南京工業大學
NANJING TECH UNIVERSITY

基于深度学习的混凝土裂缝识别技术

李延成，南京工业大学

目录

CONTENTS

01

课题背景

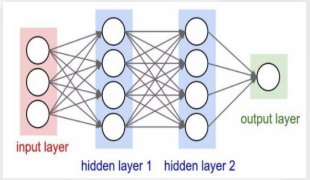
裂缝的危害

裂缝宽度超过规范规定限值的范围时，外界的空气和水进入后会使钢筋腐蚀速度变快，影响结构安全，同时使得混凝土碳化速度加快，容易出现漏水、渗水等现象，使得混凝土保护层脱落，缩短建筑的使用年限。

裂缝检测技术

为了防止事故的发生，保证混凝土结构设施的安全使用，定期检测与评估混凝土结构设施的健康状况，并针对检测结果进行预先防治是必不可少的。

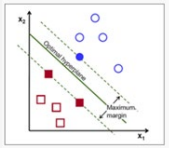
成为裂缝检测的主流方法之一



深度学习

卷积神经网络、全卷积网络

自动地学习图像特征，检测精度高

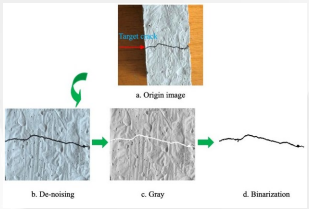


Support Vector Machine

机器学习

支持向量机、随机森林

图像特征需要人工提取，泛化能力差，易过拟合



图像处理

边缘检测、阈值法

结果依赖人工定义的特征提取算法，易受环境影响

人工检测

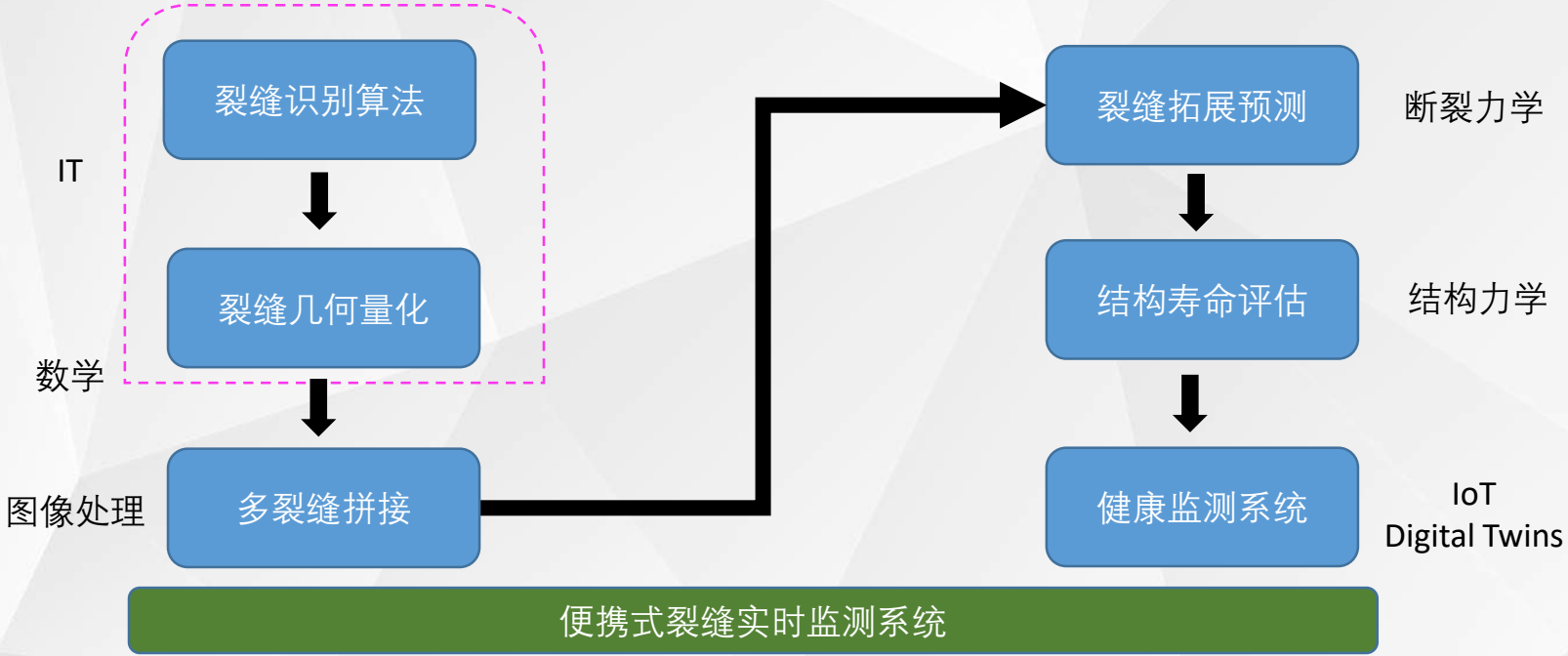
人工视觉检测

主观性强，结果不可靠，费时费力，成本高

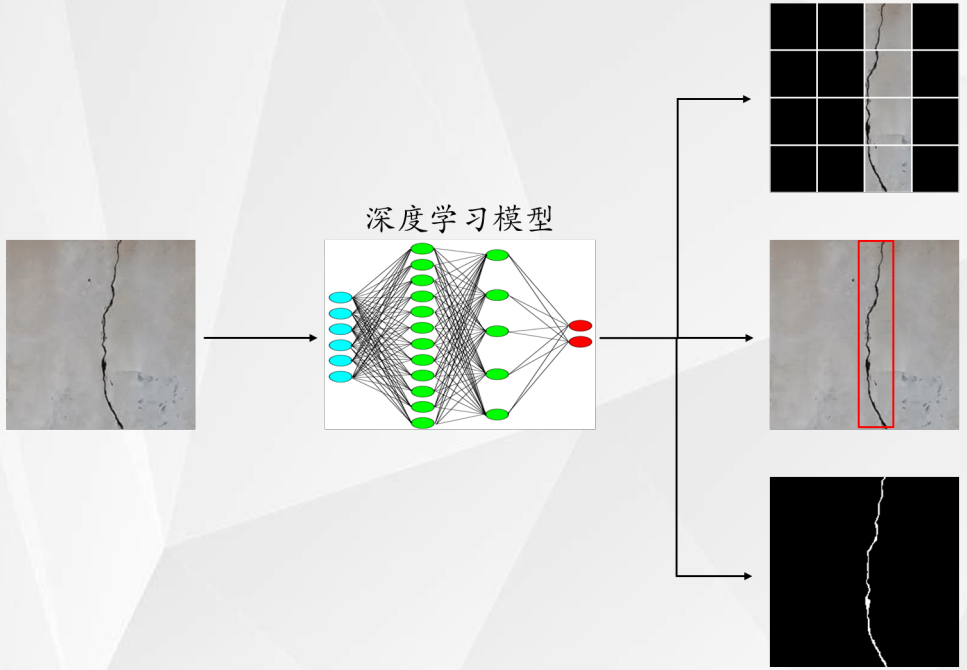
02

深度学习裂缝识别框架体系

课题框架


裂缝识别算法
裂缝几何量化
多裂缝拼接
IT
数学
图像处理
裂缝拓展预测
结构寿命评估
健康监测系统
断裂力学
结构力学
IoT
Digital Twins
便携式裂缝实时监测系统
信号处理、电子硬件、算法集成、软件

深度学习在裂缝检测中的应用



分类确定图像或者图像块中是否存在裂缝。

目标检测生成边界框从图像中提取裂缝，确定裂缝位置。

语义分割对图像中的每个像素进行分类。

裂纹识别与健康监测

有无裂缝

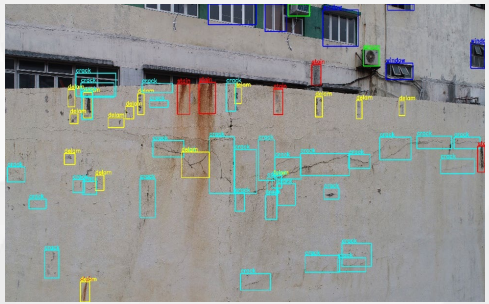
Classification

是否存在裂纹?

裂缝识别

裂缝区域

Object Detection

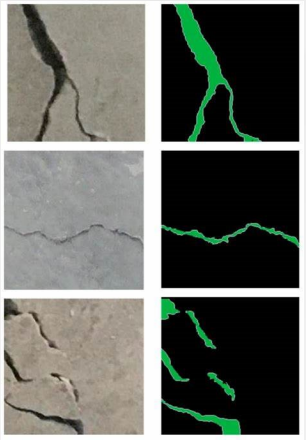


裂缝语义分割

Segmentation

- 裂缝几何信息
- 裂缝拓展预测
- 结构寿命评估
- 结构监测维护

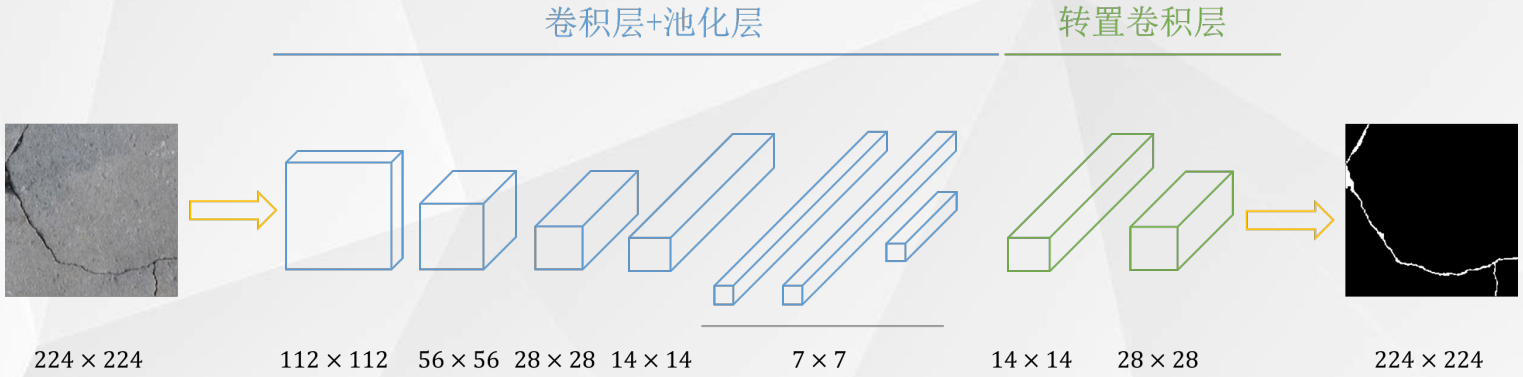
.....



→ 工程应用价值递增 →算法复杂程度剧增 →与无人机结合前景巨大

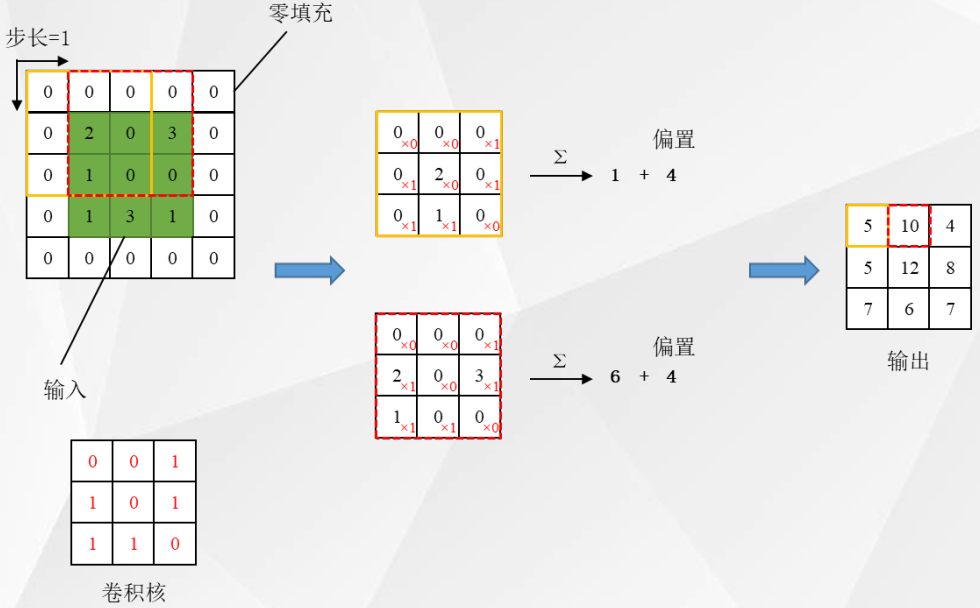
语义分割——全卷积网络

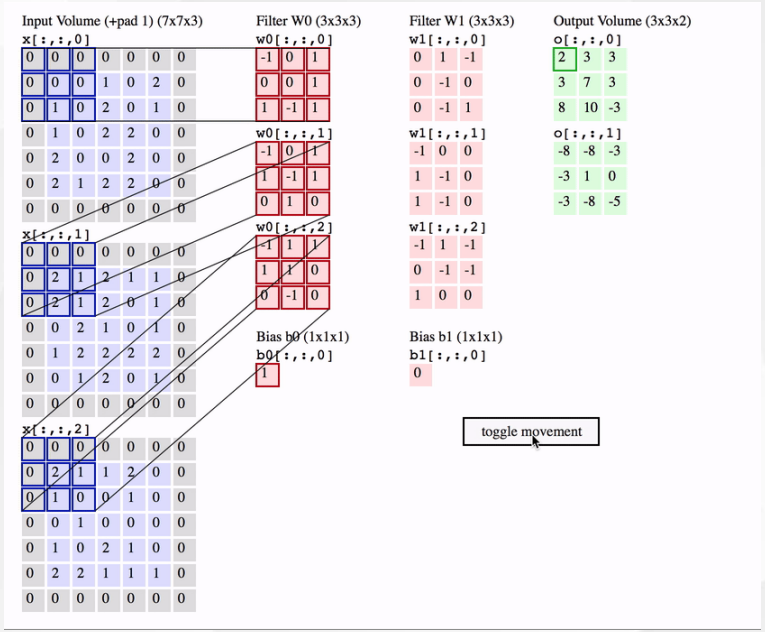
作为一个经典的语义分割模型，全卷积网络采用卷积层和池化层对图像进行下采样，提取图像特征，并缩小图像尺寸，随后采用转置卷积层对特征图进行上采样，使它恢复到与输入图像相同的尺寸，从而可以对图像中的每个像素进行分类。



