

lesson3-总结

1. 目标

1) 为从事 AI 领域各方向科研, 拥有该模型的能力.

eg: 一般来说, 如果是发论文. 确定好方向, SOTA 后. 除了模型外, 其他数据增强、预处理都是写死的. 你只需要去改模型. 这是我们本次 Hudmap lesson3 补充的.

2) 为从事 AI 业务, 需要用到分割模型. 确定能快速构建模型的能力

eg: 比如 OCR. 表格分析. 表格重建. 一般都是用分割进行处理的

3) 基本的 cv 模型模块作用分析

baselinev1: efficientNet + FPN + Unet(15) + ASPP(dilated conv)

baselinev2: swin + upernet(CNN)

baselinev3: swin + SegFormer(transformer)

青蛙哥 dicusion 几十组.

111 1 1

111 1 1

111 1 1

数据 : 2.5D. patch. muliti-view

模型 : smp.(efficent.resnet(timm) + uent/FPN/) ==> . ==> mmsegmentation

损失函数 : bce + dice

推理 trick: ensemble + tta + pl + 分水岭

2. 模型 - encoder

2.1 efficientNet

ICML 2019: <https://arxiv.org/abs/1905.11946>

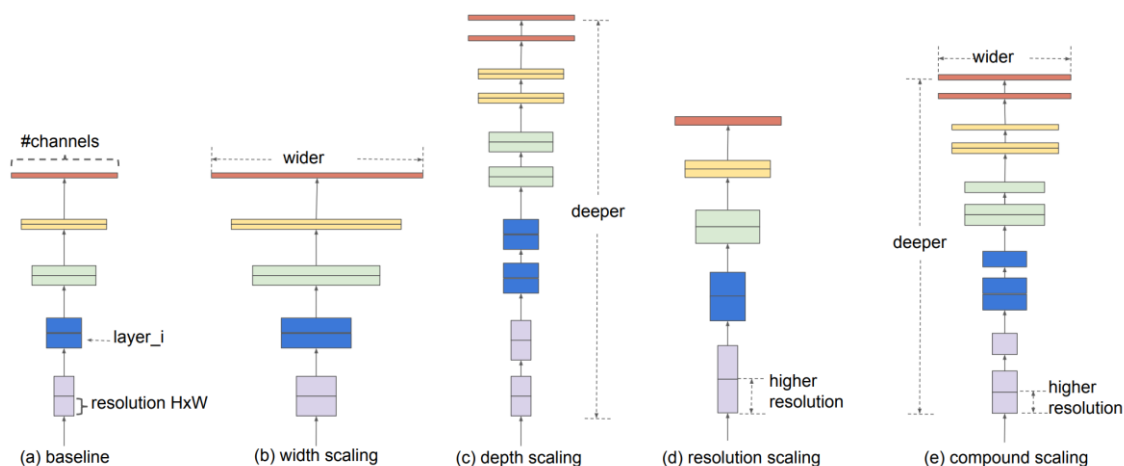


Figure 2. Model Scaling. (a) is a baseline network example; (b)-(d) are conventional scaling that only increases one dimension of network width, depth, or resolution. (e) is our proposed compound scaling method that uniformly scales all three dimensions with a fixed ratio.

设计分类网路的几个因素：输入分辨率、宽度、深度、特征的维度。

出发点：人工调整很难

baseline 网络 ==> AutoML NAS ==> 最终结果.

