目录

[人工智能在医学影像中的研究与应用 1](#_Toc183129331)

[2.1 医学成像系统简介 2](#_Toc183129332)

[CNN 复习 3](#_Toc183129333)

[卷积层&padding 5](#_Toc183129334)

[池化层 6](#_Toc183129335)

[全连接层/输出层 7](#_Toc183129336)

[图像分割算法综述总结 8](#_Toc183129337)

[什么是UNet 9](#_Toc183129338)

[Unet定义 9](#_Toc183129339)

[UNet本质 10](#_Toc183129340)

[什么是nnUNet 13](#_Toc183129341)

[重心 13](#_Toc183129342)

[网络结构 13](#_Toc183129343)

[Data analysis 13](#_Toc183129344)

[Data Preprocessing 14](#_Toc183129345)

[Augumentation 15](#_Toc183129346)

[Model training 16](#_Toc183129347)

[Model architecture 17](#_Toc183129348)

# 人工智能在医学影像中的研究与应用

大数据, 2019, 5(1): 2019004- doi: 10.11959/j.issn.2096-0271.2019004

韩冬1, 李其花2, 蔡巍3, 夏雨薇2, 宁佳1, 黄峰1

**人口老龄化问题的显现以及人们对健康与日俱增的要求，对目前有限的医疗资源和医疗技术提出了更大的挑战。（why 医疗影像）**医疗领域亟需新的技术满足这些需求。与此同时，国内外与医疗相关的[人工智能](https://www.infocomm-journal.com/bdr/article/2019/2096-0271/2096-0271-5-1-00039.shtml)技术也在飞速地发展，科研和创业项目如雨后春笋，为解决医疗领域的挑战提供了新的机遇。目前已经出现了计算机辅助诊断、智能专家系统、手术机器人、智能药物研发以及健康管理等多种产品。

在众多的医疗信息中，医学影像是疾病筛查和诊断、治疗决策的最主要的信息来源。基于医学影像的诊断和治疗是一个典型的长链条、专业化的领域，涵盖了医学影像成像、图像处理与分析、图像可视化、疾病早期筛查、风险预测、疾病辅助检测与诊断、手术计划制定、术中辅助导航、随访跟踪与分析、康复计划制定等一系列方向。目前，医院存储的信息超过90%是影像信息，影像信息已经形成了巨大的数据积累。为此，基于医学影像大数据的[人工智能](https://www.infocomm-journal.com/bdr/article/2019/2096-0271/2096-0271-5-1-00039.shtml)技术与应用就成为医疗机构、科研、产业和政府共同关注的焦点。

医学影像链可以分为成像和图像挖掘两部分。首先，作为信息源头的医学成像设备，其成像质量会对后续疾病的检测、诊断与治疗起到至关重要的作用。利用AI技术可以实现医学影像成像质量的提升，AI优化的扫描工作流可以显著提高扫描效率，并使成像质量趋于标准化，从而给整个医疗健康链条带来深远的影响，具有重要的临床与科研价值。

其次，**理解医学图像、提取其中具有诊断和治疗决策价值的关键信息**是诊疗过程中非常重要的环节。AI辅助诊断可以承担繁琐的病灶筛查工作，迅速地从海量数据中提取出与诊断相关的有价值的信息，同时避免人工阅片带来的主观性差异。AI辅助图像处理算法还可以迅速地完成分割配准等复杂功能，为用以治疗的医疗设备（例如手术导航和手术机器人）提供[精准](https://www.infocomm-journal.com/bdr/article/2019/2096-0271/2096-0271-5-1-00039.shtml)的病灶结构信息。

此外，目前AI在医学影像方面的应用还存在着诸多挑战，比如扫描成本和患者隐私问题使得医疗大数据难以被获得；医学图像的标注者需要具有一定的医学背景，获得高质量的医学图像标注甚至比获得医学图像代价更高；不同病变类型与正常的医学图像在数据量上的差距巨大。一些基于AI的非结构医学信息数据分析方法，可以结合影像和医生撰写的诊断报告，自动或者半自动地生成标注数据，扩充训练数据集。

### 2.1 医学成像系统简介

临床中常用的医学影像模态有多种，包括电子计算机断层扫描（computed tomography,CT）、磁共振成像（magnetic resonance imaging,MRI）、正电子发射计算机断层显像（PET-CT）、X射线、超声等。不同模态的医学影像适合观察不同类型的生理病理信息。医学影像设备采集人体中与特定生理病理信息相关的物理信号，并依据信号传播的数学物理模型重建生理病理信息精确的二维、三维空间分布（即医学图像）。CT图像反映的是人体内不同组织对特定谱段X射线的衰减系数分布情况，由于衰减系数与物质密度直接相关，因此CT图像间接地反映出人体组织的三维密度分布。MRI可以反映人体中质子的密度、纵向弛豫时间（T1）、横向弛豫时间（T2）、质子扩散等多种对比度。PET图像可以反映出人体内氟代脱氧葡萄糖（18F-FDG）的代谢分布情况。

三维医学图像成像过程属于逆问题（inverse problem）范畴，即通过在体外采集到的物理信号，重建能够产生该观测信号的人体内生理病理信息的分布情况。这一过程通常是不稳定的，其原因可能有多种，例如，由于扫描时间和剂量的限制而造成的采集信号量不足、原始采集信号中存在着固有噪声、原始信号中掺杂着其他干扰信号、成像设备硬件的不完美性、患者在成像过程中不自主的随机运动等。这些因素使得基于理想数学物理模型的解析重建方法通常无法得到满足临床要求的图像。为了解决这一问题，传统上需要在重建过程中加入针对图像的特定先验信息（prior information）作为约束条件，以使重建过程更加稳定。常用的约束包括图像L1范数、L2范数、全变分（total variation, TV）约束等。这些简单的约束条件具有良好的通用性，但却无法准确地反映数据的本质特征，对于特定的影像模态与成像模式无法得到最优的重建结果，因此如何设计更好的先验约束，一直是医学影像成像领域的难题之一。

近年来，AI技术的快速发展，特别是其在计算机视觉（computer vision, CV）、图像处理与分析等领域的重要突破，使得国际上的研究人员逐渐认识到将AI技术应用于医学影像成像领域的可能性。AI技术（特别是深度学习技术）摒弃了传统的人工设计的图像先验信息，采用一种完全数据驱动（data-driven）的方式，学习图像固有的深层次先验信息。这些学习到的先验信息具有更加专业化的特点，将其应用于特定领域中，效果远优于传统的浅层次先验信息。目前，AI技术与医学影像成像方法的结合已经成为领域内的研究热点，相关的研究成果呈爆发式增长。

