

第13章 计算机视觉

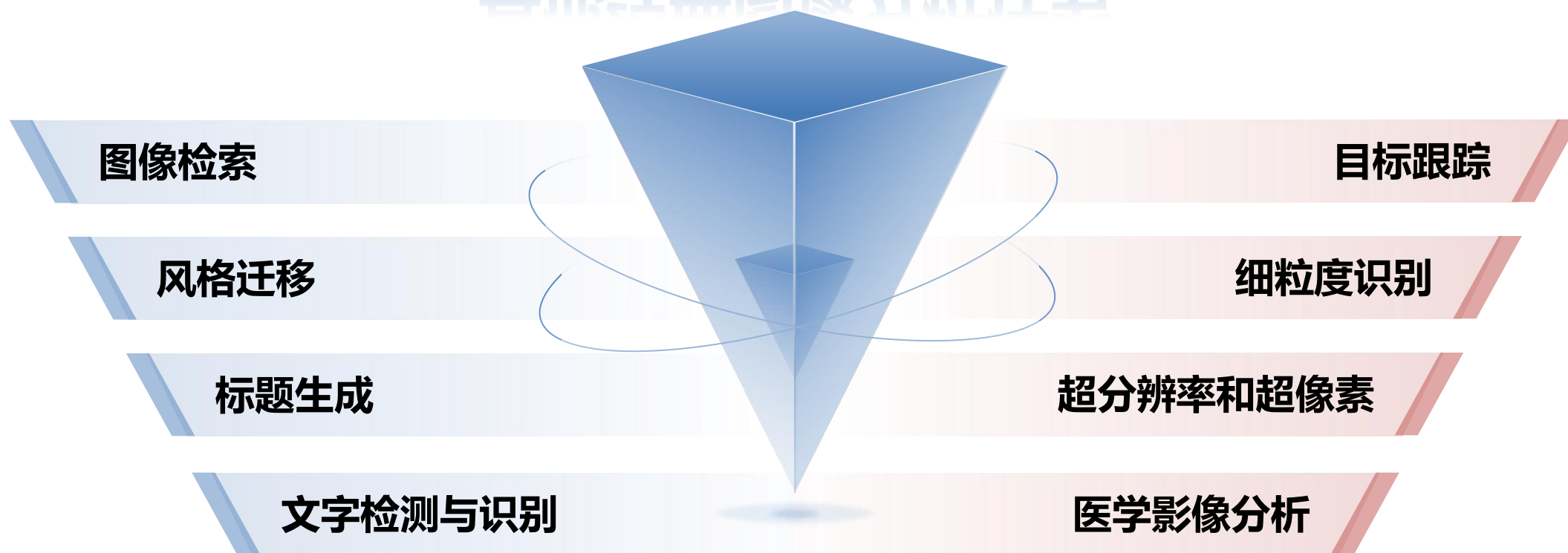
信息学院 (智能应用研究院)

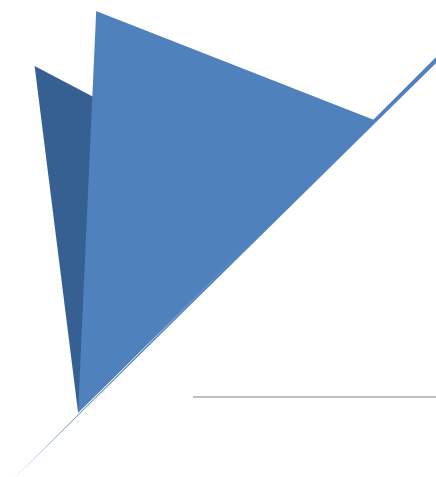
欧 新 宇



第13章 计算机视觉

其他经典图像分析任务





图像检索

基于卷积神经网络的图像检索

图像检索概述

任务描述：从海量的图像数据集中检索出和指定图片相匹配（相同/相似）的样本

基于内容的图像检索（CBIR）：根据图像的颜色、纹理和包含的物体、类别等信息进行检索。



(a)光照变化 (b)尺度变化 (c)视角变化 (d)遮挡 (d)背景杂乱

相同物体图像检索面临的挑战



(a)类内变化巨大(湖泊)



(b)类间相似性干扰

相似类别图像检索面临的挑战

图像检索的目标

- 速度快!
- 精确度高!
- 召回率高!

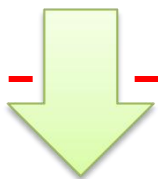
相同物体图像检索和相似类别图像检索面临的挑战

基于卷积神经网络的图像检索

图像特征的进化

手工特征

Hand Crafted Features
(颜色, 形状, 纹理, SIFT, SURF等)



Since 2012

深度特征

CNN based Features
(AlexNet, GoogleNet, VGGNet, ResNet等)

AlexNet

- 2012年ImageNet冠军, 第一次证明CNN在复杂模型下的有效性, 并且GPU使得训练在可接受的时间范围内得到结果。

VGGNet

- 2012年ImageNet的亚军, 证明了小卷积的堆叠是有用的, 并且结构简单清晰。

GoogleNet

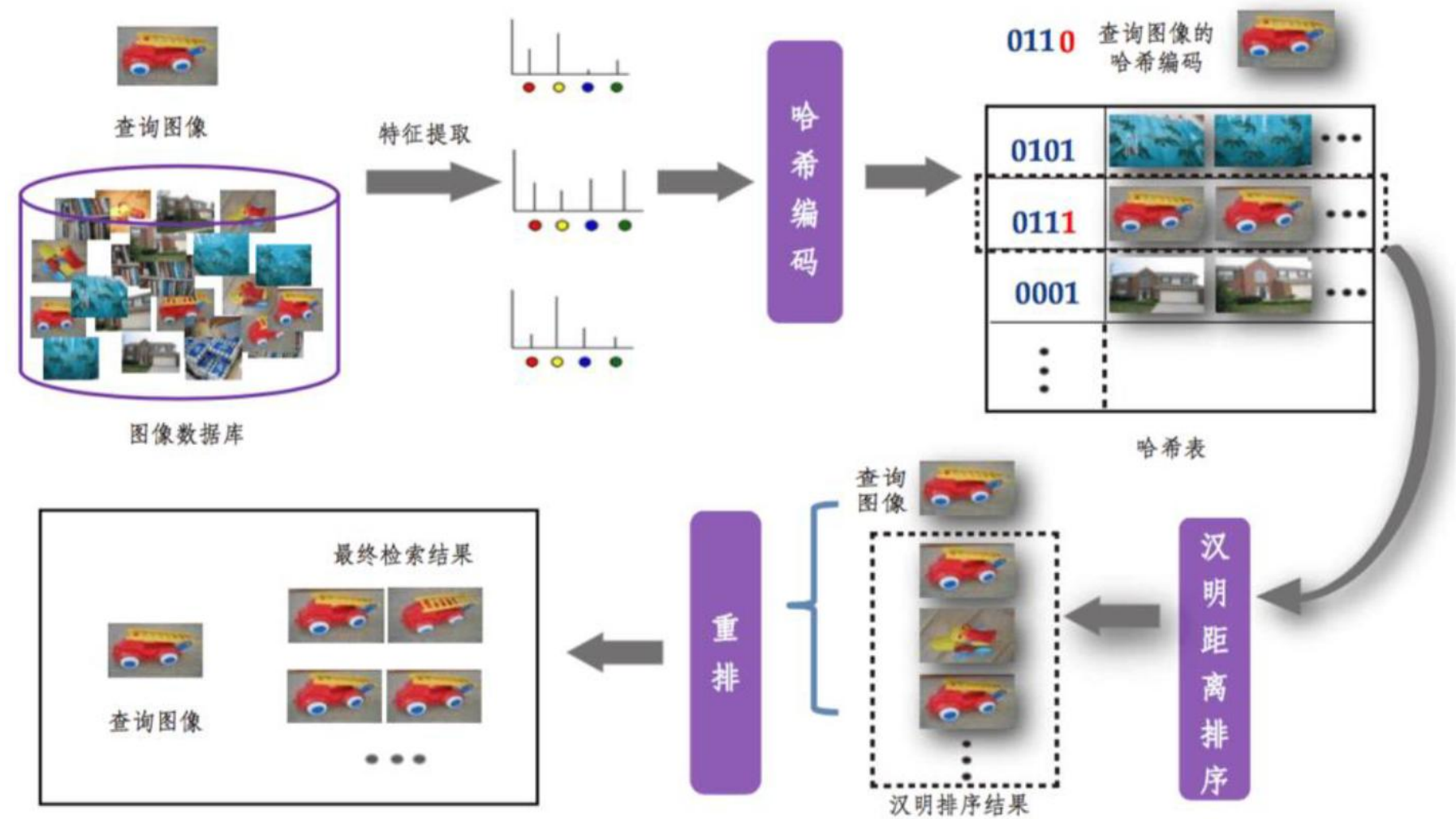
- 2012年ImageNet的冠军, 证明了更多的卷积, 更深的层次可以得到更好的结构。精心设计的网络也可以得到更好的性能。

ResNet

- 2015年ImageNet的冠军, 残差结构可以让网络变得更深

基于卷积神经网络的图像检索

基于哈希的图像检索



二进制编码：存储小、匹配计算速度快

基于卷积神经网络的图像检索

哈希方法简介

● 什么是图像的哈希编码?

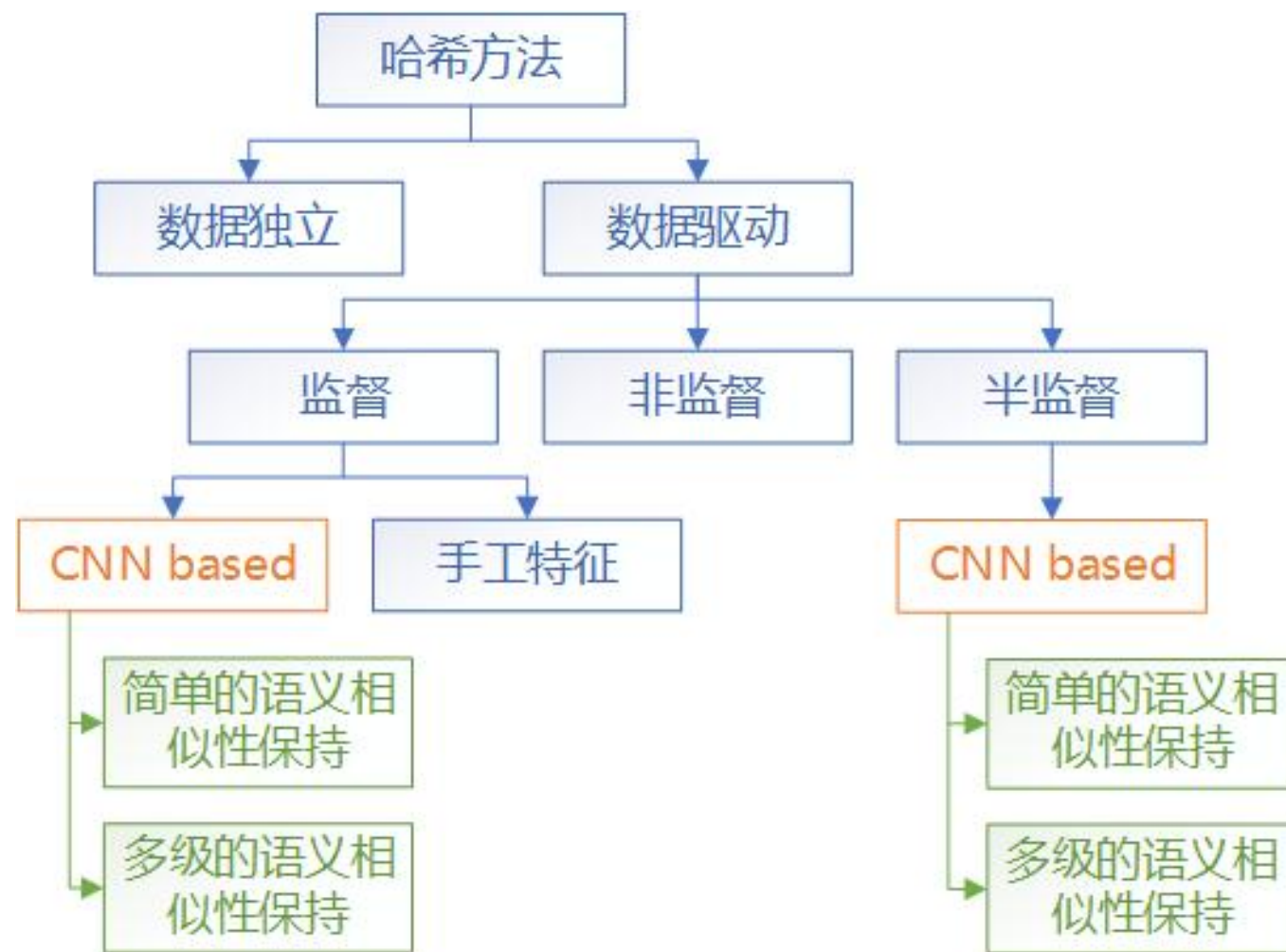
将图像的**高维内容特征**映射到**汉明空间** (二值空间) 中, 生成一个**低维的哈希序列(二值序列)**来表示图像。

● 哈希编码有什么优势?

- 降低图像检索系统对**内存空间**的需求
- 提高检索速度, 适应**海量**图片检索的需求

● 哈希编码分几个阶段?

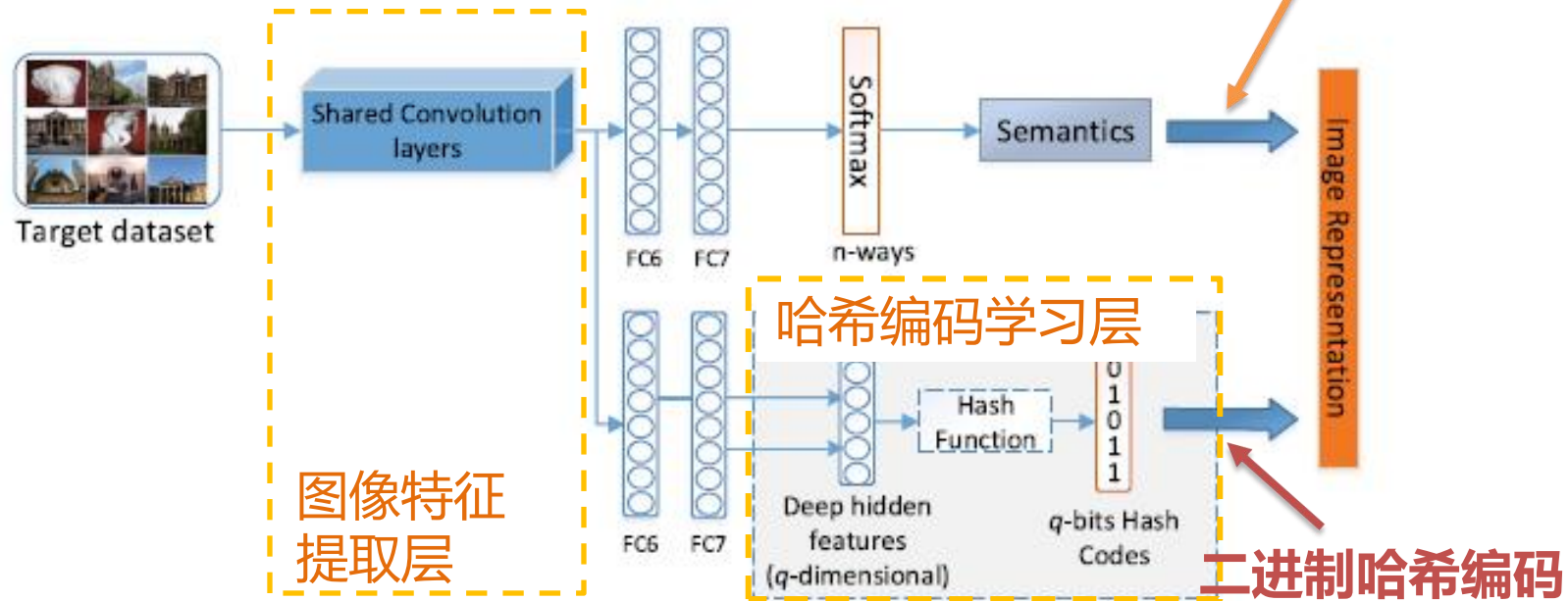
- 阶段一: 使用训练集对构造好的**哈希函数/特征**生成器进行训练。
- 阶段二: 将生成的样本的**特征**通过**哈希函数****转换**为**哈希编码**。



基于卷积神经网络的图像检索

基于卷积神经网络的哈希检索

Learning Hierarchical semantic representation on ConvNet



学习阶段

基于监督的**深度哈希编码**包括**特征提取层**和**哈希编码层**，监督学习指导整个模型的训练。与传统方法不同，深度哈希的特征提取和哈希编码学习是**同时进行的**，并且**相互影响**。

