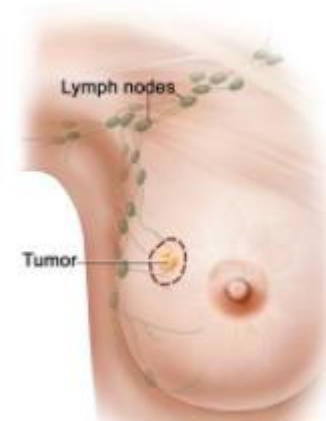


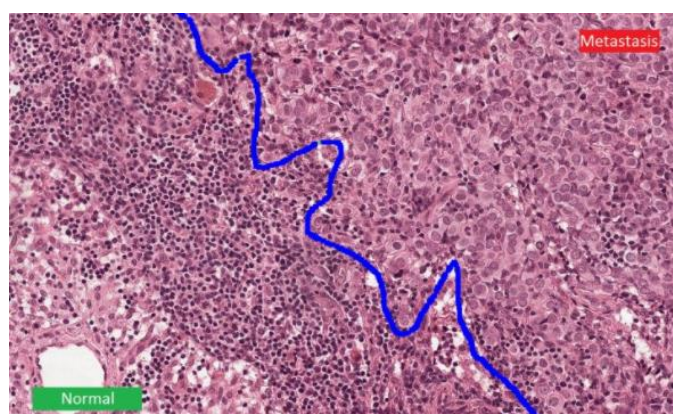
Camelyon17 挑战赛

● 任务由来

这一挑战将集中在**淋巴结中乳腺癌转移的检测和分类**上。淋巴结是过滤淋巴液的小腺体，淋巴液是通过淋巴系统循环的液体。腋窝淋巴结是乳腺癌可能传播的首位。淋巴结的转移性累及是乳腺癌最重要的预后因素之一。当癌症扩散到淋巴结时，[预后较差](#)。这就是为什么需要手术切除淋巴结并用显微镜检查。然而，病理学家的诊断程序是冗长乏味且耗时的。但最重要的是，小转移瘤很难检测到，有时甚至错过。



医生在乳腺检查中，会提取一些附近淋巴结组织做成切片，经过切片、染色、扫描等过程后，生成数字病理切片图像。癌细胞和正常细胞在颜色、纹理、大小和组织形式上都会有很多的不同，一般来说就是“核大深染”的突出特征。在大医院中，很多上了年纪的且具有很多“看片”经验的医生炙手可热，这意味着人类同样需要很多经验才能正确的进行分析判断，而年轻的或缺乏经验的医生容易出现误判



出特征。在大医院中，很多上了年纪的且具有很多“看片”经验的医生炙手可热，这意味着人类同样需要很多经验才能正确的进行分析判断，而年轻的或缺乏经验的医生容易出现误判

● 任务描述

CAMELYON17 的目标是评估新的和现有的算法，用于在组织学淋巴结切片的数字影像中进行自动检测乳腺癌转移和分类。分类结果是基于患者层面的分析，这需要将多个淋巴结转移灶的转移检测和分类结合为一个结果：[pN 阶段](#)。这项任务具有很高的临床意义，通常需

要病理学家进行广泛的微观评估。因此，评估乳腺癌患者 pN 分期的自动化解决方案有望减少病理学家的工作量，同时降低诊断的主观性

● 数据集描述

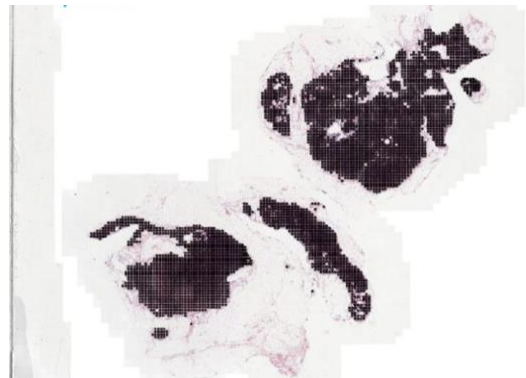
Camelyon17 一共提供了 100 个乳腺癌患者的 pN 分期等级和的带有标注的淋巴切片样本，每一病人有 5 张病理切片。其中 110 张含有癌细胞组织的切片(Tumor)，剩下则为张正常的组织切片(Normal)，癌症的区域进行了标记。看起来数据量很少，按照传统的图像处理方式应该是很简单的，但实际上数据的形式是金字塔数据形式，最大分辨率 40X 的图像矩阵大小大概是 300000×150000 ，一个样本的所占硬盘空间大小大概是 5~6G。仅从一张病理切片之中就可以提出数万张肿瘤小块 (patch)

● 算法描述

1. 预处理

在经典问题中很少有图片会这么大，但也会遇到一些，比如遥感图像等；这类图像的处理办法一般会采用分块的方式进行处理，然后再将每一块的处理结果进行汇总从而得出结果。

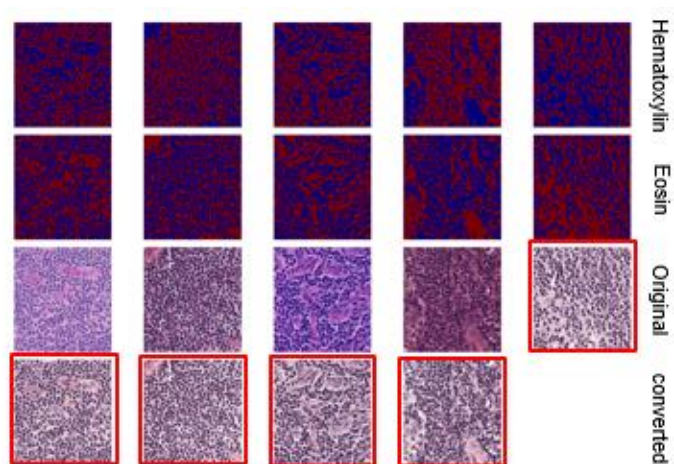
对于一张病理图像来说，只有 20%~30%的区域是有效区域 RoI，所以要首先找到 ROI。



