

映射。

在每个尺度上，编码器网络中的最后一个卷积层和解码器网络中的最后一个卷积层被连接起来，并进行叠加。

接着，叠加后的特征被输入到一个单通道的 1×1 的卷积层中，该卷积层可以将多通道的特征映射转化为单通道的特征映射。

从 1×1 的单通道卷积层输出的数据将会通过一个反卷积层对特征映射进行上采样。

反卷积后的特征映射将输入至一个 crop 层，它会将上采样的结果裁剪成输入的大小。

最后，通过一个 sigmoid&loss 层，可以计算出逐像素的预期交叉熵损失。

每个尺度的跨层融合完成以后，将对各层产生的特征映射进行叠加融合，随后将融合的结果再次输入一个单通道的 1×1 卷积层，该卷积层将输出一张预测图。

预测图通过交叉熵损失函数来判定每个像素属于裂缝的概率。

三、DeepCrack 的损失函数设计

给定一个数据集 $S = \{(X^n, Y^n), n = 1, \dots, N\}$ ，其中 N 是数据集中图像的数量。 $X^n = \{x_i^{(n)}, i = 1, \dots, I\}$ 表示原始的图像输入，而 $Y^n = \{y_i^{(n)}, i = 1, \dots, I, y_i^{(n)} \in \{0, 1\}\}$ 表示第 n 张图片的裂缝标签映射关系， I 则表示该图片的像素数量。

设网络中的卷积阶段数为 K ，那么可以得到第 k 层的特征映射 $F^{(k)} = \{f_i^{(k)}, i = 1, \dots, I\}$ 。在所有的卷积操作完成并融合后，可以得到整个融合过程的特征映射 $F^{fuse} = \{f_i^{fuse}, i = 1, \dots, I\}$ 。

裂缝检测只有两种分类，即该像素是裂缝和该像素不是裂缝。由于是裂缝的概率通常比不是裂缝的概率要小很多，即该像素为裂缝是一个小类。所以图像分割或者图像分类方法会给小类更大的权重，但是经过实验发现，在裂缝检测中，盲目的将更大的权重赋给裂缝标签这一小类会导致误判的概率升高。

因此，将损失函数定义为：

$$l(F_i; W) = \begin{cases} \log(1 - P(F_i; W)), & \text{if } y_i = 0, \\ \log(P(F_i; W)), & \text{otherwise,} \end{cases}$$

其中, F_i 是卷积层输出的特征映射的第*i*个像素, W 是网络层中的标准参数集, $P(F)$ 是标准的sigmoid函数, 能够将特征映射转化为裂缝概率的映射。

有了单个像素的损失函数, 就可以定义整个图片的损失函数:

$$\mathcal{L}(W) = \sum_{i=1}^I (\sum_{k=1}^K l(F_i^{(k)}; W) + l(F_i^{fuse}; W))$$

由于 DeepCrack 网络的编码器和解码器都有非常多的尺度, 可以提出疑问: 每个尺度的卷积操作对于最后输出的结果是否会有不同程度的影响?

实验结果证明, 若每次都移除一个跨层融合的连接 (任意尺度), 都会对降低最终结果的表现。其中, 实验结果表明, 第一个尺度的跨层融合对于最终结果的精度影响是最大的。因为第一个尺度与输入的原始图片有着相同的分辨率, 并且相较于其他尺度有最多的裂缝特征细节。

因此, 我们根据这种情况, 可以对损失函数进行重写, 得到:

$$\mathcal{L}(W) = \sum_{i=1}^I (\sum_{k=1}^K \alpha^{(k)} \cdot l(F_i^{(k)}; W) + l(F_i^{fuse}; W)).$$

其中, $\alpha^{(k)}$ 表示第*k*个尺度给定的权重, 其余变量的定义与修改之前的变量定义相同。

下表证明, 第一个尺度对于裂缝判断准确率的影响起决定性的作用, 而给其他尺度更大的权重 (大于1) 并不会显著影响结果的准确率。

$\{\alpha^{(1)}, \alpha^{(2)}, \alpha^{(3)}, \alpha^{(4)}, \alpha^{(5)}\}$	CRKWH100			CrackLS315			Stone331		
	ODS	OIS	AP	ODS	OIS	AP	ODS	OIS	AP
{1, 1/3, 1/9, 1/27, 1/81}	0.8892	0.9015	0.9209	0.8269	0.8510	0.8747	0.7737	0.8010	0.8062
{1, 1/2, 1/4, 1/8, 1/16}	0.9017	0.9110	0.9279	0.8327	0.8554	0.8770	0.8464	0.8667	0.8667
{1, 1, 1, 1, 1}	0.9095	0.9170	0.9315	0.8449	0.8671	0.8772	0.8559	0.8751	0.8883
{1, 2, 4, 8, 16}	0.9012	0.9081	0.9311	0.8328	0.8585	0.8758	0.8494	0.8726	0.8868

四、特殊情况

大多数情况下, 裂缝的颜色要比背景颜色更加暗淡, 很容易就会推断出只使用无明亮裂缝的图像数据集训练出来的网络无法识别光亮的裂缝。

实验结果表明, 将一个明亮裂缝的图片数据集交给使用无明亮裂缝的数据集训练出来的网络检测, 该网络无法对裂缝进行检测。

解决方案之一是将训练数据集的图像明亮度翻转, 将背景和裂缝的明亮度对调,

再将调整后的数据集作为训练样本交给网络学习。

训练完毕后，再将测试图片输入模型，实验表明，此时的神经网络已经可以对光亮的裂缝进行检测与识别。

五、相关内容研究

VGG16 网络：是一种用于定位和分类的卷积神经网络，其特点是通过增加卷积的深度来提高网络的性能表现；在进行工作的时候只有 3x3 的卷积操作和 2x2 的池化操作，简单直接；并且使用卷积操作代替全连接操作，可以输入任意尺寸的图片。DeepCrack 网络的解码器和编码器网络的结构设置正是参照该神经网络。

FCN 网络：Fully Convolutional Network（全卷积网络），是图像语义分割网络的开山之作。其特点是将普通的 CNN 最后的全连接层替换成卷积层，使得从原来的图像级别精细度变为像素级别精细度；再通过上采样（反卷积）对图像的尺寸进行恢复。后面几乎所有图像语义分割的网络架构均来源于 FCN。

SegNet 网络：来源于上段提到的 FCN。其改进点有两个方面：舍弃了反卷积操作，更换为 Unpooling（去池化）操作，由于 Unpooling 不需要学习和训练，其效率要高于 FCN；在解码器中，会存储一系列常用的参数和索引，使用这些参数和索引对特征映射进行 Unpooling 操作时，能够保证高频信息的完整性。

DeepCrack 网络：基于 SegNet 网络。添加了 skip-layers fusion（跨层融合），对图片细节的保留和利用相较于 SegNet 网络有着较大提升。

前面提及的神经网络均为图像语义分割网络，适用于各种图像处理场合（如自动驾驶、智能机器人等），而 DeepCrack 则是一种专门化的神经网络，仅用于裂缝检测。

在给定的数据集上，通过实验，证明了 DeepCrack 网络在裂缝检测与识别上有着

比普通语义分割网络更好的性能。