检索阶段

Retrieval via Hierarchical Deep Semantic Hashing



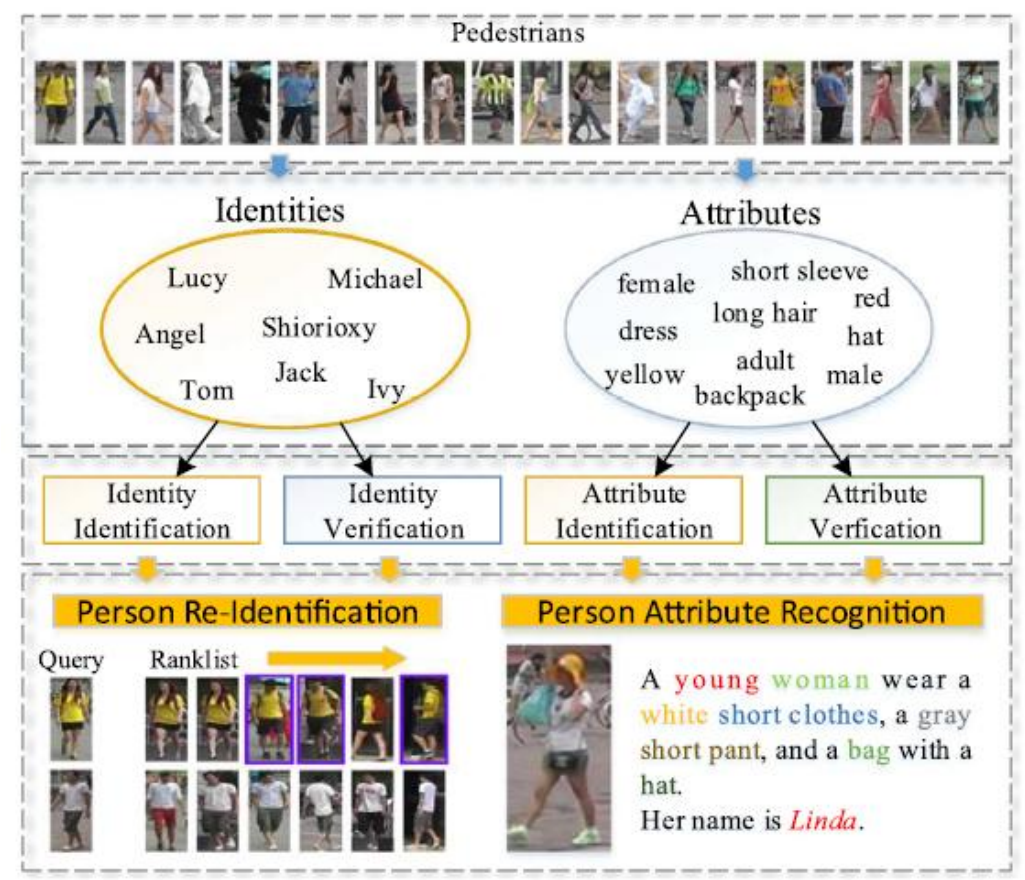
基于卷积神经网络的图像检索

行人再识别

行人再识别(Person re-identification)是一类很有挑战的任务，它需要从海量的监控视频中找到给定的待检索行人。它是包含目标检索、对象识别、属性识别等多个任务的混合任务。

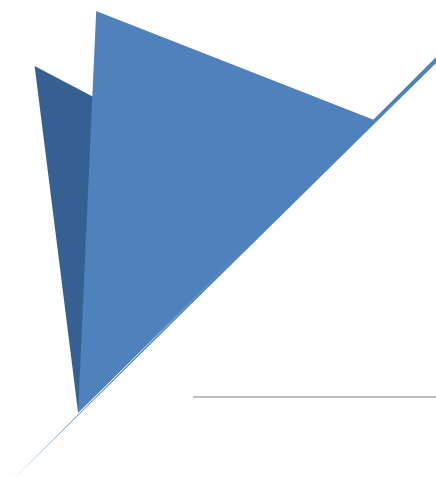


gender	male	99.98
hair	short hair	1
up	short sleeve	99.98
down	short clothing	1
clothes	pants	1
hat	no	1
backpack	yes	99.91
bag	no	1
handbag	no	1
age	teenager	1
upper-color	up yellow	94.23
lower-color	down gray	99.45





课堂互动 13.4.1

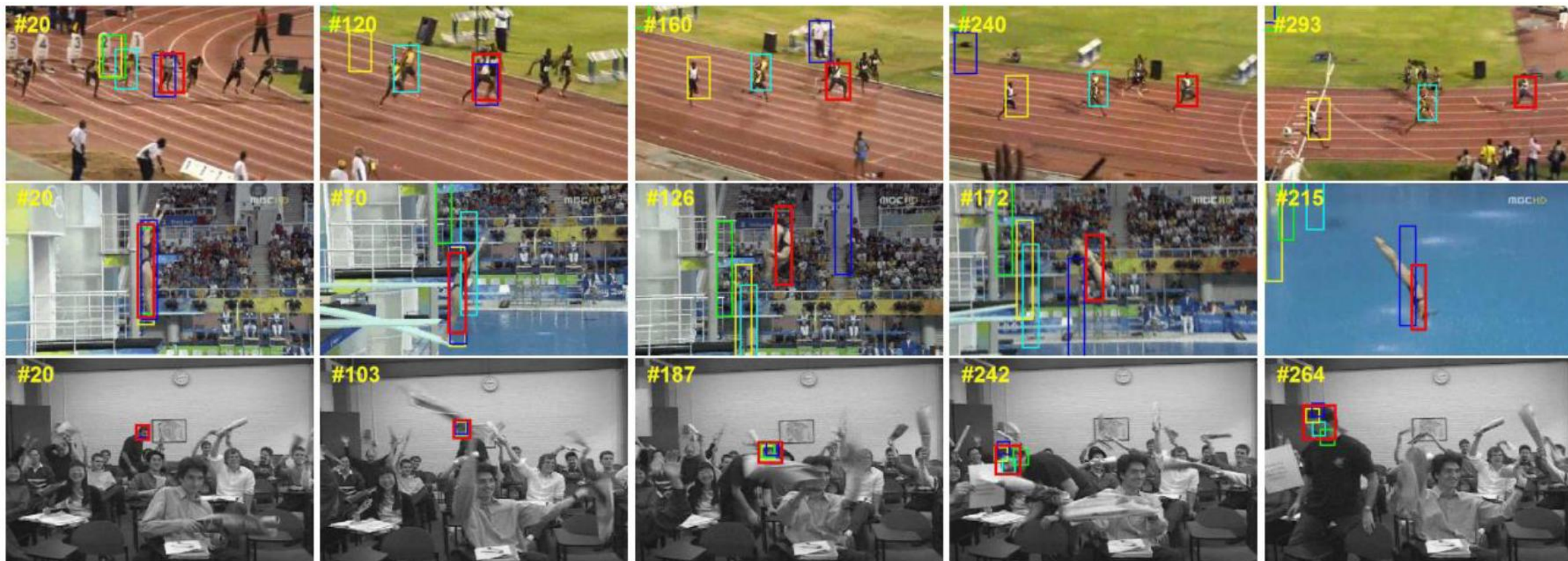


目标跟踪

基于卷积神经网络的目标跟踪

目标跟踪概述

任务描述：从**连续的视频序列**中建立**跟踪目标**的**位置信息**，得到完整的**运动轨迹**。**给定目标的初始位置**，**计算其在下一帧中的确切位置**。常见应用包括摄像机的跟踪对焦、无人机的自动目标跟踪、主目标跟踪等。



基于卷积神经网络的目标跟踪

在线跟踪方法 (MD-Net)

挑战：跟踪目标具有**特异性**。在某序列中是**目标**的物体，在其他序列中可能变为**背景**。

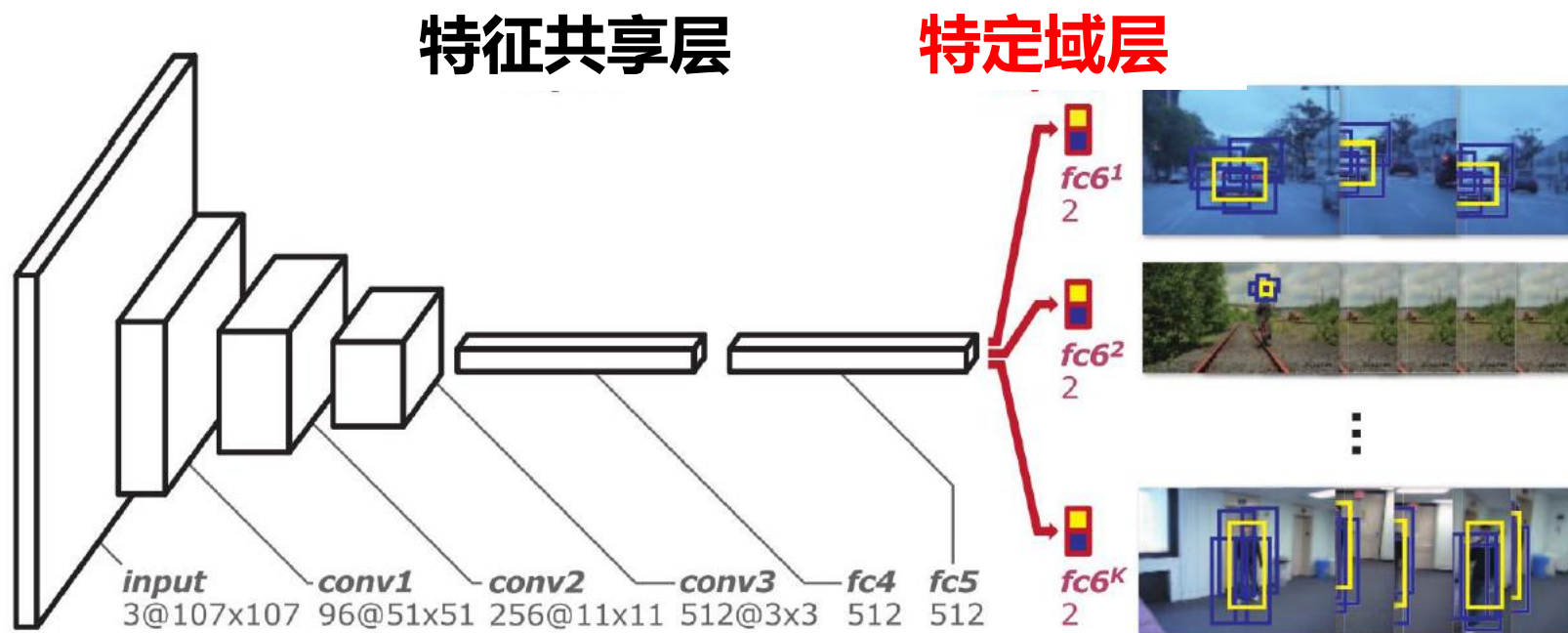
解决方案：使用**CNN主干网学习共享特征**，并使用**分支结构学习特定目标的独有特征**。

● 训练阶段

每个训练序列中提取**固定尺度**的**图像patch**，用于**同时训练CNN主干网和特定对象分支网络**

● 在线跟踪阶段

- ✓ 为在线跟踪任务新建一个fc层用于训练**二进制分类器**
- ✓ 基于**第一帧信息**，在线更新fc4-fc6 **[微调]**
- ✓ 提取多个候选跟踪区域，保留置信度最高的为检测结果

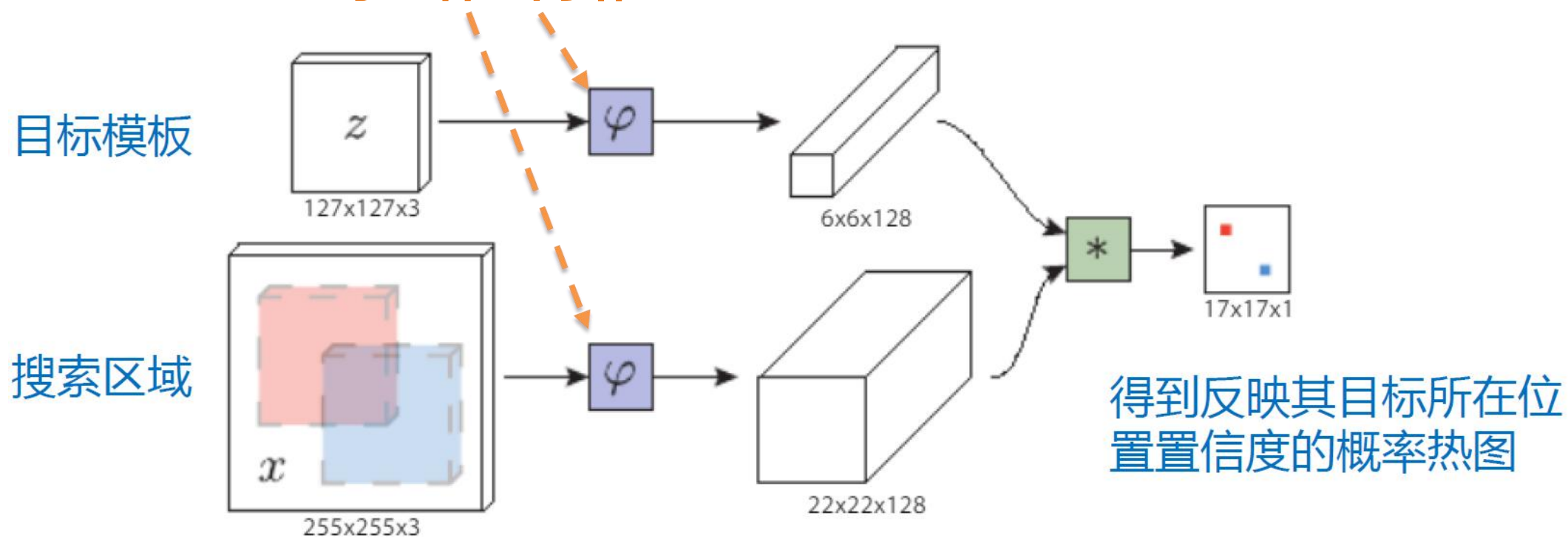


Hyeonseob Nam, Bohyung Han. Learning Multi-Domain Convolutional Neural Networks for Visual Tracking. CVPR 2016

基于卷积神经网络的目标跟踪

离线跟踪方法 (Siamese Network)

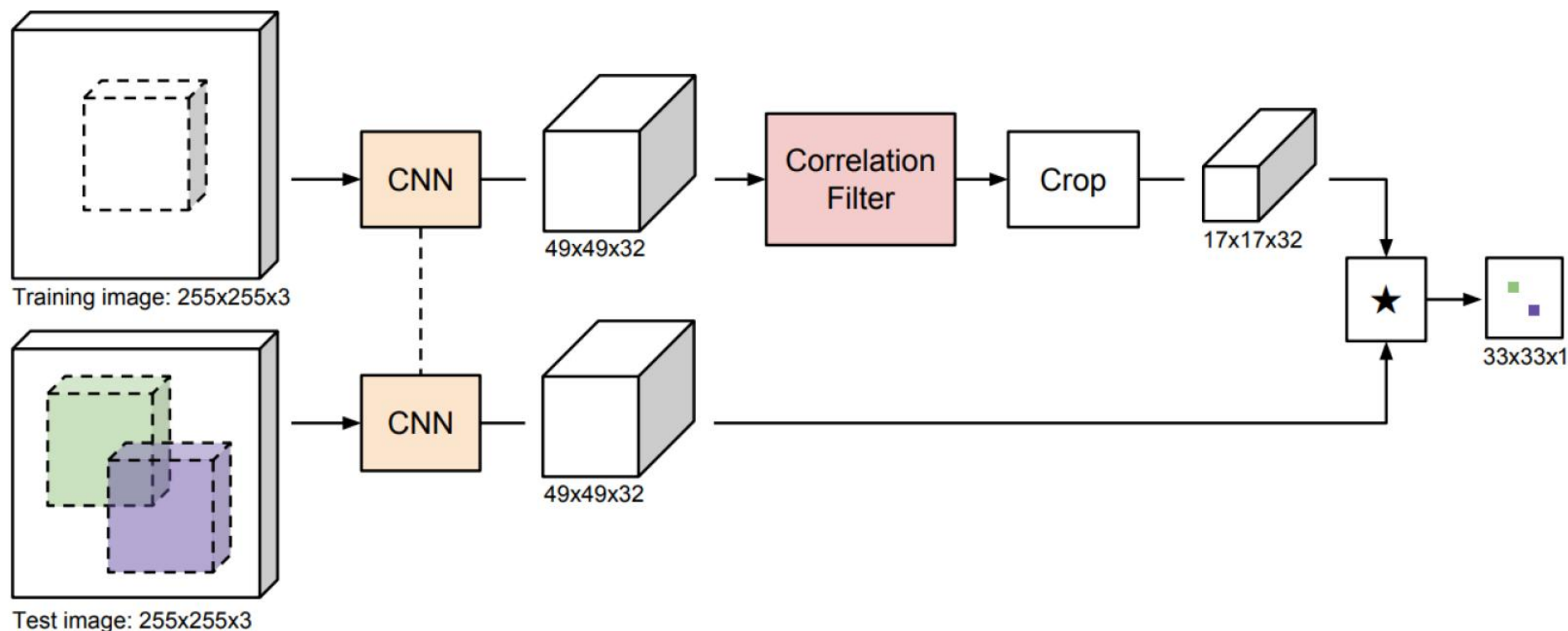
孪生神经网络 => Siamese



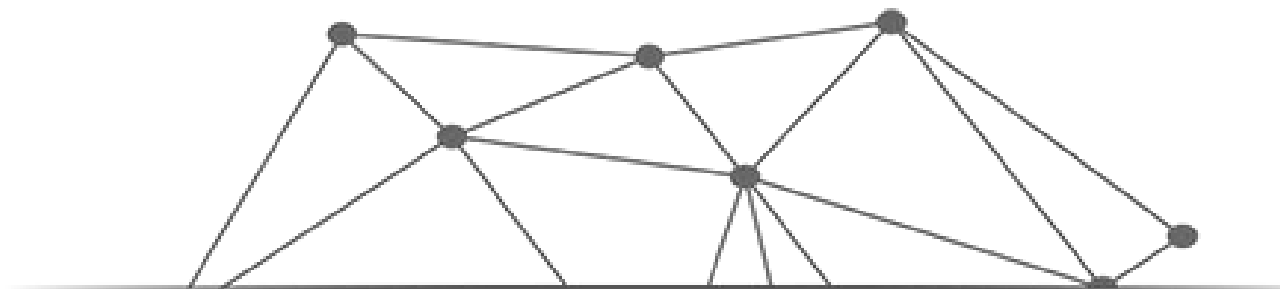
- ✓ 通过比较搜索区域与目标模板的相似度，获得概率热图
- ✓ 选取得分最高的点作为目标区域的中心点

