目标检测

李乡儒

华南师范大学计算机学院

2022 年 4 月 21 日

華南師範大學
SOUTH CHINA NORMAL UNIVERSITY

目录 I

内容概要

- 什么是目标检测
- 如何表示
- 基础知识
- Fast RCNN
- YOLO

目录 I

目标检测

- 是什么?
- 在哪里?
- 有哪些?

“是什么”意味着需要判断出找出来的目标是什么, 也就是对目标的类别做判断, 是人还是狗或者是车

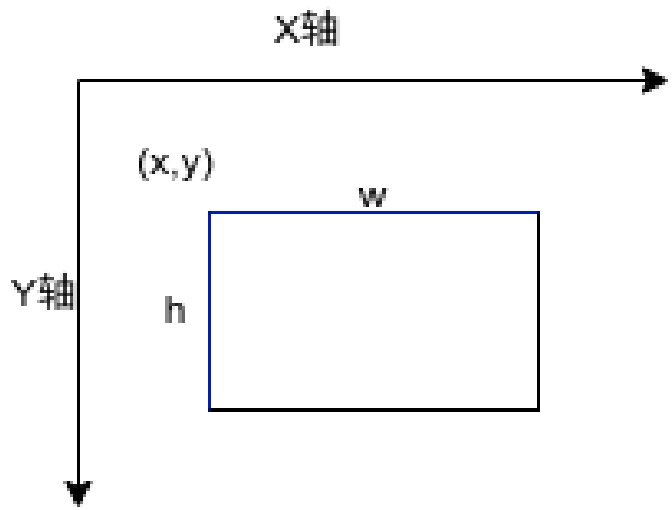
“在哪里”需要指出找到的目标在图像的那个地方, 范围是哪里

“有哪些”意味着需要将图像当中所有的感兴趣物体 (预先定义的类别) 找出来

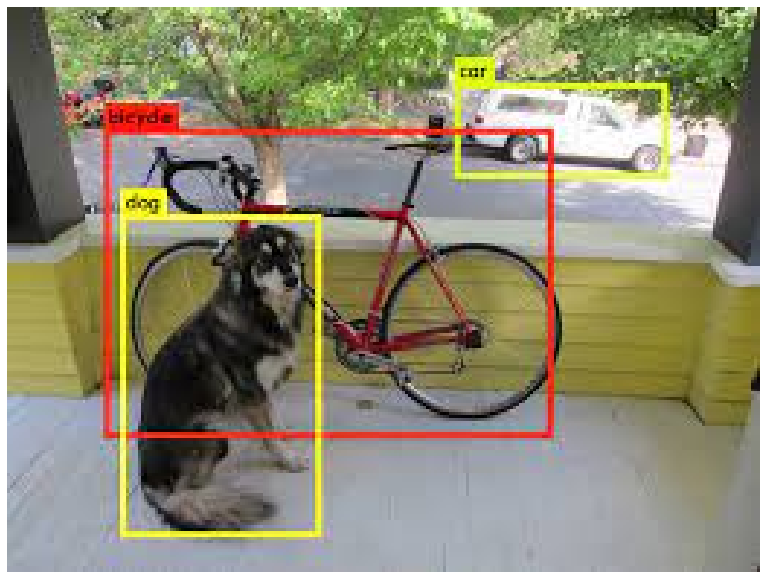
如何表示

在计算机当中需要用数字来表达 “是什么”、“在哪里”、“在哪里”

- 整数 (one-hot 向量) 表示类别
- 矩形框表示位置 (四元组 $[x, y, w, h]$)
- 堆叠所有目标组成数组 $n \times 5$



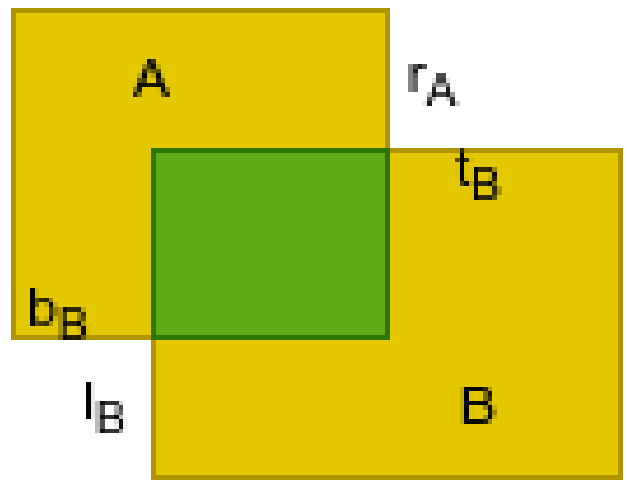
(a) box



(b) Object Detection

基础知识

- 矩形框的 3 种表示矩形框的表示形式有:
  $[x, y, w, h]$
  $[x_1, y_1, x_2, y_2]([left, top, right, bottom])$
  $[cx, cy, w, h]$
- 交并比 (IoU) 交并比是指两个矩形的交集与并集的面积之比
  $IoU = \frac{|A \cap B|}{|A \cup B|}$, 实现是采用第二种矩形表示进行实现
  $A[l_A, t_A, r_A, b_A], B[l_B, t_B, r_B, b_B]$:
  $S_A = (r_A - l_A) \times (b_A - t_A)$
  $S_B = (r_B - l_B) \times (b_B - t_B)$
  $W_{AB} = (min(r_A, r_B) - max(l_A, l_B))$
  $H_{AB} = (min(b_A, b_B) - max(t_A, t_B))$
  if $W_{AB} < 0$ or $H_{AB} < 0$ then $IoU = 0$
  else $IoU = \frac{W_{AB} \times H_{AB}}{S_A + S_B - W_{AB} \times H_{AB}}$



常见数据集

Common Objects in COntext（COCO）、PASCAL Visual Object Classes (PASAL)、The Cityscapes Dataset、The Cambridge-driving Labeled Video Database（CamVid）、Stanford Background Dataset、Barcelona Dataset、Microsoft Research in Cambridge、LITS Liver Tumor Segmentation Dataset、ISBI Challenge