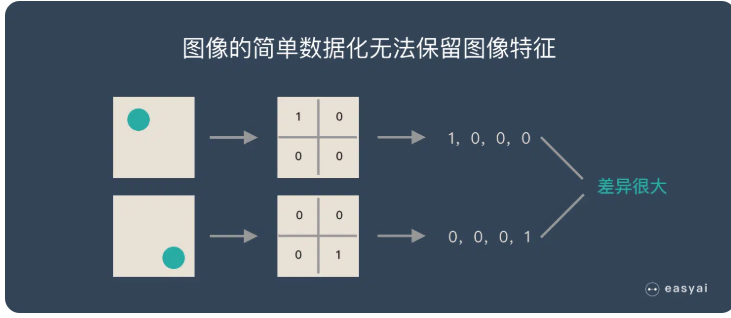
# CNN 复习

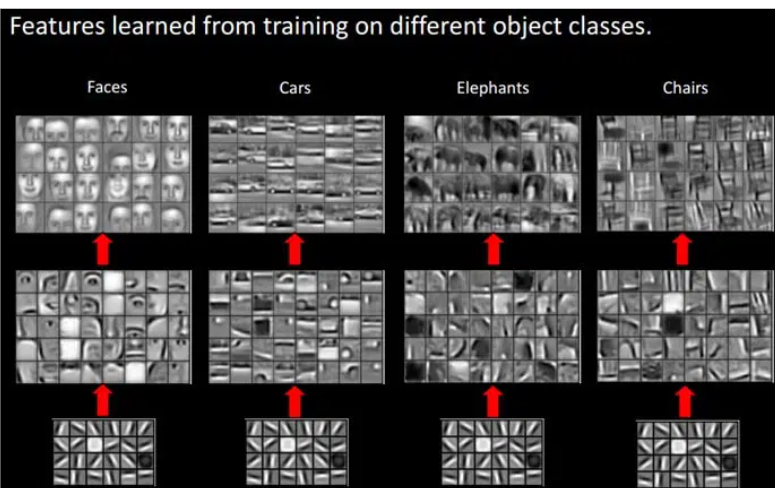
CNN 有2大特点：

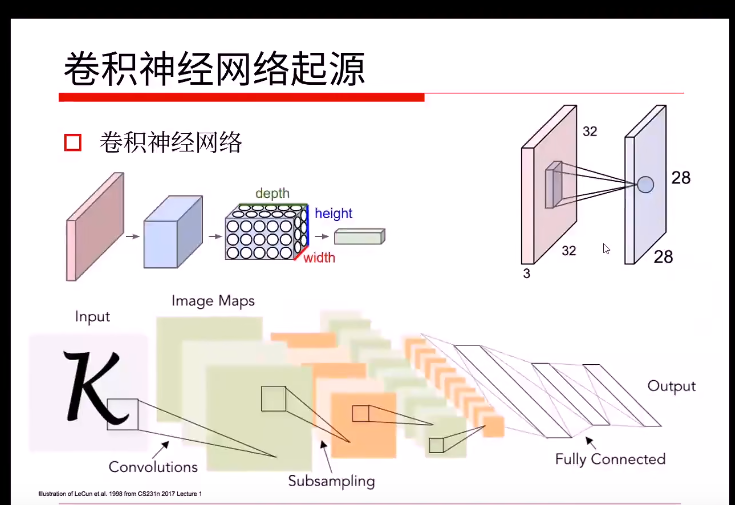
1. 能够有效的将大数据量的图片降维成小数据量
2. 能够有效的保留图片特征，符合图片处理的原则
3. 照片由nxn像素组成，每个像素有RGB三个参数表示颜色信息

灰度图像，因为其每一个像素值的范围是0~255（由纯黑色到纯白色），表示其颜色强弱程度。另外还有黑白图像，每个像素值要么是0（表示纯黑色），要么是255（表示纯白色）。我们日常生活中最常见的就是RGB图像，有三个通道，分别是红色、绿色、蓝色。每个通道的每个像素值的范围也是0~255，表示其每个像素的颜色强弱。

