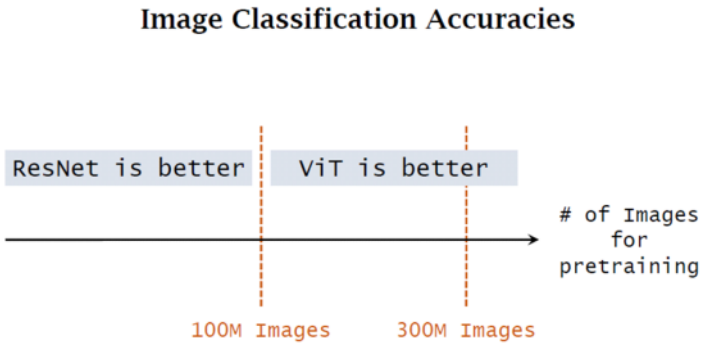


image矩阵拉伸成victor

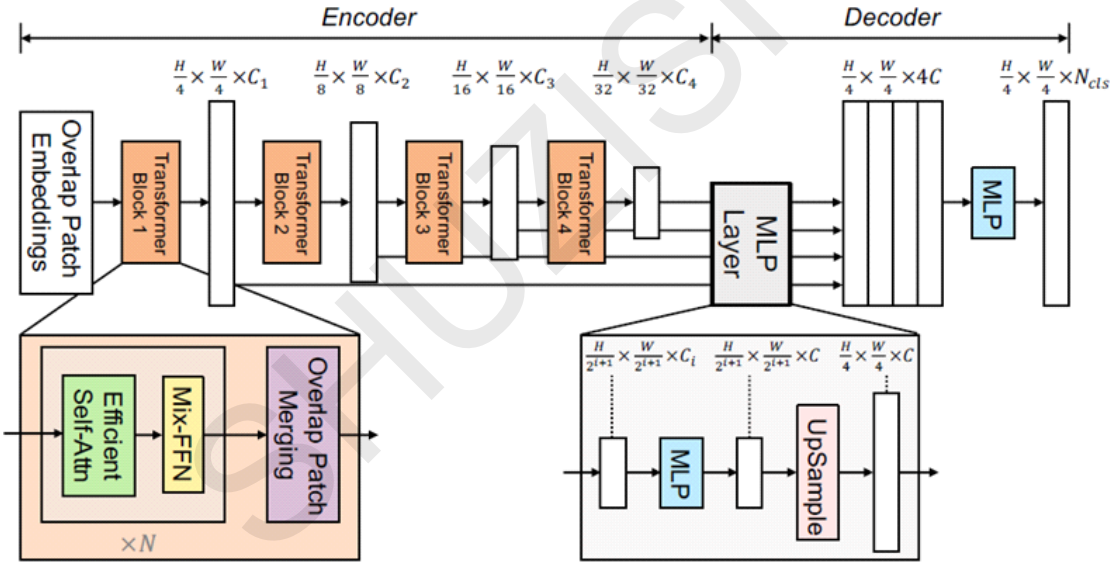




Positional encoding：transformer对位置本身没有感知，因此要加一个Positional encoding配对位置信息。



Segformer（2021）

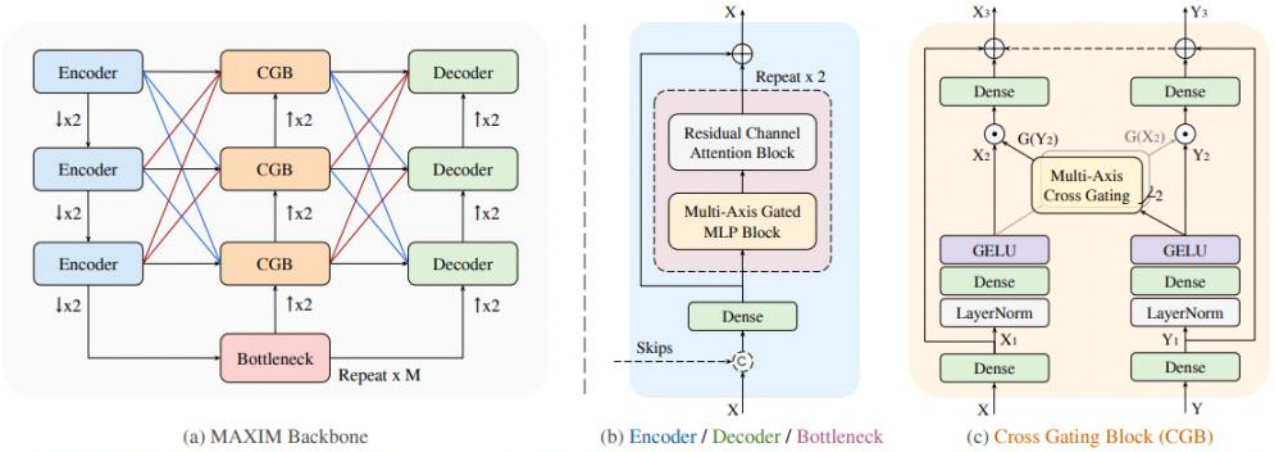


Code will be released at: github.com/NVlabs/SegFormer.（跑不了，缺了一个文件夹）

MAXIM: Multi-Axis MLP for Image Processing（2022）

论文地址：arxiv.org/abs/2201.02973

代码/模型/实验结果：https://github.com/google-research/maxim



(a) MAXIM Backbone (b) Encoder / Decoder / Bottleneck (c) Cross Gating Block (CGB)

Figure 2. **MAXIM architecture.** We take (a) an encoder-decoder backbone with each (b) encoder, decoder, and bottleneck containing a multi-axis gated MLP block (Fig. 3) as well as a residual channel attention block. The model is further boosted by (c) a cross gating block which allows global contextual features to gate the skip-connections. More detailed description can be found in Appendix A.2.

新方法太多了，个人觉得现阶段算法创新还是有难度的

论文摘取

《基于matlab的图像分割算法研究及实现》

算法设计1：基于边缘检测到图像分割法

（监测点必须是局部极值点！而非零点）

基于一阶导数边缘检测算子：Roberts,Sobel,Prewitt