卷积层

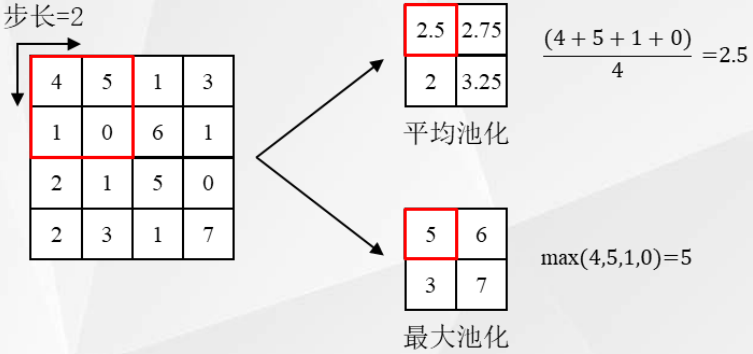
卷积层：利用特征图之间的卷积提取图像特征。为了避免由于卷积而导致特征图的尺寸减小，通常会在输入特征图的最外围采用零填充以保持相同的空间维度。

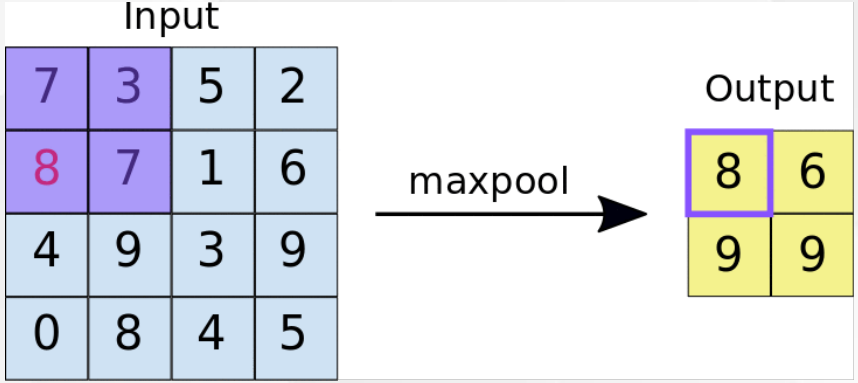




池化层

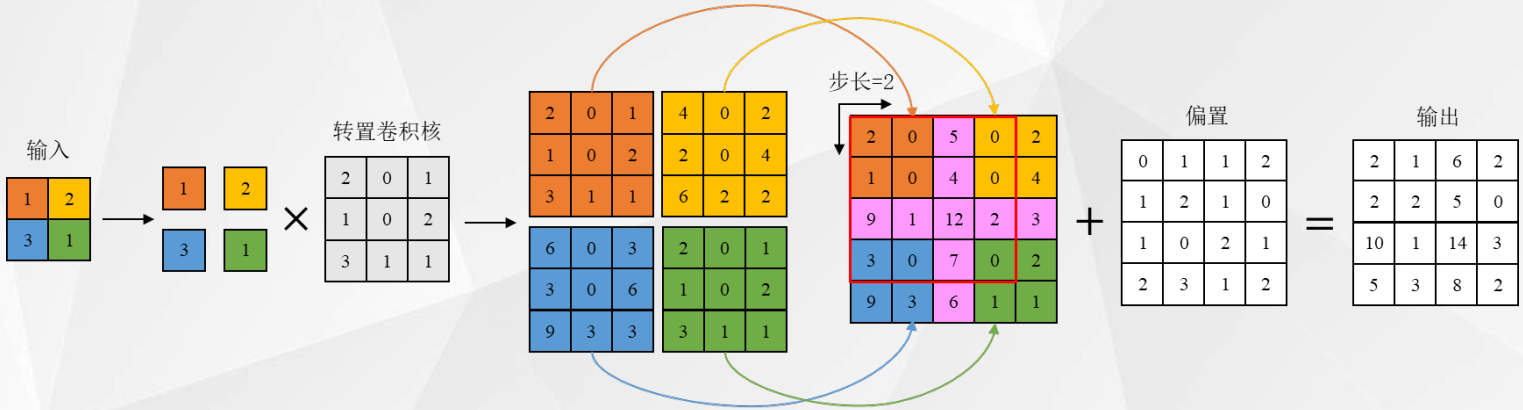
池化层：减小特征图的尺寸，保留主要特征的同时减少参数和计算量。池化层一般分为平均池化和最大池化。





转置卷积层

转置卷积层：对输入进行上采样，放大缩小后的特征图。



03

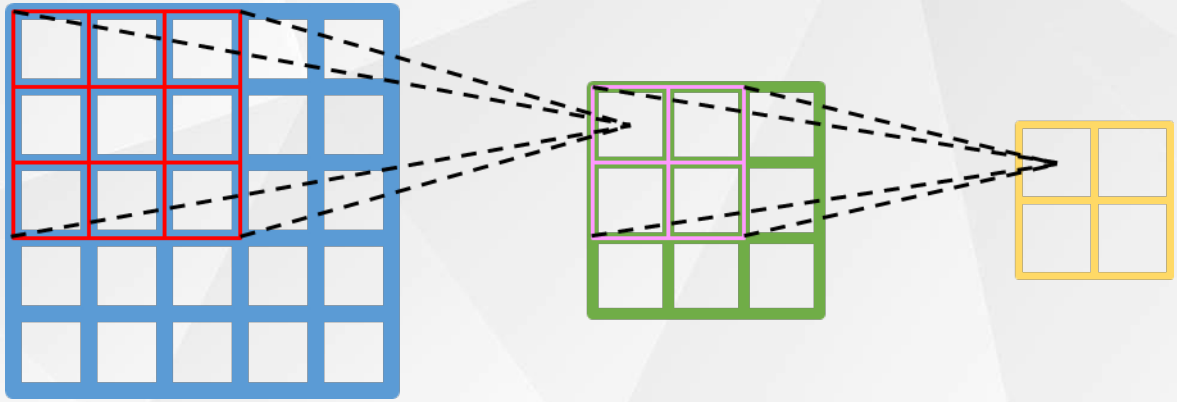
研究一：基于注意力机制的特征融合卷积网络

（此部分研究内容已被Structural Health Monitoring杂志录用）

注意力机制

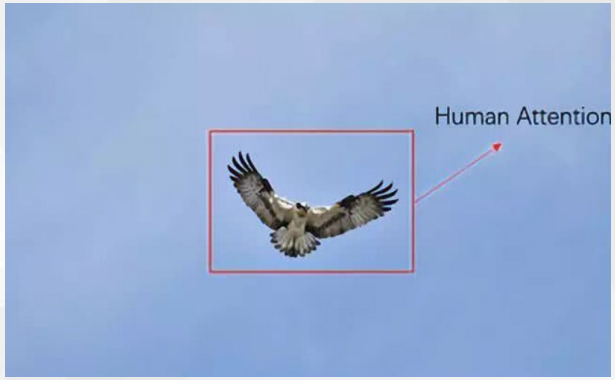
一般的语义分割模型有一个明显的缺点，就是容易忽略像素之间的关系，不能很好地聚合丰富的上下文信息，这会影响裂缝分割的可靠性。

深度学习认为图像中的每一个像素点不可能是孤立的，一个像素一定和周围像素是有一定的关系的，大量像素的互相联系才产生了图像中的各种物体。因此，一个像素与周围越多的像素有关系，模型的性能越好。



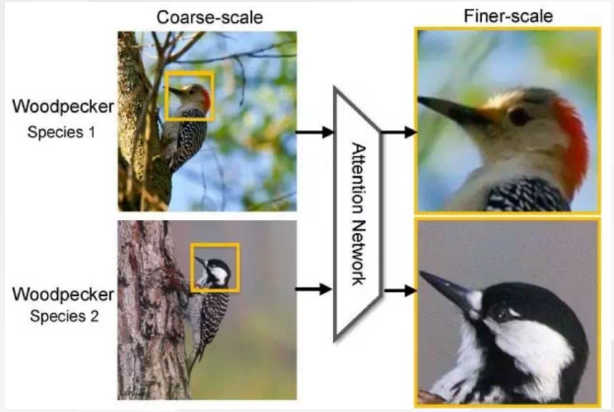
注意力机制

注意力机制的研究动机受人脑注意力的启发，人脑在感知物体的时候，一般不会将一个场景从头看到尾，而是根据需求观察特定的部分，并忽略其它不相关的部分。



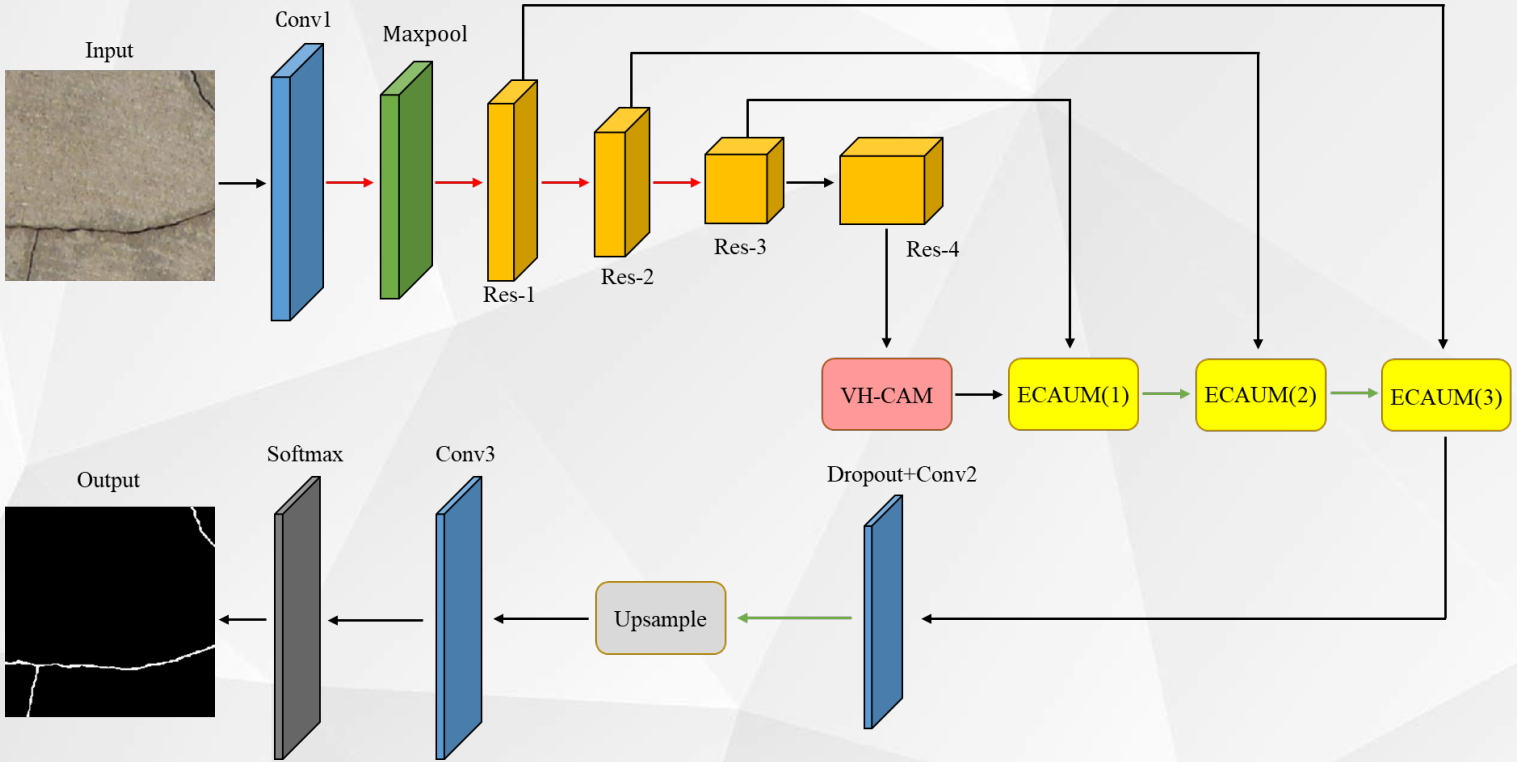
深度学习模型认为图像中所有像素的重要性是相同的。如果你不告诉模型你想关注鸟，由于整张图像中关于天空的信息更多，模型会认为这是一张有关天空的图像，而不是鸟。

注意力机制帮助模型关注重要的信息，忽略无关信息。



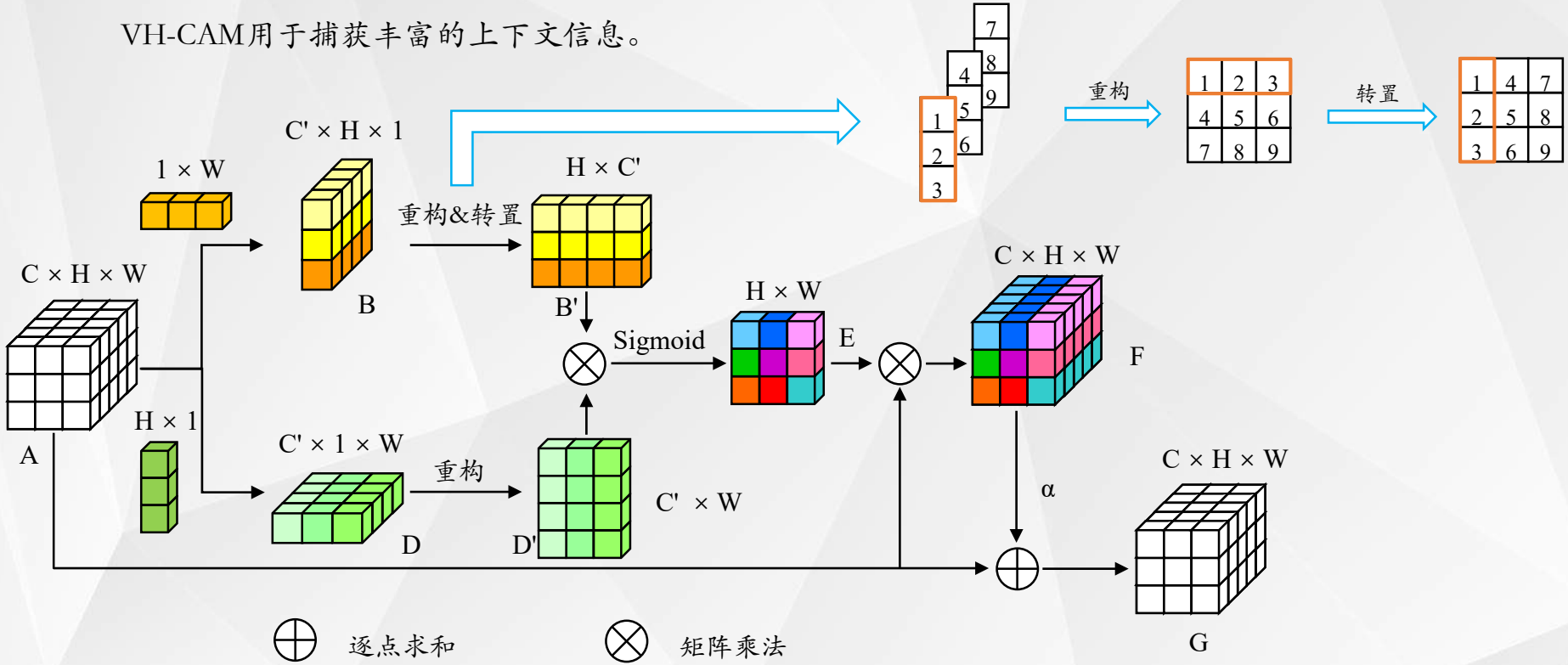
基于注意力的特征融合网络

设计了一种结合语义分割和注意力机制的模型，名为AFFNet。模型的主干网为ResNet101，注意力机制包括垂直和水平压缩注意力模块（VH-CAM）和高效通道注意力上采样模块（ECAUM）。