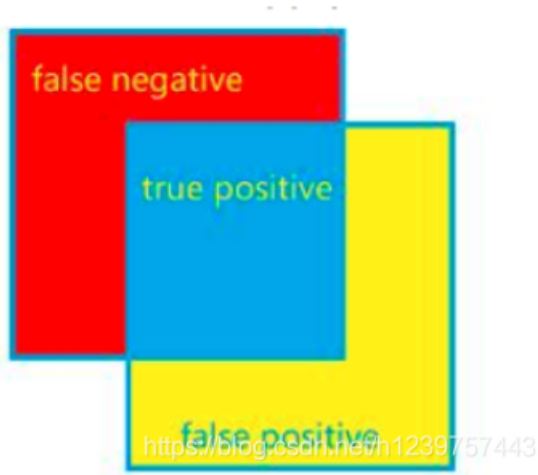
常用评价指标

- 准确率： $PA = \frac{TP+TN}{TP+FP+FN+TN}$
- Dice 系数 (Dice score, F1 分数):

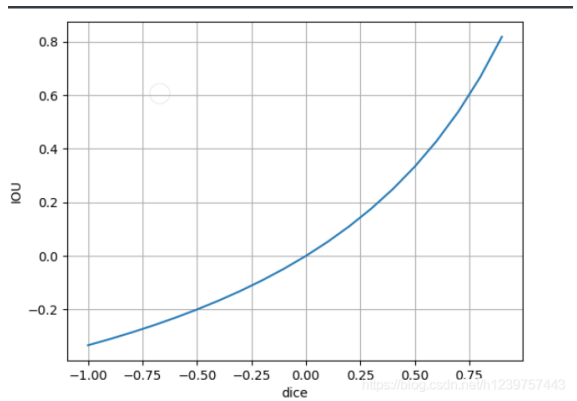
$dice(A,B) = \frac{2|A \cap B|}{|A|+|B|} = \frac{2TP}{2TP+FN+FP}$

- 雅卡尔指数（交并比）:$IoU = \frac{|A \cap B|}{|A \cup B|} = \frac{TP}{TP+FN+FP}$

$IoU = \frac{Dice}{2-Dice}$; A: 目标像素的集合； B: 算法判定为目标像素的集合。除此之外还有精确率、召回率、平均准确率、平均精确率、平均召回率和聚合雅卡尔指数等指标



(c) TP、FP、FN



(d) IoU 与 Dice

类别不平衡问题

图像当中的背景与前景常存在像素点数量不平衡的问题，这容易导致模型将所有像素点预测为同一个类别

解决方法

- 损失函数加权：$L = \sum w_i loss_i$
- 欠采样：样本多的类别只统计一部分像素点的损失
- Dice 损失函数：$L = 1 - \frac{2\sum \hat{y}_i y_i + \epsilon}{\sum(\hat{y}_i + y_i) + \epsilon}$
- Focal 损失函数：$L = -\sum(1 - \hat{y}_i)^{\gamma} log(\hat{y}_i)$

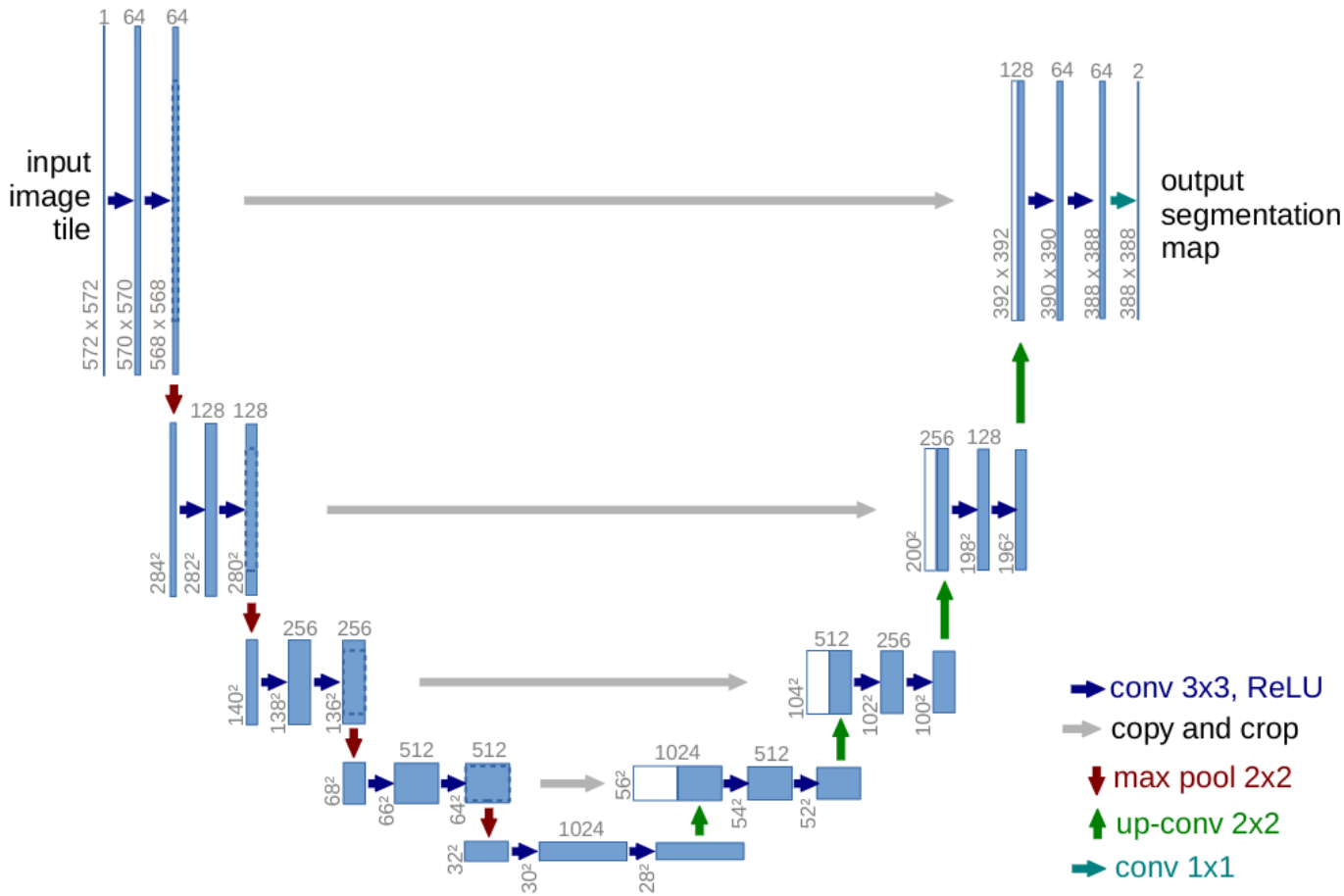
图像分割的应用

- 自动驾驶：对周围环境图像进行分割
- 医学图像病灶检测：将医学影像当中的病变部位分割出来
- 零售图像识别：对货架商品进行监控
- 人脸识别：从图像当中提取人脸区域