基于二阶导数边缘检测算子：Laplacian,Wallis,LOG,Canny（明显）

【二阶微分关心是图像灰度的突变而不强调灰度缓慢变化的区域，因此对边缘的定位能力更强】

【Laplace算子具有旋转不变性，因此存在缺点：1.没有边缘的方向信息；2.双倍加强了噪声影响】 图像边缘检测——一阶微分算子Roberts、Sobel、Prewitt、Kirsch、Robinson（Matlab实现）_ChuanjieZhu的博客-CSDN博客_一阶微分算子

因此出现了LOG算子（先进行高斯滤波再用Laplace算子算Δ，一阶导数峰值则为Laplace过零点，对过零点的精确位置进行插值估计

DOG算子：LOG算子图像进行两次高斯平滑再相减

Canny：只关注边缘法向有大变化的点，图像内容驱动 图像边缘检测——二阶微分算子（下）Canny算子（Matlab实现）_ChuanjieZhu的博客-CSDN博客_二阶边缘检测算子

算法设计2：基于阈值的图像分析算法

双峰法，迭代阈值。

（前边的算子内置算法就有用到，不多赘述）

《医学图像分割方法综述》

1.传统方法

阈值法（适用于只有目标和背景两大类的医学图像）

区域生长法（分割具有相同特征的连通区域效果较好，由于噪声和灰度不均，易产生空洞和过度分割）

阈值+区域：1.结合区域生长和水平集算法实现宫颈病灶图像分割

2.肝脏肿瘤 CT 图像进行直方图均衡化、中值滤波等预处理，再运用混合滤波策略的区域生长算法实现对肝脏肿瘤的有效分割

边缘检测法（检测区域边缘）

缺点：1.不能保证边缘的连续性和封闭性；2.在高细节区容易出现大量碎边缘，难以形成一个大区域。

聚类法（相似灰度合并）

常见方法：K 均值、模糊 C-均值算法（FCM）、参数密度估计

FCM：多数医学图像具有模糊性、图像质量低等特性，所以他是医学分割领域最常用的聚类算法。无监督算法，在一定程度上缓解了医学图像分割标签少的问题。

原理：模糊集理论+聚类算法：1.人工随机指定每个数据到各个聚类（簇）的隶属度；2.根据隶属度计算每一个簇的质心；3.接着重新进行伪划分；4.直到质心不变

改进的局部自适应模糊 C-均值算法进行肺结节分割，并验证了算法在利用肺部图像的邻域信息和灰度信息上的有效性。采用粒子群算法和遗传算法相结合的优化算法来确定初始类聚中心，再引入像素的邻域信息，克服噪声对异常值敏感的问题。

