

2021 年 秋 季学期研究生课程考核

(读书报告、研究报告)

考 核 科 目: 机 电 系 统 智 能 控 制

学生所在院(系): 机 电 工 程 学 院

学 生 所 在 学 科: 机 器 人

学 生 姓 名: 佟 鑫

学 号: 2 1 S 0 0 8 0 4 8

学 生 类 别: 学 硕

考 核 结 果 阅 卷 人

神经网络在工业线缆缺陷检测中的应用

摘要:

本文主要记述我在完成工业线缆缺陷检测项目学习中的所学知识,长期以来工业线缆的检测集中在线缆出场部分,而本文提出的前提是基于外部环境下进行检测。检测蛀牙分为预处理部分、检测部分以及损伤评估部分,在预处理部分主要涉及了图像增强的相关方法以及如何利用 Gan 网络去运动模糊以及去模糊模块的阈值设定,检测部分的主要问题在于小缺陷检测,但本文仅对图像识别的方法进行简要概述,对 One-stage 和 Two-stage 目标检测器总结概述。

关键词:工业线缆,目标识别,去模糊,RCNN,YOLO。

1. 工业线缆检测的研究意义与

背景

近年来,我国智能机器人通讯、智能家居等行业迅速发展,电缆作为一种传输电(磁)能、信息和实现电磁能转换的重要线材产品,是这些行业发展的必需品。线缆的发展成为这些行业的铺垫,其使用频率逐渐增加,这需要严格把关线缆质量。由于线缆外表主要起绝缘、防水、防潮等作用,所以线缆外表检测尤为重要,同时线缆外表也是线缆质量的一个重要指标。它的缺陷主要由于生产设备、生产方式、安装方式、人为因素等原因产生,这些缺陷会造成产品质量得不到保障、施工人员人身安全受到影响、商业经济受到损失等危害。本文主要针对已经架设好的高空线缆进行缺陷检测,相比于布匹、轨道缺陷的检测,线缆表面的图像信息不易获取。

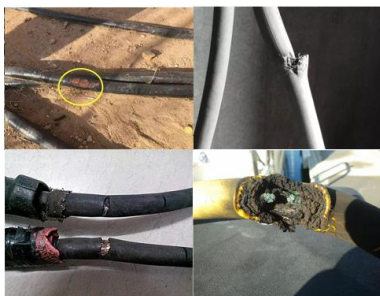


图1. 工业线缆缺陷图

目前主要的线缆检测方式仍然为

人工检测,耗时费力,且对于本文提出的高空线缆作业难度大,危险系数高,不宜采用。本文提出利用神经网络中图像识别的方法对采集的视频逐帧处理,较传统的机器视觉的方法相比由较高的识别精度以及识别速度。

我将主要从以下三个方面对神经网络在工业缺陷检测中的应用展开叙述。首先图像预处理方面,利用 DeblurGan 来对采集的图像进行去模糊;其次,利用 YOLO 对线缆表面的缺陷定位和分类;最后,对缺陷进行损伤等级评估。

2. 神经网络在图像处理的应用

神经网络的模型如下图所示

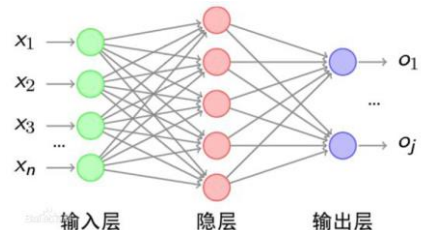


图2. 神经网络模型

输入数据由输入层输入,经过隐含层以及激活函数的处理由输出层输出,通过反向传播对网络参数进行调整,最后得到较为理想的权重值。

对于一般数据,通过调整隐含层的层数,匹配输入以及输出的特征维度,神经网络模型可以拟合出一个较为理想的效果,但对于图片处理而言(如图三所示),图片本身是一个多维度的矩阵形式,对于传统的神经网络形

式，需要对矩阵中各个维度的每一个像素值匹配一个权重值，这无疑是需要巨大的计算量并且是不显示的，这也阻碍了神经网络在图像处理中的使用。



图3. 图片像素值的表示

2012 年 AlexNet 在 ImageNet 竞赛上面大放异彩，以领先第二名百分之 10 的准确率夺得冠军，向世界展现了神经网络在图像处理方面的卓越能力。最早应用图像处理的神经网络方法由 LeNet 提出，卷积神经网络（Convolutional Neural Networks, CNN）是一类包含卷积计算且具有深度结构的前馈神经网络，是深度学习（deep learning）的代表算法之一，在图像处理方面得到了显著效果。