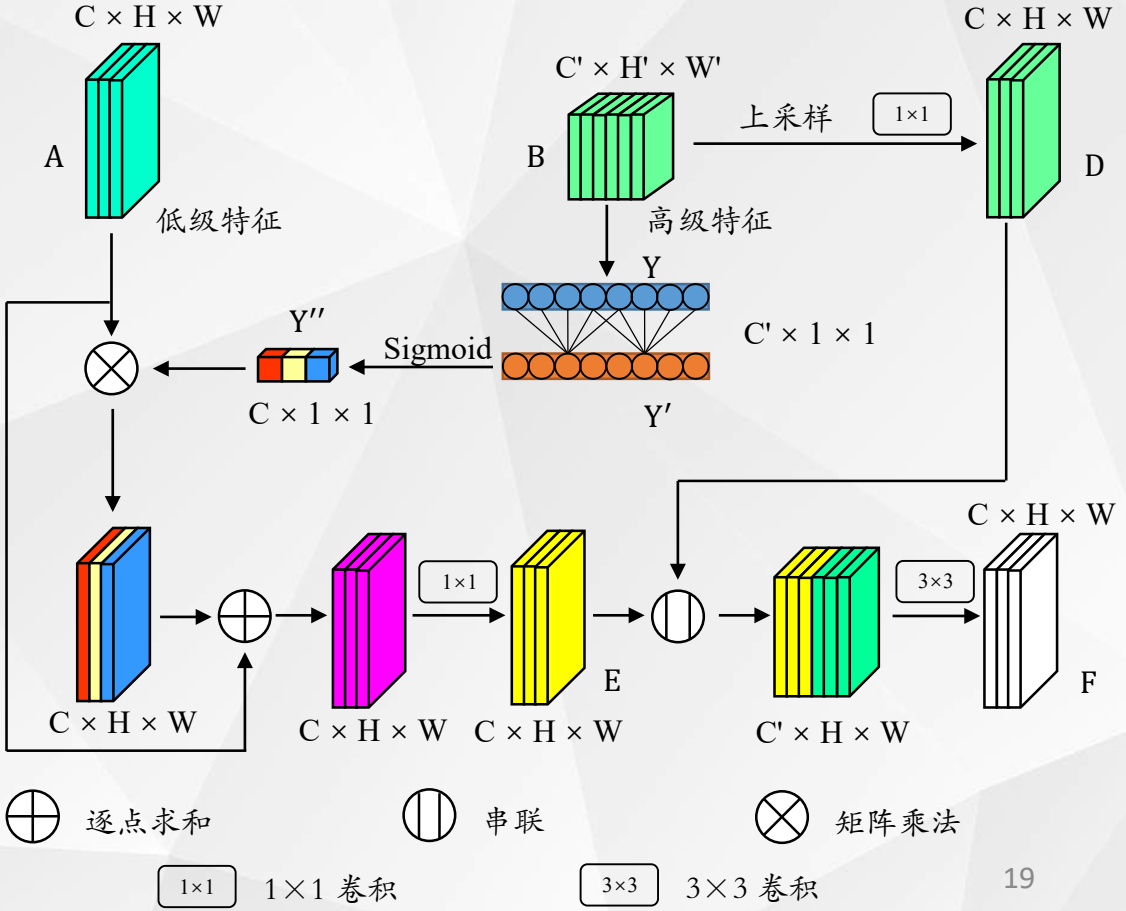
垂直和水平压缩注意力模块（VH-CAM）

VH-CAM用于捕获丰富的上下文信息。



高效通道注意力上采样模块（ECAUM）

直接的特征融合会降低裂缝分割的性能，ECAUM用于提供高级语义信息，为低级特征恢复像素定位提供指导。



评估指标

假设图像中存在100个像素，其中20个是裂缝像素，80个是背景像素。

像素精度：这是最简单的一个评估指标，只需要计算正确预测的像素数量占总像素数量的比例。

$$\mathrm{PA} = \frac{15+70}{100} = 85\%$$

平均像素精度：这是一种对像素精度改进过的测量精度，首先需要计算出每个类别中正确预测的像素数量占该类别总像素数量的比例，然后计算出所有比例的平均值。

$$\mathrm{MPA} = \frac{1}{2}\left(\frac{15}{20}+\frac{70}{80}\right) = 81.25\%$$

实际 \ 预测	裂缝	背景
裂缝	15	5
背景	10	70

平均交并比：它已经成为一个衡量语义分割模型性能好坏的重要指标，从字面上可以解释为两个交集与并集的比例的平均值，这两个集合具体表示为图像分割时的真实情况和预测结果。

$$\mathrm{MIoU} = \frac{1}{2}\left(\frac{15}{10+15+5}+\frac{70}{10+70+5}\right) = 66.18\%$$

频率加权交并比：它根据每个类别的出现频率给每个类别不同的权重。

$$\mathrm{FWIoU} = \frac{20}{100}\left(\frac{15}{10+15+5}\right)+\frac{80}{100}\left(\frac{70}{10+70+5}\right) = 75.88\%$$

在上述的四种评估指标中，MIoU因具有代表性和很好的评估效果，已经成为一个最常用的评估指标。

实验结果

为了反映AFFNet的优异性能，使用相同的数据集来训练与测试四个语义分割模型，然后将这些模型的评估指标与AFFNet进行了比较。

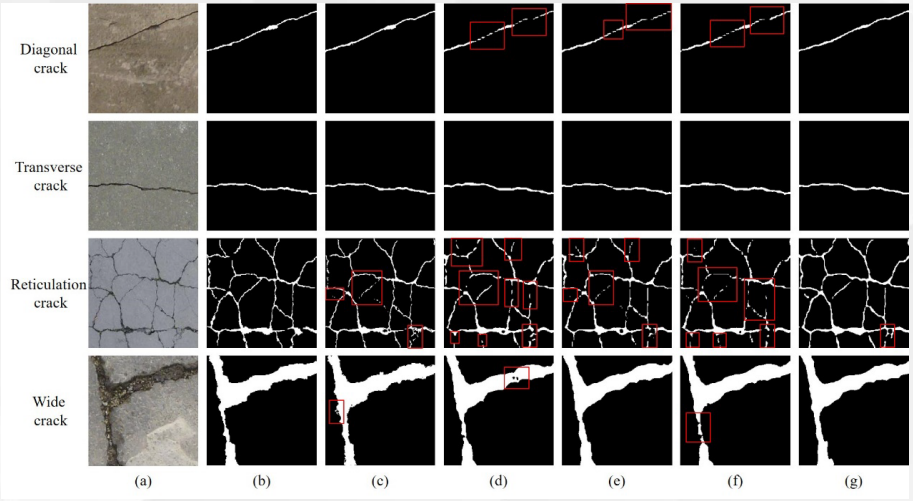
模型	PA	MPA	**MIoU**	FWIoU
U-Net	97.97	87.70	**80.85**	96.35
Dilated FCN	97.69	82.88	**76.18**	95.88
DeepLabv3+	97.87	86.42	**79.48**	96.18
PAN	97.79	84.16	**77.37**	96.04
AFFNet	98.11	91.25	**82.63**	96.65

为了体现 AFFNet 在训练时间上的优势，将AFFNet与四个模型 进行了对比。

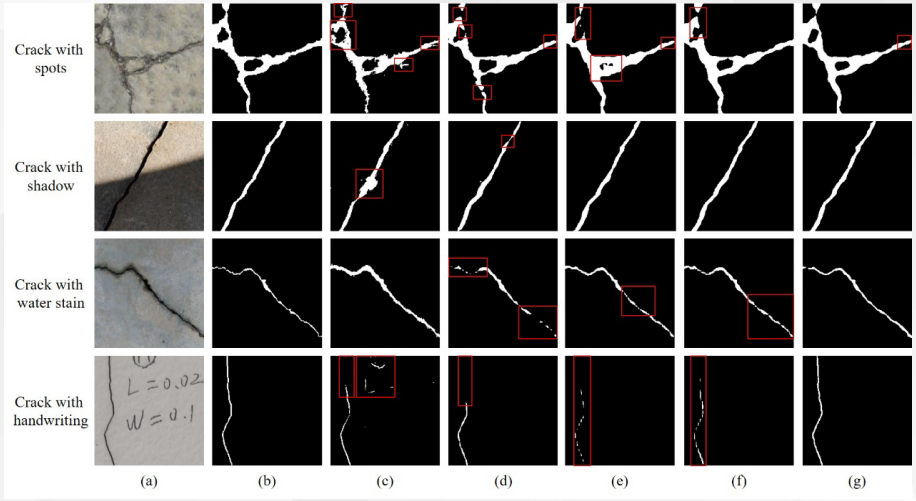
模型	主干网	时间（ms）
U-Net	-	66
DeepLabv3+	ResNet101	42
Dilated FCN	ResNet101	31
PAN	ResNet101	55
AFFNet	ResNet101	51

实验结果

无噪声裂缝的可视化结果



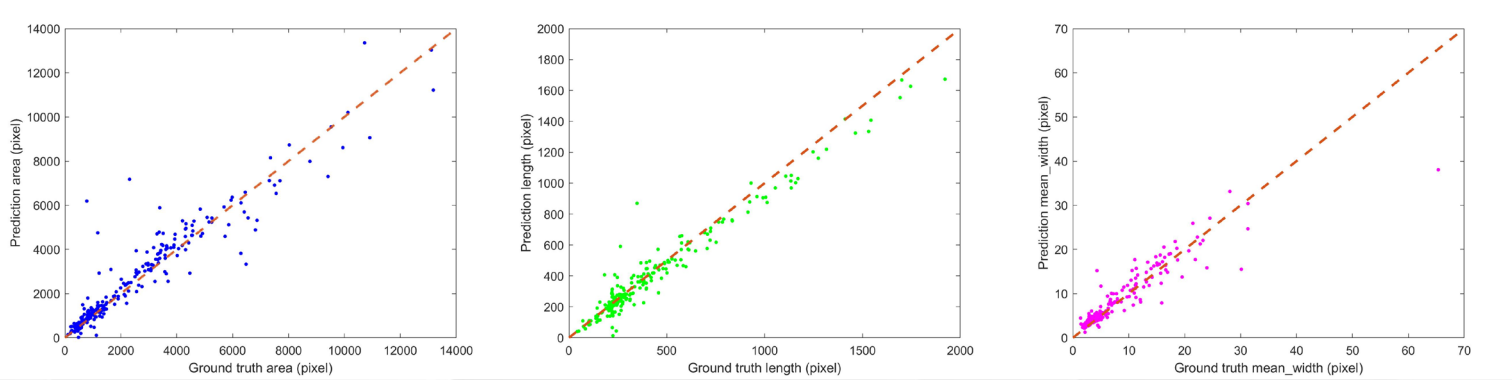
有噪声裂缝的可视化结果



（a）输入图像；（b）标签图像；（c）U-Net；（d）Dilated FCN；（e）DeepLabv3+；（f）PAN；（g）AFFNet

实验结果

在测试集中，对裂缝在像素上的三种形态特征进行量化：裂缝面积、长度和平均宽度。



(a) Crack area

(b) Crack length

(c) Crack mean width

大多数的绘制点位于诊断线附近。面积小于6000像素时，部分绘制点位于诊断线上方，大于6000像素时，部分绘制点位于诊断线下方。

大多数的绘制点位于诊断线附近。当长度大于500像素时，大多数的绘制点都位于诊断线的下方。

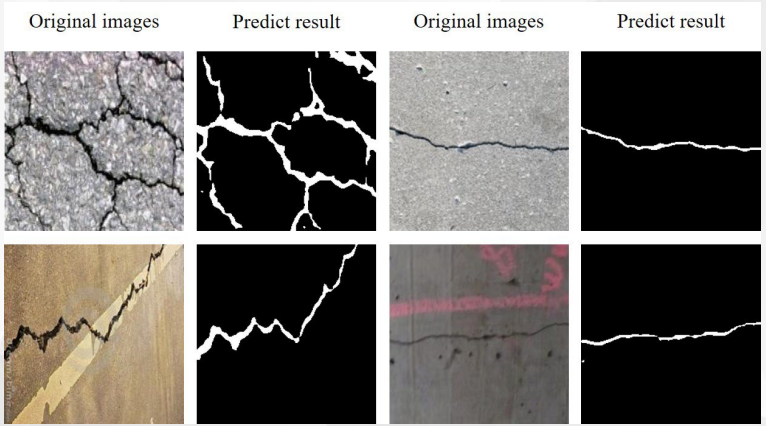
裂缝的平均宽度是裂缝面积与裂缝长度的比值。大多数的绘制点位于诊断线附近。

实验结果

为了测试AFFNet的泛化能力，使用DeepCrack和SDNET2018两个数据集进行实验。目的是测试如果不使用这两个数据集训练模型，模型是否可以检测出这些数据集中的裂缝。