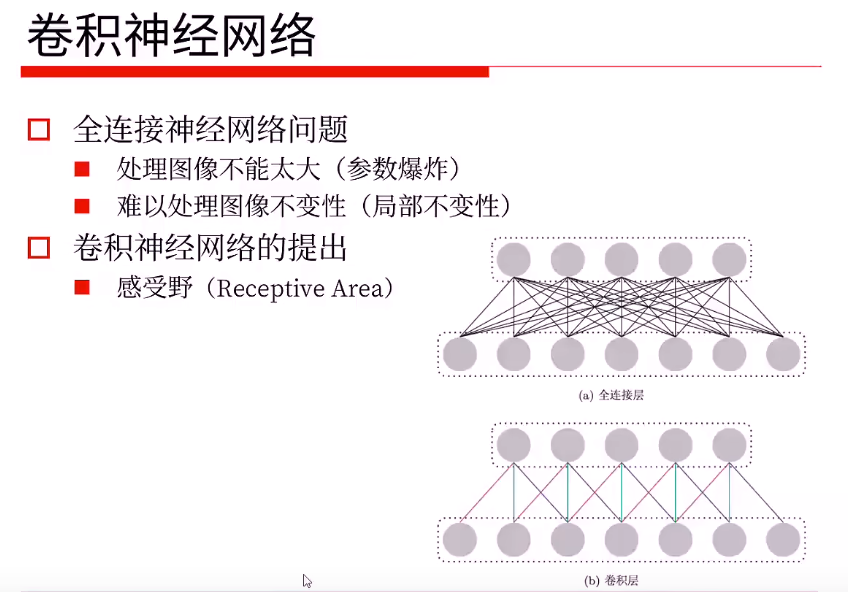
【简单问题复杂化】1000x1000 ->200x200 不影响猫狗辨别



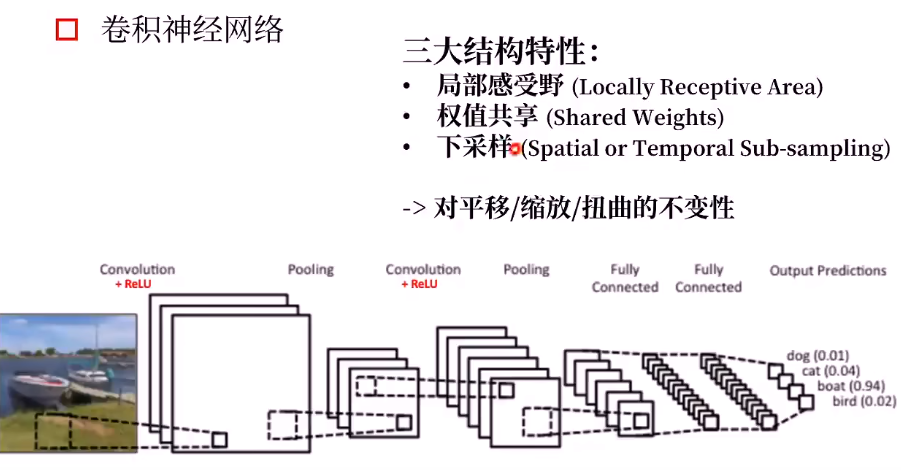




保证了空间上的不变性

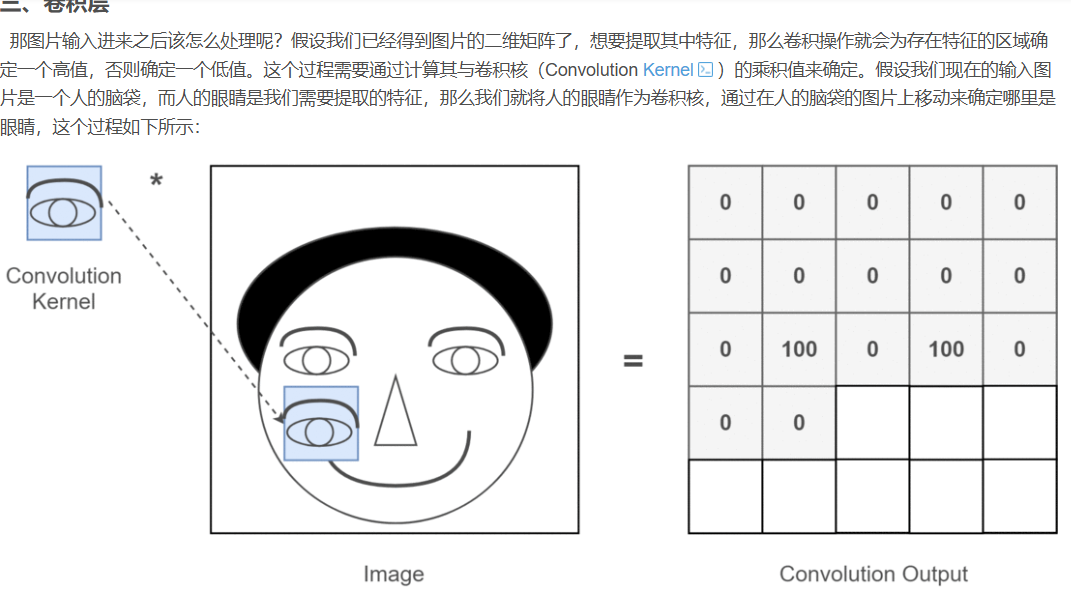


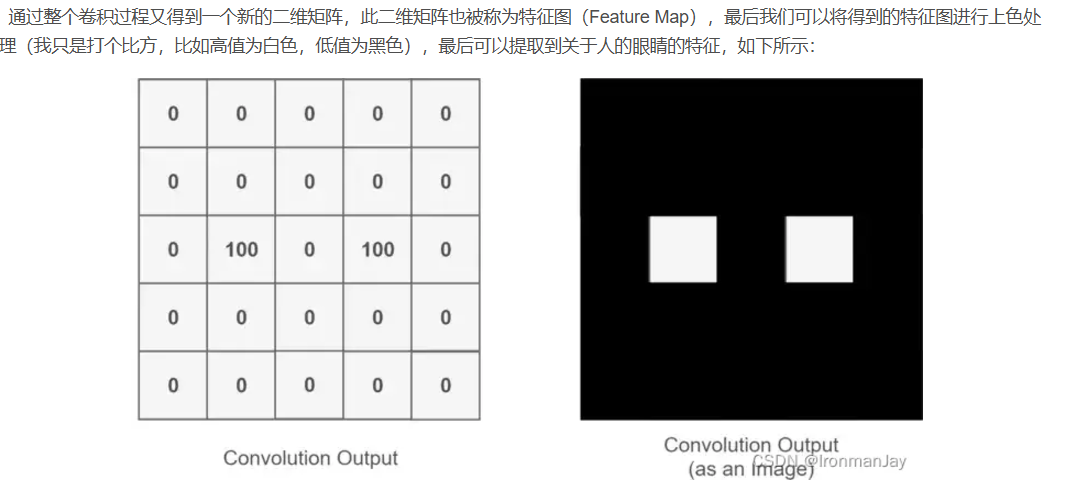




* 输入层：输入图像等信息
* 卷积层：用来提取图像的底层特征
* 池化层：防止过拟合，将数据维度减小
* 全连接层：汇总卷积层和池化层得到的图像的底层特征和信息
* 输出层：根据全连接层的信息得到概率最大的结果

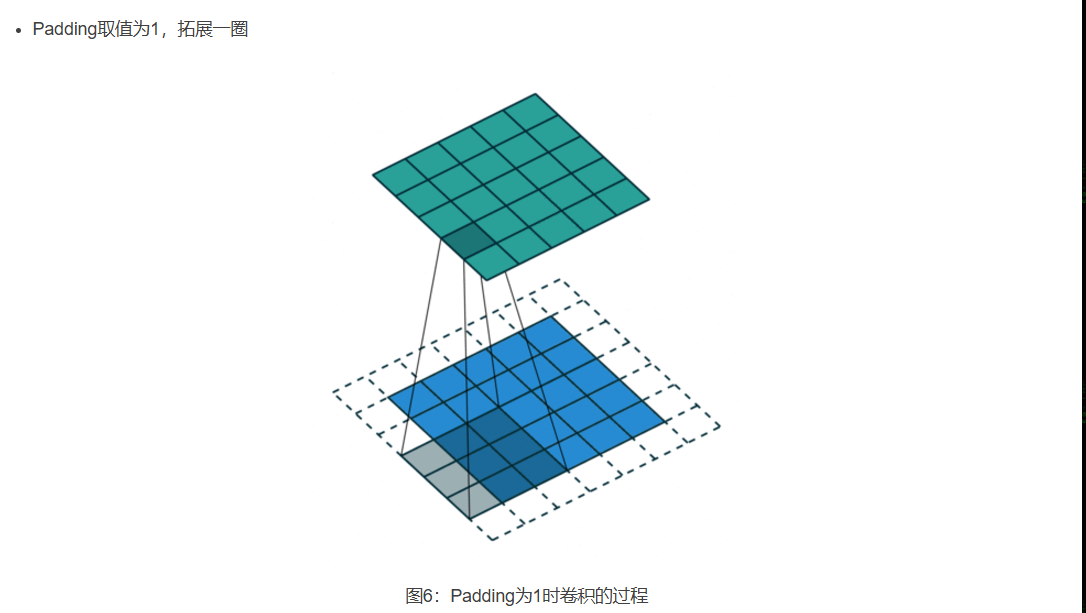
## 卷积层&padding





卷积核计算得到的新的二维矩阵称为特征图，比如上方动图中，下方移动的深蓝色正方形就是卷积核，上方不动的青色正方形就是特征图。

让我们仔细思考，如果每次计算的时候，边缘只被计算一次，而中间被多次计算，那么得到的特征图也会丢失边缘特征，最终会导致特征提取不准确，那为了解决这个问题，我们可以在原始的输入图像的二维矩阵周围再拓展一圈或者几圈，这样每个位置都可以被公平的计算到了，也就不会丢失任何特征，此过程可见下面两种情况，这种通过拓展解决特征丢失的方法又被称为Padding。



## 池化层

有几个卷积核就有多少个特征图，现实中情况肯定更为复杂，也就会有更多的卷积核，那么就会有更多的特征图，当特征图非常多的时候，意味着我们得到的特征也非常多，但是这么多特征都是我们所需要的么？显然不是，其实有很多特征我们是不需要的，而这些多余的特征通常会给我们带来如下两个问题：

1. 过拟合
2. 维度过高

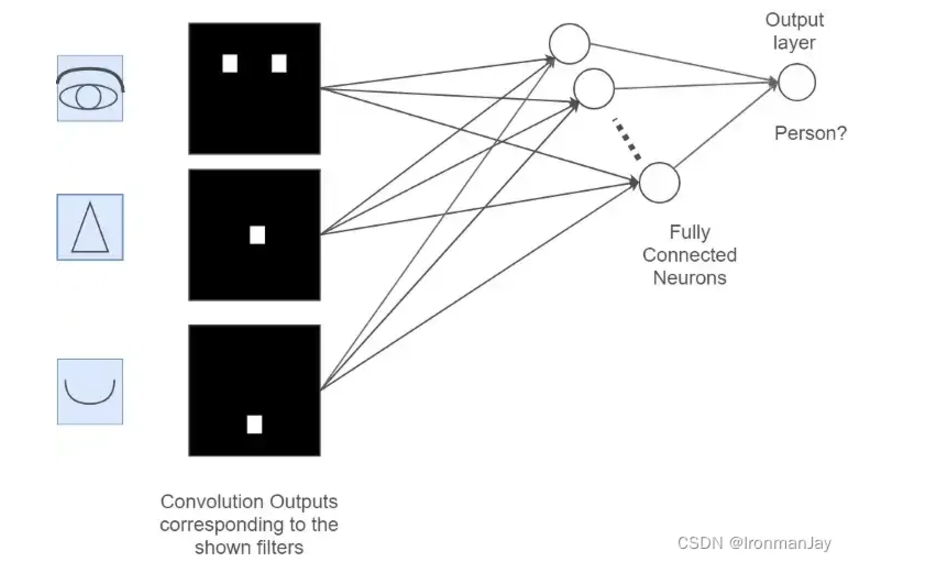
为了解决这个问题，我们可以利用池化层，那什么是池化层呢？池化层又称为下采样，也就是说，当我们进行卷积操作后，再将得到的特征图进行特征提取，将其中最具有代表性的特征提取出来，可以起到减小过拟合和降低维度的作用，

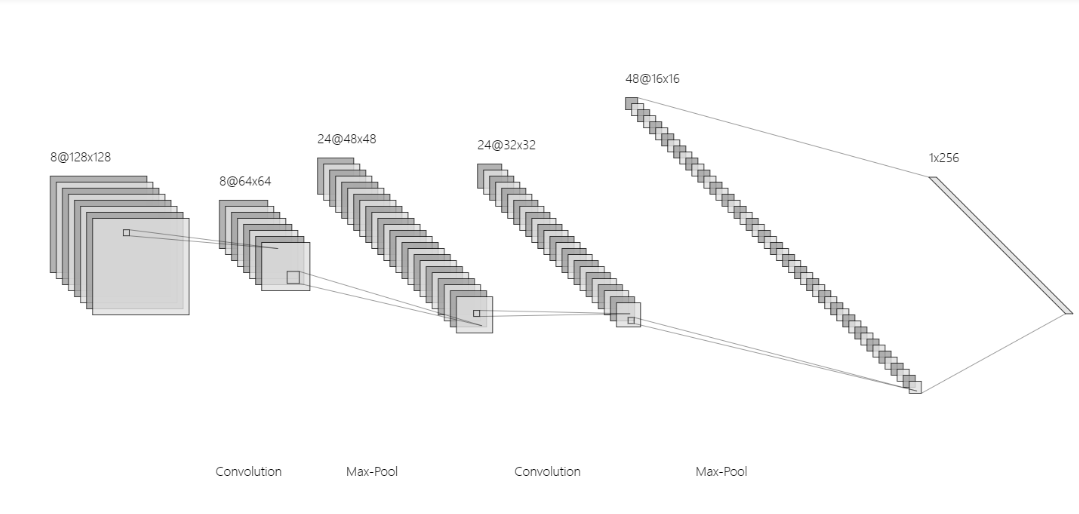
经过池化后，我们可以提取到更有代表性的特征，同时还减少了不必要的计算，这对于我们现实中的神经网络计算大有脾益，因为现实情况中神经网络非常大，而经过池化层后，就可以明显的提高模型的效率。所以说，池化层的好处很多，将其优点总结如下：