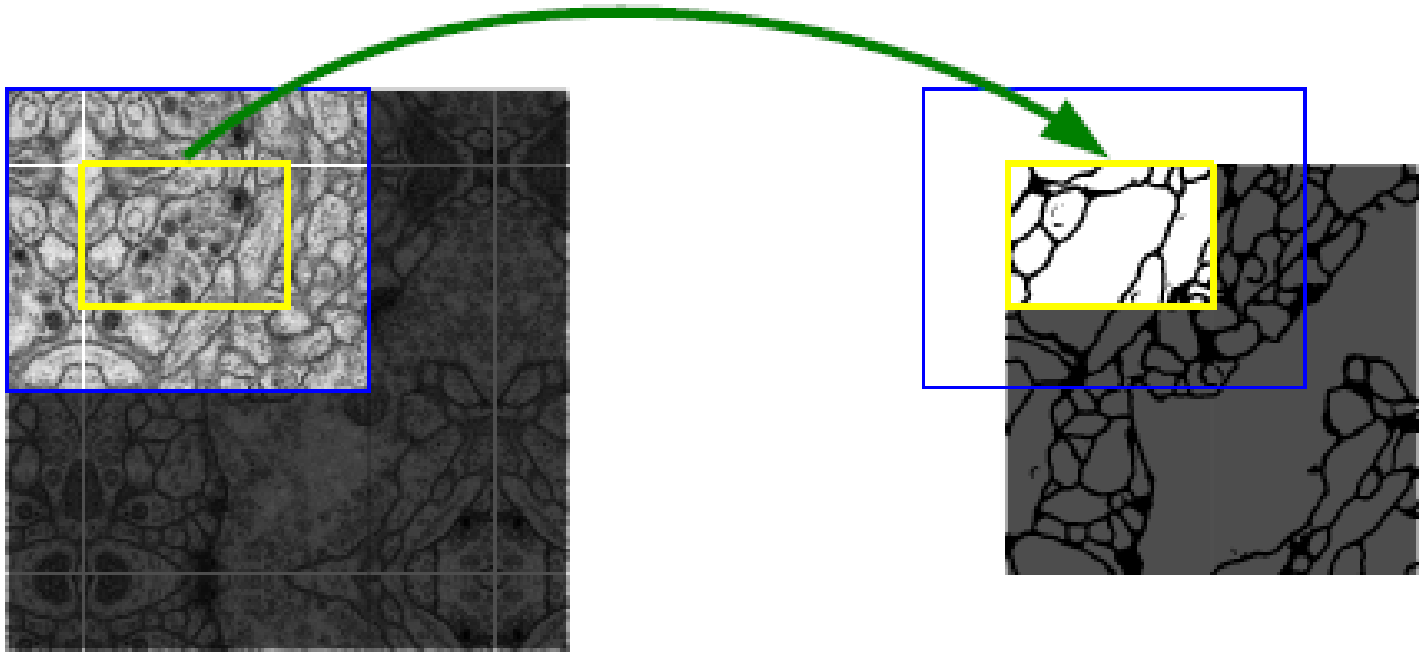
目录 I

U-net

U-net 首次提出是在 2015 年的 MICCAI 会议上，此后成为了图像分割任务的 baseline，主流的图像分割模型大多遵循 