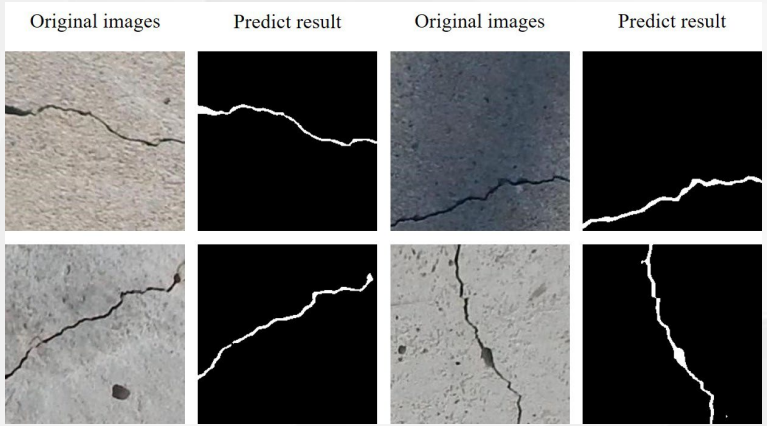
DeepCrack数据集（527张裂缝图像）

模型	PA	MPA	**MIoU**	FWIoU
U-Net	97.93	84.31	**76.42**	96.55
Dilated FCN	98.27	78.91	**74.09**	96.80
DeepLabv3+	98.47	82.21	**76.85**	97.22
PAN	98.27	79.34	**74.33**	96.79
AFFNet	98.53	88.25	**79.96**	97.43



SDNET2018数据集（50张裂缝图像）

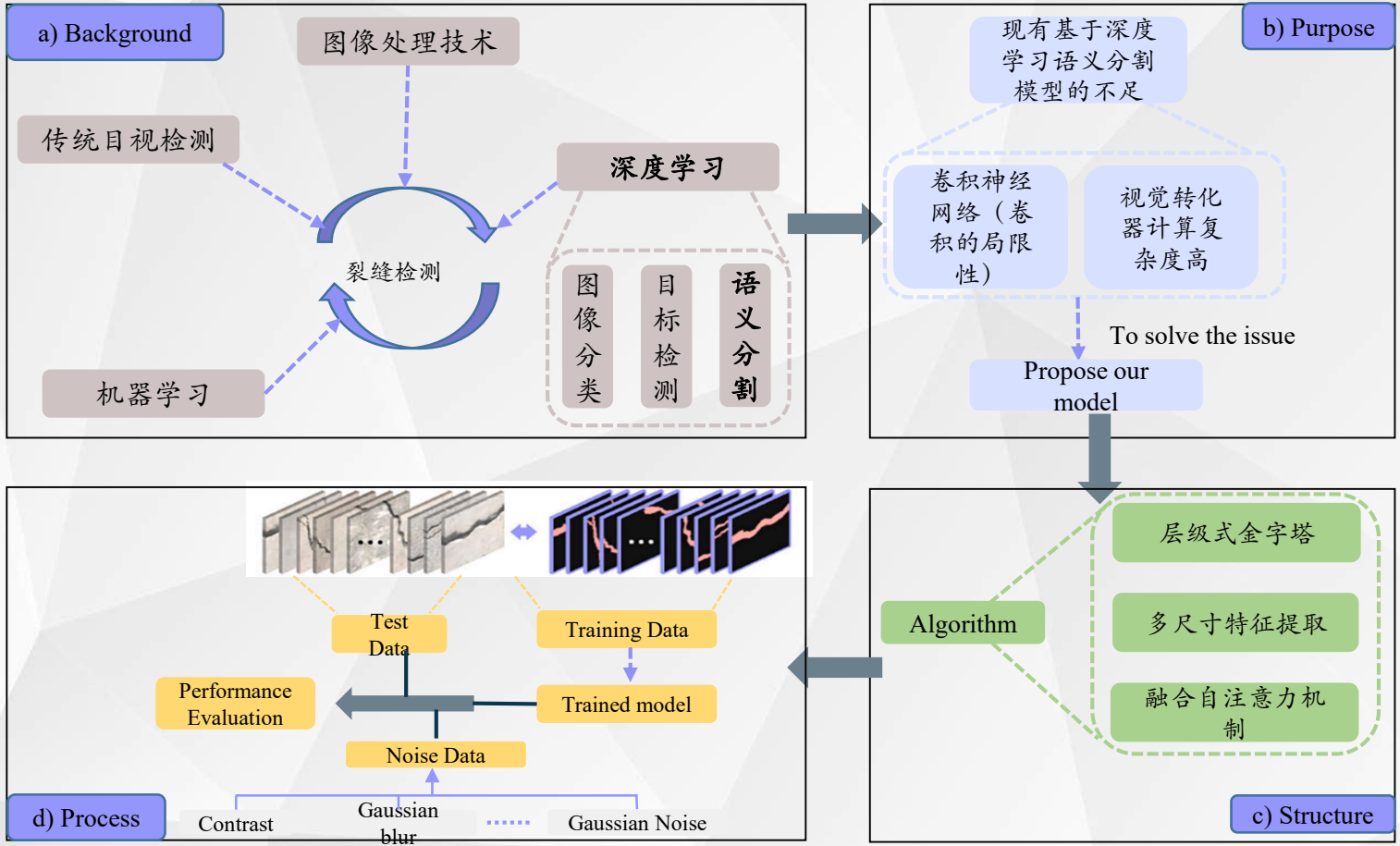
模型	PA	MPA	**MIoU**	FWIoU
U-Net	99.34	85.93	**83.98**	98.71
Dilated FCN	99.10	86.04	**80.54**	98.33
DeepLabv3+	99.27	87.30	**82.97**	98.61
PAN	99.14	85.33	**80.42**	98.39
AFFNet	99.38	93.91	**87.29**	98.86



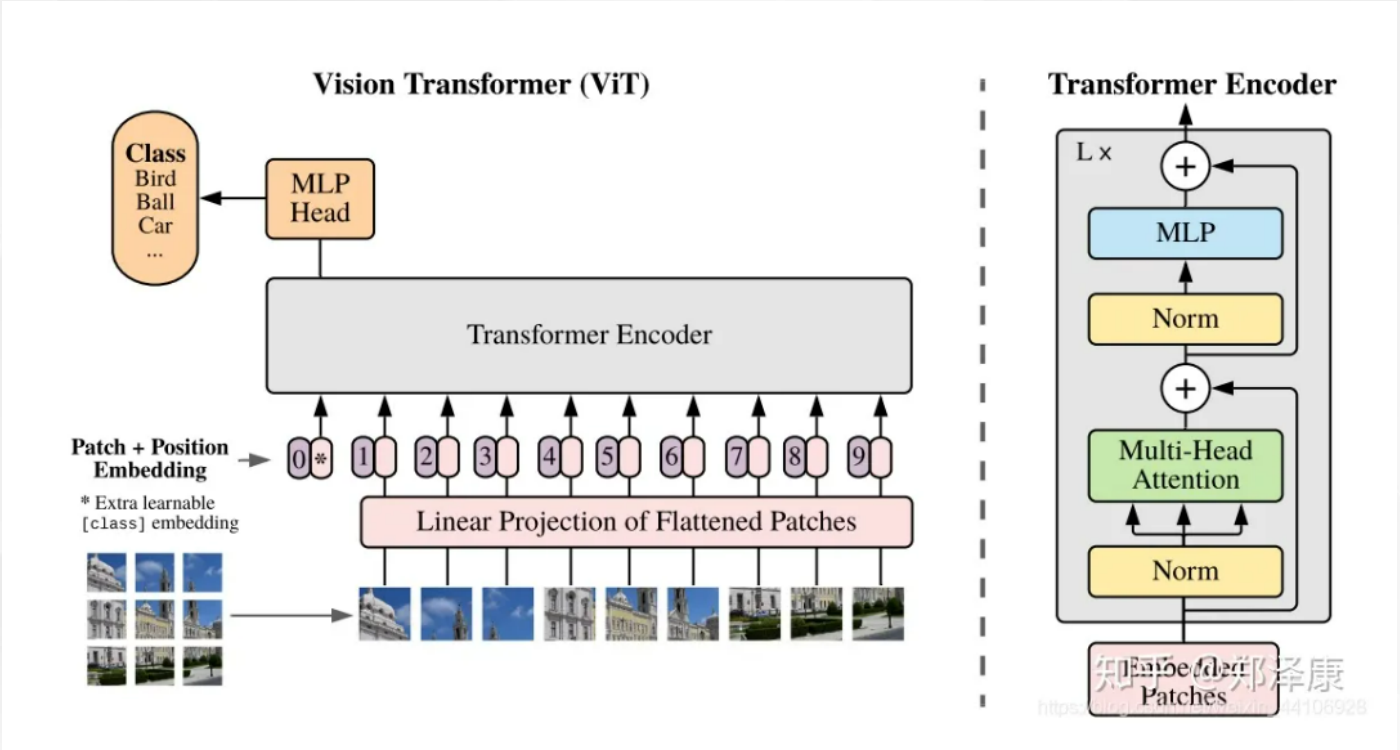
04

研究二：CrackSegFormer: 基于Vision Transformer的高效混凝土裂缝分割算法

（此部分内容论文投稿在审）


a) Background
图像处理技术
传统目视检测
深度学习
裂缝检测
机器学习
图像分类
目标检测
语义分割
b) Purpose
现有基于深度学习语义分割模型的不足
卷积神经网络（卷积的局限性）
视觉转化器计算复杂度高
To solve the issue
Propose our model
c) Structure
Algorithm
层级式金字塔
多尺寸特征提取
融合自注意力机制
d) Process
Test Data
Training Data
Performance Evaluation
Trained model
Noise Data
Contrast
Gaussian blur
Gaussian Noise

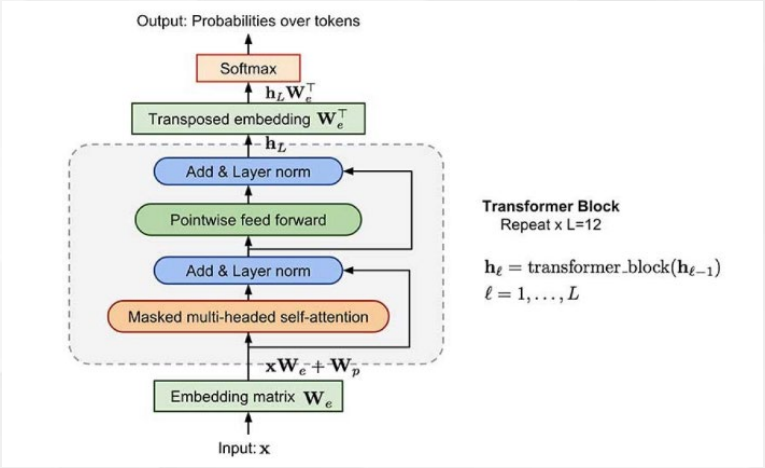
Transformer

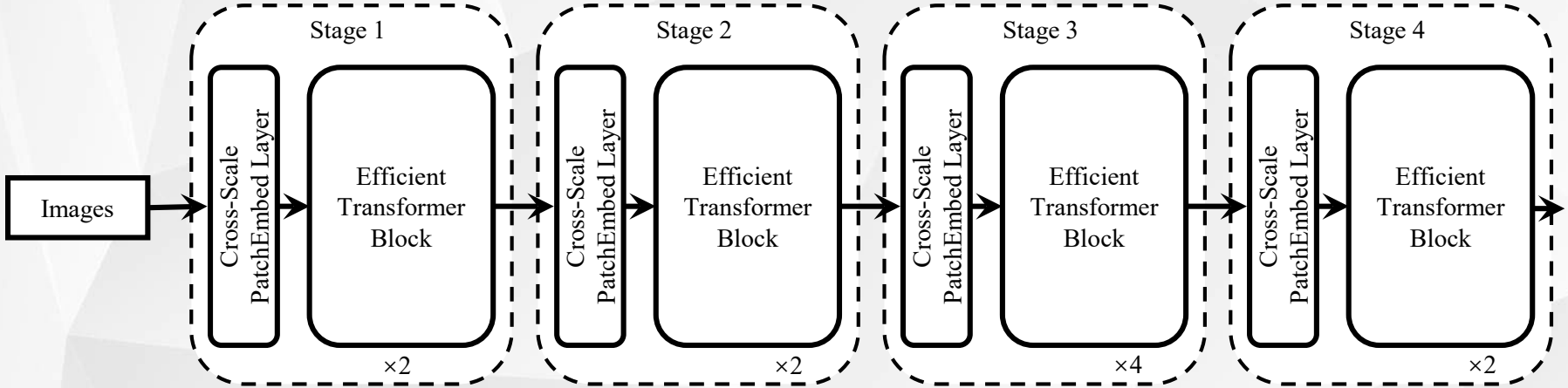


CrackSegFormer

◆Cross-Scale PatchEmbed Layer

◆Efficient Transformer block





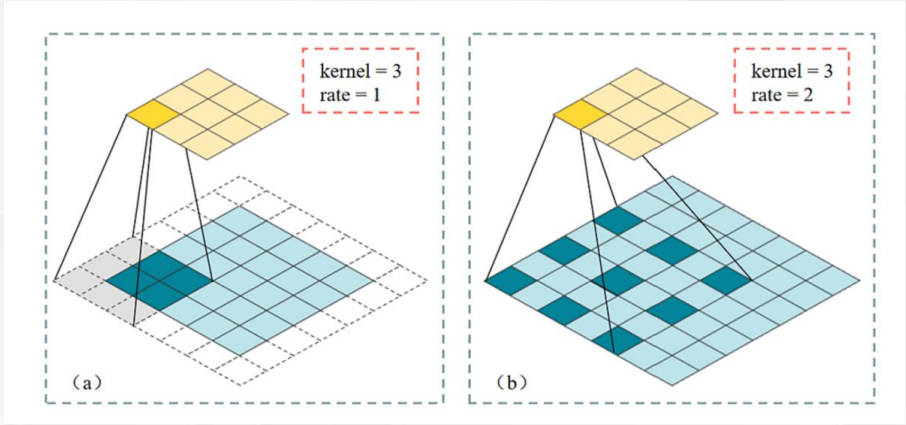
(a) Architecture of CrackSegFormer

Cross-Scale PatchEmbed Layer

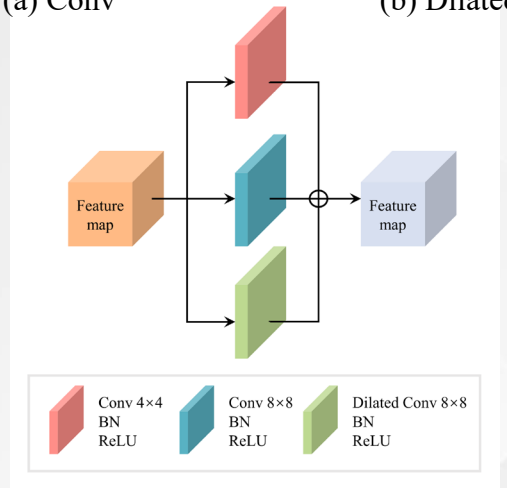
- 高效的多尺寸特征提取：普通卷积和空洞卷积
- 不同尺寸的卷积核来提取不同尺寸的特征
- 空洞卷积进一步扩大感受野