* 在减少参数量的同时，还保留了原图像的原始特征
* 有效防止过拟合
* 为卷积神经网络带来平移不变性

## 全连接层/输出层

假设还是上面人的脑袋的示例，现在我们已经通过卷积和池化提取到了这个人的眼睛、鼻子和嘴的特征，如果我想利用这些特征来识别这个图片是否是人的脑袋该怎么办呢？此时我们只需要将提取到的所有特征图进行“展平”，将其维度变为1 × x 1×x1×x，这个过程就是全连接的过程，也就是说，此步我们将所有的特征都展开并进行运算，最后会得到一个概率值，这个概率值就是输入图片是否是人的概率，这个过程如下所示：





经过两次卷积和最大池化之后，得到最后的特征图，此时的特征都是经过计算后得到的，所以代表性比较强，最后经过全连接层，展开为一维的向量，再经过一次计算后，得到最终的识别概率，这就是卷积神经网络的整个过程。

# 图像分割算法综述总结

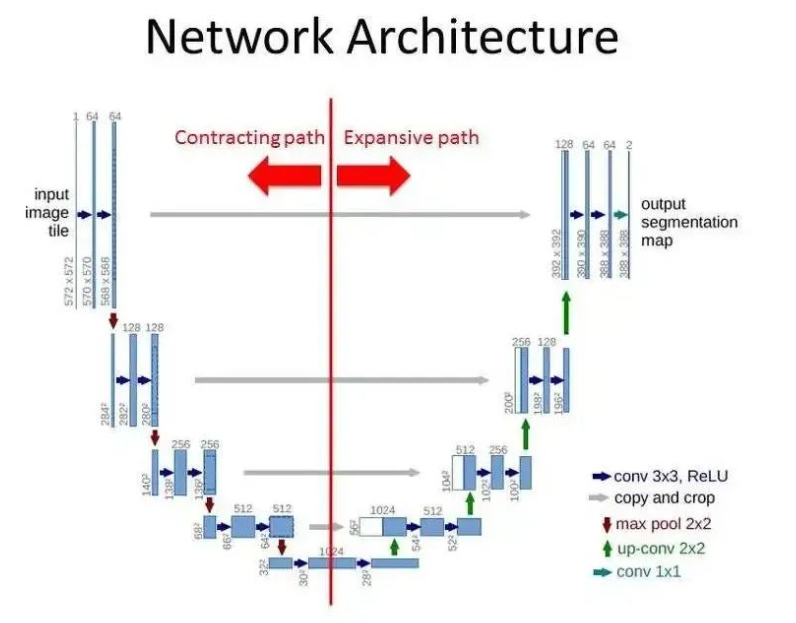
使用深度学习算法进行图像分割

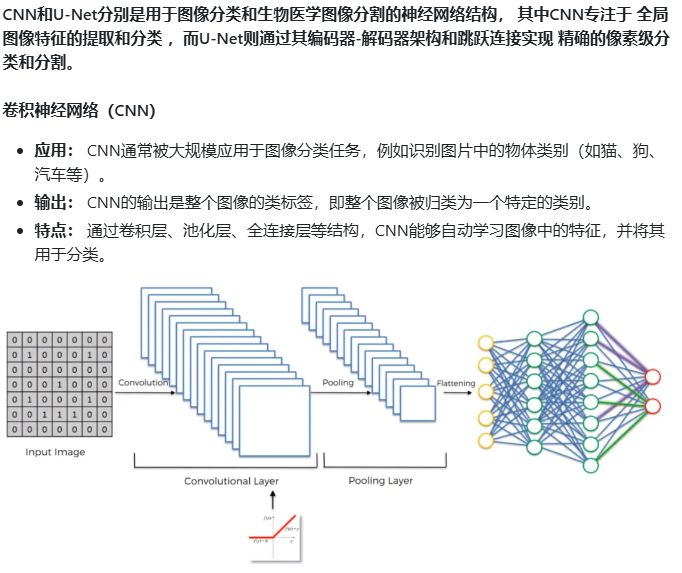
# 什么是UNet

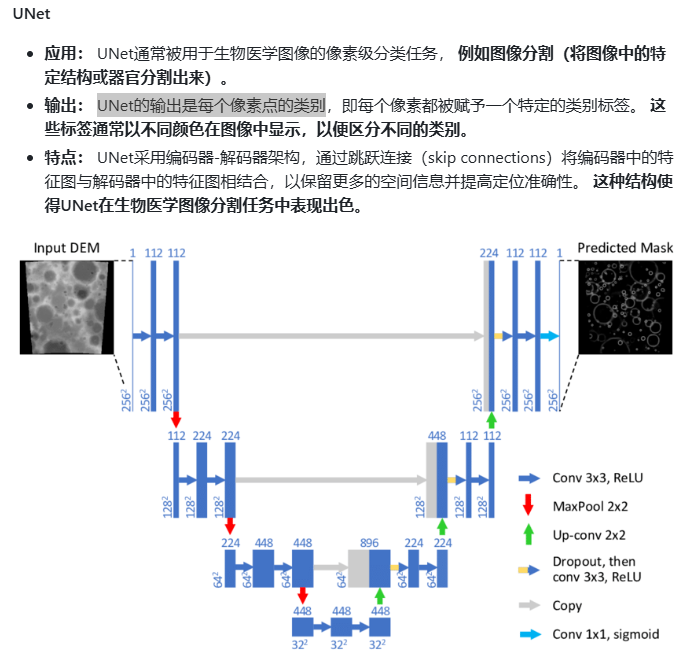
## Unet定义

**一个 基于深度学习的卷积神经网络 ，主要用于图像分割任务，特别是生物医学图像的分割。** **它由编码器（下采样路径）和解码器（上采样路径）两部分组成，形状呈U型，因此得名U-Net。**

**CNN和U-Net分别是用于图像分类和生物医学图像分割的神经网络结构， 其中CNN专注于 全局图像特征的提取和分类 ，而U-Net则通过其编码器-解码器架构和跳跃连接实现 精确的像素级分类和分割。**

****



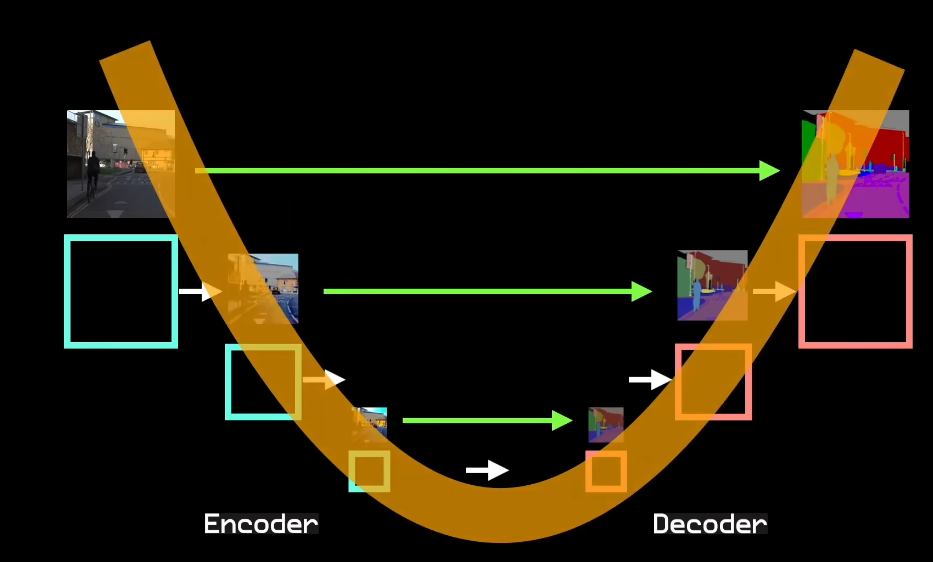


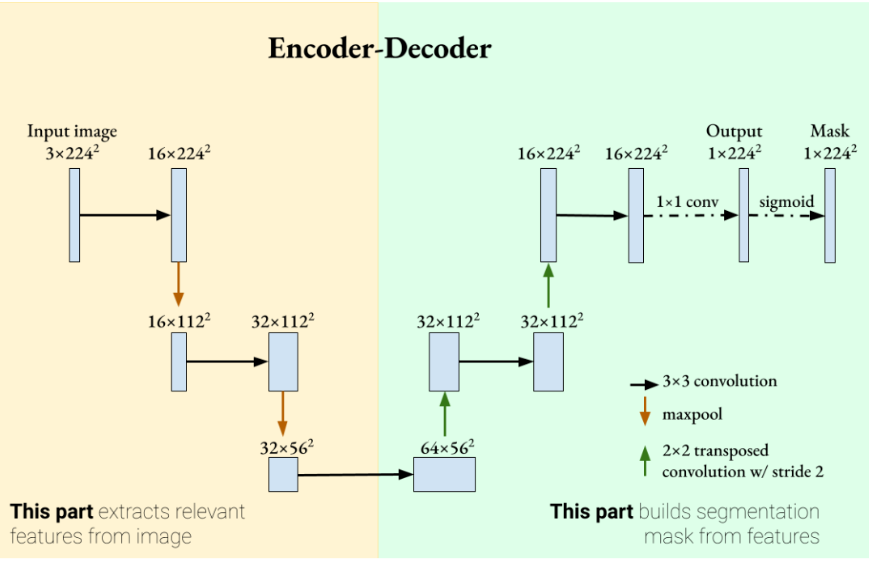
Encode : 3x3 convolution, double the channel