基于卷积神经网络的目标跟踪

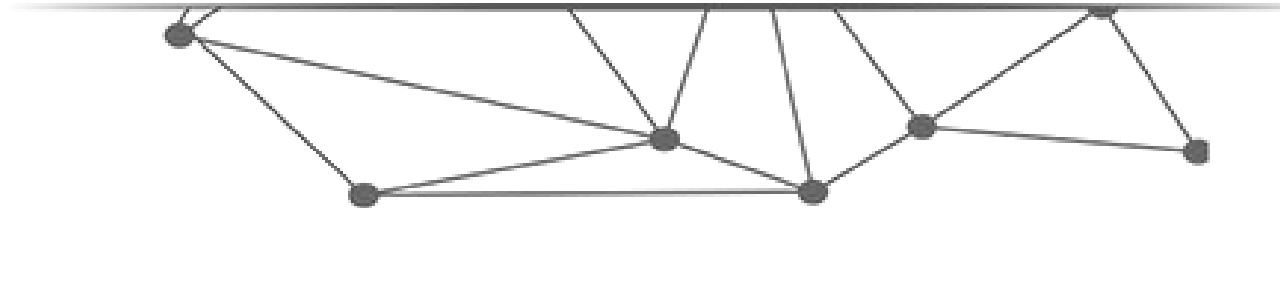
离线跟踪方法 (改进的Siamese Network)

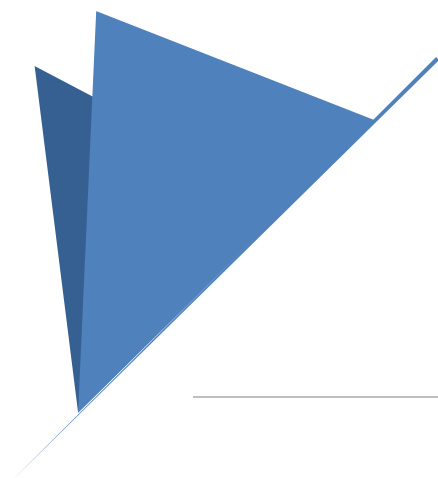


- ✓ 在**Siamese网络**上加入**CF层 (相关滤波)**，并且可以进行**端到端**的训练
- ✓ 相比**在线方法**，**离线更新系统**不需要一边测试一遍**finetune**，**速度更快**，但**性能稍差**，**相关滤波**在保证速度的同时提升了性能。



课堂互动 13.4.2





风格迁移

风格迁移

图像风格迁移

图像的信息来源于**内容(Content)**和**风格(Style)**，**内容**体现的是**结构**、**轮廓**信息；**风格**体现的是**纹理**、**样式**信息。



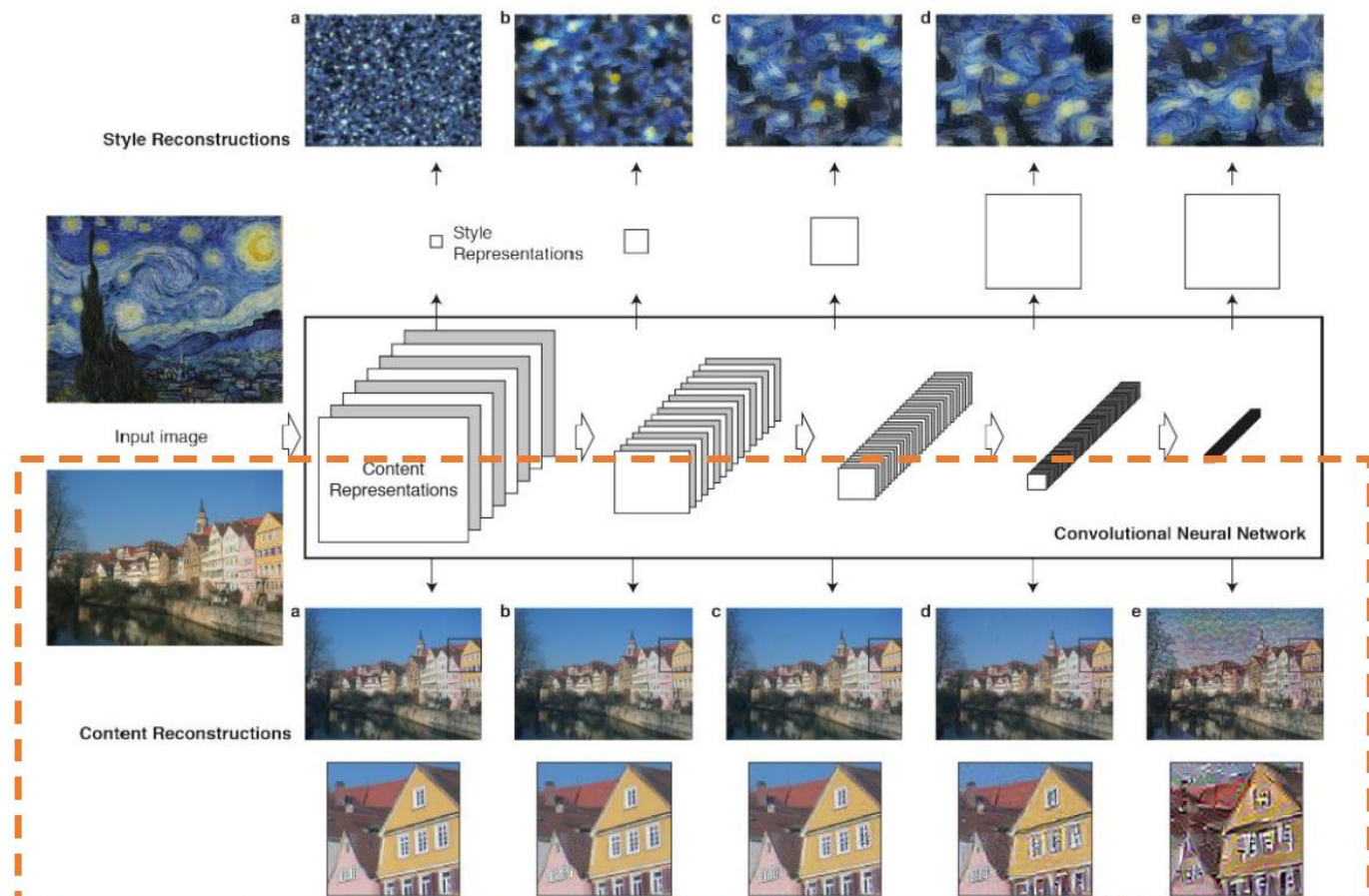
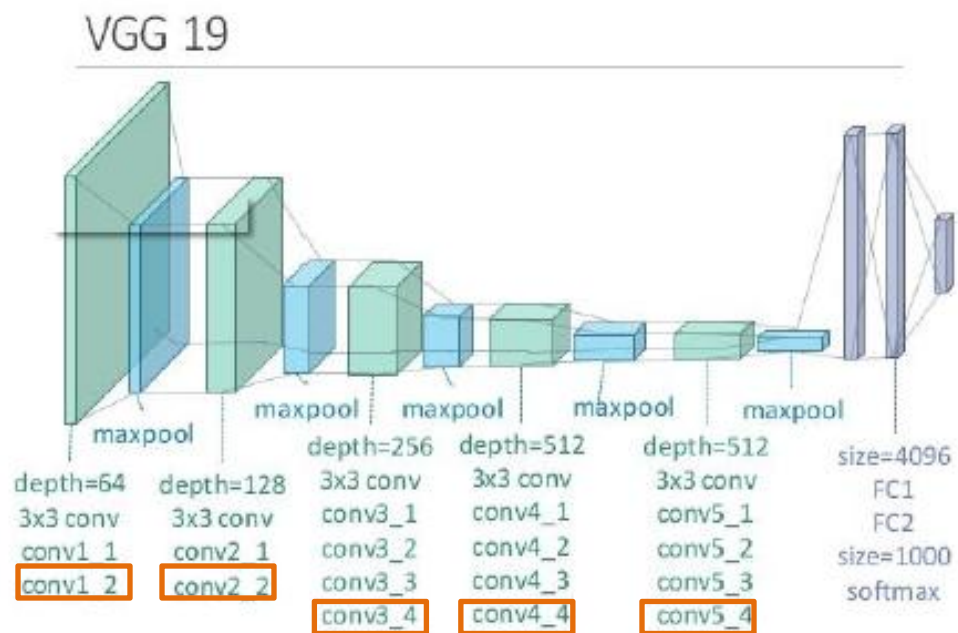
对于喜欢的绘画风格，怎么将其**风格**搬到另外一张图片上呢？

风格迁移

如何描述一张图的绘画内容

1. 确保新生成的图保留原本的内容大致不变：内容保持。

CNN的特征图是**输入**在不同层次上的表**达**，因此可以直接用来作为**内容保持**的表**征**形式。



A. Gatys, Alexander S. Ecker, Matthias Bethge. Image Style Transfer Using Convolutional Neural Network. CVPR 2016

风格迁移

如何描述一张图的绘画风格

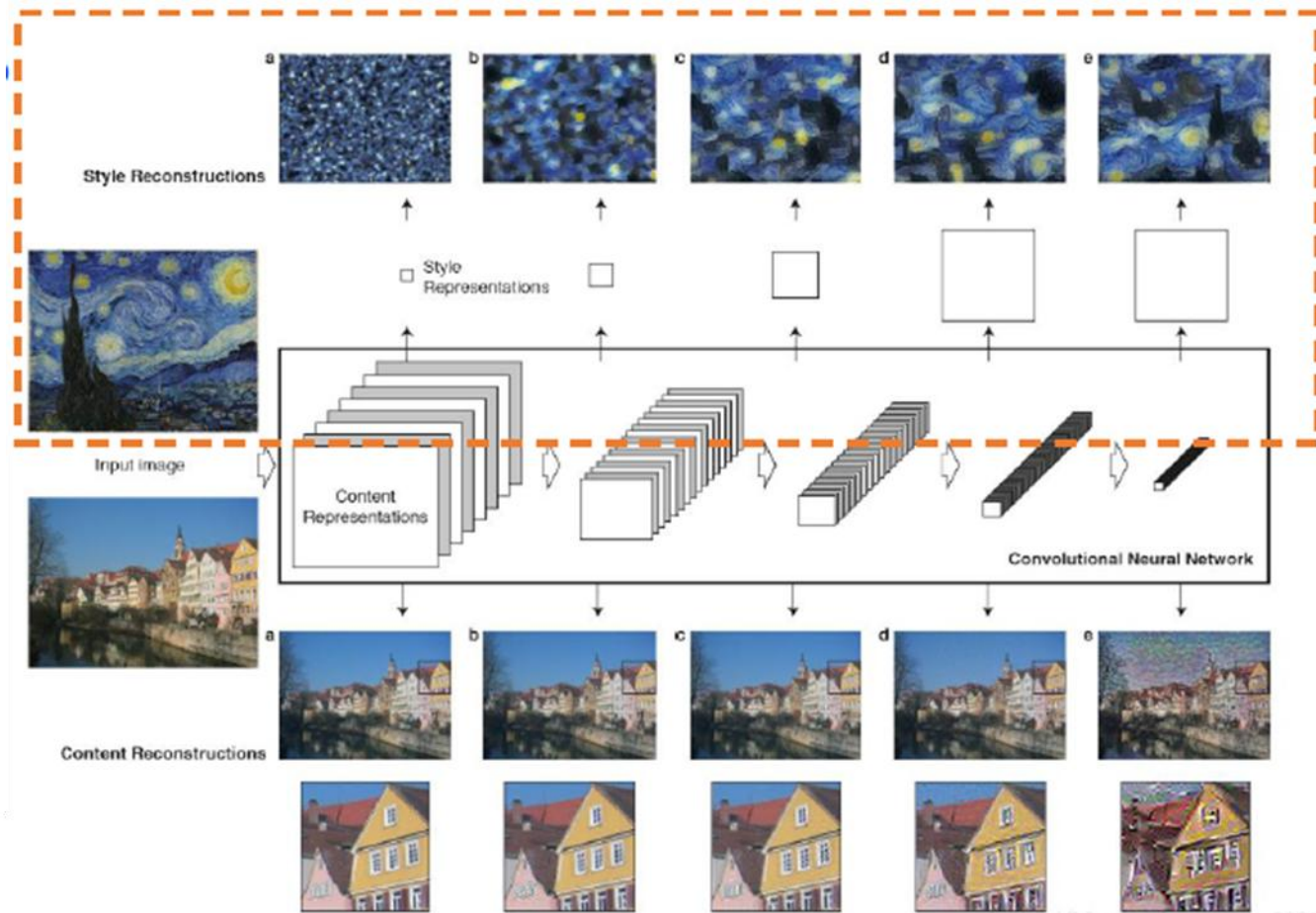
2. 确保新生成的图能够获取到参考样本的样式或风格：风格迁移

如何使用 *feature map* 获得风格呢？

在绘图中，**风格**可以理解为本的**纹理**。对于某一隐层，假设包含N个通道，每个通道的尺度为H×W。则，**两两间**的内积的和可以作为该层的**纹理(Texture)矩阵**，数学上称为**Gram 矩阵**：

$$G_{i,j}^l = \sum_k F_i^l[k] \cdot F_j^l[k]$$

其中， F 表示展开成1D的特征图。



Leon A. Gatys, A

Thomas Beyer, Image Style Transfer Using Convolutional Neural Network, CVPR 2016

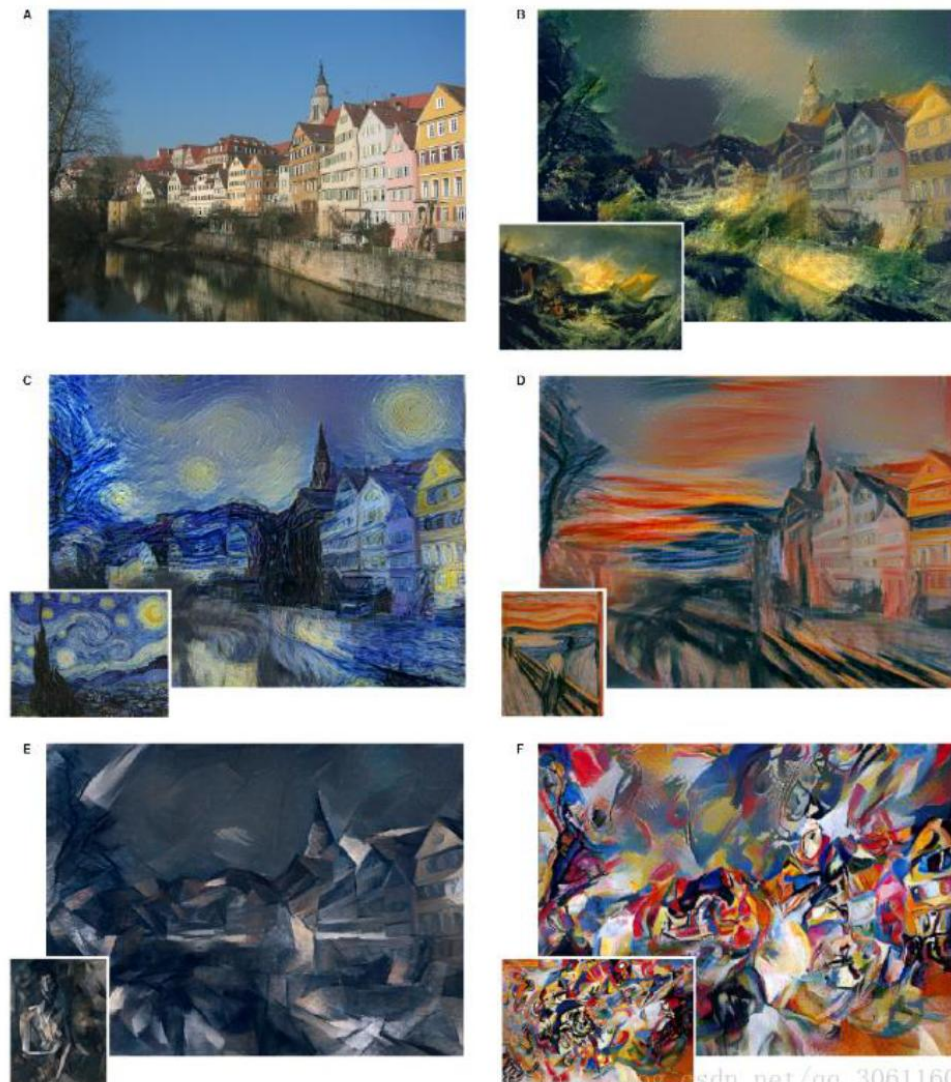
http://blog.csdn.net/qg_30611

如何权衡内容和风格

● 目标函数

$$L_{total} = \alpha L_{content} + \beta L_{style}$$

- 图像的**内容**和**风格**信息是无法完全地分开，但是可以在学习模板图像的**内容**或**风格**信息时，通过**最小化损失函数**来实现两种信息的有机融合。
- 通过调节权重，还可以**调节内容**和**风格**的**侧重点**。
- **内容**和**风格**可以来源于**同一幅图像**，也可以来源于**不同图像**。
- **不同层**具有**不同粒度**的风格