Type	Kernel	Stride	Dilation	Padding	Dim
Conv	4×4	4×4	/	/	dim/2
Conv	8×8	4×4	/	6×6	dim/4
Dilated Conv	3×3	4×4	6×6	6×6	dim/4

(c) Specific parameters of Cross-Scale PatchEmbedding Layer



(a) Conv (b) Dilated Conv



(d) Architecture of Cross-Scale PatchEmbed Layer

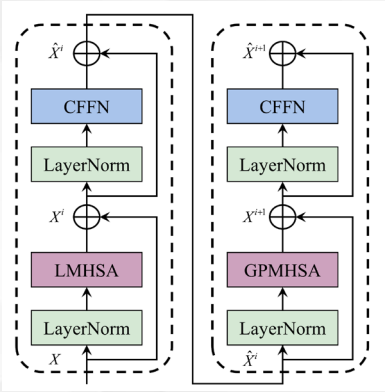
Efficient Transformer block

Local-Multi-Head Self-Attention 局部多头自注意力机制

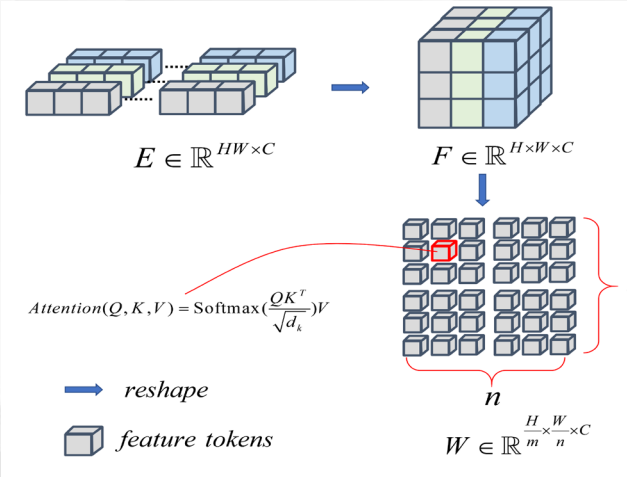
- 在子窗口上进行自注意力计算，充分提取局部特征
- 大幅降低计算复杂度

Global Pooling-Multi-Head Self-Attention

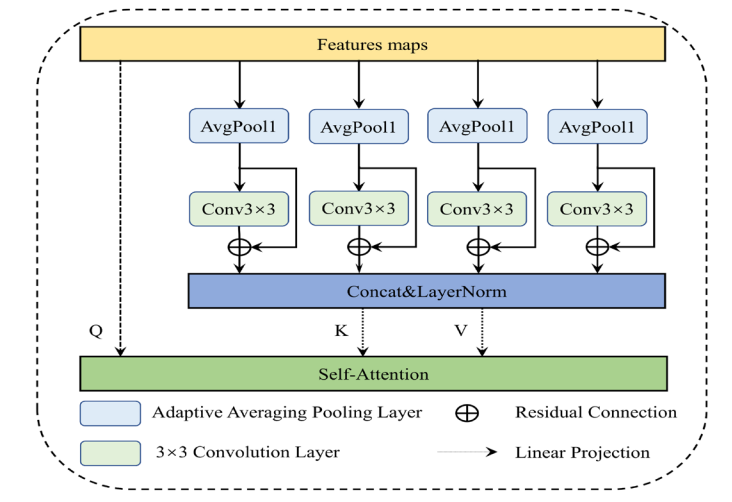
- 使用Pooling和Conv 3×3进一步提取有效特征信息
- 全局特征和局部特征之间的信息得到交互



(a) Architecture of Efficient Transformer block

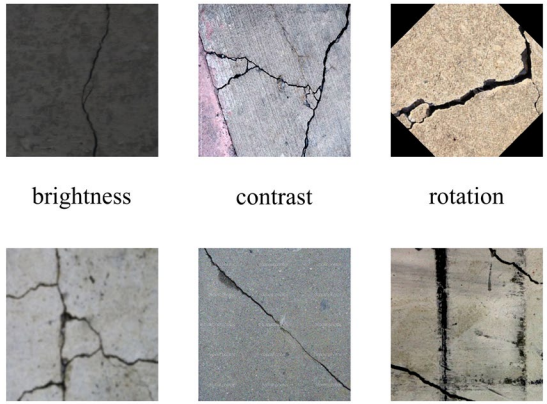


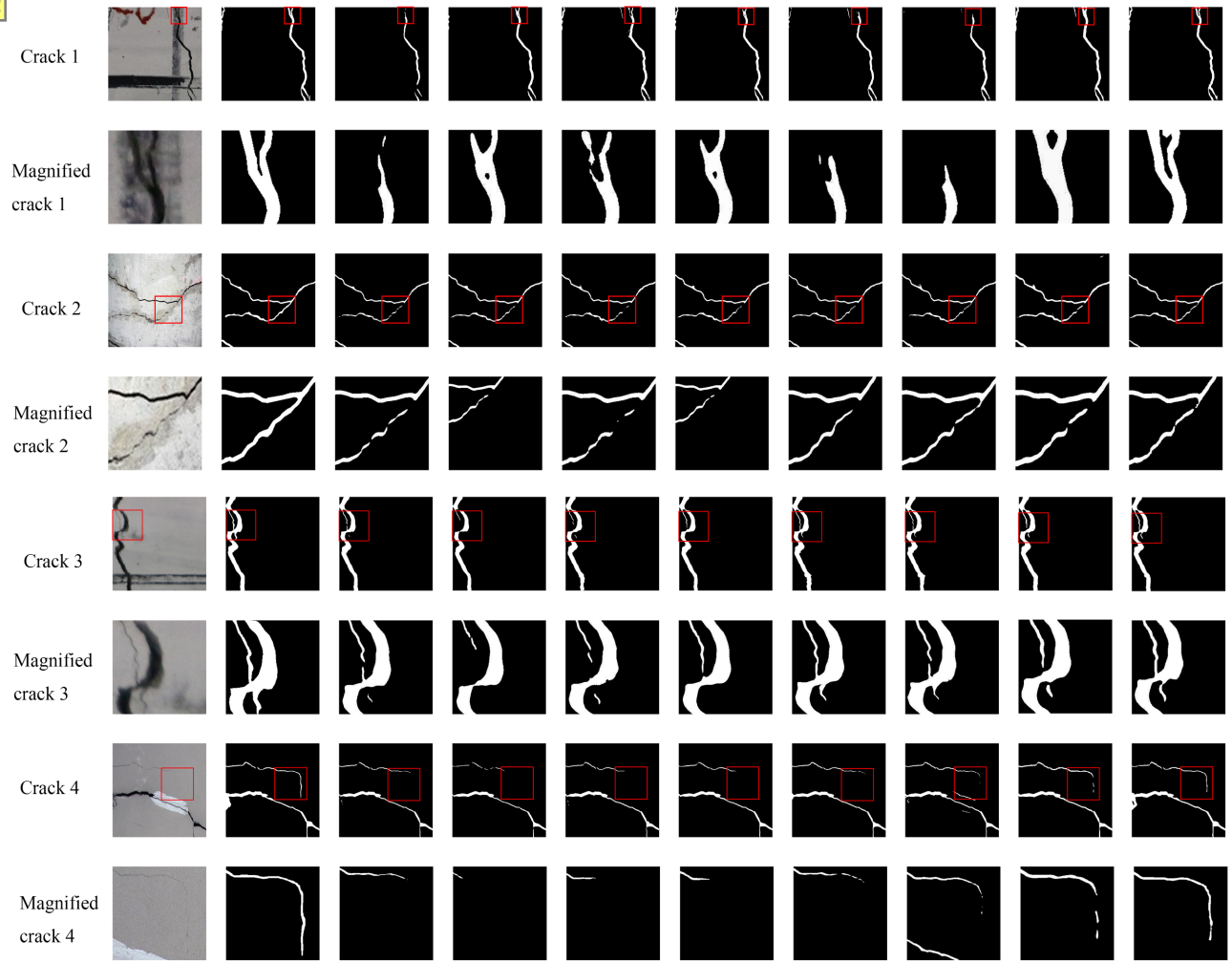
(b) Architecture of Local-Multi-Head Self-Attention



(c) Global Pooling-Multi-Head Self-Attention

- Dataset： 5985 crack images, 80% for training, 20% for validation.
- Data augmentation: brightness, contrast, rotation, Gaussian blur, Gaussian noise, and pepper-salt noise.
- Performance metrics: precision, recall, F1-score, and mIoU.
- State-of-the-art (SOTA) model: Deeplabv3plus, Pspnet, U-Net, FCN, PVT, PVTv2-b5, Swin Transformer-small.





(a) Input image.

(b) Ground truth.

(c) Deeplabv3plus.

(d) Pspnet.

(e) U-Net.

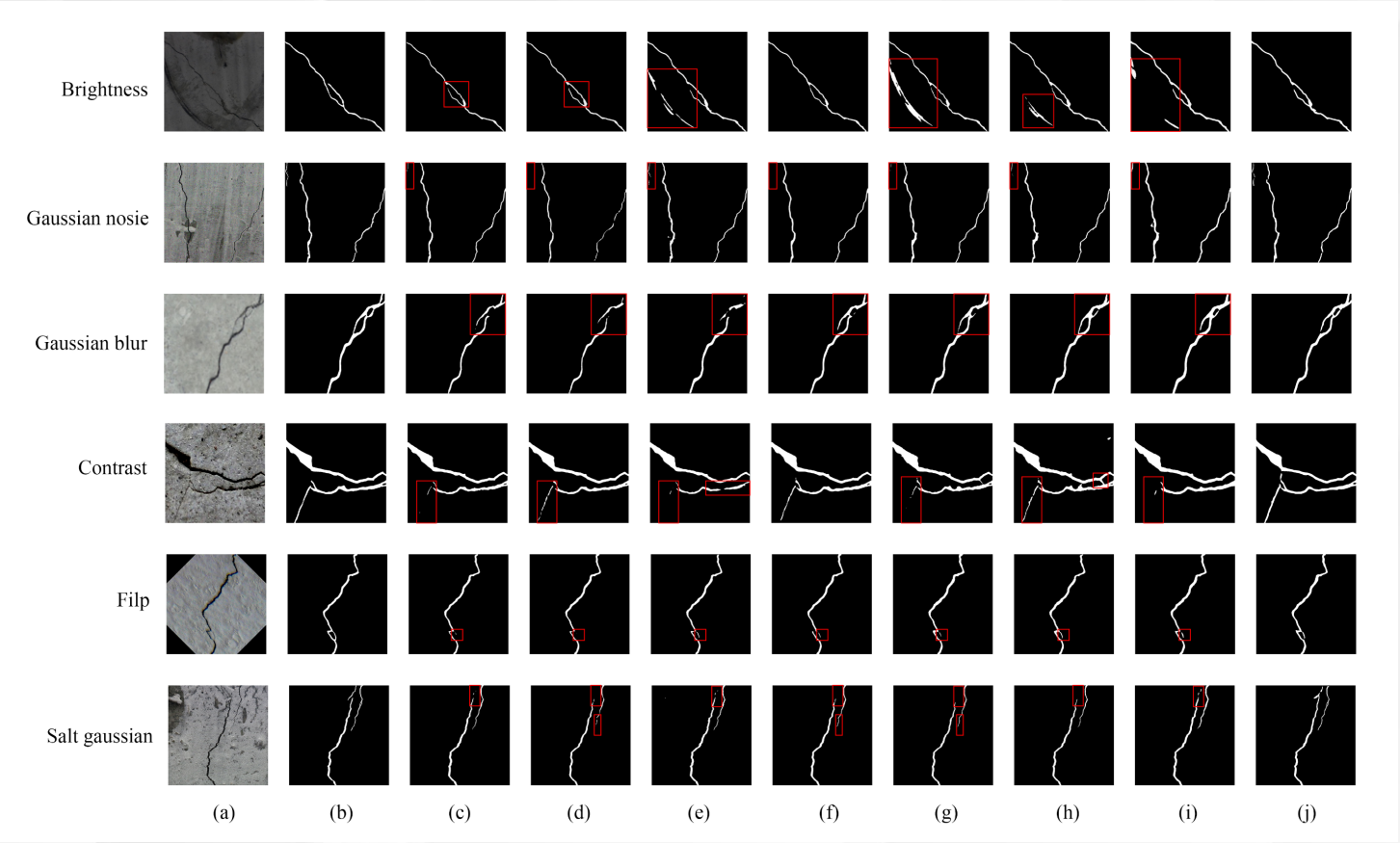
(f) FCN.

(g) PVT.

(h) PVTv2-b5.

(i) Swin Transformer-small.

(j) The proposed CrackSegFormer.



(a) Input image.

(b) Ground truth.

(c) Deeplab v3 plus.

(d) Pspnet.

(e) U-Net.

(f) FCN.

(g) PVT.

(h) PVTv2-b5.

(i) Swin Transformer-small.

(j) The proposed CrackSegFormer.

(a) The validation results of comparative models.