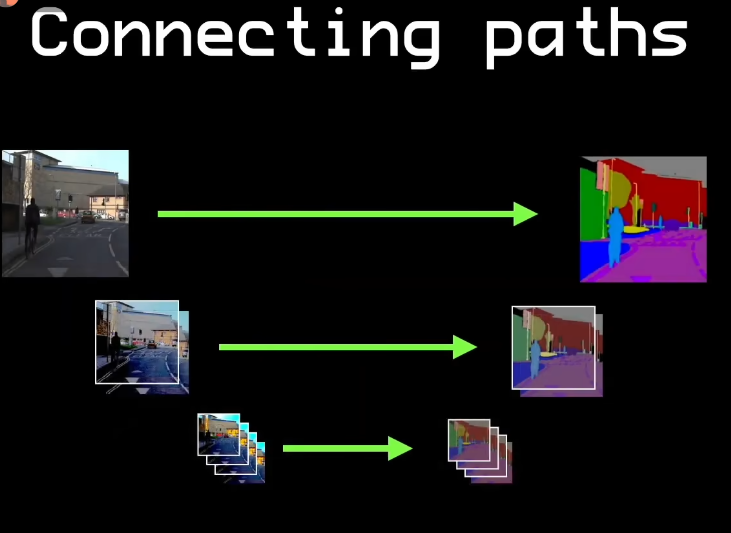
Decode: up-convulution, decrease the channel

## UNet本质

**一个基于全卷积网络（Fully Convolutional Network, FCN）的对称编解码结构（Encoder-Decoder）的图像分割网络**

* **编码器-解码器结构：UNet采用了编码器-解码器结构， 其中编码器逐渐减小特征图的空间尺寸以捕获上下文信息，而解码器则逐渐恢复空间尺寸以精确定位每个像素的类别。**
* **跳跃连接：UNet中的跳跃连接（skip connections）将编码器的特征图直接传递给解码器，从而保留了更多的空间信息，有助于提高定位准确性。**
* **端到端训练：UNet可以直接对整个图像进行训练，而无需使用滑动窗口方法。这大大减少了计算量，提高了训练效率。**
* ****



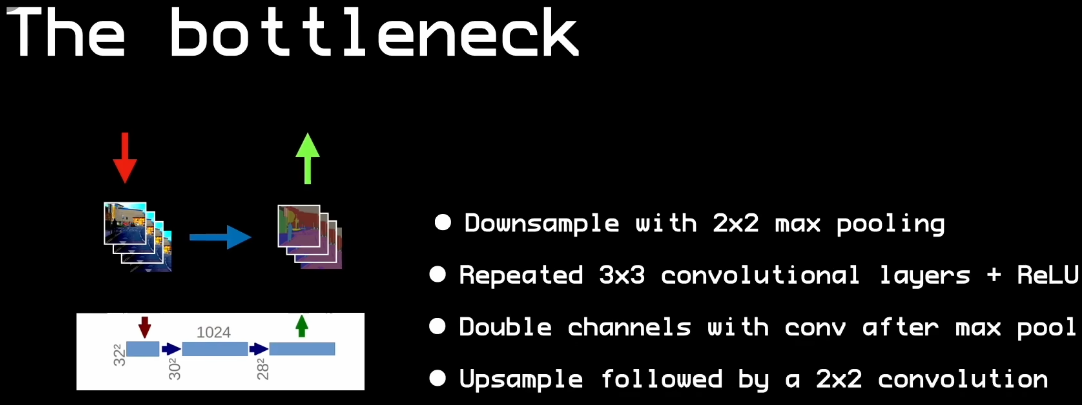


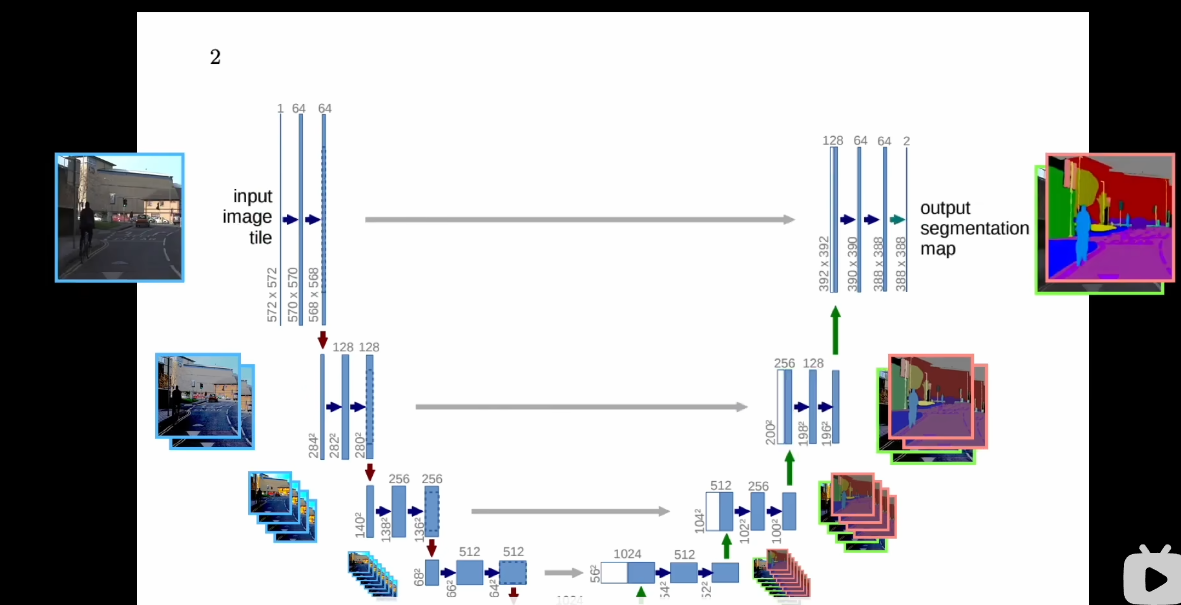
结合encode和decode的特征

Decoder包括更多语义特征-> bike

Encoder 包括更多空间特征 -> 某一部分的像素是单车

复制对应阶段，，后续convulsion可以同时操作decoder和encoder的特征，获得像素级别分割





同时upsampling同时传递encoder的信息（单车 单车 单车）

# 什么是nnUNet

记录：2023年前53 task是winner

Method： 1. Data analysis

2. Data preprocessing

## 重心

把重心放在：预处理（resampling和normalization）、训练（loss，optimizer设置、数据增广）、推理（patch-based策略、test-time-augmentations集成和模型集成等）