三个切入思路：

1.优化某类算法；

2.基于某类型图像（or某种病理图）做其图像识别综述并（或）优化其算法；

3.基于某类型图像（or某种病理图）对比其图像识别各个方法的异同作综述。

2.深度学习方法

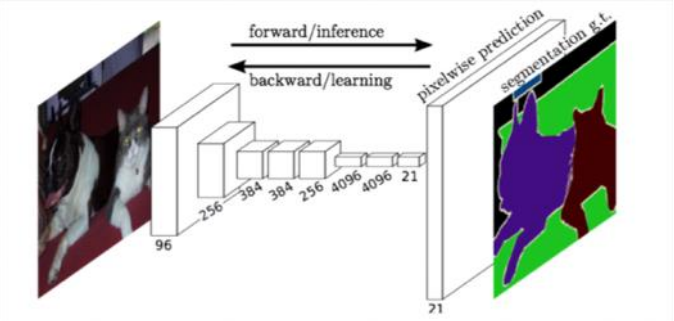
全卷积神经网络（FCN）

[全卷积网络FCN进行图像分割_大村chen的博客-CSDN博客_全卷积网络 图像分割]

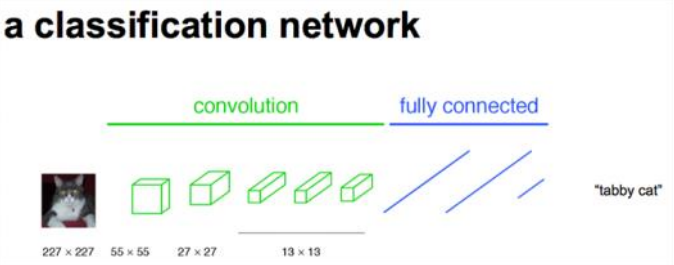
CNN：能对物体分类（缺点：1.存储开销很大。2.计算效率低下。3.像素块大小限制了感知区域的大小。）

FCN：能识别特定部分的物体（缺点：1.得到的结果还是不够精细。上采样的结果还是比较模糊和平滑，对图像中的细节不敏感。只采用一次上采样操作。2.对各个像素进行分类，没有充分考虑像素与像素之间的关系。忽略了在通常的基于像素分类的分割方法中使用的空间规整（spatial regularization）步骤，缺乏空间一致性。）

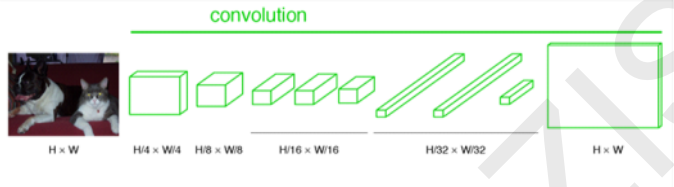
FCN结构示意图：



CNN分类网络示意图：



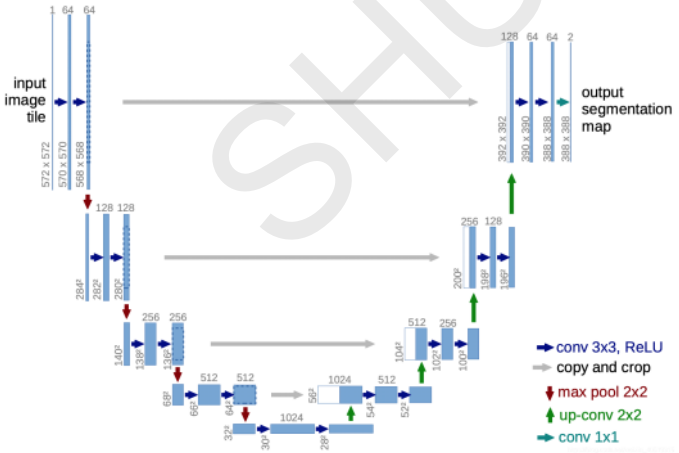
FCN卷积后示意图：



FCN用的是加操作（summation），U-Net用的是叠操作（concatenation）

U-net网络结构

[论文解读U-Net+与FCN的区别+医学表现+网络详解+创新_薛定谔的炼丹炉！的博客-CSDN博客]