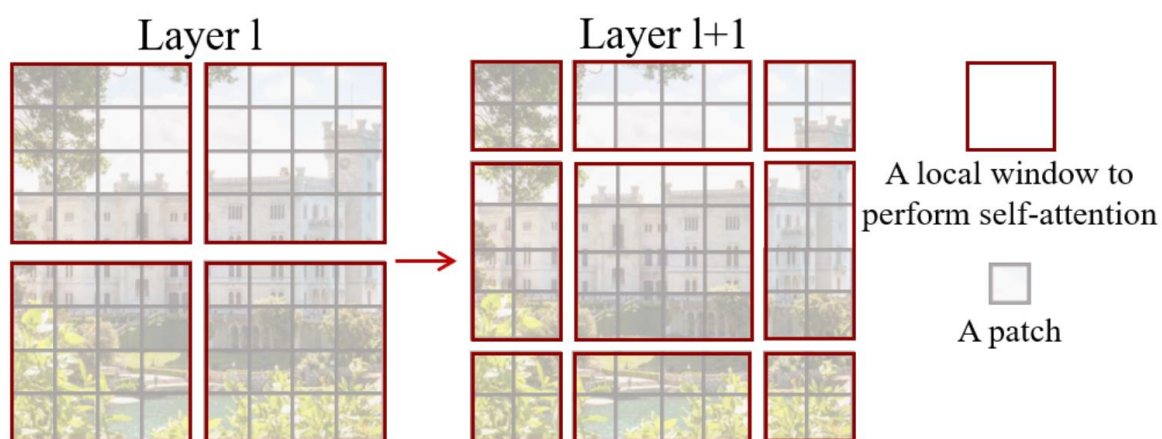
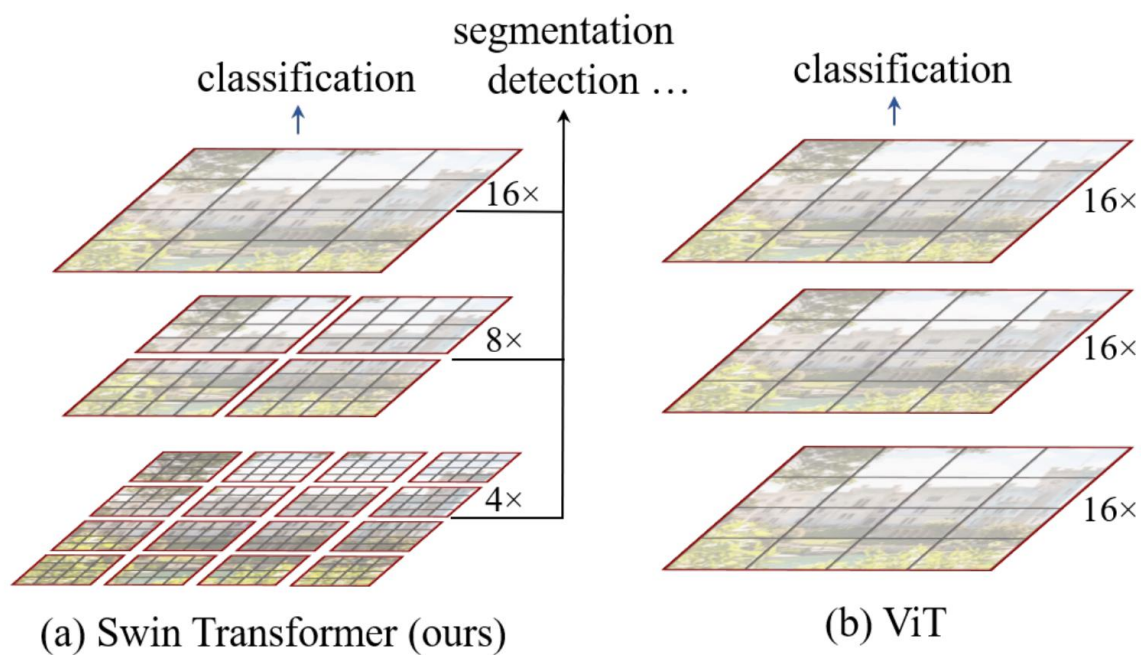
设计检测网路的几个因素：多尺度融合（multi-scale feature fusion）、模型复合扩张

FPN: 无区别的去特征相加. 不同分辨率特征包含的信息是不同都得..