## 网络结构

基础版UNet：2D UNet，3D UNet，UNet级联（第一级对下采样低分辨率图像进行粗分割，第二级结合第一级的结果进行微调，两级都用UNet）

微小修改：

（1）ReLU换 leaky ReLU（neg.slope 1e-2）；

（2）Batch Norm换Instance Norm

默认参数设置：

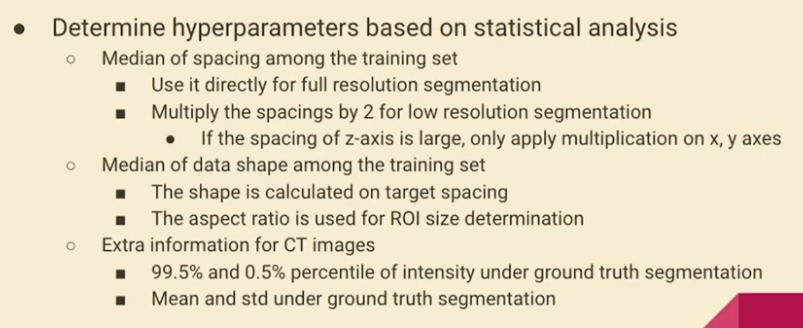
2D UNet：crop-size<=256x256（中值尺寸小于256时，采用中值尺寸）; batch-size<=42; base-channel=30; pooling to size>=8; pooling\_num<6

3D UNet: crop-size<=128x128x128（中值尺寸小于128时，采用中值尺寸）; batch-size>=2; base\_channel=30; pooling to size>=8; poolingnum<6

## Data analysis

统计：

1. mid spacing among the training set [resample目标的值（target shape），得到较小失真]
2. mid data shape – resample到目标大小size，aspect ratio【长宽高比例】，[ROI](https://so.csdn.net/so/search?q=ROI&spm=1001.2101.3001.7020) 通常指的是感兴趣区域（Region of Interest），通常情况下，我们只对图像中的一些特定区域（ROI）感兴趣，而不是整个图像
3. CT影响 – CT像素与密度有关（骨头or水，水肿、正常区域），找到目标区域。和处理MR不太一样



## Data Preprocessing

数据格式的转换，**裁剪crop，重采样resample以及标准化normalization。**

1. Normalization：Normalization的目的是**让训练集中每张图像的灰度值都能具有相同的分布**。nnUNet提供了两种normalization的策略，一种单独针对CT图像，一种应用于其它非CT图像。相同的地方是都使用z-scoring(即减去均值除以标准差)。不同的地方有两点。第一点，CT图像做normalization用的是整个训练集前景的均值和标准差，而非CT图像normalization时仅使用单张图像的灰度信息计算均值和方差。策略不同原因在于，CT图像中，***强度信息HU值能反映不同组织的物理性质***，用整个训练集前景的统计信息，可以有效的利用HU值的额外信息。第二点，CT图像中经常会有异常大的孤立值和异常小的孤立值，需要先将图像HU值clip到前景HU值[0.5, 99.5]百分比范围之间，而非CT图像没有clip的必要。值得注意的一点是，如果在crop阶段图像大小减小了1/4以上的时候，nnUNet会选择仅在nonzero region进行normalization
2. Resampling：resample 到目标大小，crop掉黑色区域【recall 影响黑色部分，不会对精准度有伤害但是会加快计算】

重采样目的是解决在一些三维医学图像数据集中，不同的图像中单个体素voxel所代表的实际空间大小spacing不一致的问题。因为卷积神经网络只在体素空间中进行操作，会忽视掉实际物理空间中大小信息。为了避免这种差异性，需要对不同图像数据在体素空间进行resize，保证不同的图像数据中，每个体素所代表的实际物理空间一致。

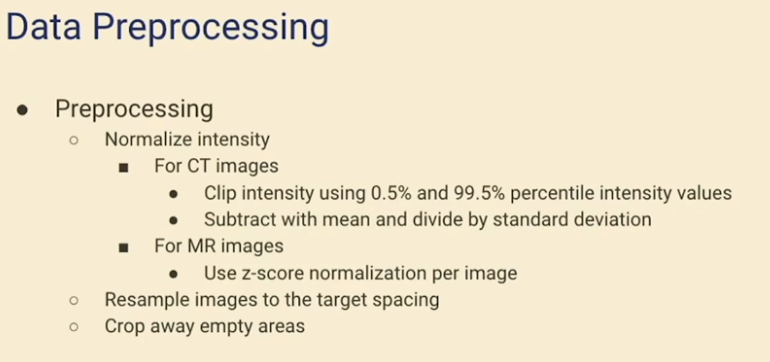
**具体需要将整个数据集resample到多大的spacing，即目标空间大小target\_spacing应该多大呢？**nnUNet给出的建议是在大多数时候使用数据集各个图像不同spacing的中值，但是在各向异性（最大坐标上的spacing÷最小坐标上的spacing>3）的数据集中，取数据集10%分位点的spacing值作为spacing最大坐标的目标空间大小会是更好的选择。重采样的步骤可以简单分成3步。

**第一步是确定重采样的目标空间大小**。在之前数据格式转换的时候，每个数据的spacing信息存储在对应的pickle文件中，需要依次进行读取，然后一起存放在一个列表spacings当中。之后调用numpy中函数统计每个维度spacing的中值即可。

接下来根据中值spacing进行判断，数据集是否存在各向异性的问题。nnUNet设定的判断标准是，中值spacing中三个维度，是否有一个维度spacing大于另一个维度spacing的3倍，并且，该维度的中值size小于另一个维度中值size的1/3。如果存在各向异性，对spacing特别大的维度，取数据集中该维度spacing值的10%分位点作为该维度的目标空间大小。

**第二步根据target\_spacing确定每张图像的目标尺寸**。每张图像, spacing和shape之间的乘积为一个定值,代表整个图像在实际空间中的大小。

**第三步调用skimage库中的reisze函数对每张图像进行resize即可**，但在nnUNet中会根据图像是否存在各向异性进行不同的resize策略。如果不存在各向异性，对整个三维图像进行3阶spline**插值**即可。如果图像存在各向异性，设spacing大的维度为z轴，则仅在图像的xy平面进行3阶spline插值，而在z轴采用最近邻插值。而对于分割的标注图像，无论各向异性与否, 在三个维度上都采用最近邻插值。下面代码中，用do\_seperate\_z表示是否存在各向异性，用axis表示各向异性图像中spacing最大的轴。