U-Net对称性好，且FCN的decoder相对简单，只用了一个deconvolution的操作，之后并没有跟上卷积结构。

《基于深度学习的语义分割综述》

传统方法：阈值分析法、边缘检测法、区域法、马尔可夫随机场模拟算法

马尔可夫随机场

卷积神经网络语义分割算法：FCN、PSPNet、U-Net、DeepLab 系列

PSPNet：

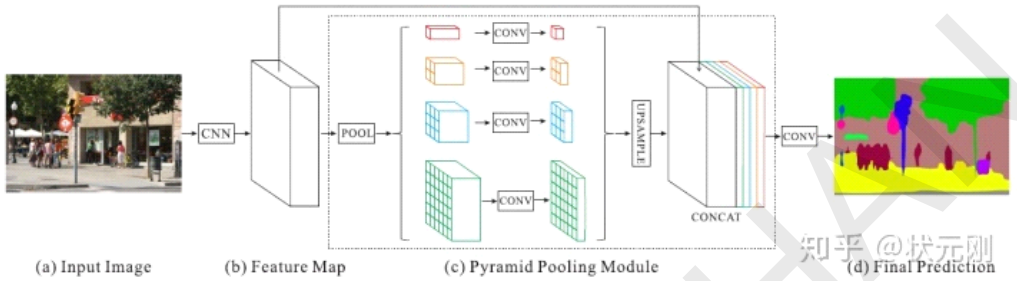
https://zhuanlan.zhihu.com/p/72845837

FCN基础上加上

1、增大分隔层的感受野。
空洞卷积（dilated convolution）：这是在deeplab算法上成功应用的实现方式
全局均值池化操作：PSPNet的全局均值池化操作也是增加感受野的一种方式
2、深层特征和浅层特征的融合，增加浅层特征的语义信息，这样在浅层进行分割时就有足够的上下文信息，同时也有目标的细节信息，这种做法早在FCN中就有了，但是包括融合策略和分割层的选择都有一定的优化空间。

SPP即空间金字塔池化（Spatial Pyramid Pooling）主要做图像分类和目标检测
PSP即金字塔场景解析（Pyramid Scene Parsing）主要做语义分割，还需要上采样，进行逐像素的分类，讲究对不同的场景进行解析

PSPNet模块：
1.金字塔池化模块
全局平均池化（GAP global average pooling）:通常用于图像分类任务
金字塔池化（SPP）：全局先验模块是为消除CNN进行图像分类时需输入固定尺寸图像的这一约束而设计的。为了进一步避免丢失表征不同子区域之间关系的语境信息，提出了一个包含不同尺度、不同子区域间关系的分层全局信息。将该金字塔池化模块的输出作为深度神经网络最终的特征图，并称其为全局场景先验信息。
2.PSPNet网络结构



PSPNet为像素级场景解析提供了一个有效的全局语境信息，会比全局池化所得的全局信息更具代表性。
在计算成本方面，PSPNet并没有比原来的扩展FCN网络增加很多

DeepLab 系列
https://blog.csdn.net/fanxuelian/article/details/85145558

数据集：
PASCAL VOC2012 数据集
PASCAL VOC2012 是 Pascal 系列语义分割方向最常用的数据集
训练集，验证集，测试集分别为1464，1449，1456张，经过数据增强后的数据集其训练集达到了10582张。
有 20 个类别和 1 个背景类。
图片包括原图 jpg 格式，图像分类分割 png 格式和图像物体分割 png 格式
图像分类分割中，物体分割颜色是特定，而图像物体分类分割中，图像生成由不同物体生成不同轮廓，颜色随机填充。
VOC2012 数据集文件夹包含 5 个文件夹，Annotations 文件夹对应其图像的 xml 信息，ImageSets 文件夹中 的 Segmentation 的三个 txt 文件为标记的用于图像分割的图 像，JPEGImages 文件夹为原图,SegmentationClass 文件夹包含 png 图像用于图像分割分类,SegmentationObject 文件夹的 png 图为不同物体的分割。