CVPR2020: EfficientDet: Scalable and Efficient Object Detection

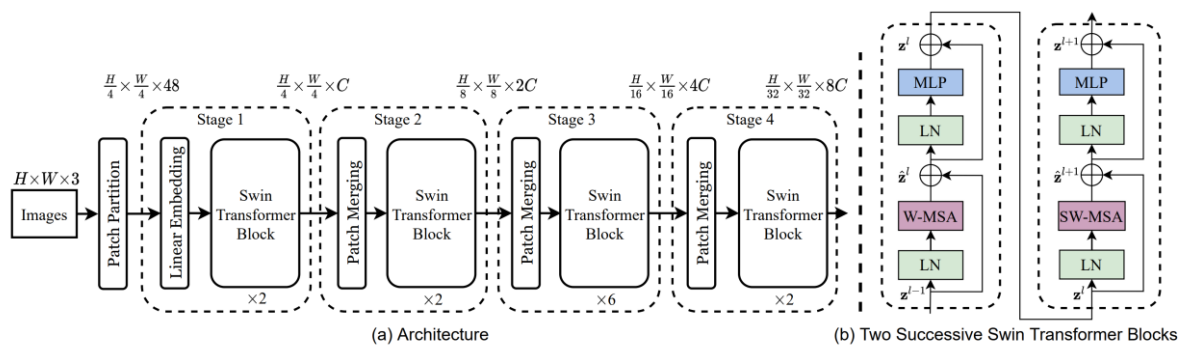
<https://arxiv.org/abs/1911.09070>

2.2 Swin



CVPR2020: <https://arxiv.org/abs/2103.14030>

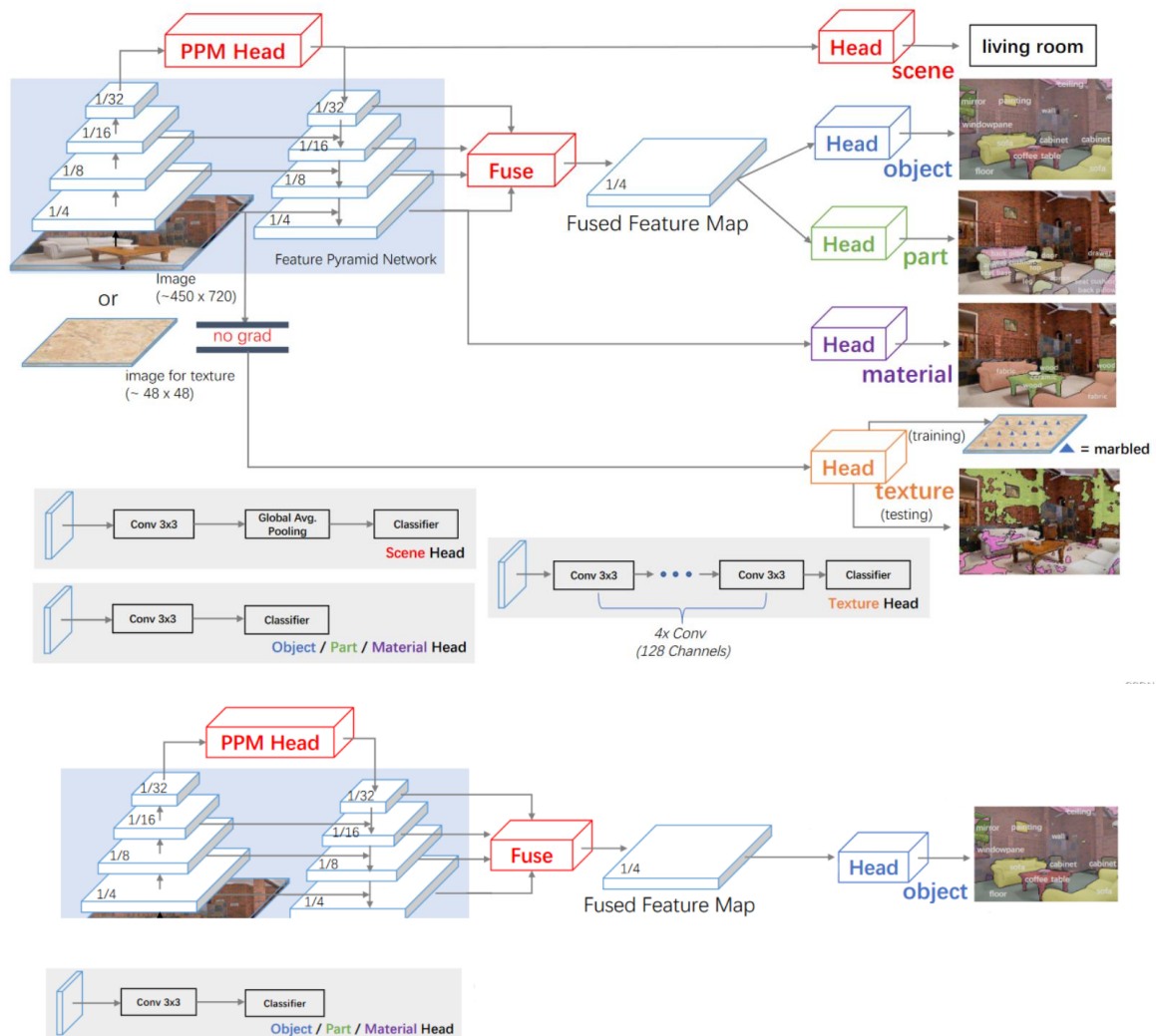
出发点：基于全局的 attention 计算了巨大
层次化感受野的扩大。



====> Moco

3. 模型 - decoder

3.1 Upernet(CNN)



基于 FPN,(Feature, Pyramid Network) & FPN(Pyramid Pooling Module)

每个 task 都有不同的检测头.

3.2 Segformer(transformer)

NIPS21: <https://arxiv.org/pdf/2105.15203.pdf>

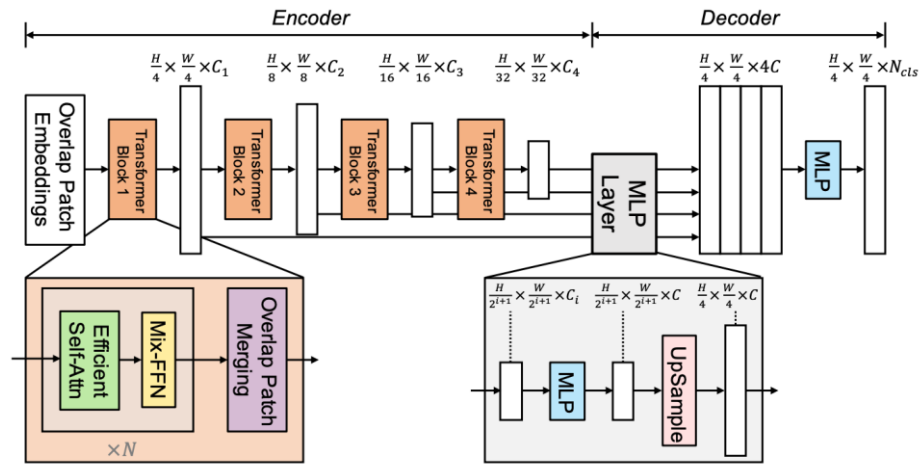


Figure 2: **The proposed SegFormer framework** consists of two main modules: A hierarchical Transformer encoder to extract