U-net 的基本框架。



无缝分割策略



处理流程

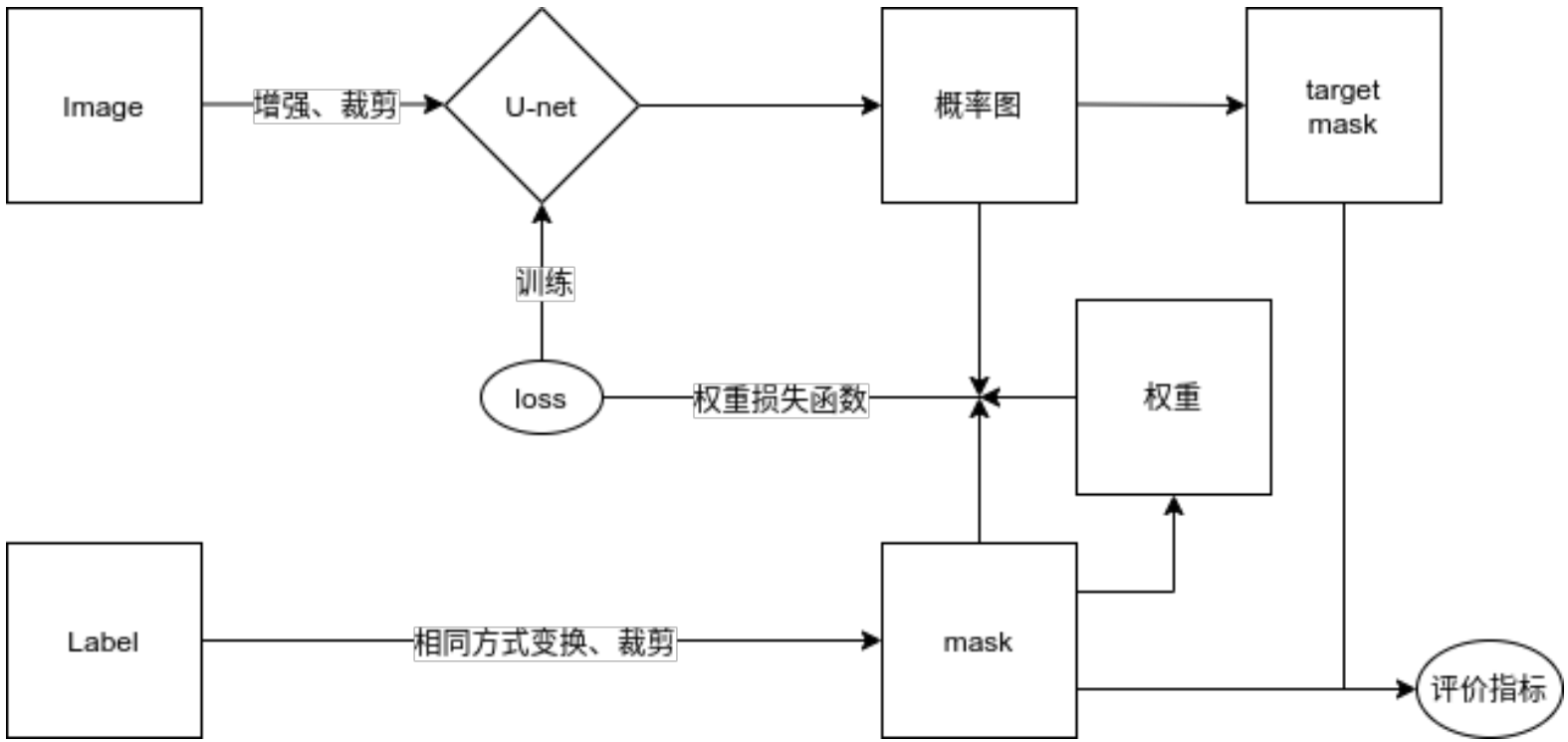
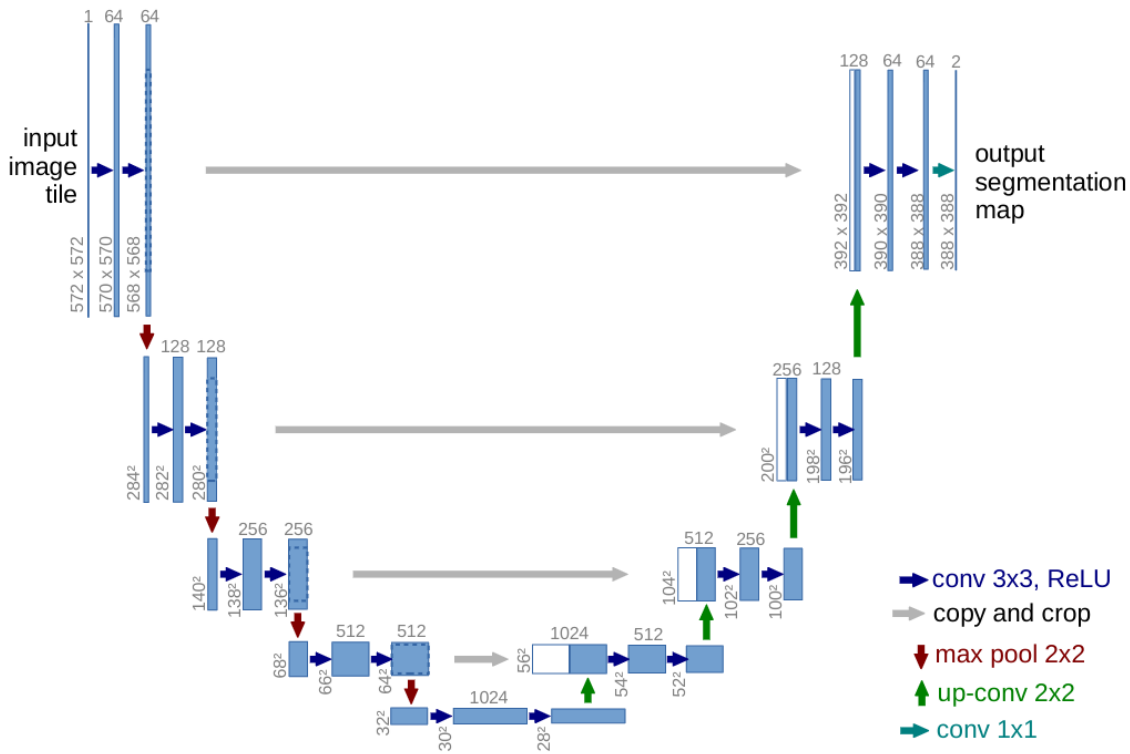