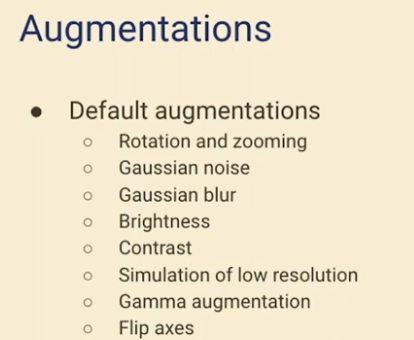


## Augumentation

Simulation of low resolution：对于模糊照片的处理

Flip axes：对xyz的处理

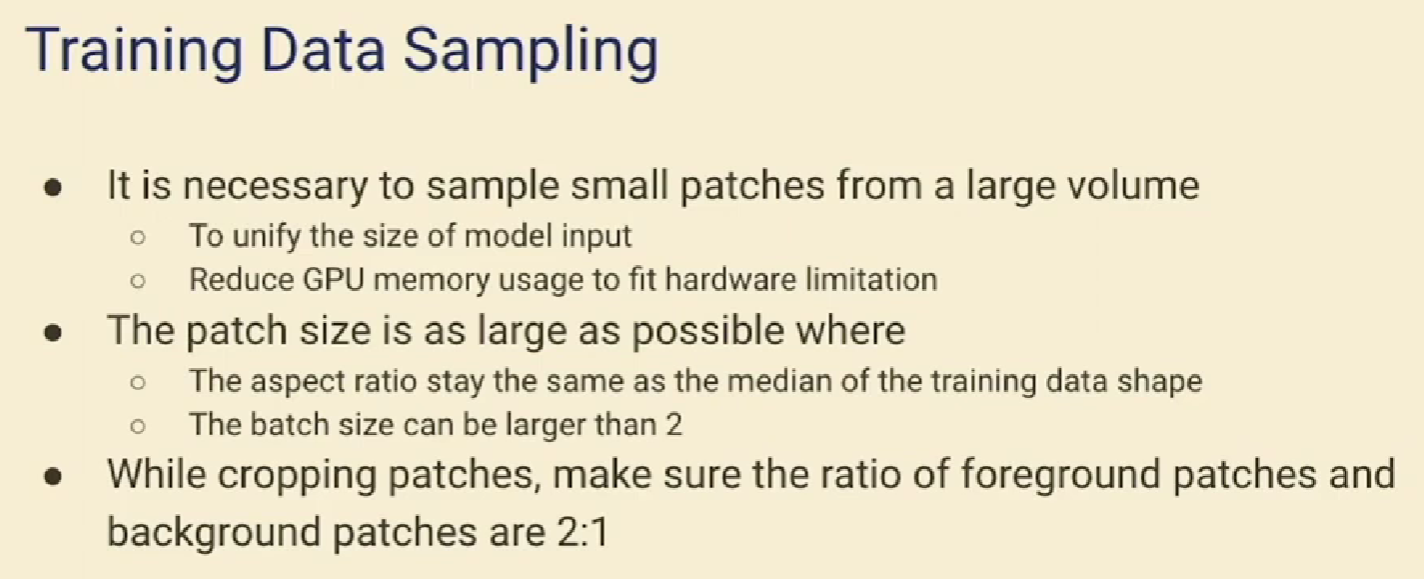
高斯噪声、模糊、亮度变换、反转图像、对比度增强、随机降采样、连续的两次gamma变换等等，另外还包含空间变换的方法镜像变换、随机旋转、缩放、平移，维度变换的方法针对二进制操作符的随机应用、随机移除一部分与分割标签连接、标签变换的方法对数据和标签的处理、使用掩码进行数据归一化



## Model training

维持原本长宽高比让patch size越大越好

Sampling保证 有vs没有 到2:1



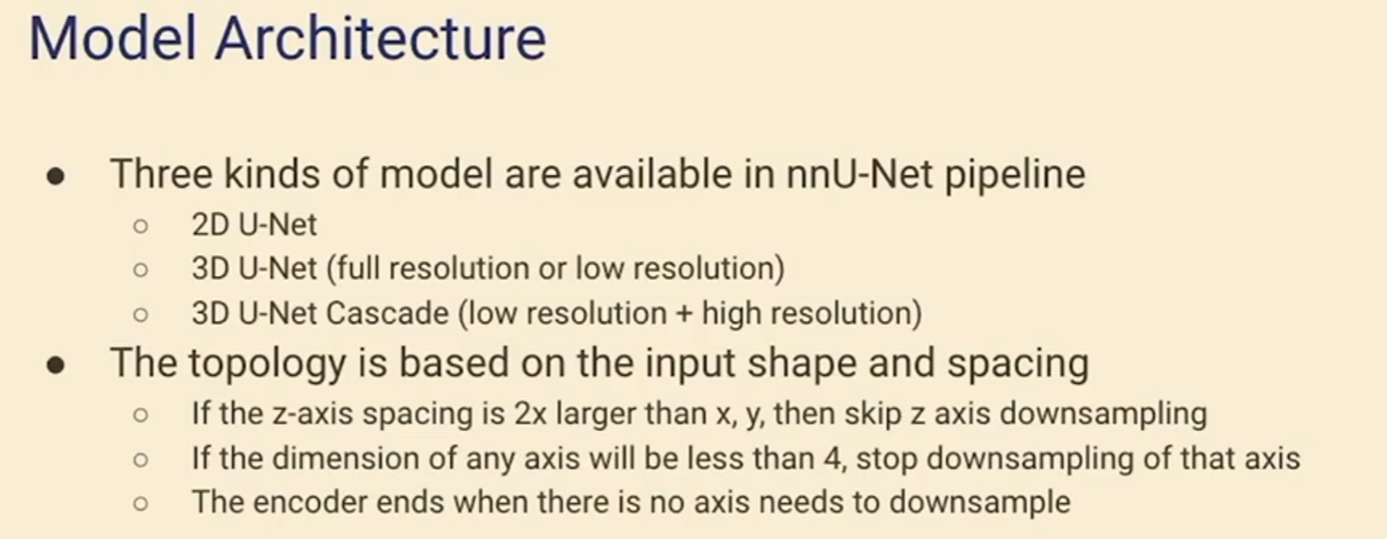
## Model architecture



Down sample终止：任一feature=4

目标xyz spacing一致

假如z比xy大很多！！！！！会跳过z！！



## Training configuration: Loss function

Dice loss + cross entropy

Dice loss = 1 – dice score

计算dice loss是一整个batch！！！ 而不是sample

E.g 3d 有 5 个dimension【batch size ， channel，x, y, z】

Sample dice求得是xyz上的平均【会梯度爆炸 performance烂掉】

Batch 是算三坨class上的dice loss，更能真实模拟数据整体distribution

不过还是看gpu性能啦

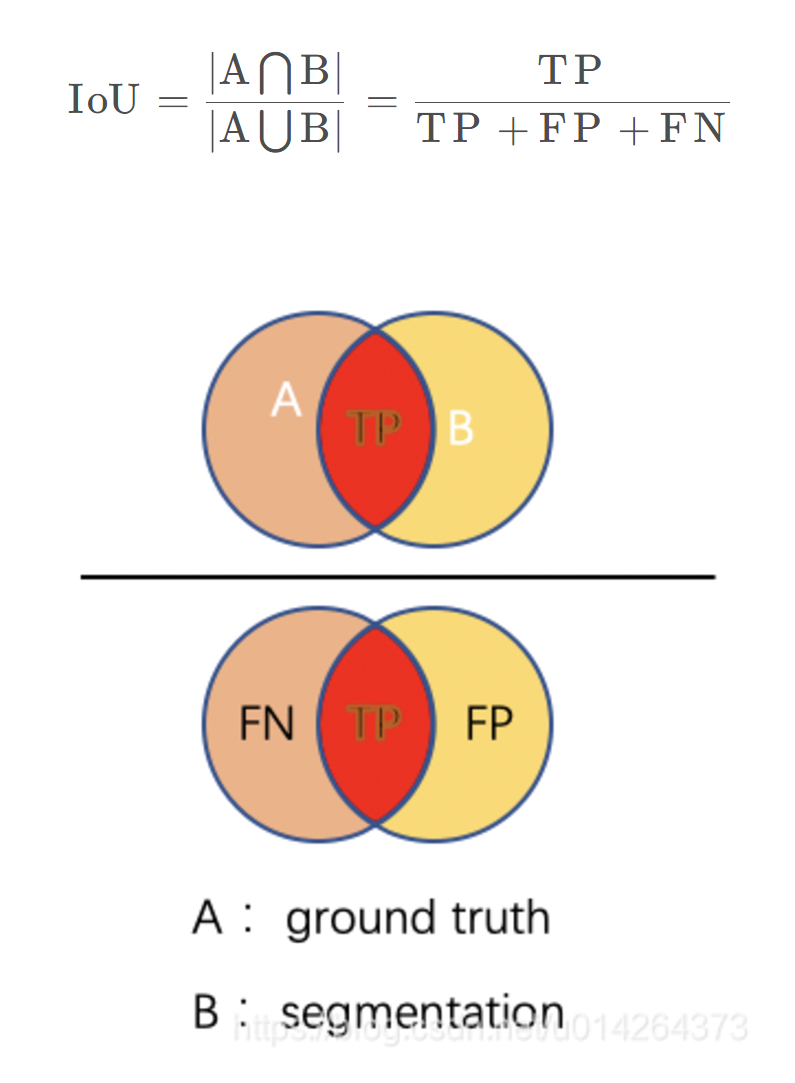
##### 要把batch dice = true 【gradient会变得很稳定】，论文没写

### 评估指标

#### IOU

Intersection-Over-Union (IoU)，也称为 Jaccard 指数，是语义分割中最常用的指标之一……这是有充分理由的。IoU 是一个非常简单的指标，非常有效