coarse and fine features; and a lightweight All-MLP decoder to directly fuse these multi-level features and predict the semantic segmentation mask. “FFN” indicates feed-forward network.

分割：增大感受野、之前 ASPP(空洞卷积) ==> self-attention

dilated SA= > [400, 400] ==> [400, 200]

20 年 ==> CNN + transformer ==> multi-branch

pdb ==> C++

3.2 MaskFormer(transformer)

ECCV - paper list

MM - paper list

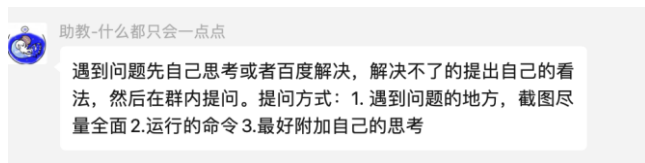
==> AAAI. CVPR

4. 答疑

4.1 混合精度的目的 - AI 推理芯片.

<https://pytorch.org/docs/stable/amp.html>

4.2 遇到问题如何解决呢？ - debug !!!

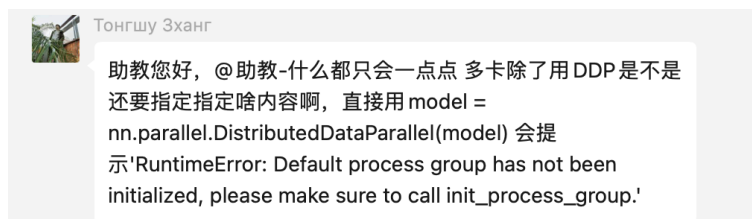


4.3 训练收敛问题：

DETR code ==> conditional DETR (wang jingdong)