Image
增强、裁剪
U-net
概率图
target
mask
训练
loss
权重损失函数
权重
Label
相同方式变换、裁剪
mask
评价指标

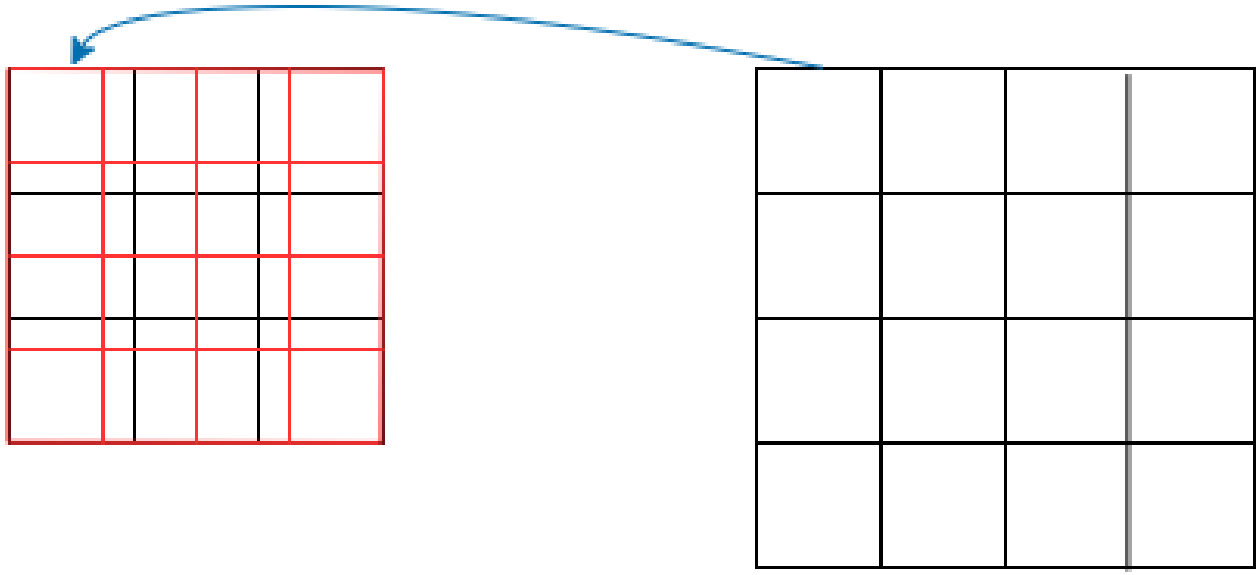
U-net 为什么有效

- 跳跃连接保留了更多细节
- 多层次特征图能够学习语义特征
- 多层次特征融合既能保留细节又能学习语义



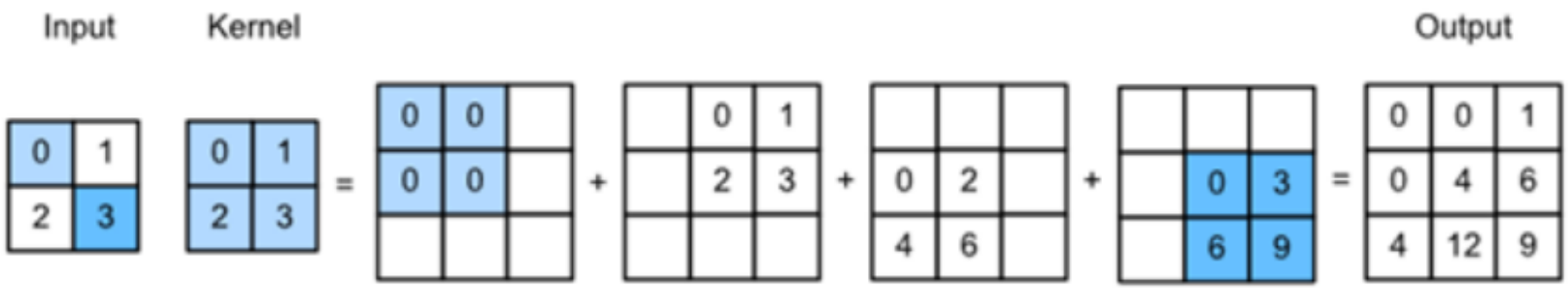
上采样方法

- 采样（线性插值、三次插值、最临近等）

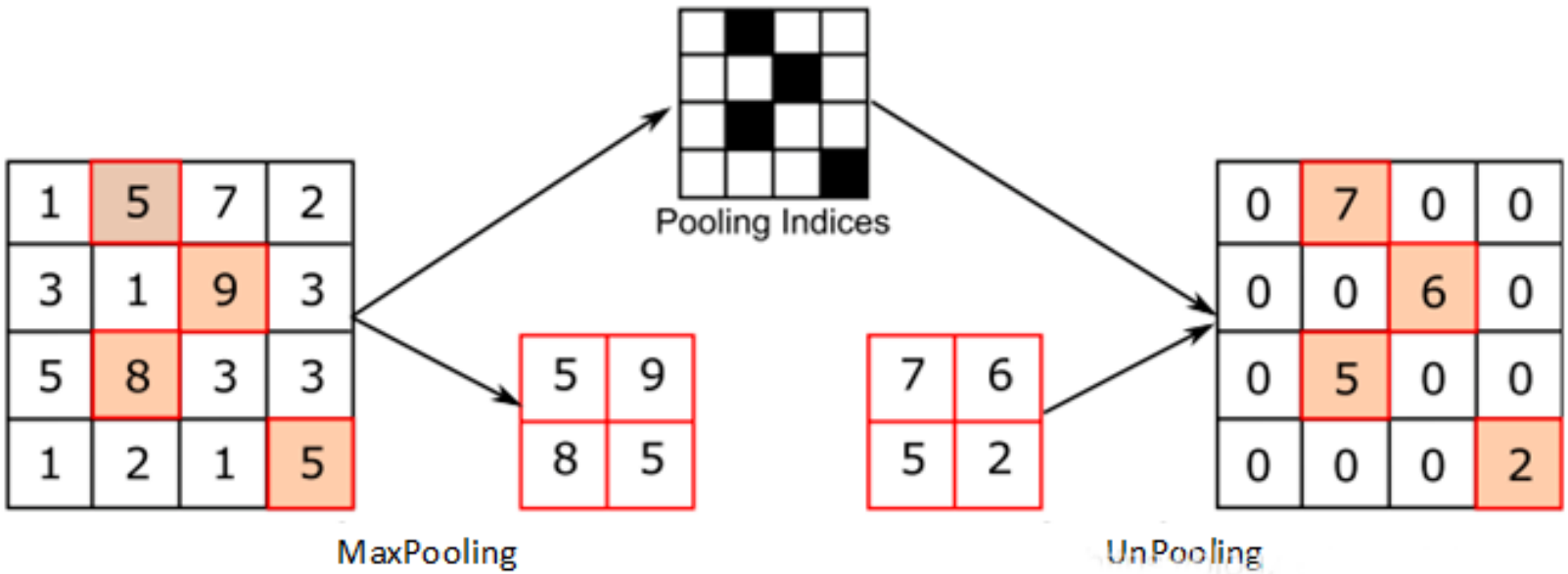


原图像3X3　　目标图像4x4

- 转置卷积（反卷积）



- 上池化



在 pytorch 的 MaxPool 操作当中，可以选择输出选取的位置，而 MaxUnPool 操作需要特征图和位置作为输入

```
1  #采样
2  torch.nn.functional.interpolate()
3  torch.nn.functional.grid_sample()
4  torch.nn.functional.upsample()
5  torch.nn.UpsamplingBilinear2d
6  torch.nn.UpsamplingNearest2d
7  torch.nn.Upsample
8  #转置卷积(反卷积)
9  torch.nn.ConvTranspose2d
10 torch.nn.functional.conv_transpose2d()
11 #上池化
12 torch.nn.MaxUnpool2d
13 torch.nn.functional.max_unpool2d()
```