Microsoft COCO 数据集
Microsoft COCO 数据集用于场景理解的任务中，图像来 源于复杂背景下的生活场景。COCO 数据集有 91 个物体类别，它含有 32.8 万张图片，标注实例有 250 万个，是目前为止 最大的语义分割数据集。这个大规模数据集专注于解决图像场景理解中的三个关键问题：目标分类，目标检测和场景语义标注。COCO 数据集其特点是每一张图片平均由 3.5 个类别 和 7.7 个实例组成的，评估标准比PASCAL VOC严格，大家乐意用它来测评模型质量。

Cityscapes 数据集
Cityscapes数据集是对城市街道场景的语义理解，从 50 多个不同的城市中根据不同的季节，良好的天气情况手动选出视频中的帧，由大量动态对象，各种场景布局和变化的背景所记录产生的大规模数据集，有 5000 张高质量标注的图片，2 万 多张粗糙的标记图片。它提供了 8 种类型 30 个类别的语义化，实例化和密集像素的标记，8 种类型为平面、人、车辆、建 筑、物体、自然、天空、虚空，用于训练深度神经网络。

Github 、kaggle上面有很多各方面的开源训练集和代码，我们选题的时候也可以参考这些。

《基于深度学习的图像语义分割技术综述》

Table 2 Semantic segmentation methods based on deep learning

表 2 基于深度学习的语义分割方法

年份	事件	结构创新点
2015	Jonathan Long 等提出 FCN	创造性地使用上采样代替全连接层且接受任意大小的输入图片
2015	Olaf Ronneberger 等提出 U-Net	网络框架采用左右对称的 U 形字母结构，其池化层被上采样层取代，可运行小批量图片，在医疗图像处理中有更大作用
2015	Badrinarayanan 等提出 SegNet	基于编码器-解码器网络结构，利用上采样方式恢复图像尺寸，去掉全连接层，大大减少了参数，提升了网络运行速度
2017	Zhao 等提出 PSPNet	使用空洞卷积改善 ResNet 结构，并添加了一个金字塔池化模块
2017	Lin 等提出使用链式残差连接的 RefineNet	对解码器结构进行改进，形成 Long-range 残差连接，能通过上采样方式融合底层和高层语义特征
2018	Chen 等提出 DeepLab v3+	V3+版本使用改进版的 Xception 作为基础网络，加强了图像边缘分割效果
2018	YU 等提出双向分割网络 BiSeNet	包含 Spatial Path 和 Context Path，分别用来解决空间信息缺失和感受野缩小的问题
2019	何恺明等提出全景 FPN	将 FCN 和 Mask R-CNN 相结合，使用丰富的多尺度特征，可同时解决语义分割与实例分割任务

这个综述多少有些敷衍了...

《全卷积神经网络图像语义分割方法综述》

全卷积神经网络图像...

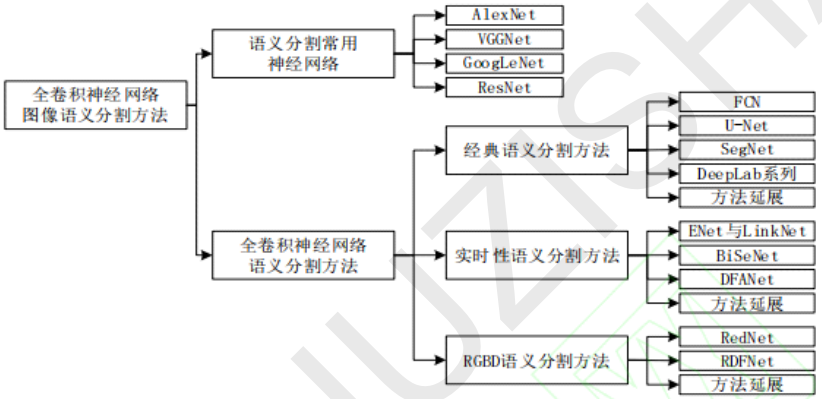


图 1 全卷积神经网络图像语义分割方法分类

Fig.1 Classification of semantic segmentation methods for fully convolutional neural network images