Model		Precision (%)	Recall (%)	F1-score (%)	mIoU (%)
CNN-based	Deeplabv3plus	96.61	83.17	89.02	81.83
	Pspnet	96.4	83.89	89.38	82.32
	U-Net	95.32	87.67	91.12	84.79
	FCN	96.28	89.49	92.65	85.09
Transformer-based	PVT-Medium	89.73	89.62	92.46	86.8
	PVT-Large	90.31	89.84	92.47	86.81
	PVTv2-B4	89.51	89.34	91.87	85.9
	PVTv2-B5	89.52	89.55	92.08	86.23
	Swin-tiny	93.69	92.68	91.73	87.13
	Swin-small	94.36	93.2	92.11	87.94
	CrackSegFormer	96.89	93.77	94.8	90.53

(b) The comparison of models' computational efficiency

Model	Parameters (M)	FLOPs (G)	mIoU (%)
PVT-Medium	43.3	103.25	86.8
PVT-Large	60.47	152.34	86.81
PVTv2-B4	62.04	157.12	85.9
PVTv2-B5	81.44	182.10	86.23
Swin-tiny	27.5	69.97	87.13
Swin-small	48.79	136.74	87.94
CrackSegFormer	30.05	65.55	90.53

05

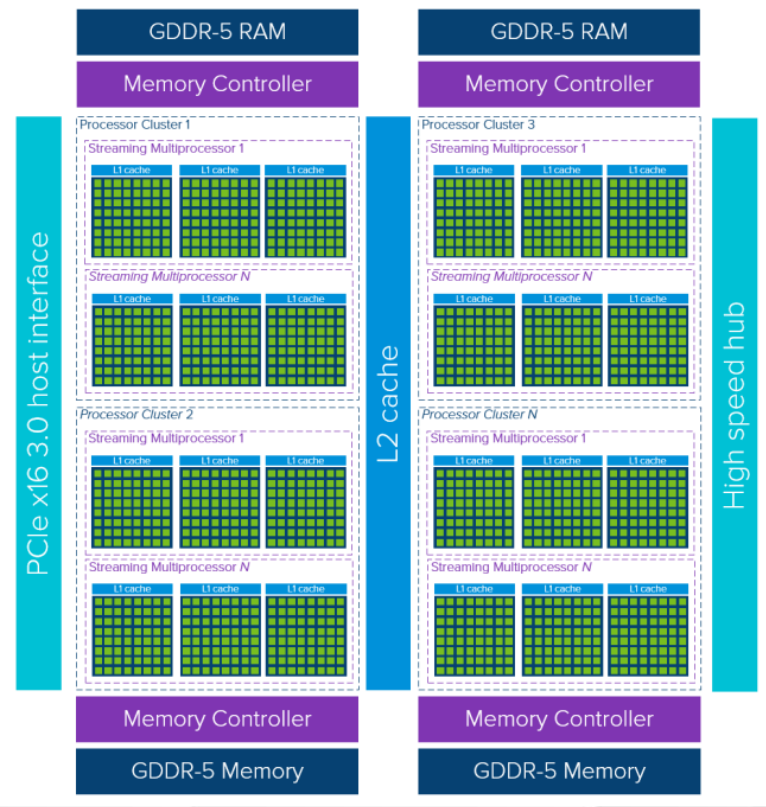
研究三：轻量化裂缝识别算法设计准则和实例

（此部分内容论文撰写中）

研究目的

- 复杂算法面临的挑战
  - 对GPU、AI芯片内存容量和算力需求大
  - 多适用于离线后处理
- 轻量化算法的必要性
  - 便携式系统大势所趋
  - 一体化处理系统：无人机+裂纹识别+结果展示
  - 降低时间和硬件成本

亟需开发参数少、算力需求相对较弱且识别精度高的轻量化算法



2. 轻量级神经网络——技术路径

南京工业大学 NANJING TECH UNIVERSITY

轻量级模型设计原则：

1 改进底层实现方式

如sigmoid函数在实现过程中采用近似的方式，不仅很复杂，还会导致精度损失，因此MobileNet_v3中提出了h-swish非线性激活函数，这个函数的特点就是底层实现很简单，也不会导致推理阶段的精度损失。

2 减少参数量

① 使用深度可分离卷积；
② 使用分解卷积；
③ 使用分组卷积；
④ 使用特征表示能力强的结构；
⑤ 使用1x1卷积代替3x3卷积。

3 减少计算量

2中的五种方式都同时具有减少计算量的作用，
此外，特征复用也具备减少计算量的作用。

4 降低实际运行时间

① 使用等通道宽度卷积；
② 控制好组卷积的组数；
③ 降低碎片化程度；
④ 减少元素操作。

3. 轻量级神经网络——设计过程

南京工业大学 NANJING TECH UNIVERSITY

1 采用深度可分离卷积

引入了深度可分离卷积作为传统卷积层的有效替代，大大减少计算量。

2 调整卷积层顺序

将depthwise conv上移。用1×1 conv替代大尺寸卷积核进行通道数扩展，可以有效降低计算量。

3 调整卷积核大小

将depthwise conv的卷积核大小由3×3改成了5×5，过程中尝试了3、 5、 7、 9，取到5时准确率就达到了饱和。

4 激活函数

尝试了多种激活函数组合，最终确定第一层模块采用ReLU，后三层采用GELU。此时，效果最佳。

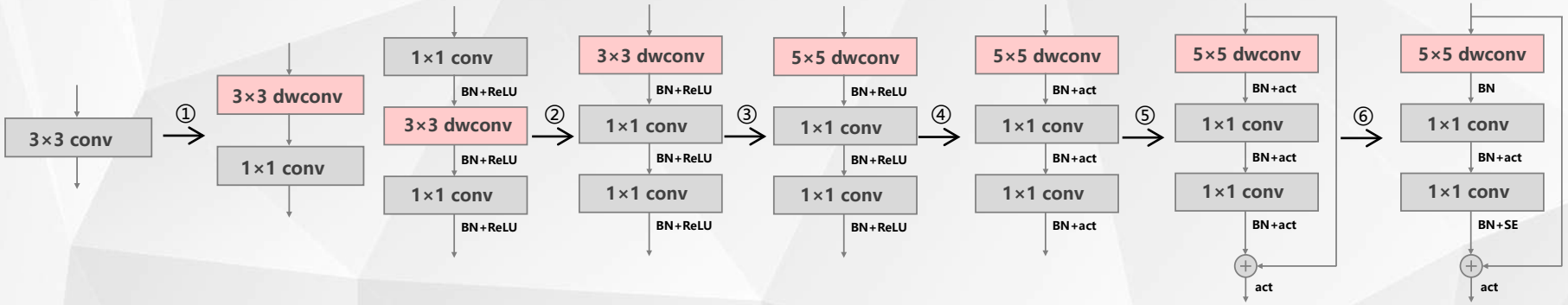
5 残差连接

采用ResNet的残 差 连接模式，将残差路径与快捷路径相加，然后发送到激活层，性能更好，更适合裂纹图像分割。

6 调整激活函数数量，加入注意力

减少激活函数的使用，准确率增长。

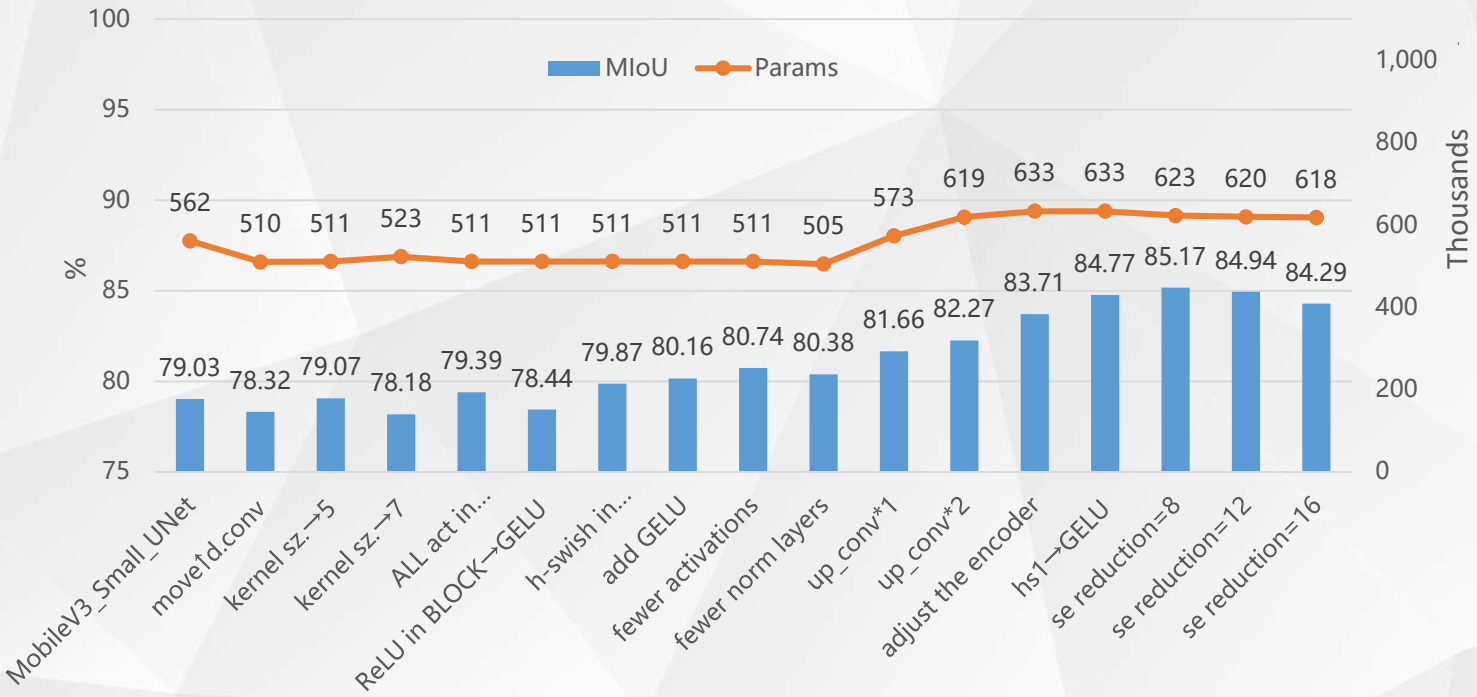
接着加入SE注意力，提高特征提取能力。



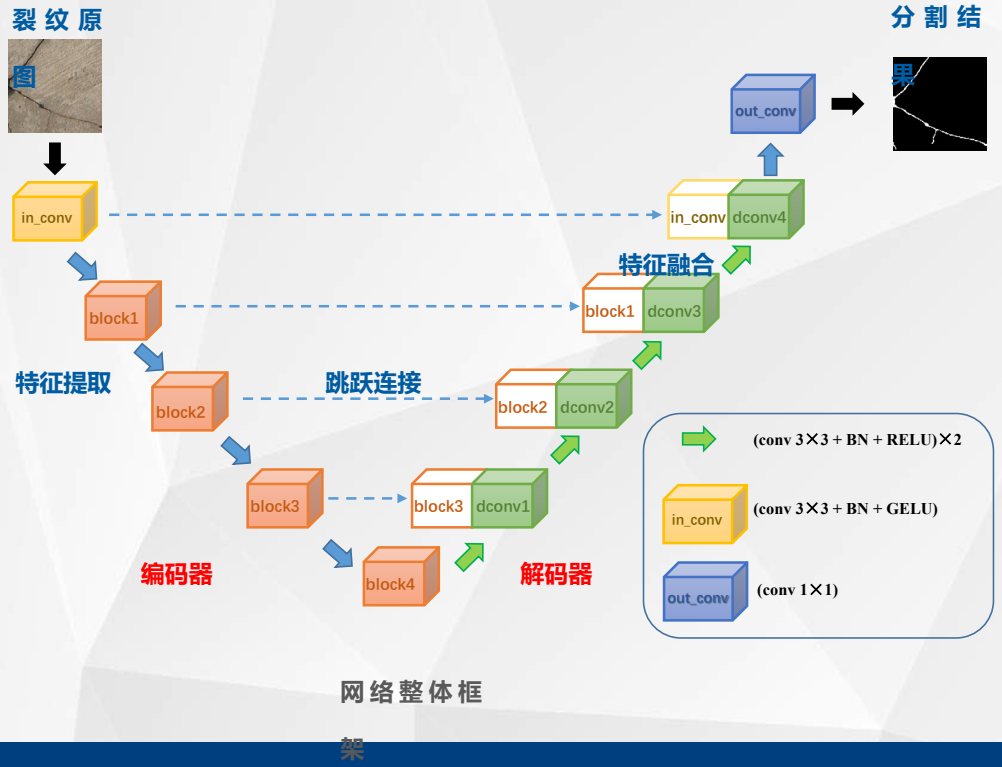
3. 轻量级神经网络——设计过程

南京工業大學 NANJING TECH UNIVERSITY

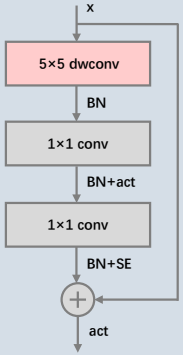
模型设计过程中准确率和参数量的变化：



4. 算法框架



网络整体框架



bneck 结构图

表1：模型参数

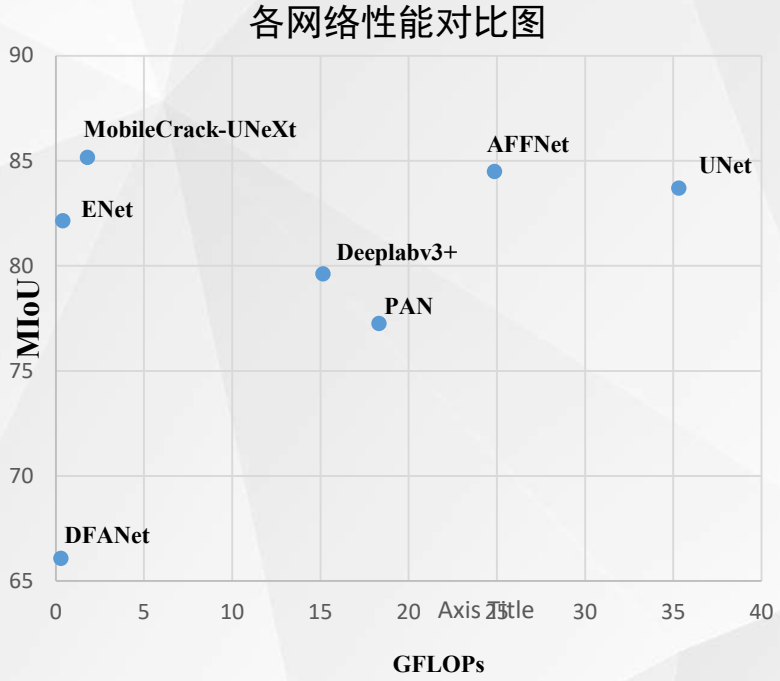
	input	operator	exp size	output	SE	act	s
in_conv	224*224*3	conv 3*3	-	16	-	GE	2
block1	224*224*16	bneck 5*5	72	24	-	RE	2
block1	112*112*24	bneck 5*5	88	24	-	RE	1
block2	112*112*24	bneck 5*5	96	40	√	GE	2
block2	56*56*40	bneck 5*5	240	40	√	GE	1
block3	56*56*40	bneck 5*5	120	48	√	GE	2
block3	28*28*48	bneck 5*5	144	48	√	GE	1
block4	28*28*48	bneck 5*5	288	96	√	GE	2
block4	14*14*96	bneck 5*5	576	96	√	GE	1