Leon A. Gatys, Alexander S.Ecker, Matthias Bethge. I

sdn.net/qq_30611601

风格迁移

如何获得风格转换后的图像

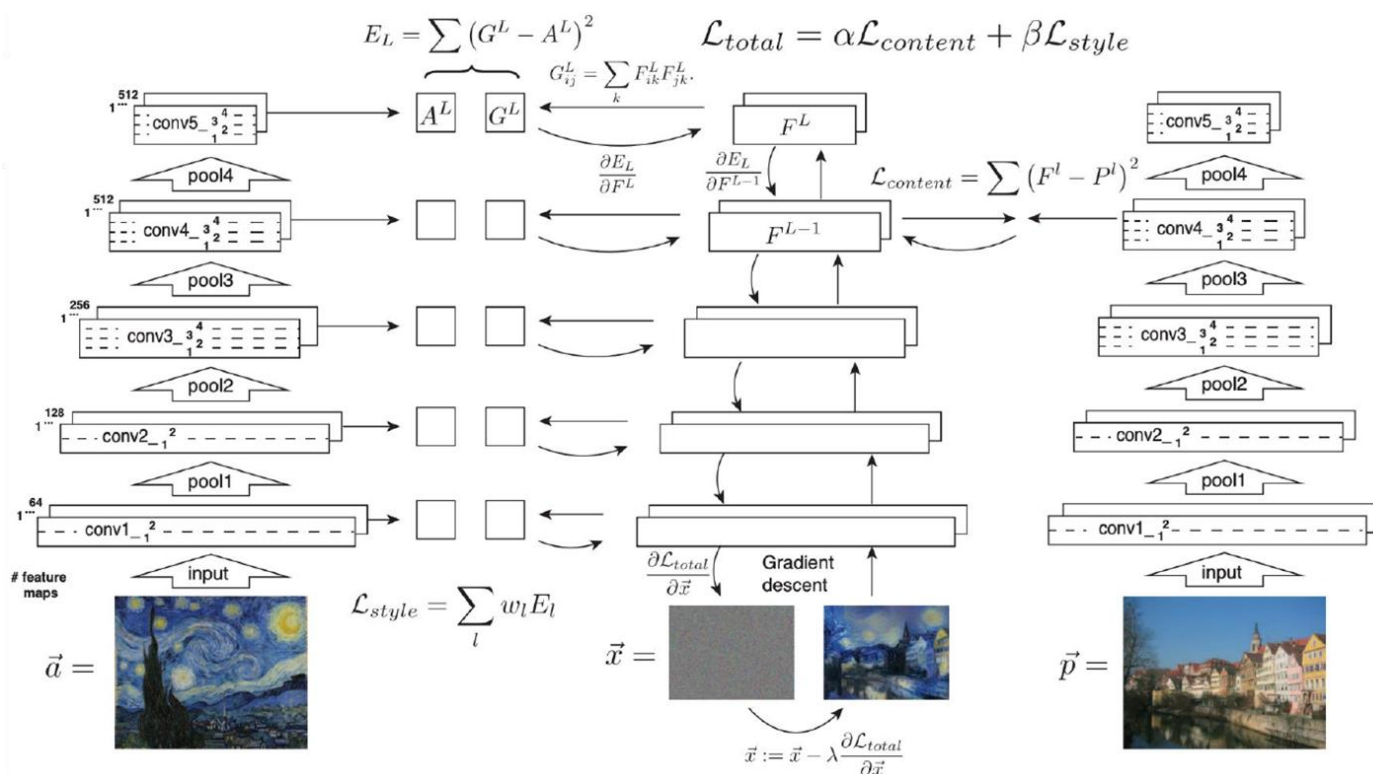
● 内容分支

内容图像的每个层都会生成 feature map, 通常经验选取最清晰层的卷积特征图与生成图像计算均方误差, 获得内容损失。

● 风格分支

风格图像的每个层也会生成 feature map, 通常会混合多个层的特征和生成图像之间的Gram矩阵来计算风格损失。

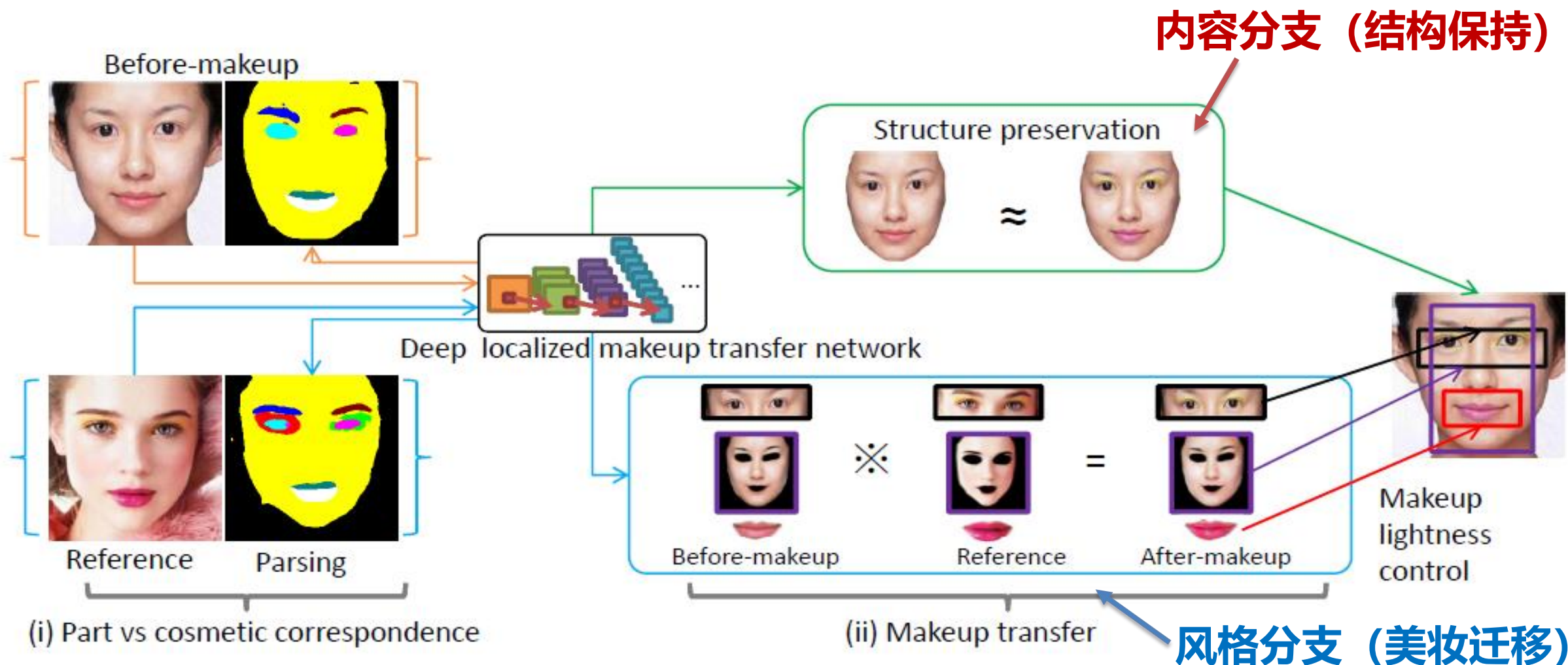
风格转换是一种生成算法, 它从**高斯白噪声**开始, **逐渐**生成一张合成图像 (风格化图像)。



Leon A. Gatys, Alexander S. Ecker, Matthias Bethge. Image Style Transfer Using Convolutional Neural Network. CVPR 2016

风格迁移

风格迁移的应用 – 美妆迁移



Si Liu, Xinyu Ou, Ruihe Qian, Wei Wang, Xiaocun Cao. Makeup like a superstar: Deep localized makeup transfer network. IJCAI 2016
 Xinyu Ou, Si Liu, Xiaochun Cao, Hefei Ling. Beauty eMakeup: A deep makeup transfer system. ACM MM 2016

风格迁移

$$A^* = \operatorname{argmin}_{A \in \mathbb{R}^{H \times W \times C}} \lambda_l (R_{up}(A) + R_{low}(A)) + \lambda_f R_f(A) + \lambda_s R_s(A) + \lambda_e (R_l(A) + R_r(A)) + R_{V\beta}(A)$$

唇彩和粉底迁移:

$$A^* = \operatorname{argmin}_{A \in \mathbb{R}^{H \times W \times C}} \sum_{l=1}^L \left\| \Omega_{ij}^l(A(s'_b)) - \Omega_{ij}^l(R(s'_r)) \right\|_2^2$$

人脸结构保持和眼影迁移:

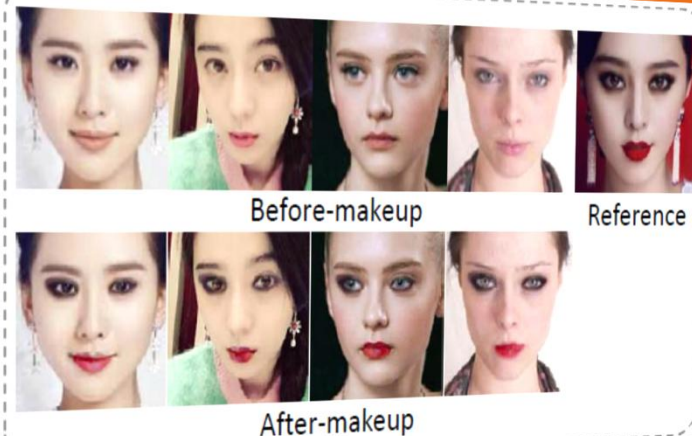
$$A^* = \operatorname{argmin}_{A \in \mathbb{R}^{H \times W \times C}} \left\| P \left(\Omega^l(A(s'_b)) \right) - P \left(\Omega^l(R(s'_r)) \right) \right\|_2^2$$

总方差:

$$R_{V\beta}(A) = \sum_{i,j} \left((A_{i,j+1} - A_{ij})^2 + (A_{i+1,j} - A_{ij})^2 \right)^{\frac{\beta}{2}}$$

Si Liu, Xinyu Ou

风格迁移

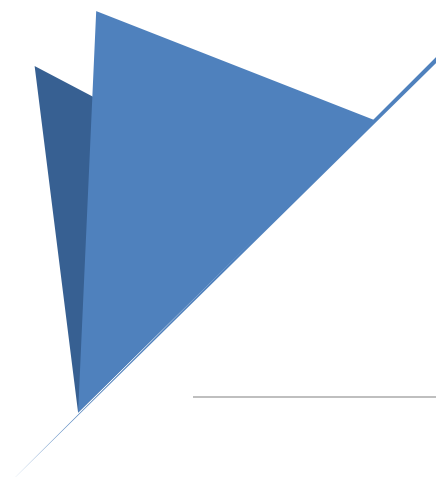


不同的女孩使用相同的妆容



同一个女孩使用不同的妆容

like a superstar: Deep localized makeup transfer network. IJCAI 2016
 Xinyu Ou, Si Liu, Xiaohui Hu, Ruifeng Wang, Xinyu Ou, Si Liu, Xiaohui Hu, Ruifeng Wang. Beauty eMakeup: A deep makeup transfer system. ACM MM 2016



细粒度识别

细粒度识别

什么是细粒度识别



细粒度识别

细粒度分类模型 Mask-CNN

Caltech-UCSD Birds-200

数据集包含:

- 类别标注
- 部件标注(15parts/image)
- 总共322种二进制属性。



(a) Part annotations



(b) Part rectangles



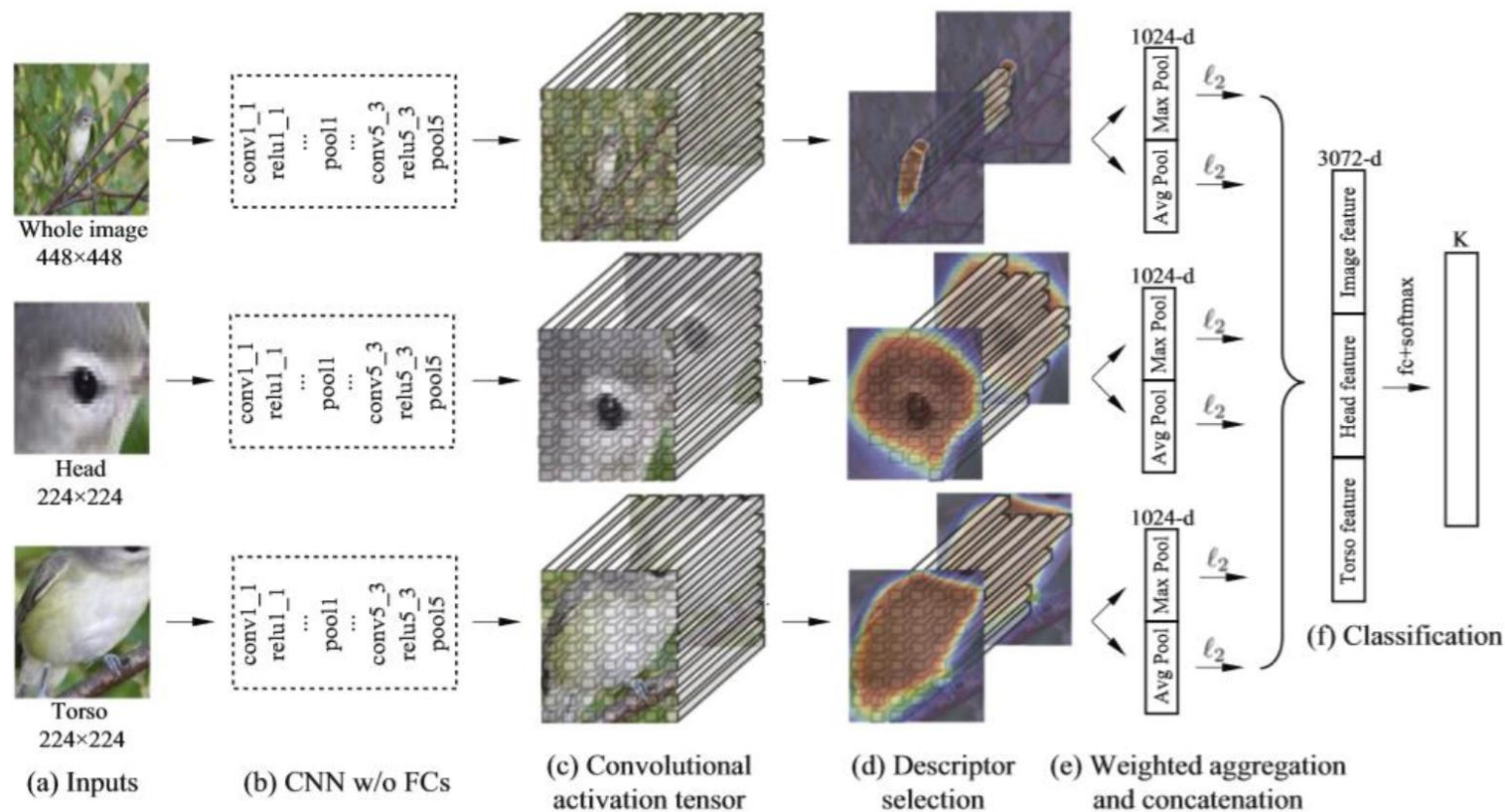
细粒度识别

细粒度分类模型 Mask-CNN

Mask-CNN

- 将部件标注转换为边界框（头部/躯干），并作为 mask-GT
- Mask-CNN分为三个流，分别处理原图、头、躯干
- 三个分支分别使用VGG进行特征提取，最后组合后作为样本的总特征