特征图融合方法

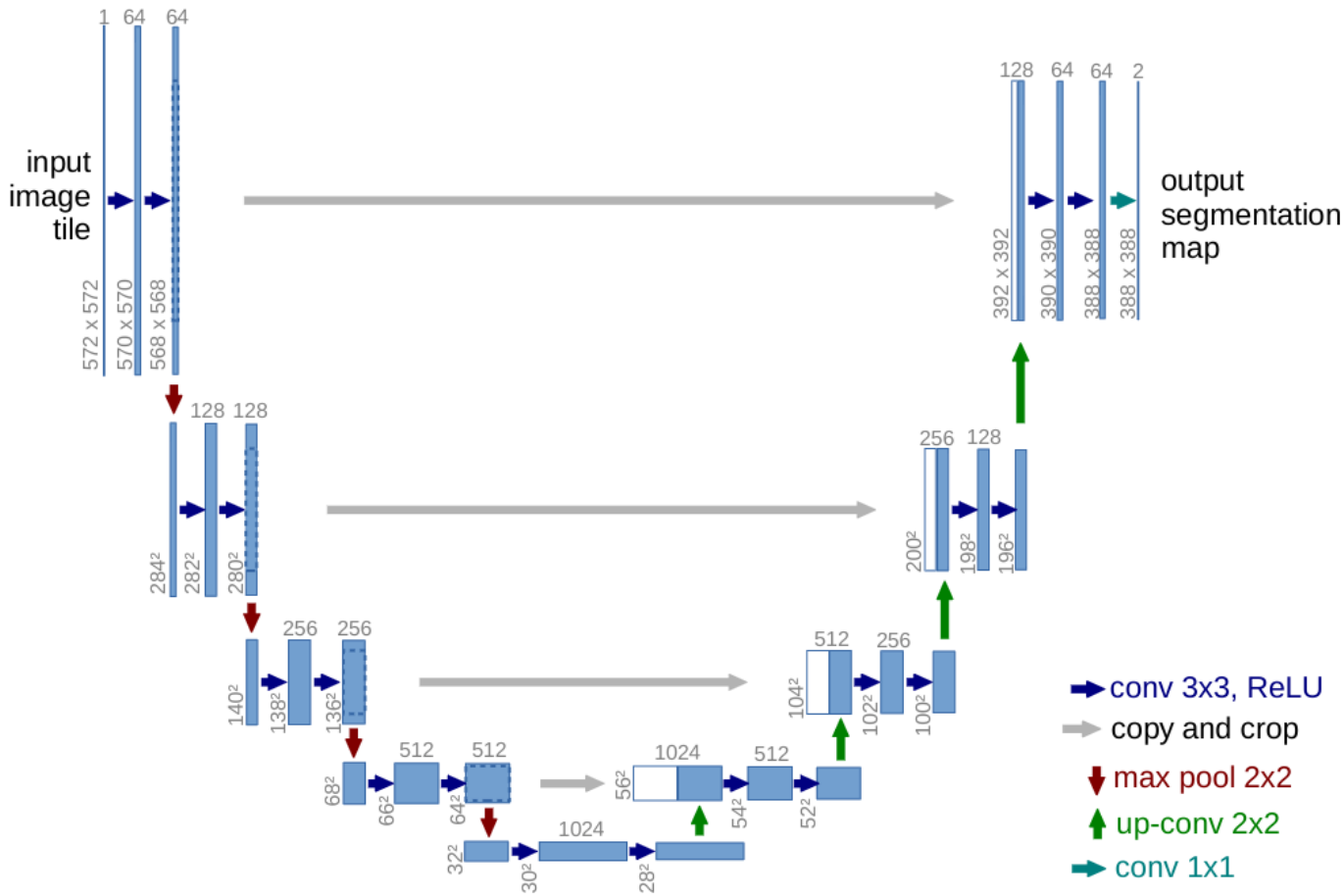
- 拼接：在通道维度上将特征图拼接起来
- 相加：把相同形状的特征图直接相加

区别：拼接不需要通道数一致但需要图尺寸相同，而相加需要通道数和图尺寸保持一致

联系：相加可以看作是特殊的拼接，因为通常在拼接后会使用 1×1 卷积进行加权

特征图融合方法

U-net 首次提出是在 2015 年的 MICCAI 会议上，此后成为了图像分割任务的 baseline，主流的图像分割模型大多遵循 U-net 的基本框架。



目录 I

实现网络模型

U-net 的实现网络模型并不复杂，使用到的操作有：卷积，转置卷积、池化、激活函数、批归一化、中心裁剪、拼接

```
nn.Conv2d(c_i,c_o,3)                    #卷积
nn.ConvTranspose2d(c_i,c_o,2,2)         #反卷积
nn.functional.max_pool2d(x,2,2)         #最大池化
nn.ReLU()
nn.BatchNorm2d(c_o)                     #批归一化
x = x[:,:,c:-c,c:-c]                    #中心裁剪
torch.cat([x1,x2],dim=1)                #拼接
```

目录 I

损失函数与目标函数

交叉熵 (Cross Entropy)。对于一个样本 $\mathbf{x}$ 的类别估计

$$\begin{aligned} h_\theta(\mathbf{x}) &= (P(y_1 = 1|\mathbf{x}), P(y_2 = 1|\mathbf{x}), \cdots, P(y_K = 1|\mathbf{x})). \\ &= (p_1(\mathbf{x}), p_2(\mathbf{x}), \cdots, p_K(\mathbf{x})) \end{aligned}$$

和参考信息 $\mathbf{y} = (y_1, y_2, \cdots, y_K)$, 它们之间的交叉熵是

$$H(\mathbf{y}, h_\theta(\mathbf{x})) = -(y_1 \log P(y_1 = 1|\mathbf{x}) + \cdots + y_K \log P(y_K = 1|\mathbf{x})) \tag{1}$$