典型图像实例分割算法分类：

基于二阶段目标检测的实例分割

基于直接掩码生成的实例分割

基于一阶段目标检测的实例分割

《图像实例分割综述》

不仅具有语义分割的逐像素分类特点，同时具有目标检测的特点，即定位图片中所有不同实例并分配给它们各自掩码 。

表 1　主流方法比较

网络名称	出现年份	测试数据集	算法精度
SDS	2014	PASCAL VOC	$mAP^{0.5}$:52.6
MNC	2015	PASCAL VOC	$mAP^{0.5}$:63.5
InstanceFCN	2016	PASCAL VOC	$mAP^{0.5}$:61.5
FCIS	2017	MS COCO	AP:29.2
Mask R-CNN	2017	MS COCO	AP:37.1
PANet	2018	MS COCO	AP:36.6
YOLACT	2019	MS COCO	AP:29.8
PolarMask	2019	MS COCO	AP:32.9
RDSNet	2019	MS COCO	AP:36.4
MS R-CNN	2019	MS COCO	AP:39.6
SOLO	2019	MS COCO	AP:40.4
BlendMask	2020	MS COCO	AP:41.3

常用数据集：还是上边那三个

常用性能评价标准：

Precision-Recall（P-R）曲线：Precision代表了 查准率，recall代表查全率，当查准率和查全率发生变化则可以绘制出P-R曲线

平均精度（Average-Precision，AP）：AP即 为P-R曲 线下面的面积

平均精度均值(Mean Average Precision，MAP)： 将各个类别的AP值取平均值就即可得到MAP值

均交并比(Mean Intersection over Union，MIoU)： 真实值与预测值两个集合的交并比

实例分割未来探索：（也可以为我们的研究方向提供参考）

轻量化；3D图像实例分割；弱/无监督（减少算法对人工标注的依赖）

《基于深度学习的实例分割研究综述》

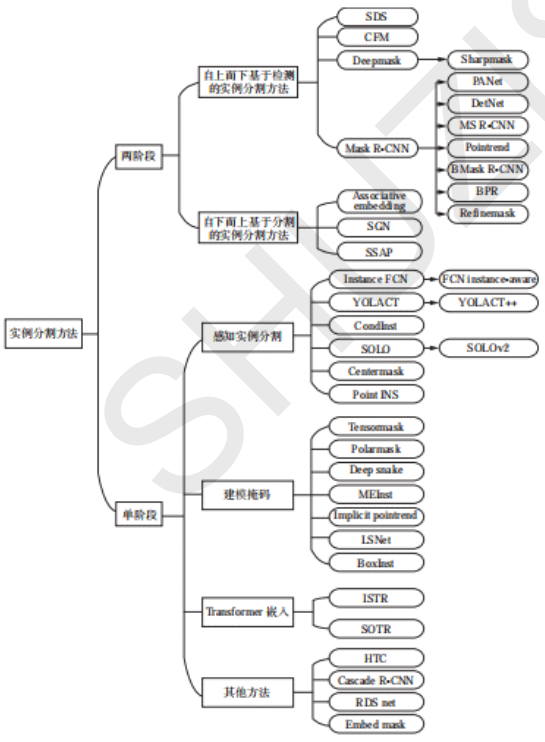


图 1　本文涉及的实例分割方法

Fig. 1　Paper focuses on the instance segmentation methods

补充：

数据集：

Mapillary Vistas 数据集

新建立的，大场景的街景数据集，用于图像语义分割以及图像实例分割，旨在进一步开发用于视觉道路场景理解的先进算法。

与 Cityscapes 相比，Mapillary Vistas 的精细注释总量大了 5 倍，并包含来自世界各地在各种条件下捕获的图像，包括不同天气，季节和时间的图像。

LVIS 数据集

Facebook AI Research 于 2019 年建立的大型词汇实例分割数据集。目前公布的实例分割数据集的目标类别还是较少，与实际应用场景下存在大量（未知）类别相违背。

相比于 COCO数据集，LVIS 人工标注掩码具有更大的重叠面积和更好的边界连续性，更精确的掩码。并且在数据成长尾分布 (类别种类多而单类的实例个数少) 时仍有很好的训练效果。

虽然我对实例分析算法的了解较为浅薄，但单从这几篇综述性文章来看，我们可以利用已有的评价指标测试某类场景（或者特定的数据量/数据类型）各个模型的准确度。