缺点：需要部件和掩膜标注



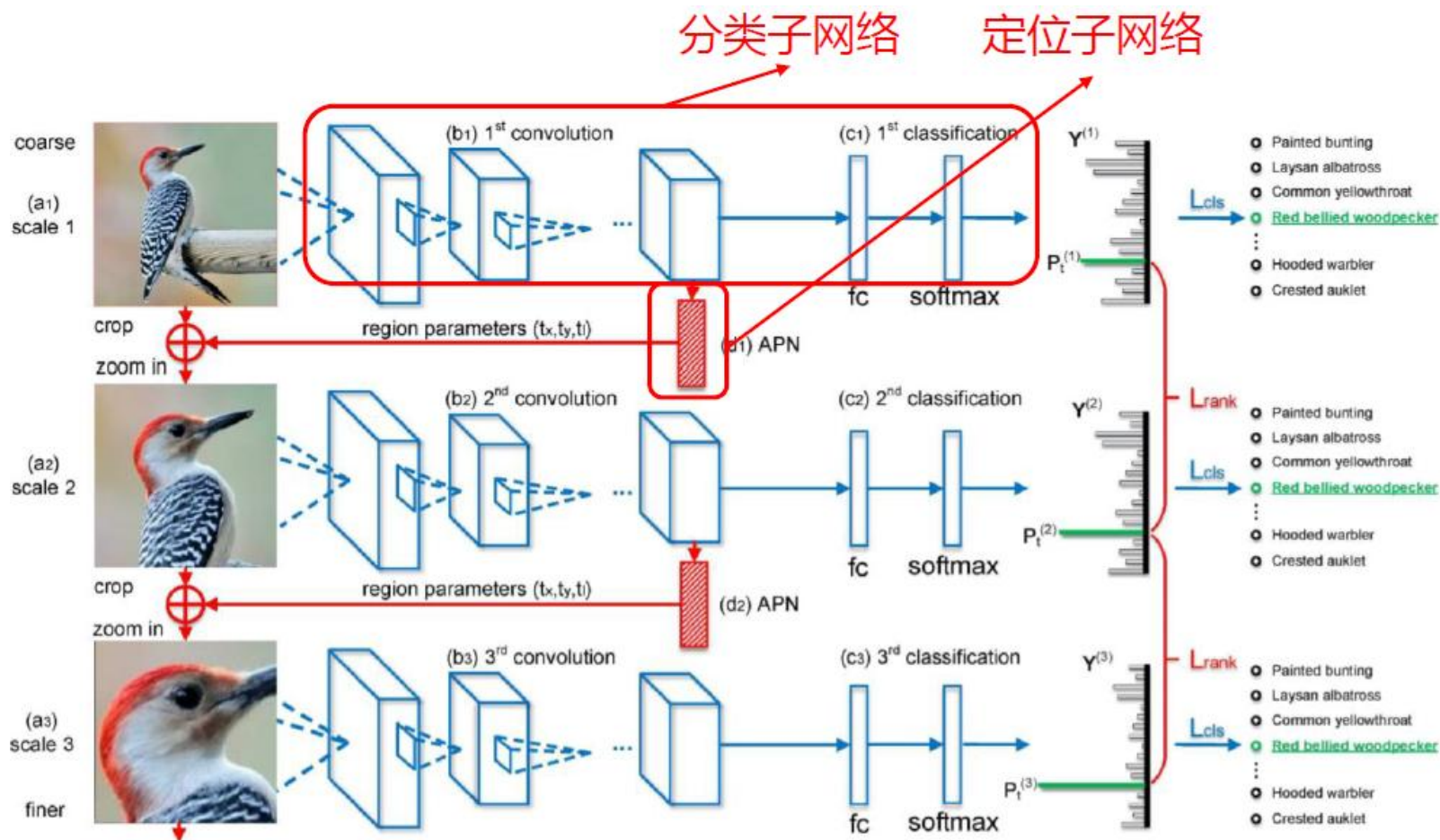
Xinshen Wei, Chenwei Xie, Jianxin Wu, Chunhua Shen. Mask-CNN: Localizing parts and selecting descriptors for fine-grained bird species categorization. Pattern Recognition 2018

细粒度识别

细粒度分类模型 RA-CNN

RA-CNN

- 使用**弱监督方法**逐层精化**注意力区域**，并使用多损失函数提高识别性能
- **注意力建议网络(APN)**，用于提取**注意力(Attention)**区域
- **排序损失(Rank Loss)**，提取置信度最高的类别

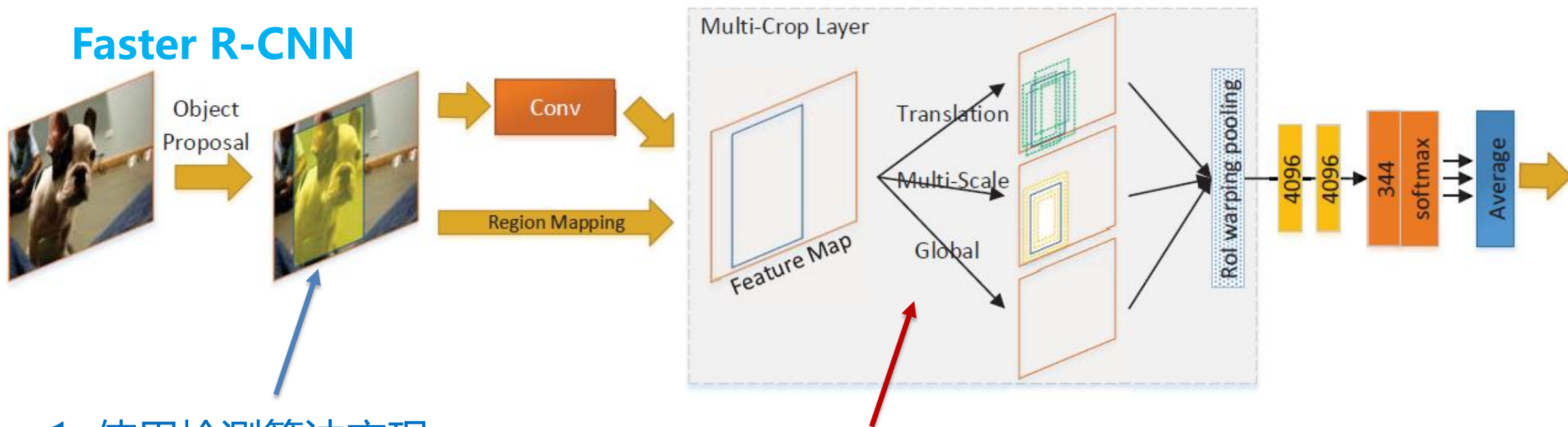


Jianlong Fu, Heliang Zheng, Tao Mei. Look Closer to See Better: Recurrent Attention Convolutional Neural Network for Fine-grained Image Recognition. CVPR 2017

细粒度识别

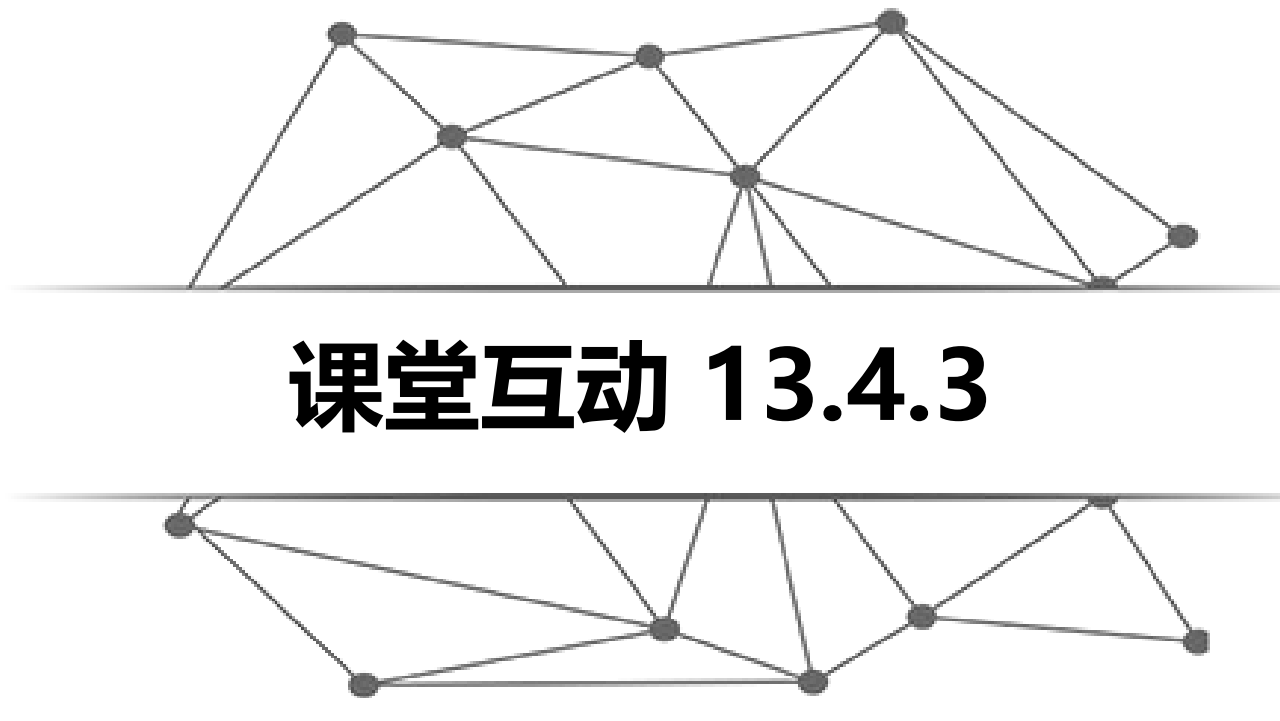
细粒度分类模型 Deep Multi-Context Network

Faster R-CNN

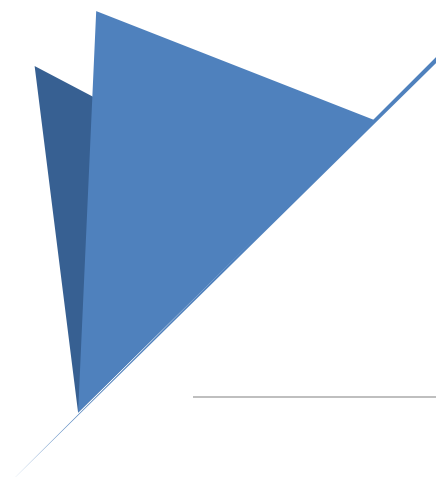


1. 使用检测算法实现
特定目标的获取

2. 使用多上下文方式
提取特征



课堂互动 13.4.3



标题生成

什么是图像标题生成 (Image Caption)

输入： 图片

输出： 对图片进行描述的文本

步骤：

1. 理解图片的内容和含义
2. 在内容和语言之间建立语义关联
3. 基于语义生成文字描述



Two people are walking down at river in a wooded area

MS Captivator

基本方法

1. **Detect words**: 使用目标检测算法和属性识别算法实现对象和属性的识别, 构造实体词;
2. **Generate Sentences**: 使用多实例学习 (Multiple Instance Learning, MIL) 的弱监督算法进行造句
3. **Re-rank Sentences**: 重整句子结构

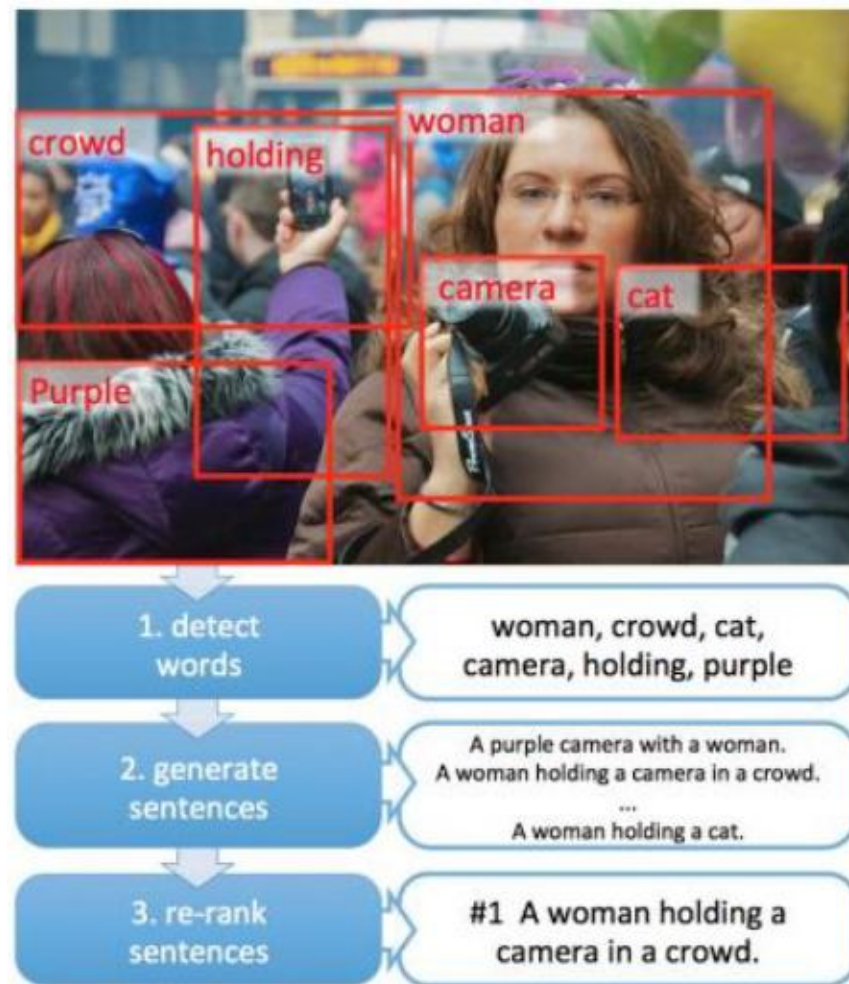
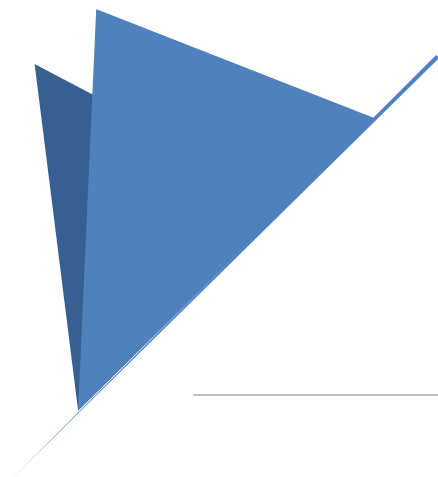


Figure 1. An illustrative example of our pipeline. 7383



图像超分辨率和超像素

图像超分辨率和超像素

什么是图像超分辨率



图像超分辨率是指由一幅/多幅低分辨率图像或图像序列恢复出高分辨率图像

图像超分辨率和超像素

常见的图像超分辨率应用



监控视频优化



老电影修复



智能插帧

常见的超分辨率方法包括插值算法和生成学习算法。相对于传统的双三次插值及单纯的CNN模型SRResNet，基于生成对抗网络的SRGAN网络可以重建出**细节更出众**的超分辨率图像。

图像超分辨率和超像素

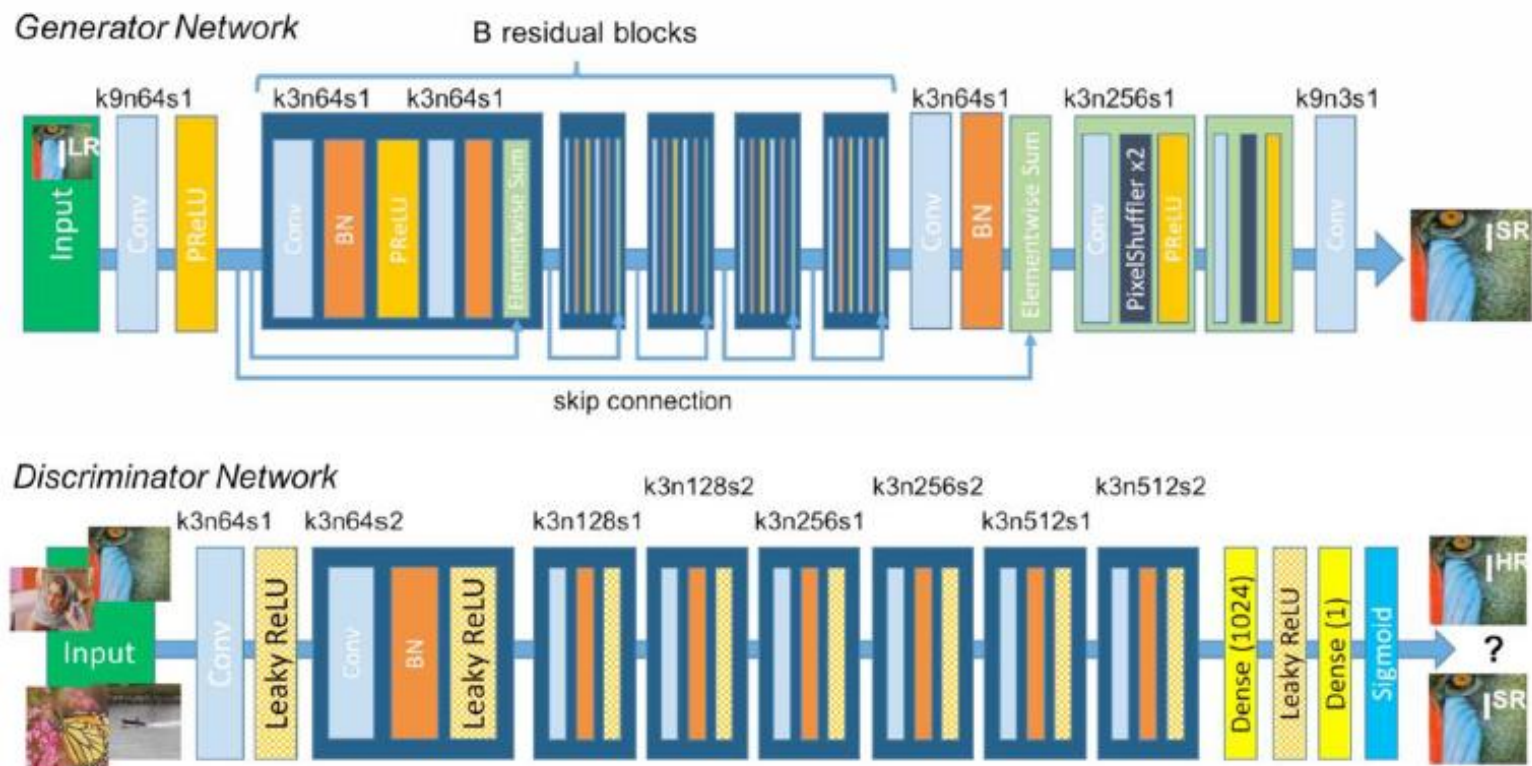
基于GAN重建的图像超分辨率

SRGAN

● 训练阶段

- ✓ 生成器将输入的低分辨率图像生成高分辨率图像。
- ✓ 判别器判别输入图像是生成器生成的高分辨率图像，还是原始的高分辨率图像
- ✓ 生成器和判别器通过交替迭代，最终使生成器生成的图像更接近于原始高分图像