Human-Readable

Pet
Cat
Dog
Turtle
Fish
Cat

Machine-Readable

Cat	Dog	Turtle	Fish
1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	1
1	0	0	0

$$\begin{aligned} H(\mathbf{y}, h_\theta(\mathbf{x})) &= -\log P(y_{l(\mathbf{x})} = 1|\mathbf{x}) \\ &= -\log p_{l(\mathbf{x})}(\mathbf{x})) \end{aligned} \tag{2}$$

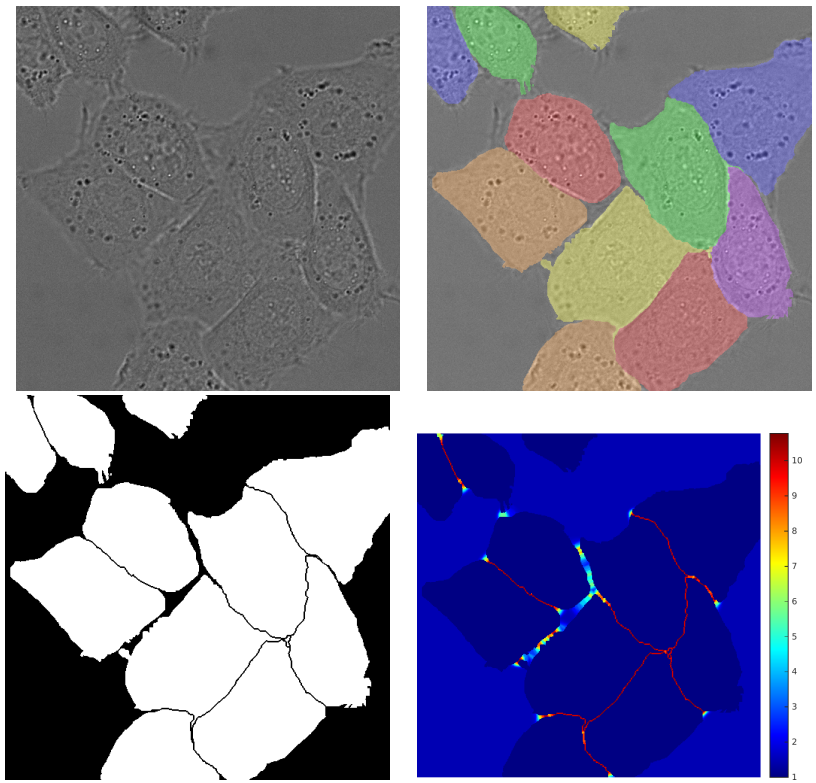
损失函数与目标函数

$$\begin{aligned} H(\mathbf{y}, h_\theta(\mathbf{x})) &= -\log P(y_{l(\mathbf{x})} = 1)|\mathbf{x}) \\ &= -\log p_{l(\mathbf{x}))}(\mathbf{x})) \end{aligned} \tag{3}$$

$$E = -\sum_{\mathbf{x}\in\Omega} log(p_{l(\mathbf{x}))}(\mathbf{x}))) \tag{4}$$

$$E = -\sum_{\mathbf{x}\in\Omega} w(\mathbf{x}) log(p_{l(\mathbf{x}))}(\mathbf{x}))) \tag{5}$$

权重函数

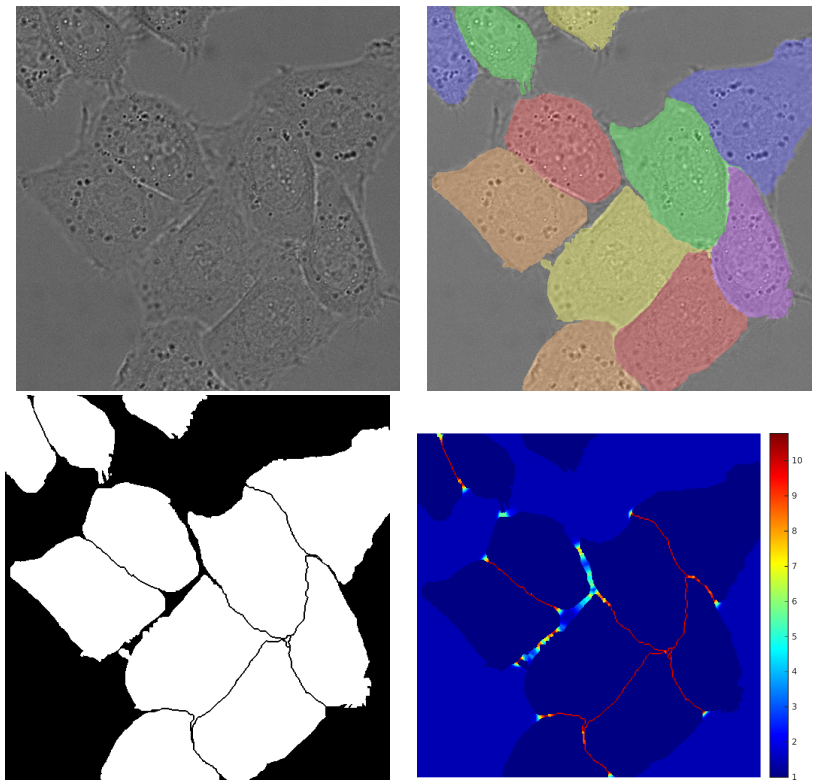


用 DIC（差分干涉对比度）显微镜记录玻璃上的 HeLa 细胞。(a) 原始图像。(b) 参考分割结果。不同的颜色表示 HeLa 细胞的不同实例。(c) 生成的分割掩码（白色：前景，黑色：背景）。(d) 像素级的损失权重，以促使网络模型更好地学习识别边界像素。图像数据来自文献?