2. 染色均衡

由于比赛数据来自于不同的医院或研究机构，所采用了不同厂家的扫描仪，因此图片在色调上存在着很大的差异，这对模型来说，会造成 confused 现象，因此在做成数据集之前，对所有的图片进行了染色均一化过程。其主要的思想是通过将 RGB 变换到 HSV 色域表征，

通过调节色调分布使得不同图像进行染色均一化。



3. 模型构建

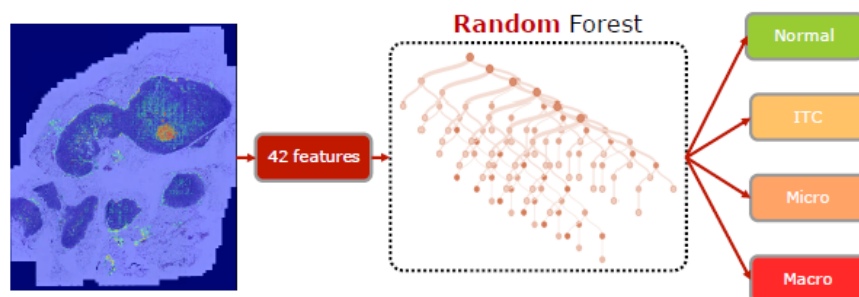
分别尝试了 InceptionV3 googleNet ResNet 等模型



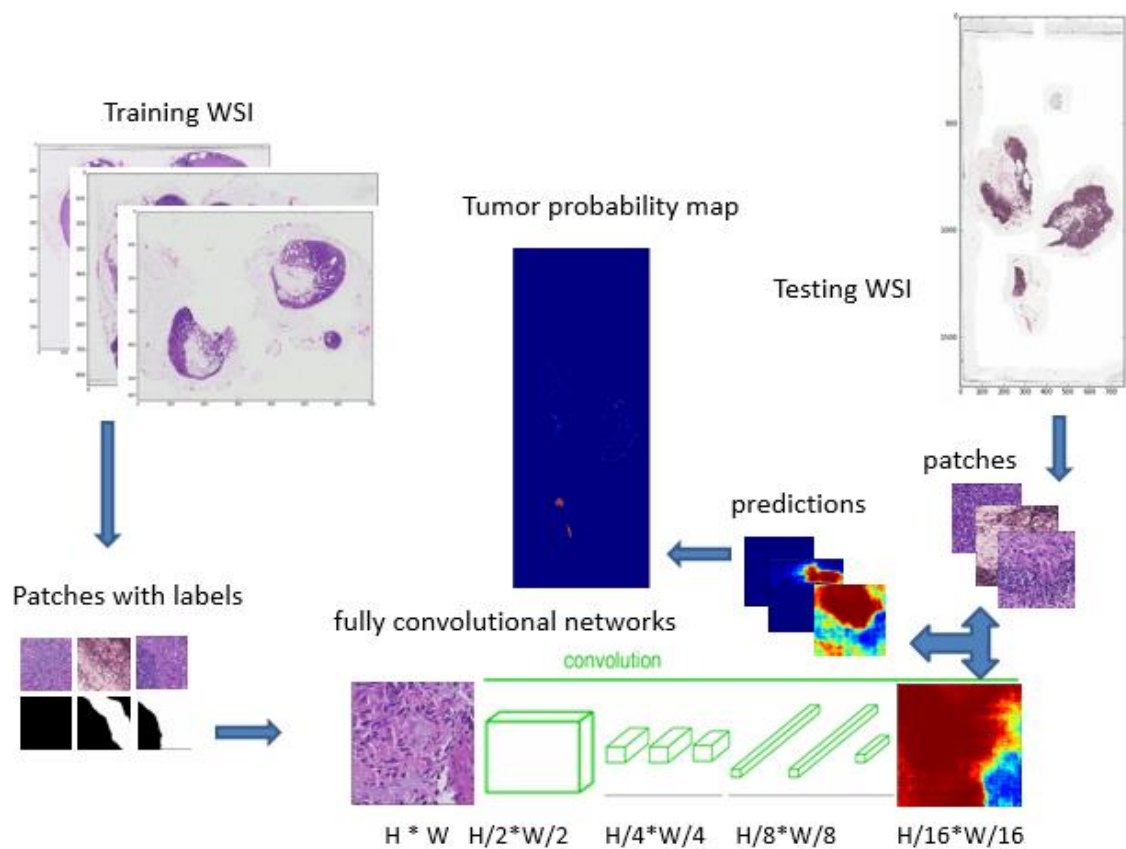
4. 后处理 (生成热力图)



5. 提取特征用机器学习算法分类



整个流程



● 实验结果

ConvNet Model	Accuracy on Camelyon17 (patch level)
GoogleNet [2]	98.4 %
Inception v3	99.5 %

Our method on Camelyon17 Train set shows 0.92 kappa score in patient-level