- 测试阶段，生成器将实现对输入的低分辨率图像的超分辨率重建

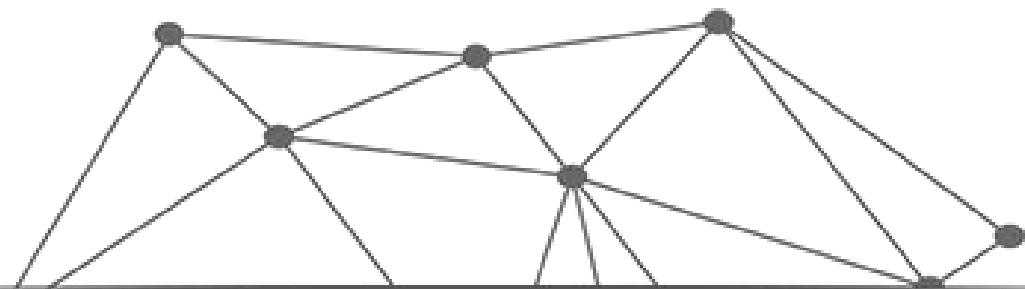


图像超分辨率和超像素

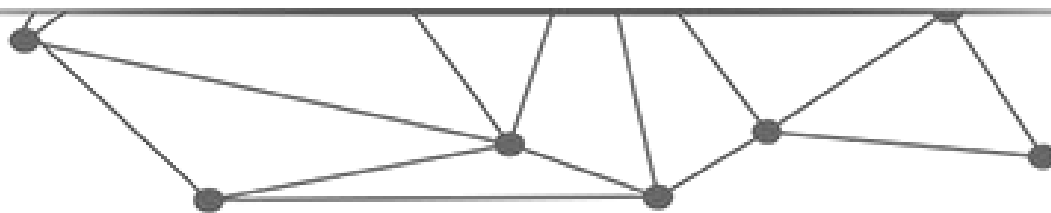
超像素

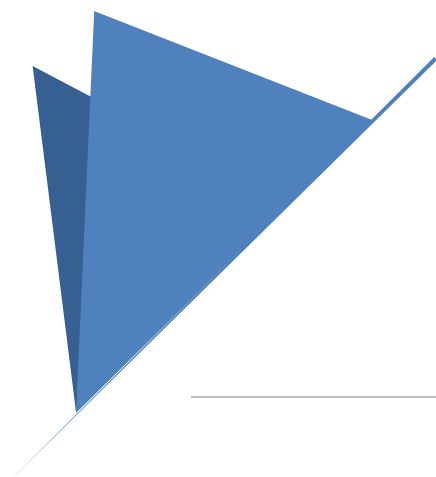


- **超像素(Superpixels)**是一系列像素的集合，这些像素具有**相似**的颜色、纹理等特征，距离比较近。
- 超像素**区域保留了**进一步分割的**有效信息**，且**不破坏**图像中物体的**边界信息**。
- 超像素大大**减小**图像处理的**规模**，带来计算上的优势，常被应用到计算机视觉的**前期处理**中。



课堂互动 13.4.4





医学影像分析

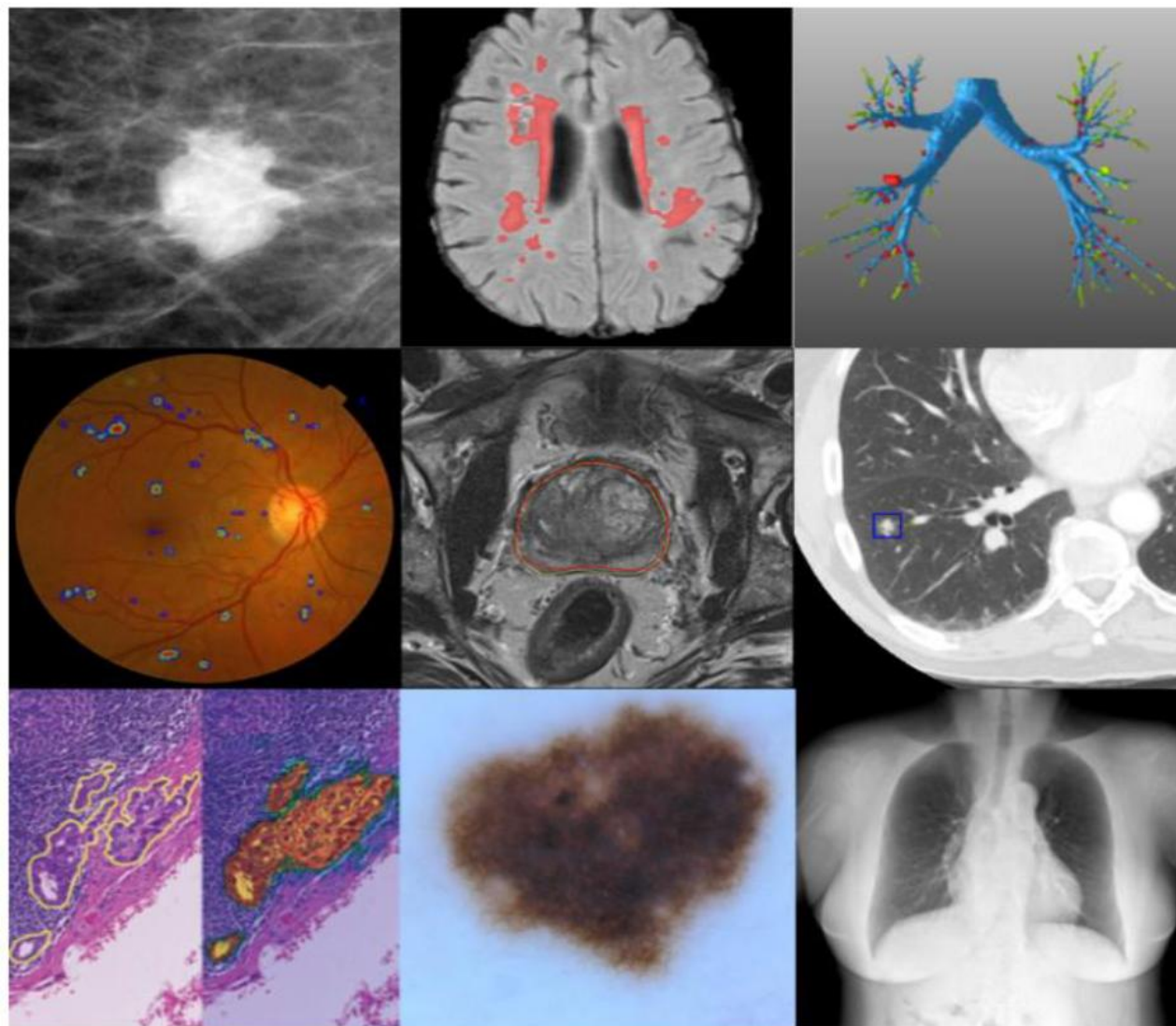
医学影像分析

深度学习在医学影像领域的应用

**深度学习常规任务（分类、检测、分割）
在医学影像领域的应用：**

例如：乳腺肿块分类任务、脑损伤部位的分割、气道树病变部位分割检测，糖尿病视网膜病变分类，前列腺分割，结节分类，淋巴结中的乳腺癌转移检测，皮肤病变分类，x射线骨髓抑制检测等

可见，深度学习技术在医学影像分析领域应用十分广泛。

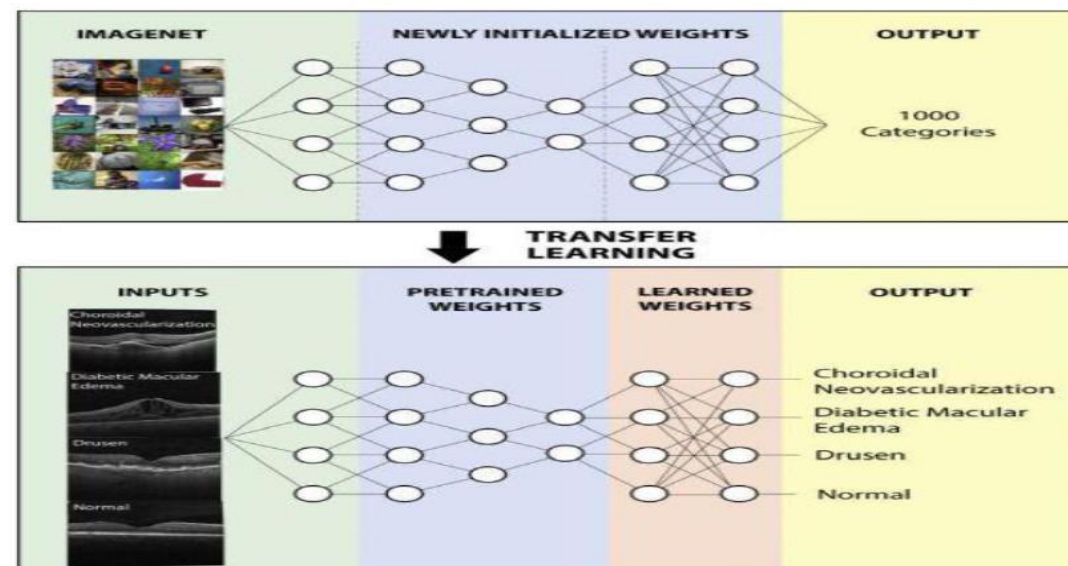
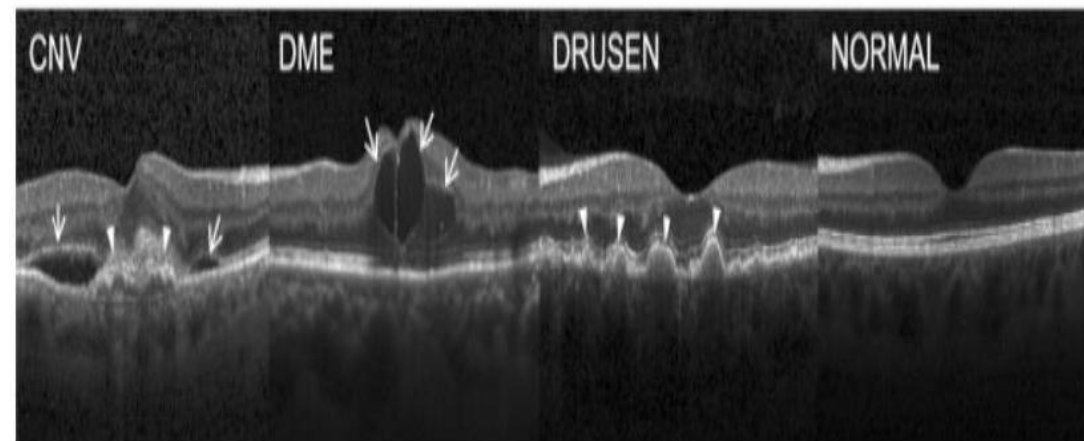


医学影像分析

实例1 – 眼底病变分类

在美国，近1000万人患有年龄相关性黄斑变性，近75万名年龄在40岁以上的人患有糖尿病性黄斑水肿，视网膜OCT图像对于指导治疗方案至关重要。

右图给出了脉络膜新生血管(CNV)，糖尿病性黄斑水肿 (DME)，玻璃疣 (DRUSEN) 与正常 (Normal) 四种情况下的OCT图像。



2018年，加州大学圣迭戈分校的张康课题组在顶级期刊Cell上发表了用于精确诊断致盲性视网膜疾病的论文，实现了图像精准分类并荣登封面。

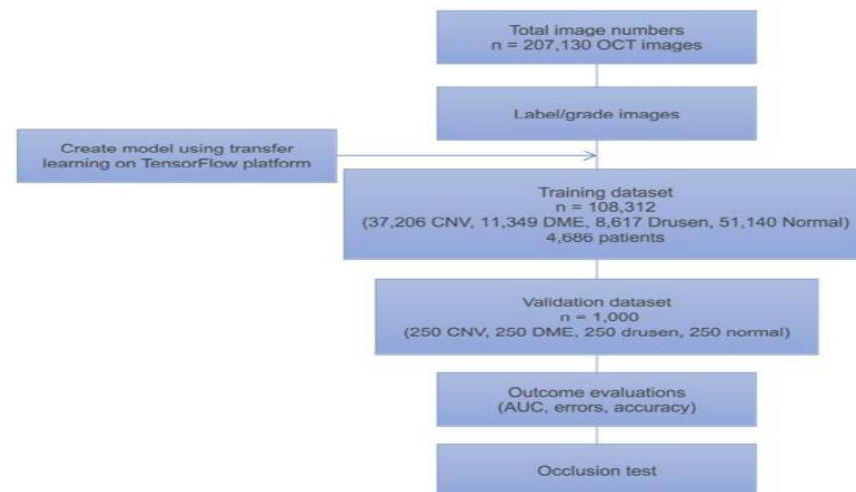
Kermay DS, Goldbaum M, Cai W, et.al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Cell, 2018, 172(5):1122-1131

实例2 – 肺炎胸片鉴别

方法流程

1. 采集OCT图像数据，并请专家进行标注
2. 训练CNN分类模型，并在验证集上进行评价
3. 对预测图像进行预测

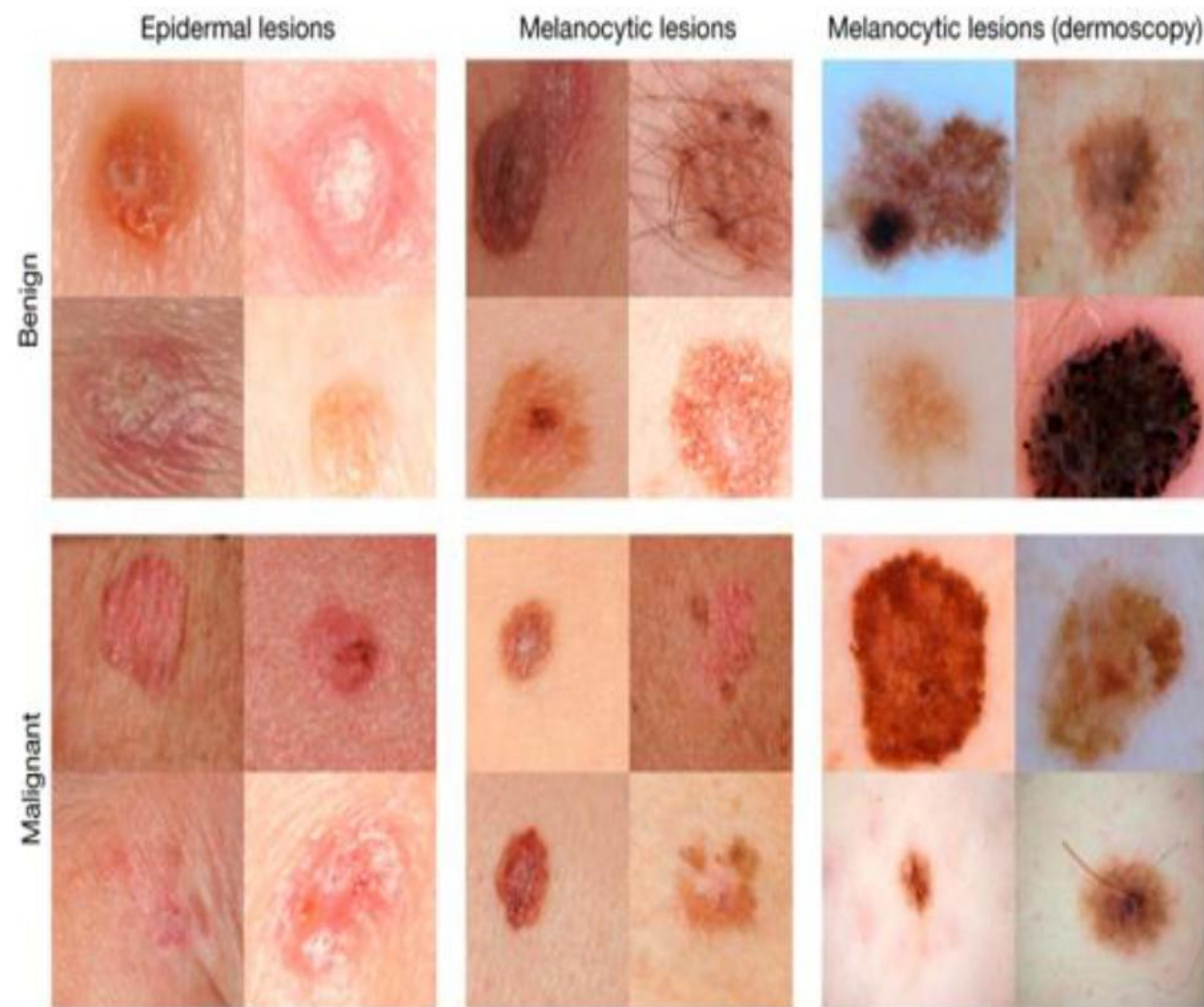
该方法与人类专家观察结果对比，发现在诊断眼底病变分类的评估指标上**高于人类专家**，且更稳定。该方法迁移到肺炎胸片的鉴别准确度高达**90%以上**。