$$w(\mathbf{x}) = w_c(\mathbf{x}) + w_0 \cdot exp(-\frac{(d_1(\mathbf{x}) + d_2(\mathbf{x}))^2}{2\sigma^2}) \tag{6}$$

d_1 和 d_2 分别表示一个像素距离最近和第二近细胞的边界的距离。

目录 I



用 DIC（差分干涉对比度）显微镜记录玻璃上的 HeLa 细胞。(a) 原始图像。(b) 参考分割结果。不同的颜色表示 HeLa 细胞的不同实例。(c) 生成的分割掩码（白色：前景，黑色：背景）。(d) 像素级的损失权重，以促使网络模型更好地学习识别边界像素。图像数据来自文献**?**

数据量与经验

平移、旋转、形态

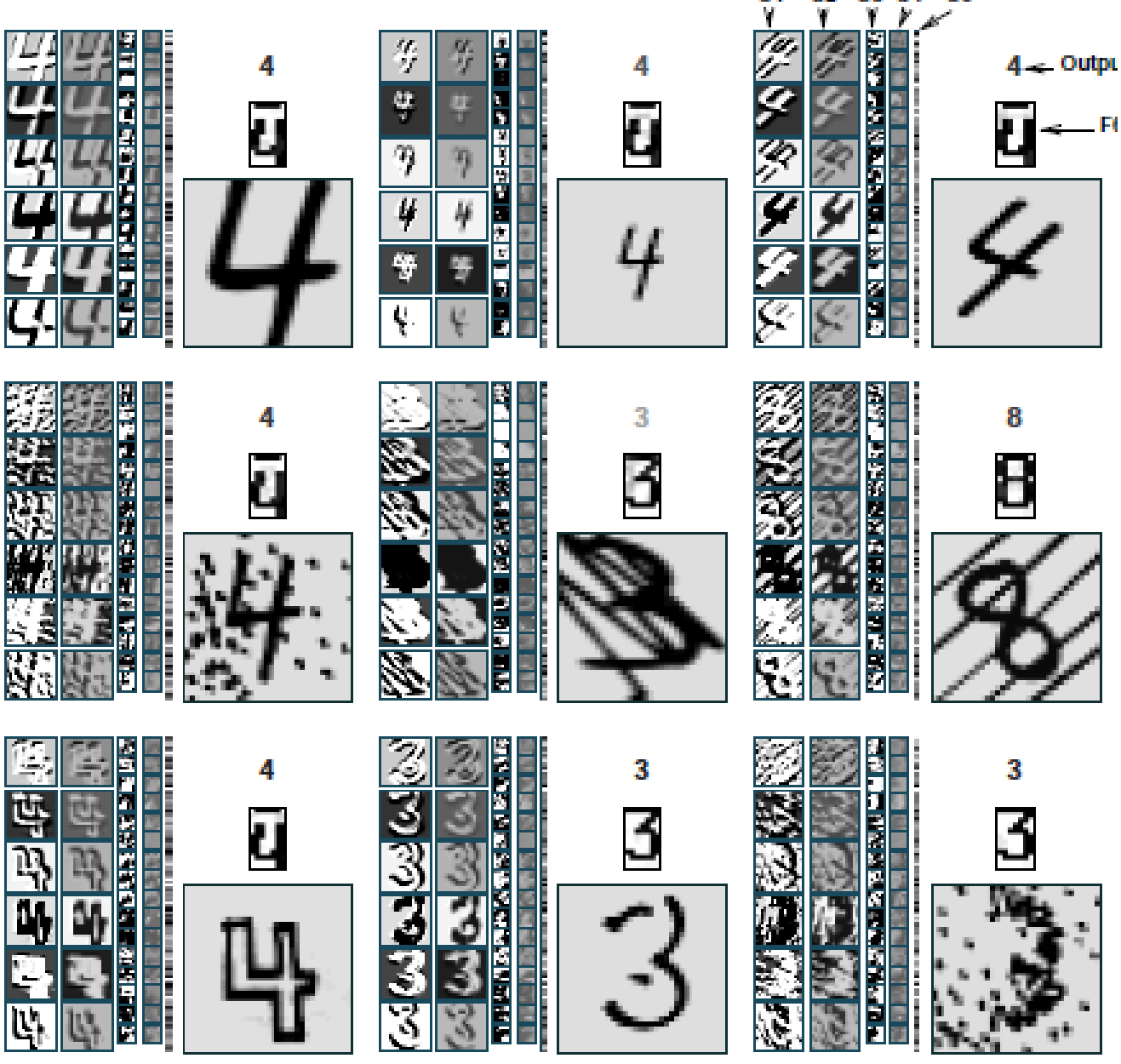
advantageously in specific contexts and circumstances (Blair & Raver, 2012; Ellis & Del Giudice, 2014).

Defining and Measuring Executive Functions

Stated simply, *executive functions* refer to aspects of cognition that are called on in situations when brain and behavior cannot run on automatic. More specifically, executive functions describe interrelated cognitive abilities that are required when one must intentionally or deliberately hold information in mind, manage and integrate information, and resolve conflict or competition between stimulus representations and response options. In this process of integration and control, it is generally agreed that executive functions include *working memory*, defined as the active maintenance or updating of information over a relatively short time period; *inhibitory control*, defined as the activation of specific information and inhibition of automatic but nonoptimal or incorrect responses; and *cognitive flexibility* or *attentional set shifting ability*, defined as the ability to shift the focus of attention or cognitive set flexibly and to adjust behavior accordingly. In general, these aspects of cognition are important for planning, future-directed thinking, and monitoring of behavior, all of which are aspects of cognitive experience encompassed by definitions of executive functions.

Given the presence of [illegible]rking memory, inhibitory control, and attentional set shifting compon[illegible]nction, researchers have been interested in the

数据量与经验
平移、旋转、形态



数据量与经验

平移、旋转、形态

目录 I

探讨与扩展阅读

* U-Net: Convolutional Networks for Biomedical Image Segmentation (**?**)
* nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation (**?**)
* nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation (**?**).

目录 I

思考题 (选做)

9.1 请参考文献? 和网上调研，在课程提供的 U-Net 参考代码基础上添加数据增强模块，并测试和撰写简要说明。

9.2 请尝试在课程提供的 U-Net 参考代码基础上添加公式 (6) 中的权重，并测试和撰写简要说明。

参考文献 I