注：SE表示该块中是否使用SE注意力；act表示使用的激活函数类型：GE表示GELU激活函数，RE表示ReLU激活函数；s表示步距（stride）。

5. 与经典语义分割网络对比

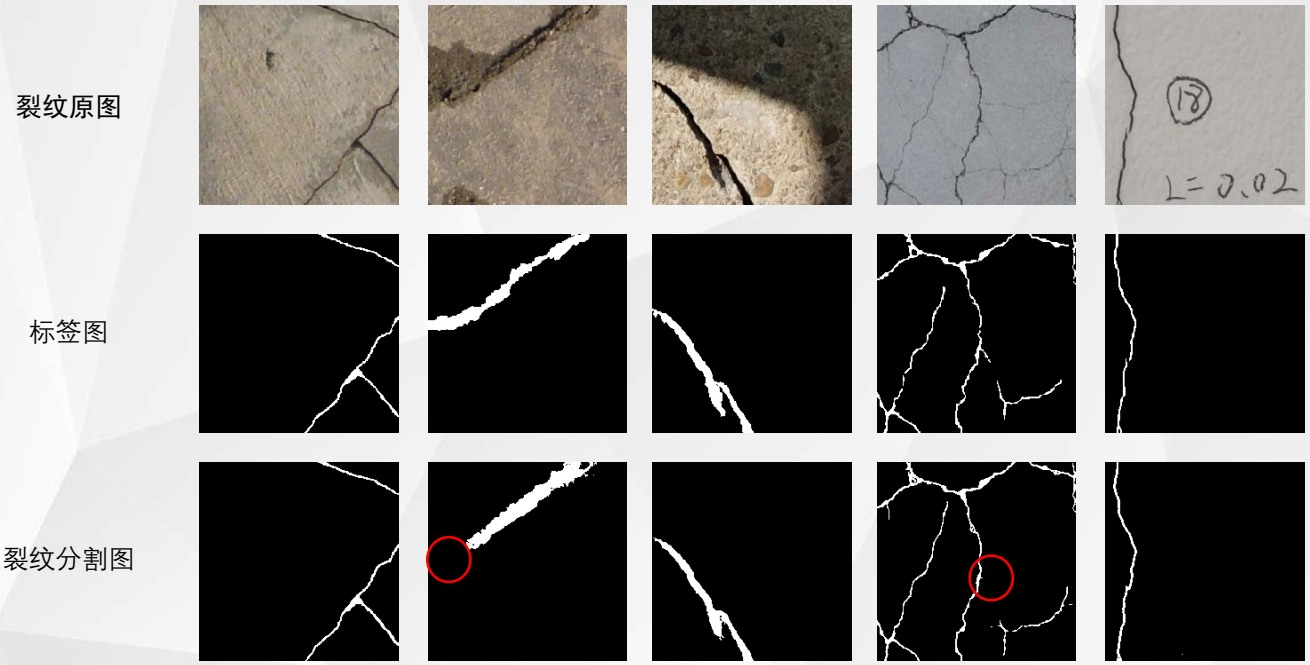
Models	PA	MPA	MIoU	FWIoU	GFLOPs	Params
UNet	98.23	90.29	83.71	96.78	35.32G	31.03M
Deeplabv3+	97.80	86.57	79.62	96.04	15.16G	58.81M
PAN	97.60	84.07	77.26	95.68	18.32G	83.82M
AFFNet	98.36	92.01	84.49	97.07	24.88G	82.93M
MobileCrack-UNeXt	98.39	91.36	**85.17**	**97.07**	1.8G	**0.6M**
ENet	98.11	90.66	82.15	96.64	0.4G	0.3M
DFANet	95.56	76.46	66.08	92.83	0.3G	2.15M

输入为 [1, 3, 224, 224]



- 与经典网络UNet、Deeplabv3+等相比，所提出网络计算量和参数量大幅度减少
- 与实时语义分割网络ENet、DFANet等相比，所提出网络准确率更高
- 在训练速度上：更快，ENet、DFANet每轮训练为20s，所提出网络仅需13s

6. 裂纹分割效果

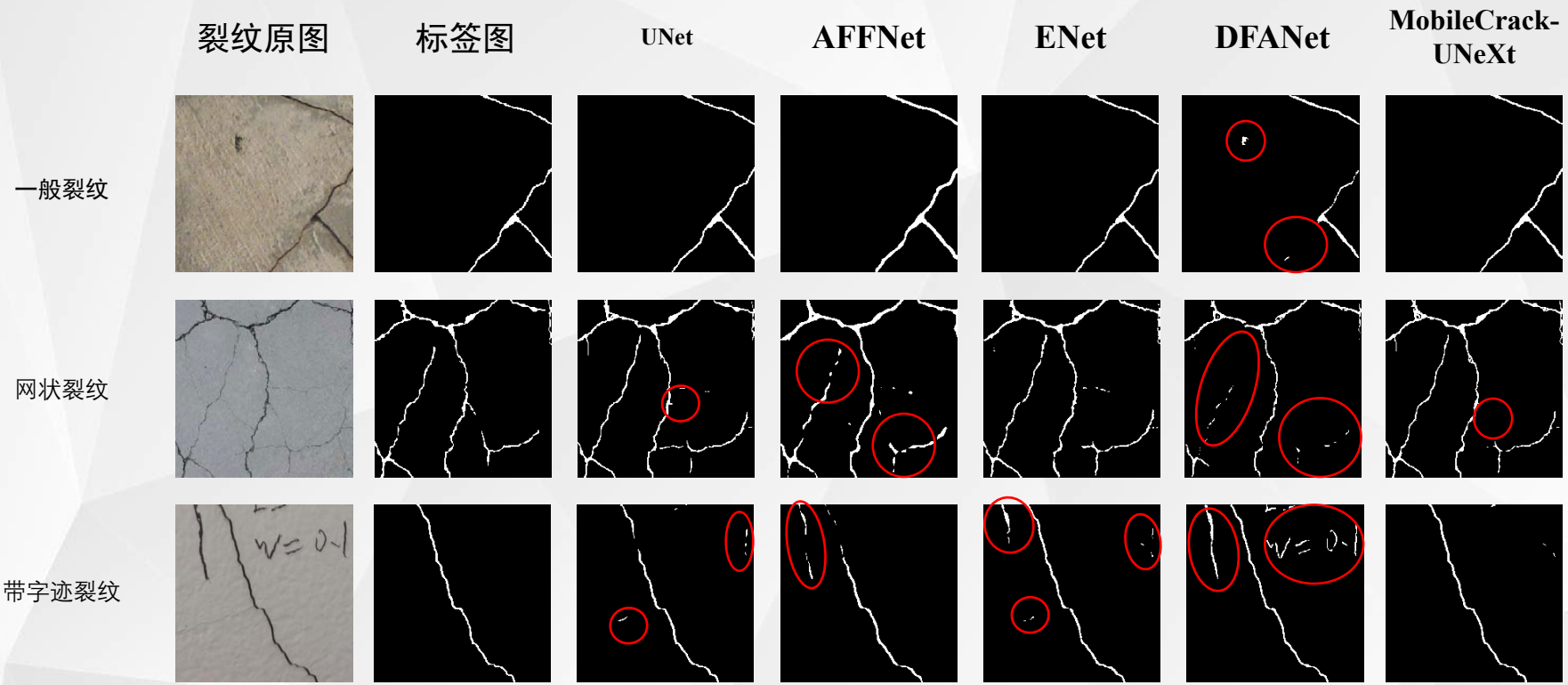


整体可以看出检测效果较好

其中：

- 水渍处因形状不似裂纹而未被检测出
- 网状裂纹的极细小处检测时有所丢失
- 一般裂纹，及手写和阴影裂纹检测效果未受影响

6. 裂纹分割效果


裂纹原图
标签图
UNet
AFFNet
ENet
DFANet
MobileCrack-UNeXt
一般裂纹
网状裂纹
带字迹裂纹

06

研究四：复杂裂缝量化框架

（此部分内容论文撰写中）

研究目的

- 现有裂缝识别算法的局限性：可按像素标记裂缝，裂缝几何信息识别效率低
- 裂缝扩展预测的需求：三个裂缝的宽度、长度及随时间变化数据

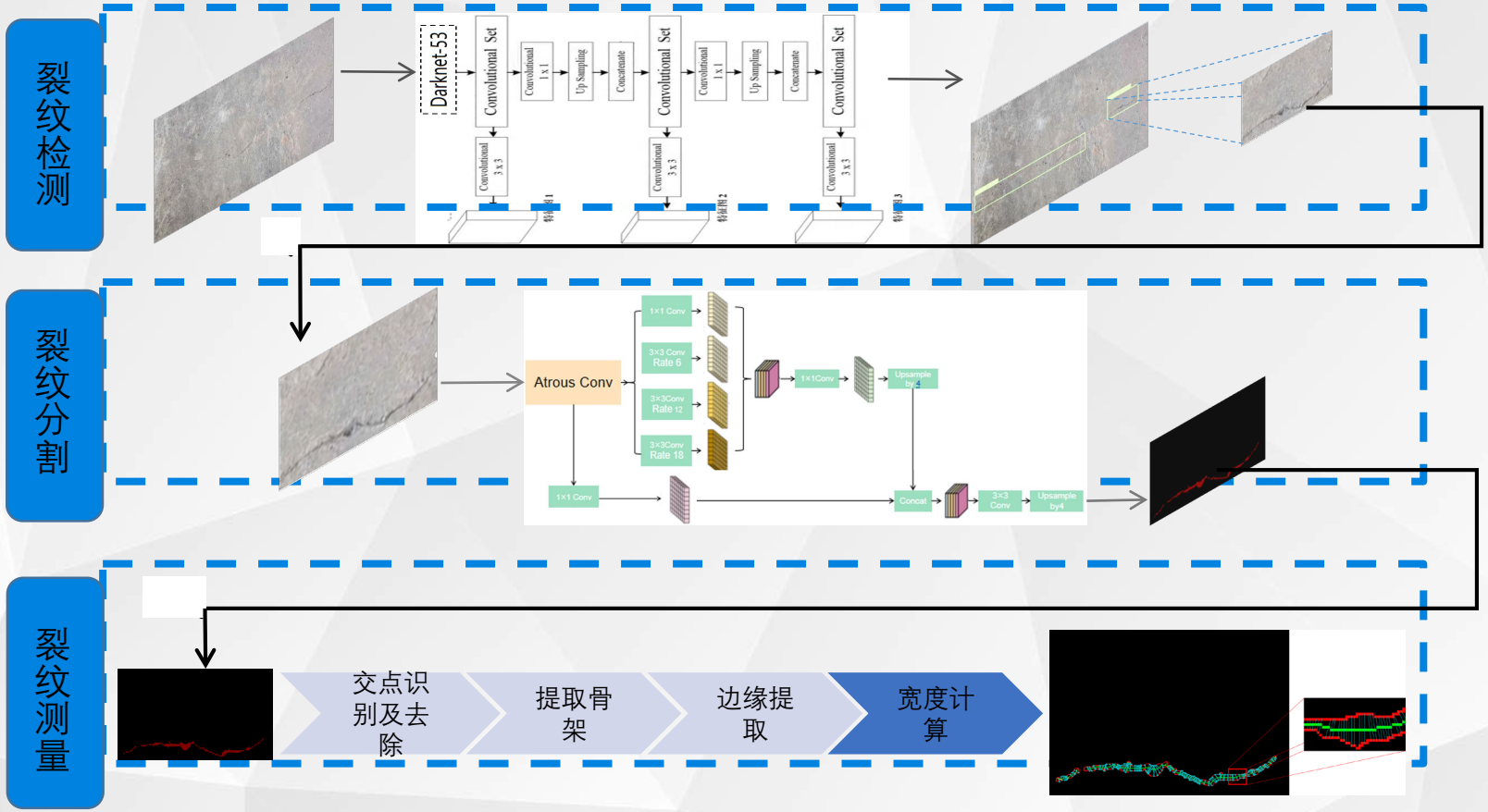
提出一种面向应用的裂纹自动评估量化系统，目的是统一裂缝识别、分割和宽度测量的过程，沿着并针对实际应用场景提高识别性能，我们提出的裂纹自动评估量化框架，包括三个步骤：

（1）基于YOLOv3的裂缝检测
（2）基于改进的deeplabv3+的像素级裂缝分割
（3）裂缝量化

算法量化期望的裂缝

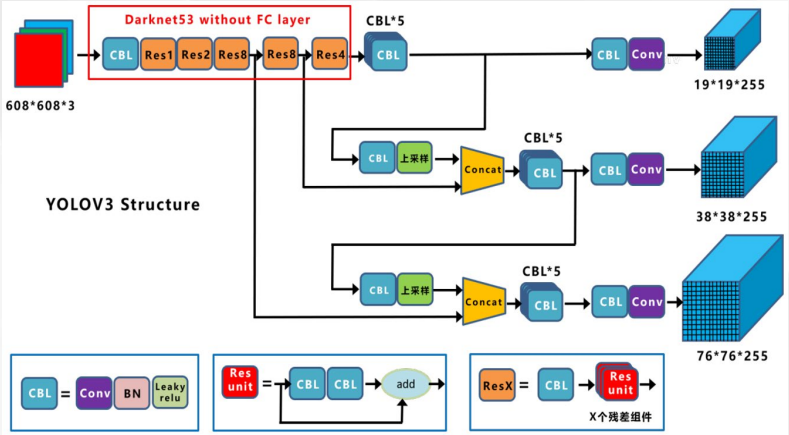
工程中遇到的裂缝

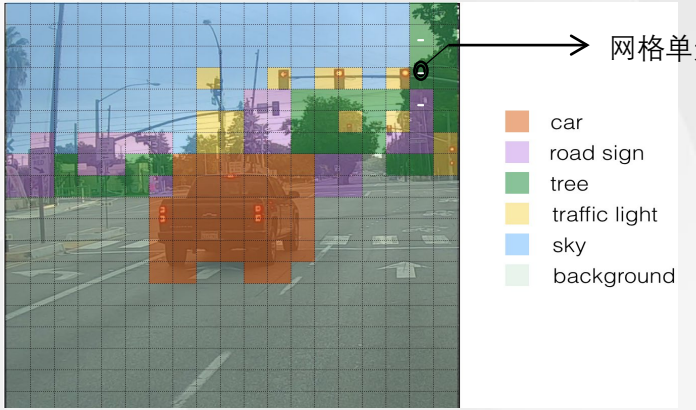
裂纹评估框架


裂纹检测
Darknet-53
Convolutional Set
Convolutional 1 x 1
Up Sampling
Concatenate
Convolutional 3 x 3
裂纹分割
Atrous Conv
1×1 Conv
3×3 Conv Rate 6
3×3Conv Rate 12
3×3Conv Rate 18
1×1Conv
Upsample by 4
Concat
3×3 Conv
裂纹测量
交点识别及去除
提取骨架
边缘提取
宽度计算