实例2 – 皮肤病变分类

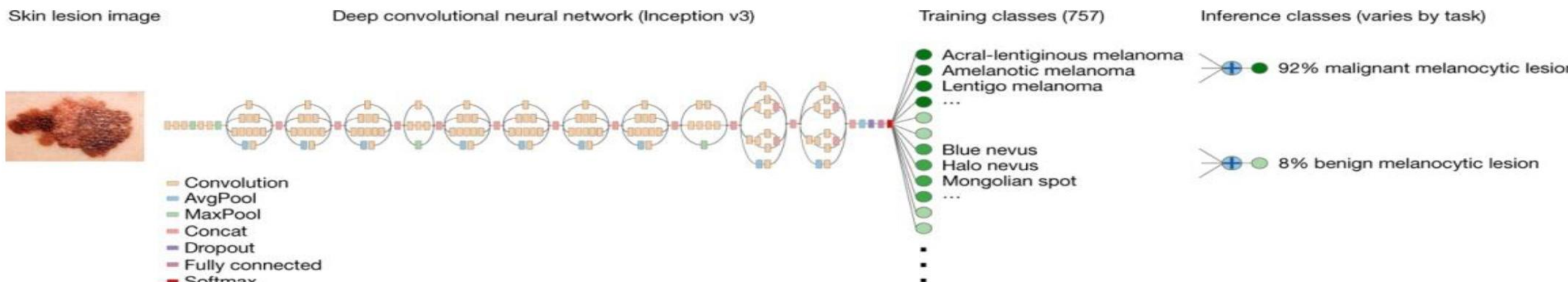
美国每年有540万新皮肤癌病例，其中五分之一被诊断为恶性肿瘤。其中黑色素瘤占5%，但致死率高达75%，仅美国每年死于该病的就超过10000人。早期检测对于治愈该病至关重要，如果早期检测到，则五年死亡率从99%降低到14%。

Andre Esteva团队基于GoogleNet v3设计了一个病变分类器，其准确度笔皮肤科医生更优越，论文发表在国际顶级期刊Neture上。

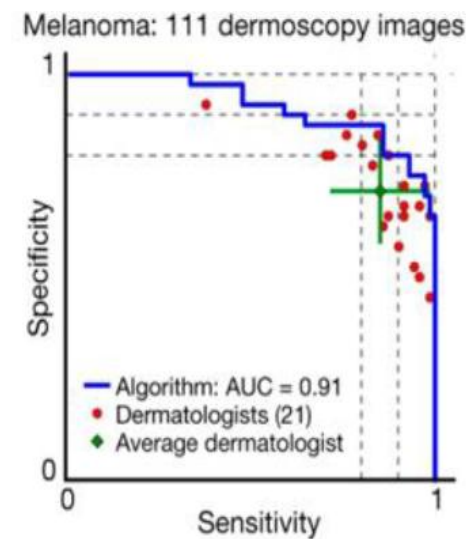
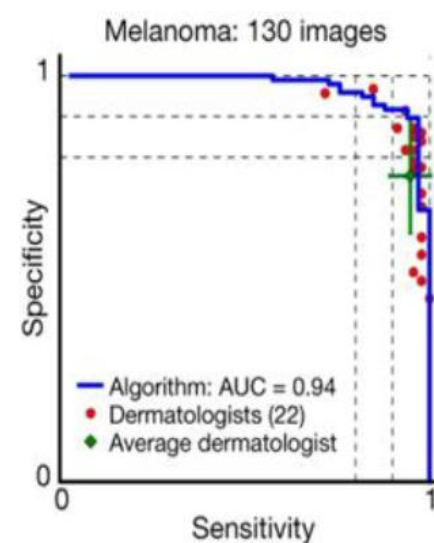
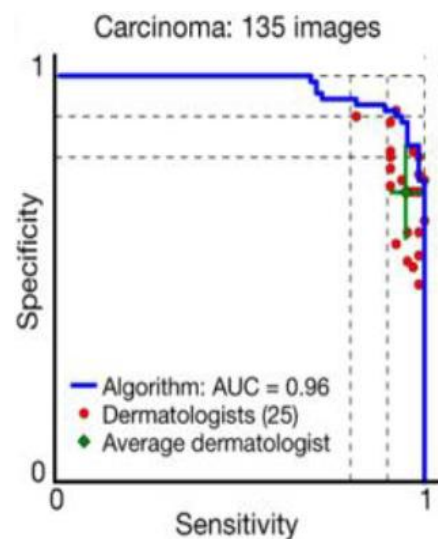


Esteva A, Kuprel B, Novoa R A, et. Al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017, 542(7639):115-118

实例2 – 皮肤病变分类



AUC是CNN对性能的对量，最大值为1。若敏感性/特异性点位于蓝色曲线下（大多数情况），则CNN的性能优于皮肤科医生。

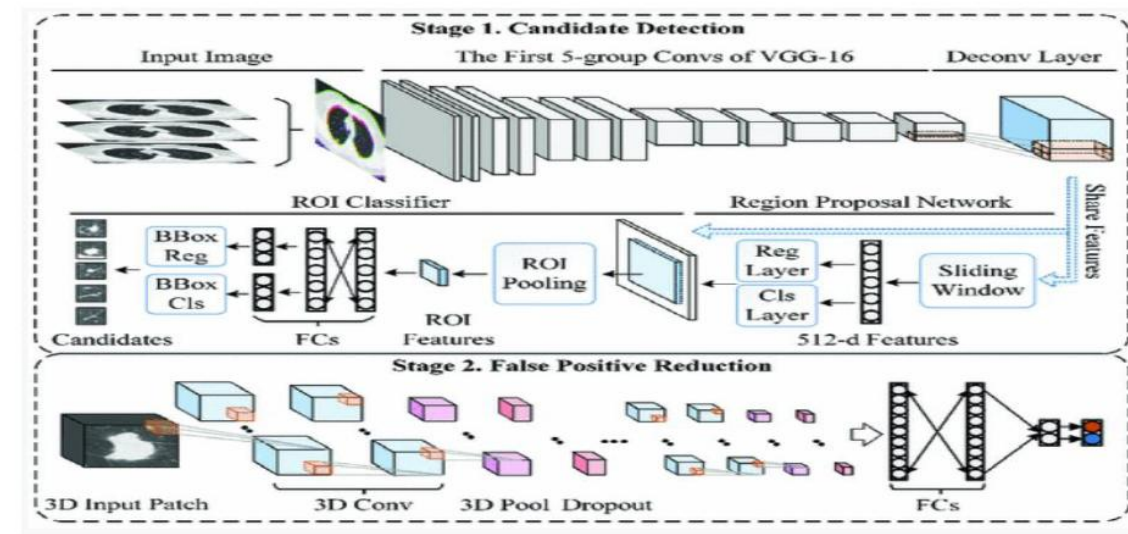


实例3 – 肺结核检测

肺癌早期发现是提高患者生存机会最有希望的方法。计算机断层扫描(CT)图像中准确的肺结节检测是诊断肺癌的关键步骤。

贾丁等人在阿里天池肺结节检测比赛中提出一种基于深度卷积神经网络的方法，该方法获得了极佳的性能。基本步骤如下

1. 使用Faster RCNN进行病变区域检测
2. 使用3DCNN进行假阳性样本减除



System	Sensitivity
ISICAD	0.856
SubsolidCAD	0.361
LargeCAD	0.318
M5L	0.768
ETROCAD	0.929
Baseline(w/o deconv)	0.817
Baseline(4 anchors)	0.895
Ours	0.946

Ding J, Li A, Hu Z, et. al. Accuracy Pulmonary Nodule Detection in Computed Tomography Image Using Deep Convolutional Neural Network. 2017.

医学影像分析

实例4 – 乳腺癌病理转移检测

Camelyon 2016/2017挑战赛

● 任务

- ✓ 根据全片扫描图像H&E染色(WSI), 判断乳腺癌是否转移
- ✓ 定位乳腺癌向邻近淋巴结扩散的位置

● 数据集

- ✓ **Camelyon16**, 训练集270张, 测试集130张, 数据量700GB
- ✓ **Camelyon17**, 训练集50张, 测试集500张, 数据量2.25TB

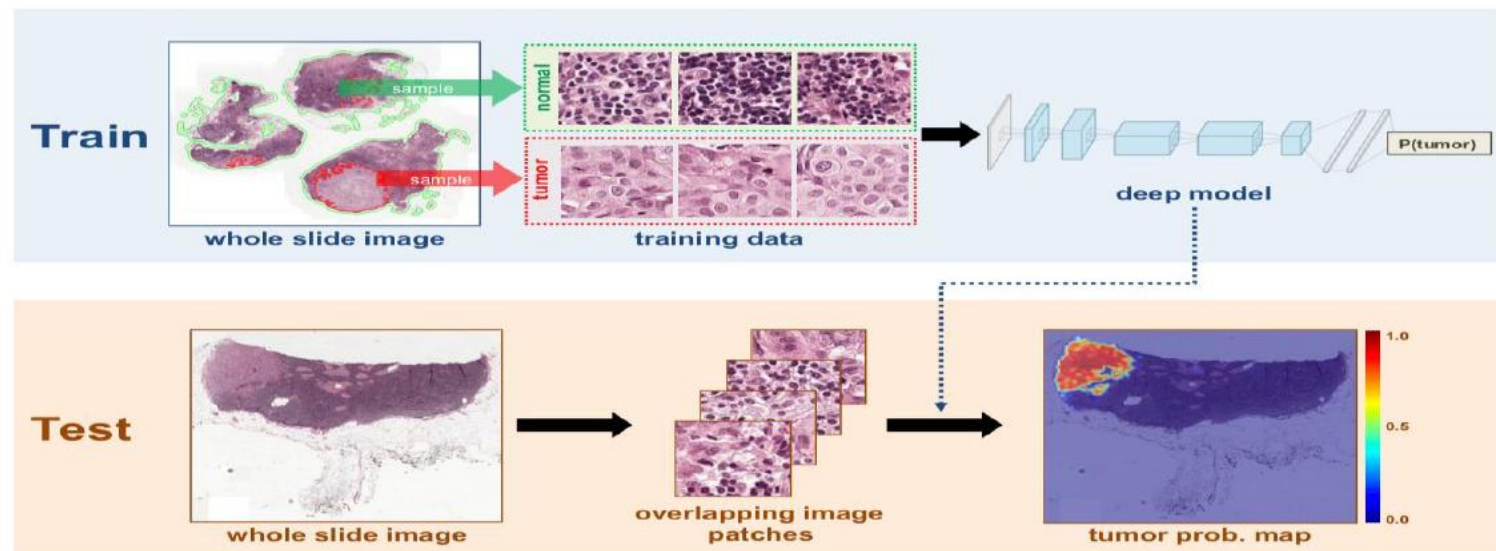


医学图像数据集的特点：样本少，单样本文件大

实例4 – 乳腺癌病理转移检测

哈佛医学院采用基于深度学习的方法屡获Camelyon的**冠军**

- 训练一个卷积神经网络对图像的patch进行分类(tumor or normal)
- 整合patch级预测结果构建肿瘤可能存在的热图(hotmap)
- 预测全片并定位肿瘤区域

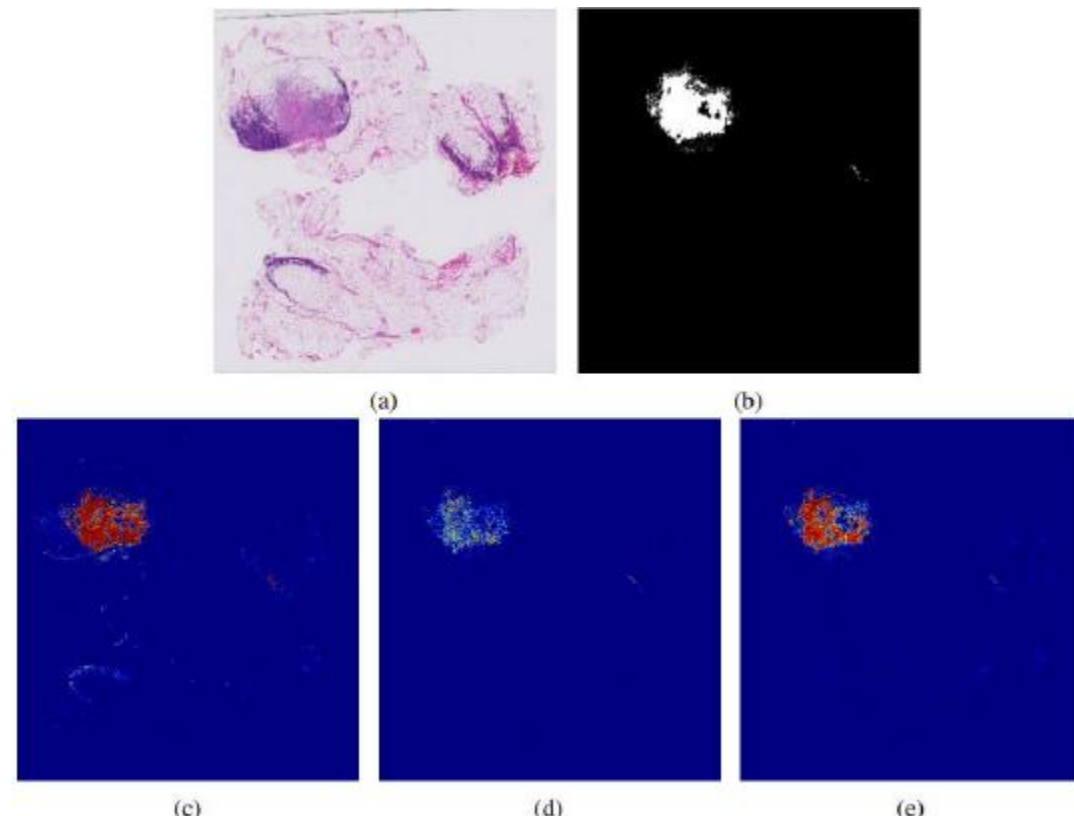
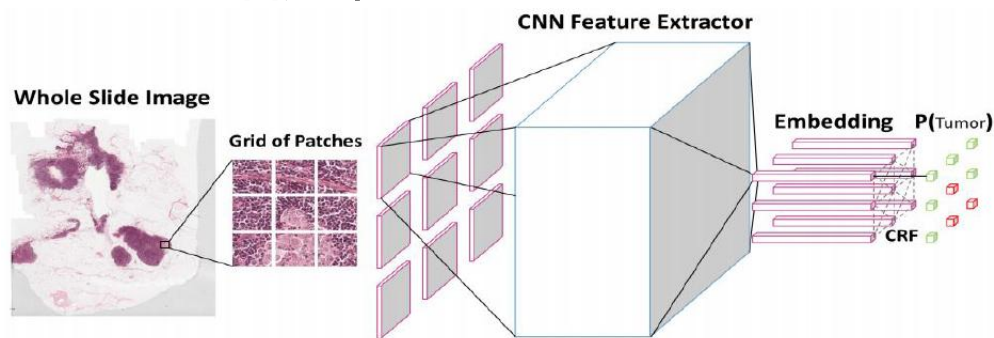


将病理学诊断和深度学习系统预测结果相结合，可以将病理学家的AUC得分提高至0.995，这意味着将减少近85%的人为误判。

实例4 – 乳腺癌病理转移检测

2018年百度提出NCRF

- 基于resnet18和resnet34架构
- 整合CNN和条件随机场建模相邻patch之间的空间相关性

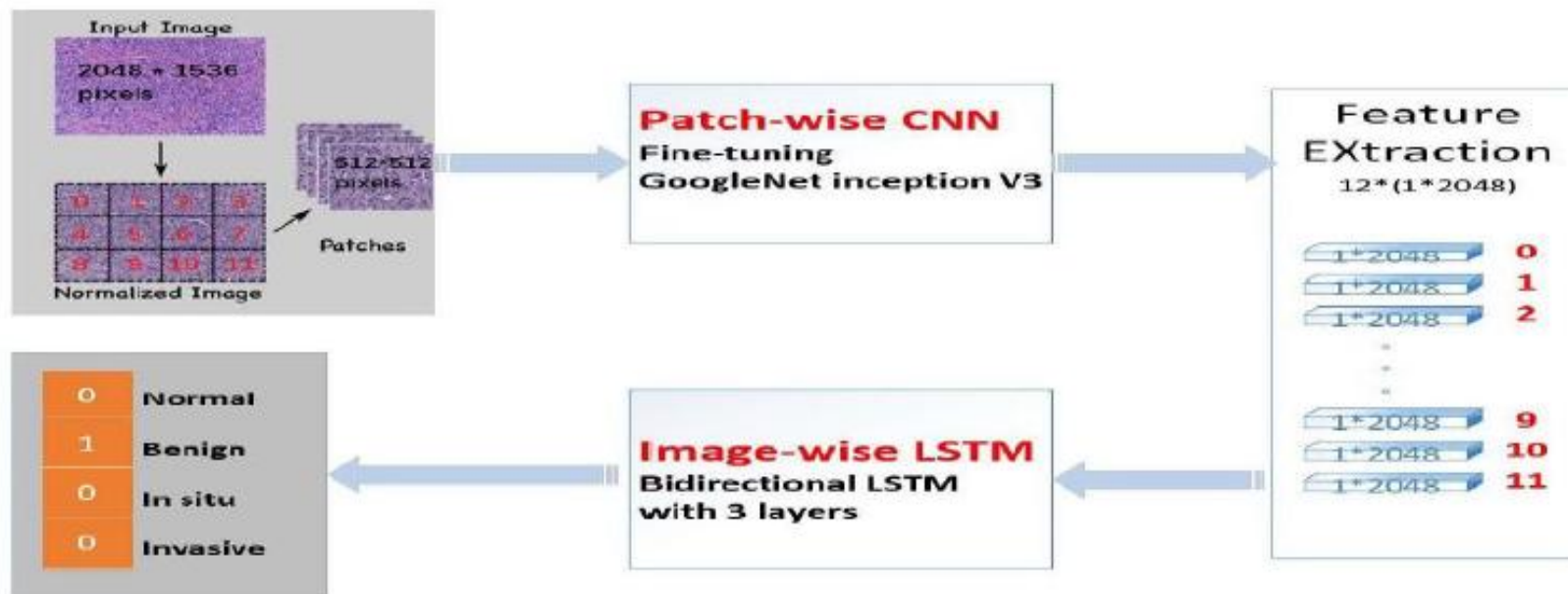


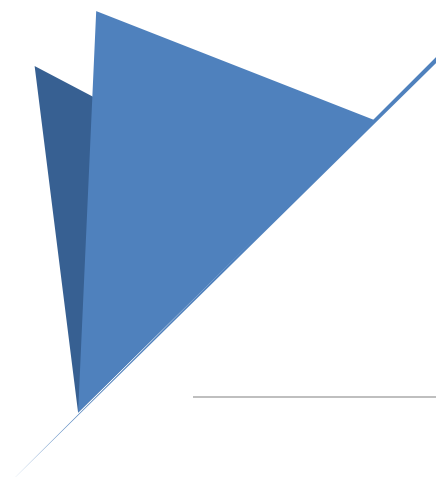
NCRF获得了更好视觉质量的patch预测的可能性热图。同时，NCRF在Camelyon16的检测任务上超过Baseline(人工识别)。