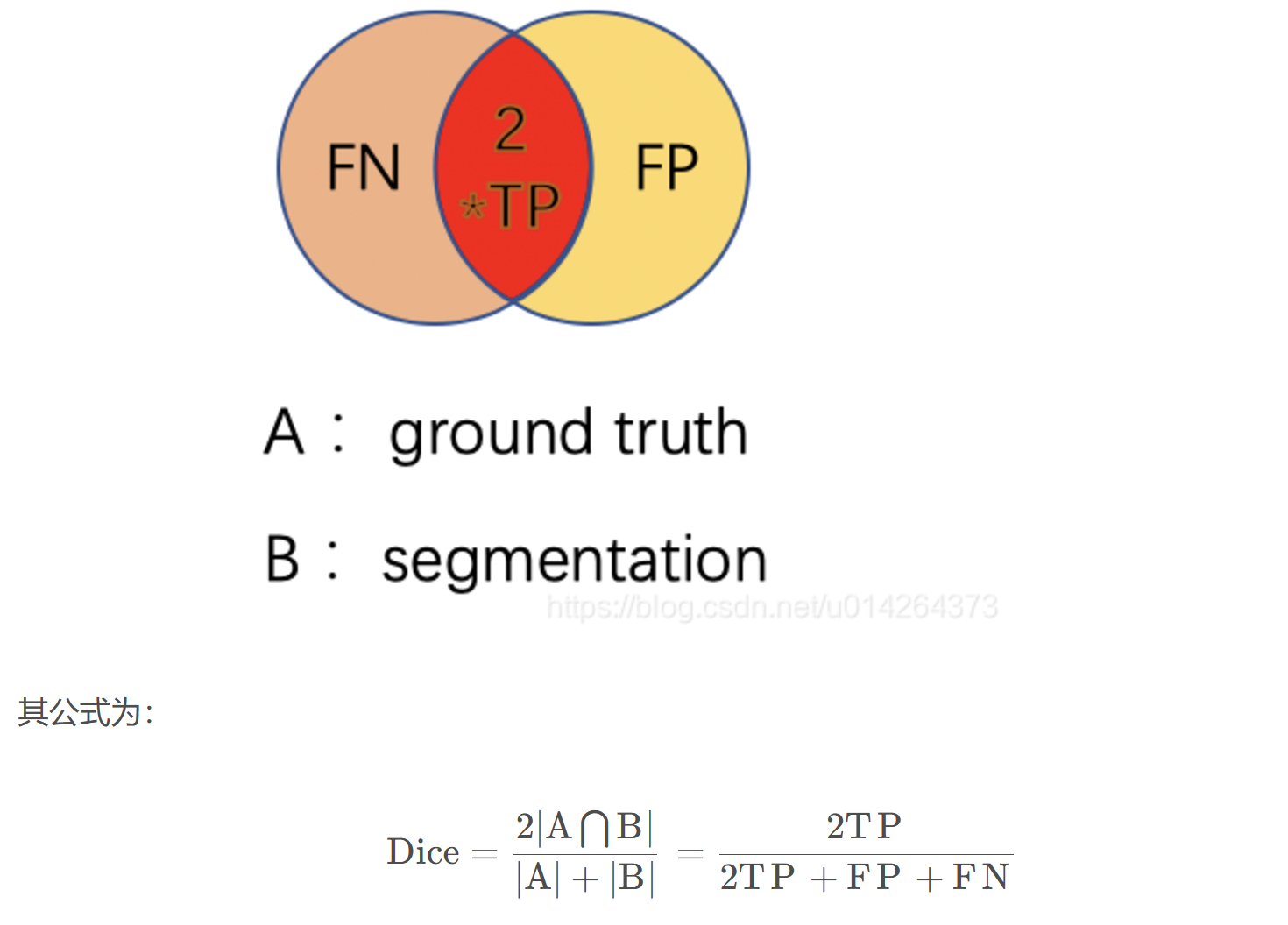
**是预测分割和标签之间的重叠区域除以预测分割和标签之间的联合区域（两者的交集/两者的并集），如图所示。该指标的范围为 0–1 (0–100%)，其中 0 表示没有重叠，1 表示完全重叠分割。**

****

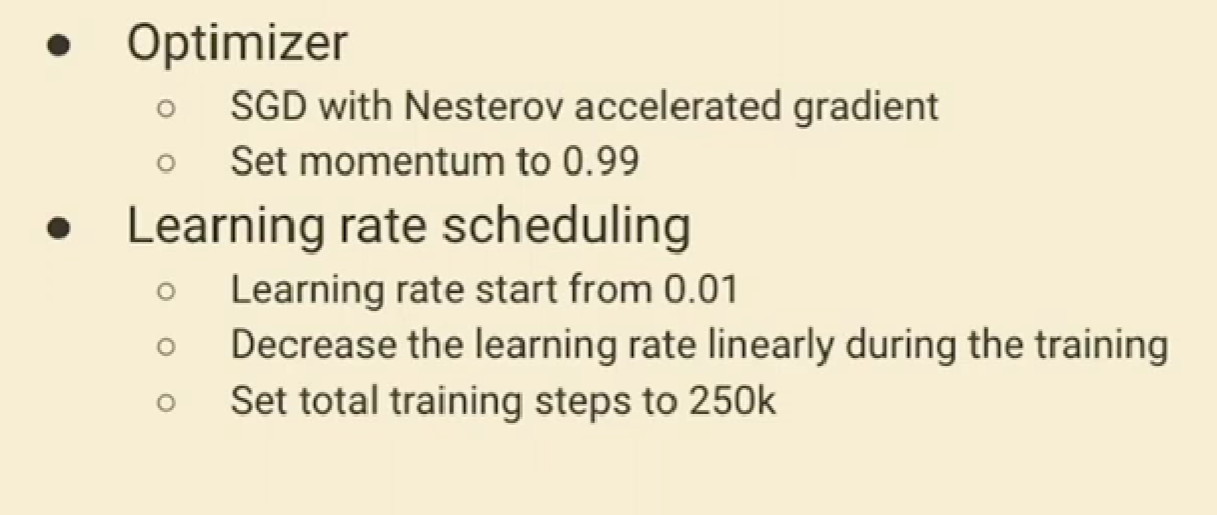
#### Dice score【和IOU很像???】= 1 – dice loss

Dice系数定义为两倍的交集除以像素和，也叫F1 score。Dice 系数与 IoU 非常相似，它们是正相关的。这意味着如果一个人说模型 A 在分割图像方面比模型 B 更好，那么另一个人也会这么说。

与 IoU 一样，它们的范围都从 0 到 1，其中 1 表示预测和真实之间的最大相似度。



### Optimizer & learning rate scheduling



1000步decade一次

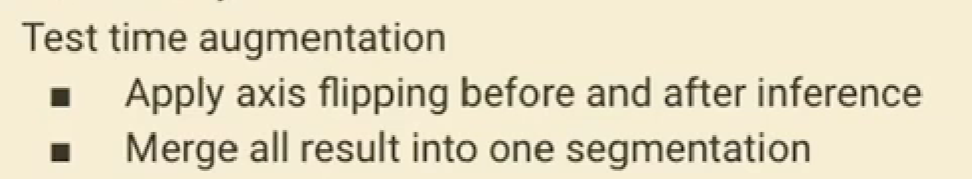
## Inference

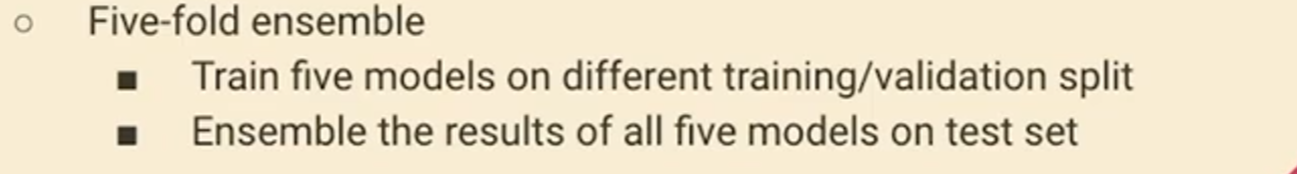
特别之处：

Test time augmentation: inference 的时候会对影像做flip，对flip前后的分别做inference，再把计算结果merge

优点：获得更robust结果

缺点：两次inference， 两倍时间





Five fold获得5个model

nnU-Net会忽略每个单独折叠上的后处理，因为它需要为整个交叉验证找到一个最优的后处理配置,单独的折叠后处理结果只是为了开发!