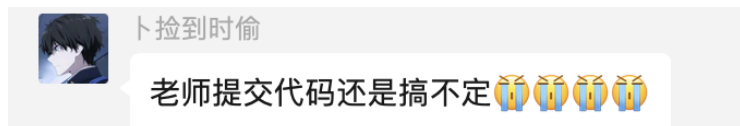
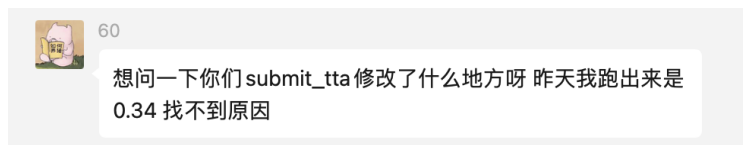
4.4 多卡问题：



4.5 为什么只有 0.35 & 提交代码

mmsegmentation 0.78 ==>

可视化. ==>



4.6 thr

kaggle 经验：对于一些启发式值. 尽量的 ensemble

eg. thr. 数据切分方式

tta: dicusion. 青蛙哥 ==> 求和平均



金立

老师 我想问一下关于 thr 调整有没有什么套路啥的



Yingshu

@老师-frank

1. 这次也尝试了 tta, 尝试了加不同的翻转的方式及组合, 但是加了 tta 基本上没有提升, 有时还会掉点。想问一下, 一般 tta 在那种情况下加以及加 tta 需要注意些什么。
2. 对于 mask 的阈值 Th 的设置, 我们发现按照之前的阈值 0.4 成绩不是很好, 如果 TH 调低, 比如 0.2 左右, 会上一两个点。对于阈值 TH 的设定, 有没有什么理论上的调参依据。因为这次比赛公榜和私榜的数据类型不一致, 我们考虑如果 TH 调低得到在公榜上的高分会不会造成在私榜上掉分很多。

4.7 segformer 模型怎么改



高远

老师, 我看讨论区有人说 segformer 不让用于商业化, 那这个是指获得奖金, 还是只要获得奖牌也算 🤔

4.8 高效调参



LeBelief

老师 我有几个问题 1 我们后续需要自己调参吗? 如果需要调参 怎么更高效合理地完成调参呢? 2 如何融合模型? 是把预测输出拿出来按投票制吗? 3 怎样融合 patch 与 full image 结果? 也是将预测拿出来投票吗?

同一个图片. ==> mask real pred(prob) ==> mask binary pred

patch ==> mask real pred(prob) ==> 0.8.. 0.2.. 0.9

full ==> mask real pred(prob) ==> 0.8.. 0.2.. 0.2

4.9 预训练

比赛 (多融合模型) . 业务 (单模型)



Ceazer

@老师-frank

1. 对于这次比赛数据训练集来自全部 HPA, public test 来自 HPA 和 HuBMAP, private test 全部来自 HuBMAP, 这样的数据分布不同的情况, 可以谈一谈有什么好的解决办法或者解决思路吗?
2. HPA 的数据都是 3000*3000 的, HuBMAP 的数据是 4500*4500 到 160*160 的范围的, 针对这种情况, 从什么方面考虑数据增强呢?
3. HPA 数据 (DAB) 和 HuBMAP 数据 (PAS) 产生方式不同, 看评论区有人获得了额外数据, 我们可以从哪些渠道获得 HuBMAP 的数据看一下呢?

4.10 :

image. mask 转置问题 ==> 对齐其他人的推理代码.

kaggle ==> code ==> inference

不同 mask2rle. rle2mask. shape

可视化看一下 :



60

@老师-frank 老师, 可不可以说一下提交代码并取得正常分数案例运用, 或者是 debug 提交代码的注意事项?
我根据 swin+upernet 的 Net() 类中的 forward 部分, 修改了 submit_tta 代码里面的 dataloader, 但是提交上去的 score 非常低。
我 debug 了 infer 中 model 输入输出的 shape、mask 的 shape 等等都是符合预期的, 找不到具体的问题, 但是代码同样 work...现在没有思路了, 所以卡在了提交这块

4.11



LeBelief

奇怪 swim 第二个 fold 不到 400 轮就 0.8 了...这也太不科学了 我第一个 fold 最好才 0.7 🤔🤔🤔



LeBelief

每个 fold 差距应该不大才对

4.12



wky

老师，想问下为啥 cv 和 lb 差距这么大。蛙哥从 3000 直接缩放到 768 来训练，推理的时候用 768 推理，然后放到 3000 来生成 mask，是不是丢失了太多信息。@老师-frank