实例4 – 乳腺癌病理转移检测

2018年中科院计算所提出CNN+LSTM的方案

- 提出Inception-v3和LSTM的混合架构
- 将CNN提取的特征送入到3层双向LSTM
- 同时发布了一个包含1568个乳腺癌病例图像的数据集





文字检测

文字检测

文字检测简介

文字检测是文字识别的前提

● 任务

给定一张图片，找出图片中文字出现的所有位置。

● 自然场景下文字检测的挑战

- ✓ 文字存在多种分布
- ✓ 文字排布形式多样
- ✓ 文字存在多个方向
- ✓ 文字形变严重
- ✓ 多种语言混合

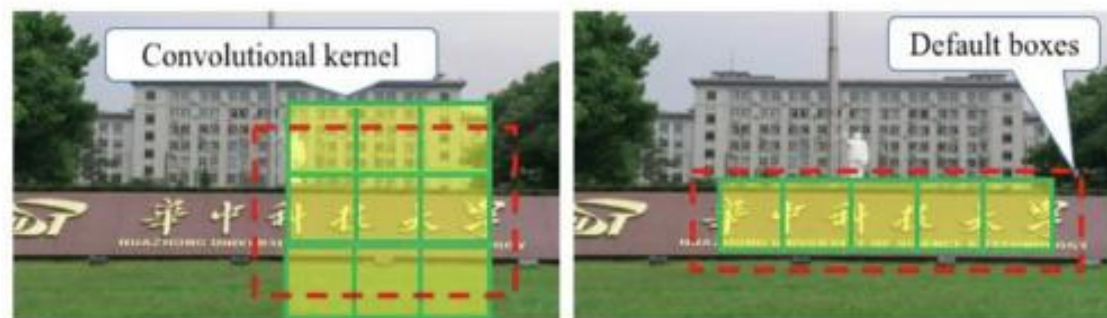
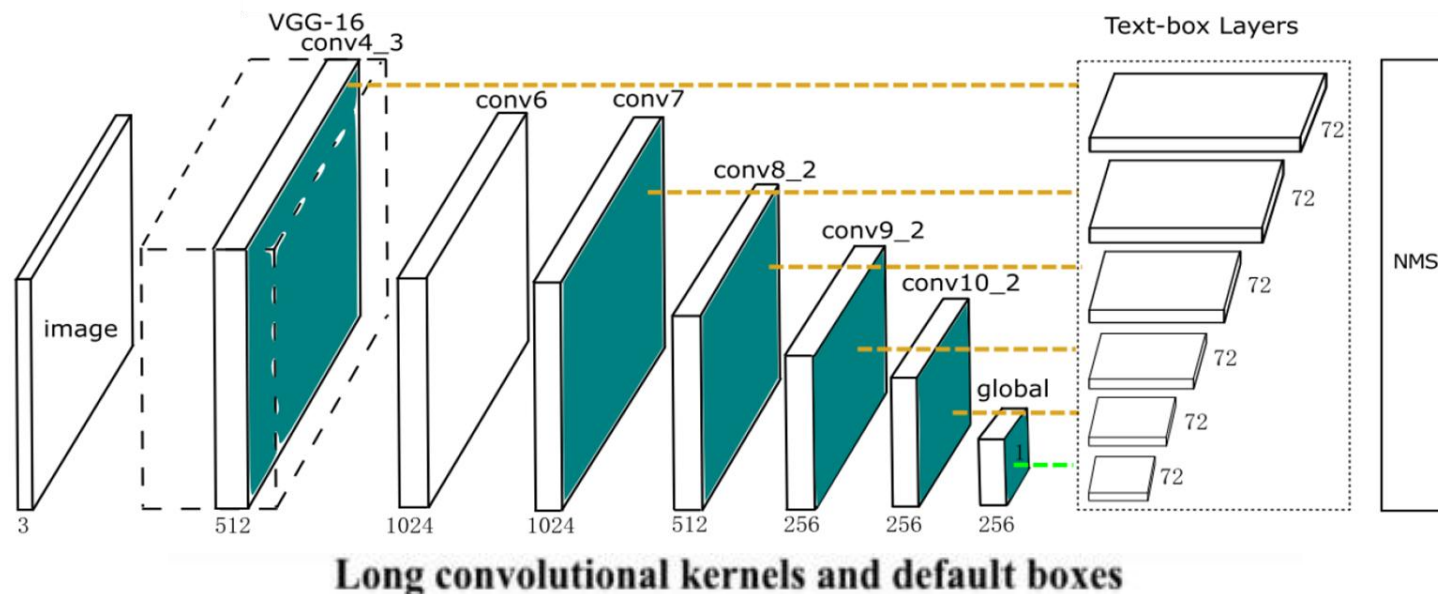


文字检测

TextBoxes与SSD类似，主干网络继承自VGG16。

1. 保留conv1_1-conv4_3用于图像特征提取;
2. 通过多尺度的方法提取6个不同尺寸的特征，送入TextBox Layers; 其中卷积核尺度改为 1×3 ，以适应长文本
3. 输出若干个可能存在文字的区域候选框。
4. 经过NMS得到最终的预测区域。

TextBoxes



SSD: 3×3 conv



TextBoxes: 1×5 conv

Liao M, Shi B, Bai X, et al. TextBoxes: A Fast Text Detector with a Single Deep Neural Network. AAAI 2017.

TextBoxes++

● TextBoxes存在的问题

Default box (anchor) 的水平框，不能很好检测倾斜文字（如右图所示）。

● TextBoxes++的改进

- ✓ 边界框的表示改为四边形 (x1,y1,x2,y2,x3,y3,x4,y4)
- ✓ 卷积尺度改变 $1*5 \Rightarrow 3*5$ ，适应倾斜文本
- ✓ 级联NMS和OHEM策略的应用
- ✓ 多尺度训练以适应不同尺度的目标

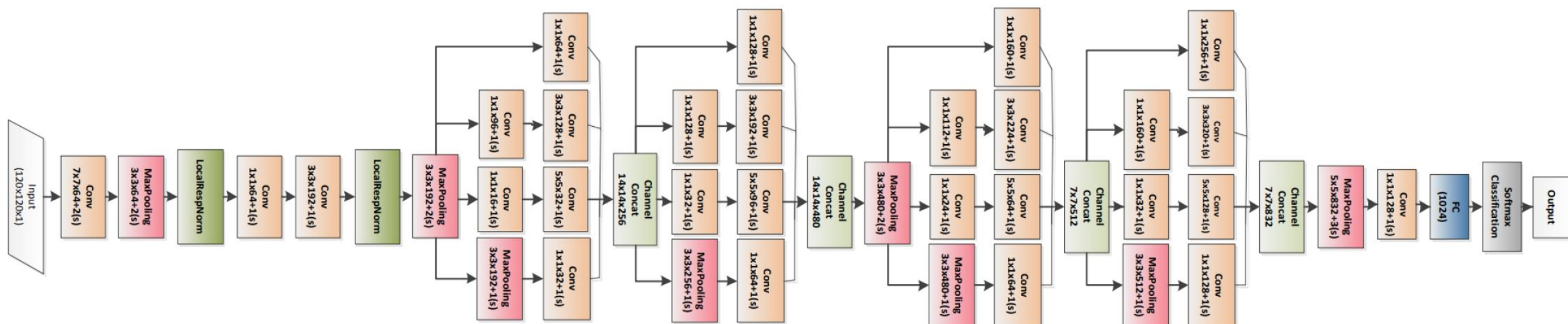


在后续的研究中，有人提出形变网络，大大改进了倾斜字符的识别能力

文字识别 – CNN单字识别

● HCCR-GoogLeNet

在ICDAR2013离线手写汉字识别竞赛数据集上达到96.35%的准确率。HCCR模型共19层，由Inception模块组成，该模型可以识别单个文字及文字片段，并输出识别分数。



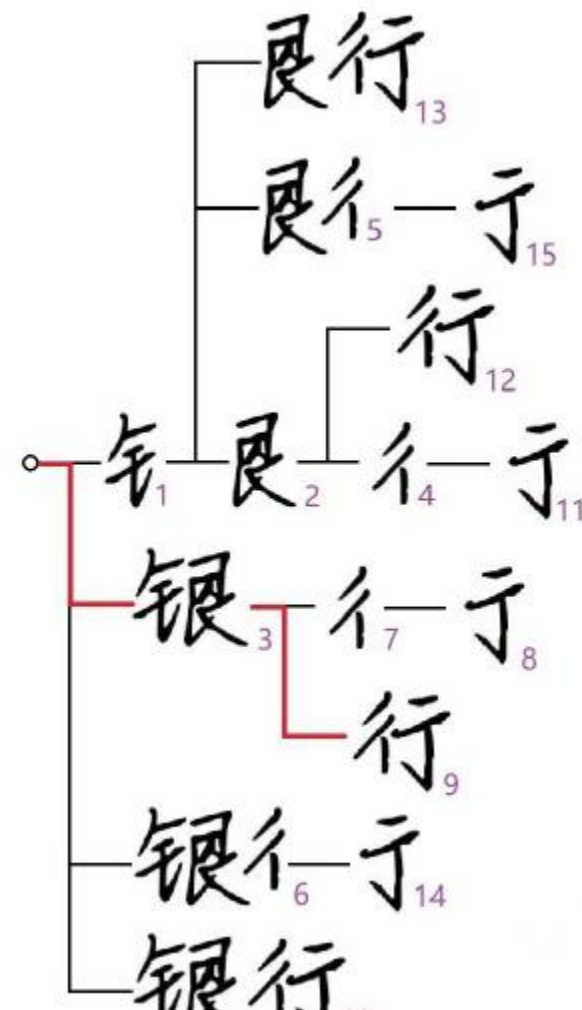
文字识别 – 集束搜索

● 集束搜索

使用集束搜索，将过分隔后的文字片段组合成正确的识别结果。

● 集束搜索的依据

- ✓ 单个文字片段的识别分数
- ✓ 前后文字片段的语义关系
- ✓ 文字片段的最大组合数



文字检测

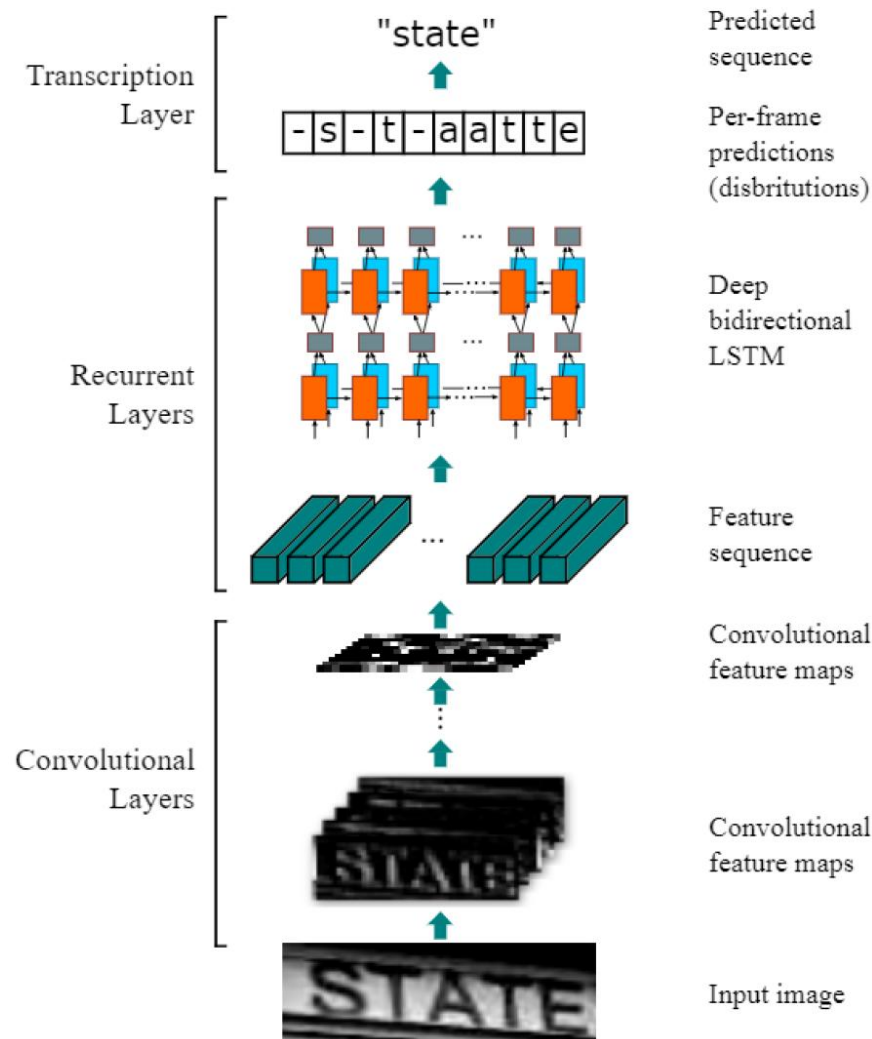
文字识别 – 无分割方法

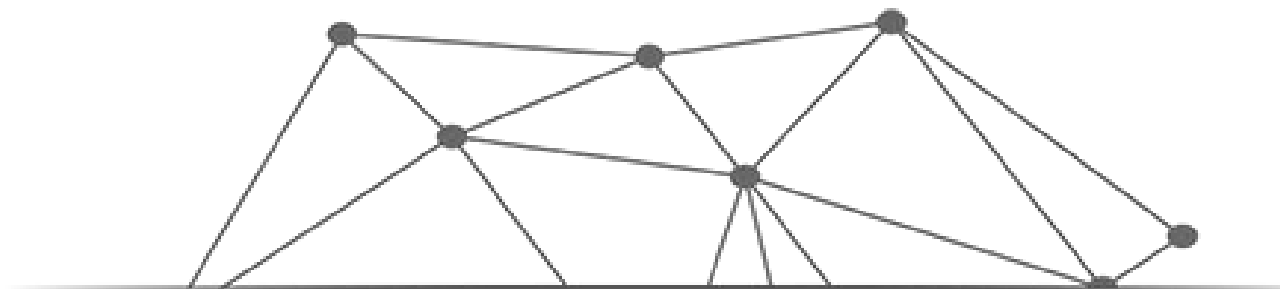
CRNN

端到端识别文本行图像，无需分割过程。

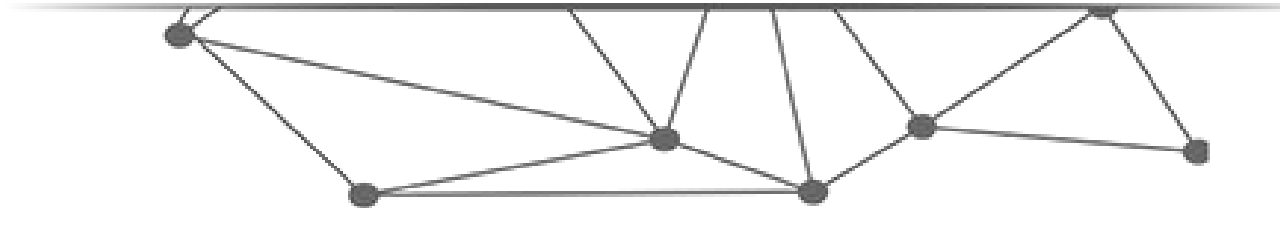
基本步骤

1. **卷积层**：从输入图像中提取特征序列表示
2. **循环层**：使用LSTM预测特征序列中每个特征向量的概率分布
3. **转录层**：使用CTC将预测的概率序列转换成最终识别结果





课堂互动 13.4.5



读万卷书 行万里路 只为最好的修炼



QQ: 14777591 (宇宙骑士)

Email: ouxinyu@alumni.hust.edu.cn

Website: <http://ouxinyu.cn>

Tel: 18687840023

地址: 安宁校区 诚远楼201

南院 智能应用研究院A306-2