目标检测模块

YOLOV3模型框架：



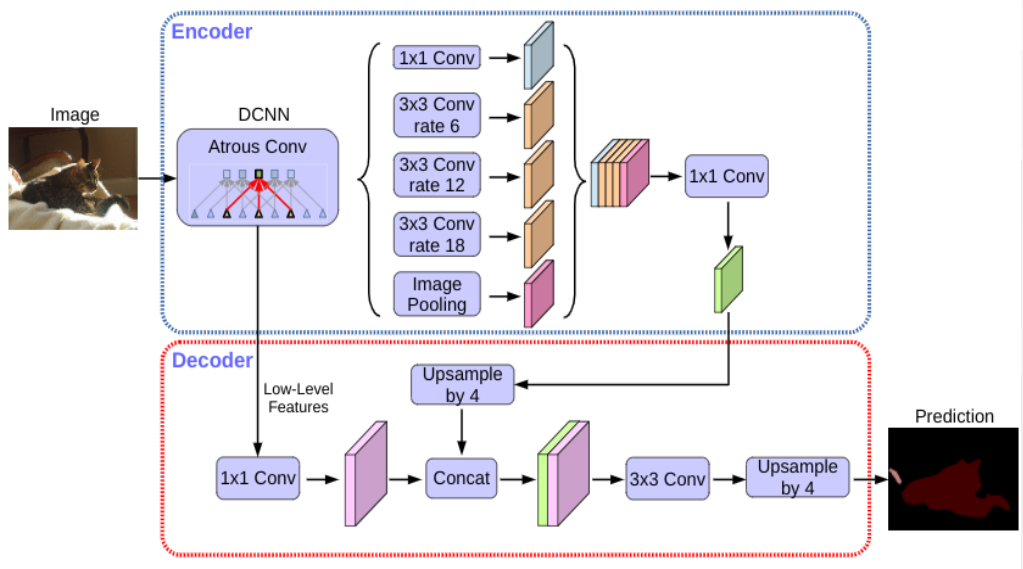


YOLO将输入图像划分为多个具有相同高度和宽度的网格，如果对象的中心属于网格单元，则允许检测对象。

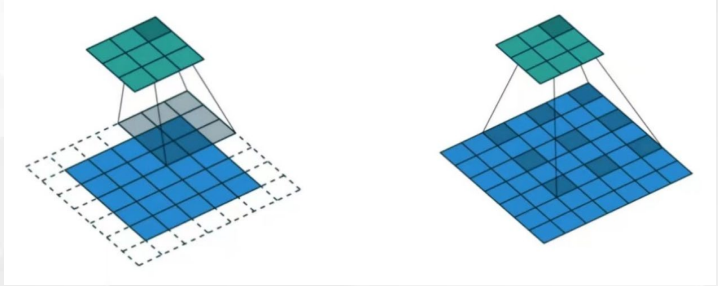
裂缝检测效果：

语义分割模块

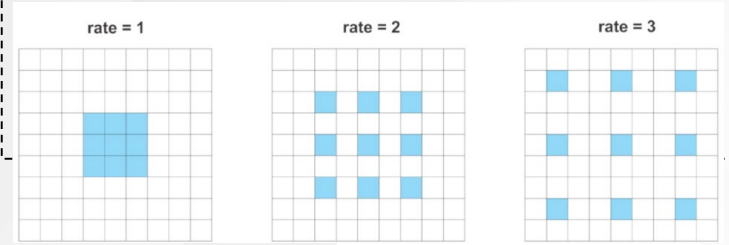
DeeplabV3+模型框架：



核心点：



传统卷积　　空洞卷积

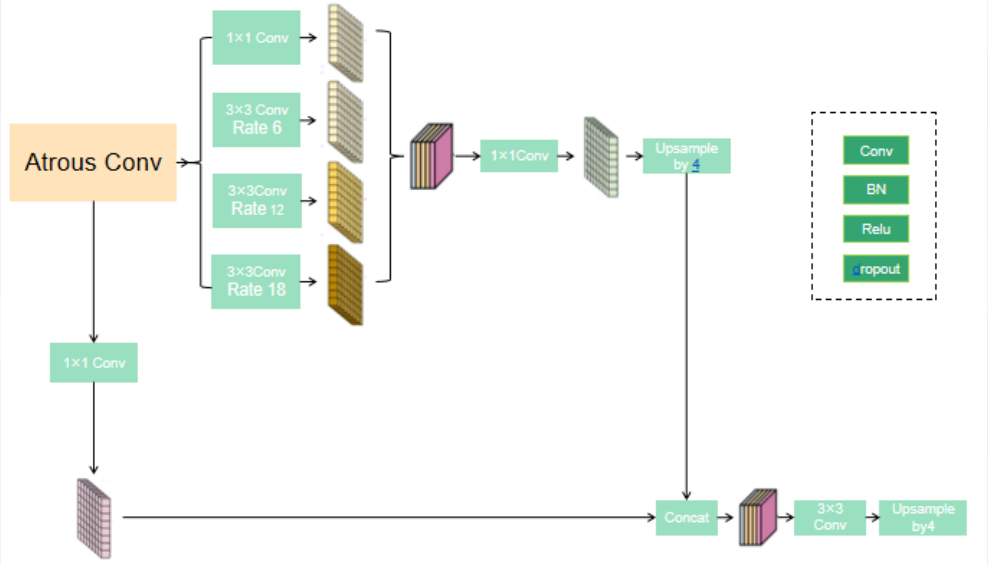


等效卷积核大小：$k' = k + (k-1) \times (d-1)$

语义分割模块--改进后

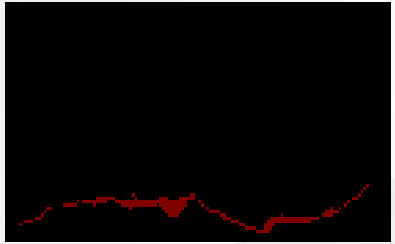
模型优化后效果

改进后的DeeplabV3+模型框架：



	miou	MPA	Precision	Recall
deeplab V3+	79.28	88.48	83.51	87.48
6e-4	81.15	88.56	88.02	88.09
plus dropout	82.04	90.02	87.98	90.78
remove pooling	83.45	90.98	89.52	91.13

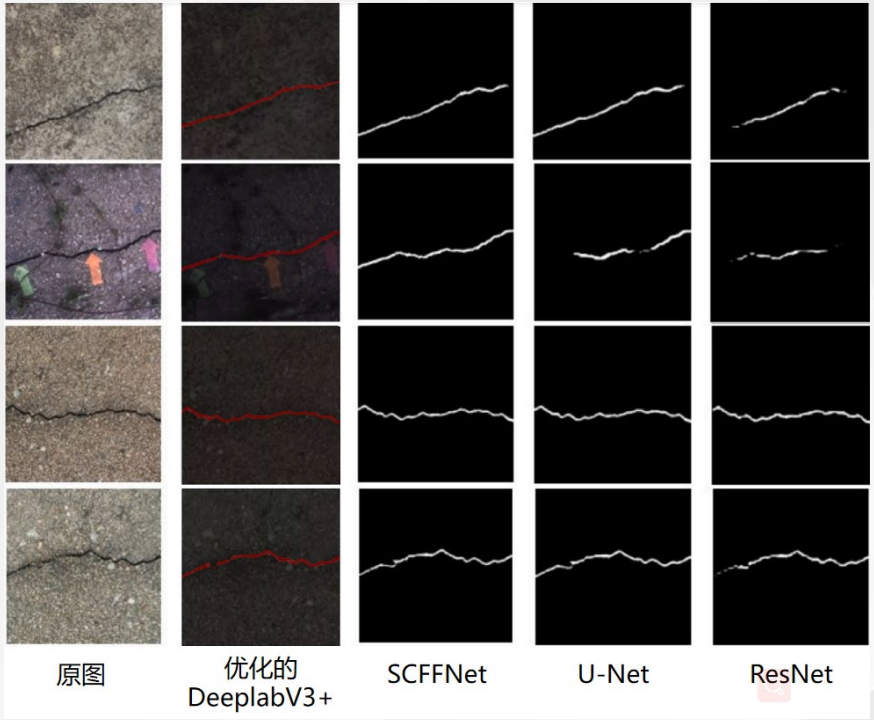
裂缝分割效果：



语义分割模块--对比

与常用裂纹分割算法的比较

	PA	MPA	MIoU	FWIoU	batch size	epoch
U-Net	0.97952	0.87162	0.80393	0.96288	4	100
ResNet	0.97792	0.84735	0.77841	0.96032	4	100
SCFFNet	0.98122	0.90612	0.82409	0.96669	4	100
优化的 deeplab V3+	**0.98925**	**0.9098**	**0.83450**	**0.96843**	**4**	**100**



分割效果的视觉比较

裂纹量化模块

01 02 03 04

交点识别及去除

Bresenham画圆算法

提取骨架

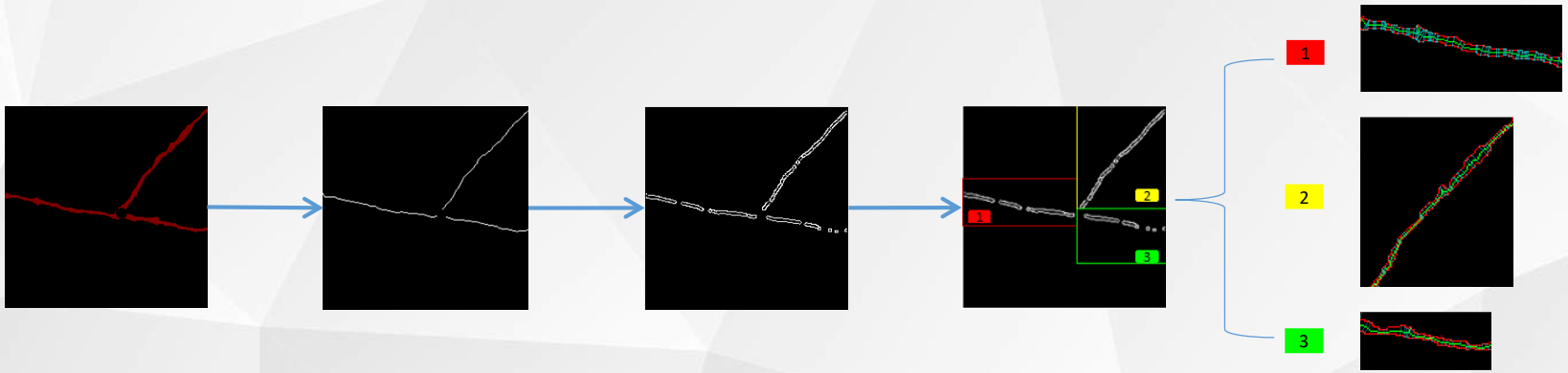
中轴变换

边缘识别

canny算子

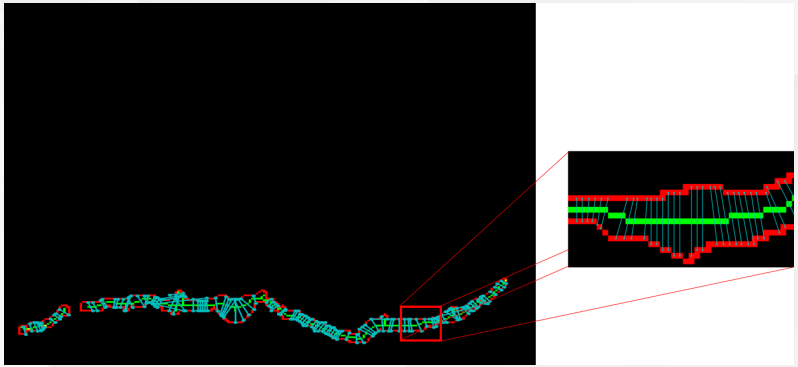
裂纹宽度

正交投影法



裂纹量化模块-- 实际宽度换算

裂缝量化效果:



实际宽度:

$$W_c = \frac{D_w}{P_c L_f} W_p$$

P_C 是所用相机传感器的尺寸
D_w 为工作距离（mm)
L_f 为相机焦距（mm)
W_p 为裂缝像素宽度

所提框架的优势

内容 / 框架	裂纹检测	裂纹分割	裂纹量化
Framework 1	√	×	√
Framework 2	×	√	√
Our framework	√	√	√

Concrete crack detection and quantification using deep learning and structuredlight. Construction and Building Materials.2020.

YOLOv3-tiny检测裂纹和激光区域

根据比例推算裂纹宽度

- 适用的裂缝宽度有限制
- 基于处理良好的数据的理想条件

Automatic pixel-level crack detection and evaluation of concrete structures using deep learning.Struct Control Health Monit.2022.

Crack-FPN分割裂纹

正交线法测量裂纹

- 局限于简单裂纹模式（无分支\均匀裂缝）
- 平均宽度无法准确量化开裂程度

YOLOv3检测裂纹

DeeplabV3+分割裂纹

量化模块测量裂纹

- √ 统一裂缝识别、分割和宽度测量的过程
- √ 更适合实际工程环境
- √ 更好的识别性能
- √ 裂纹量化结果更准确

THANK YOU