2.1 卷积层

卷积层的功能是对输入数据进行特征提取，其内部包含多个卷积核，组成卷积核的每个元素都对应一个权重系数和一个偏差量（bias vector），类似于一个前馈神经网络的神经元（neuron）。卷积层内每个神经元都与前一层中位置接近的区域的多个神经元相连，区域的大小取决于卷积核的大小，在文献中被称为“感受野（receptive field）”，其含义可类比视觉皮层细胞的感受野。而权值共享就是说，给一张输入图片，用一个卷积核去扫这张图，卷积核里面的数就叫权重，这张图每个位置是被同样的卷积核扫的，所以权重是一样的，也就是共享。这一理论的提出大大减少了将神经网络应用于图片处理领域中所需要训练的参数数量。感受野的尺寸一般选定为奇数，使运算变得简洁、高效，能够在超大规模数据集上运算，举例说明使一个 $m \times m$ 的卷积核在图像上扫描进行特征提取，如果 channels 为 k

的话，那么参数总量为 $m \times m \times k$ 。

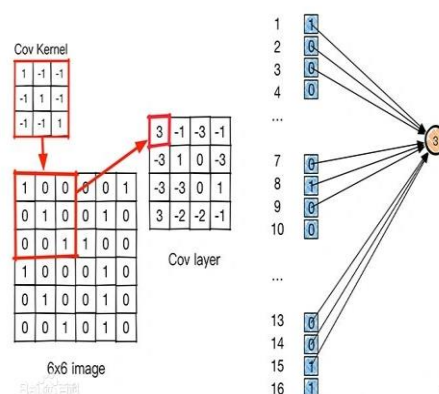


图4. 卷积神经网络的特征提取

通过反向传播计算误差，修正卷积核的参数，可以实现多种用途的图片处理，而在训练过程中卷积层的各项参数同样对训练结果起着不容忽视的作用，包括卷积核大小、步长和填充，卷积核的大小即为感受野的大小，对于大感受野而言，训练参数较多，不易使用，目前较为多的为 3×3 的卷积核，随着深度学习的进一步发展，还可以将 3×3 的卷积的进一步分解为 3×1 和 1×3 的卷积核，进一步将训练参数值从 9 减少到了 6。填充的作用即保证在隐含层对图片做一次特征提取之后，输入图片与输出图片在空间上具有相同的尺寸。

2.2 激活函数

激活函数在图像处理方向的作用是引入非线性，由于卷积操作对于图片而言是线性操作，通过引入激活函数，我的理解是，可以将低纬度的线性不可分的数据点映射到高纬度从而线性可分(即支持向量机的原理)。

传统的激活函数有 Sigmoid 以及 Tanh，但是这两个激活函数在图像处理方面没有取得较好的效果，Sigmoid 函数在神经网络层数过多时会造成梯度消失，由于神经网络在训练过程中采用的梯度下降策略，每更新一次卷积核参数需要计算一次梯度，梯度消失会导致训练终止，为了解决这一问题，AlexNet 首次使用了 ReLU 激活函数(图 5)，又称为线性整流函数。

$$f(x) = \max(0, x),$$

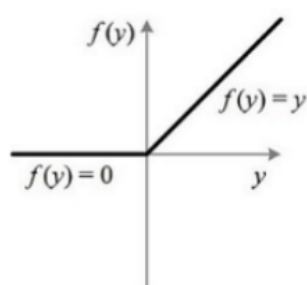


图5. ReLU 激活函数

它可以极大程度的加速随机梯度下降的收敛速度，并且计算简单，同时引入非线性，但其是一个不可逆的过程，因此在实际使用的过程中需要注意。

在之后又引入 Leaky ReLU 激活函数，已一个小参数乘以 y 小于 0 的部分，这样一定程度上弥补了 ReLU 激活函数的缺陷，之后又针对不同的特定的问题提出了许多激活函数如 Mash，这里不再一一介绍。

3. 优化策略

其实神经网络在图像处理方面仍然有许多细节值得讨论，这里再针对神经网络的优化函数分享一下我的学习心得。

Pytorch 提供了多种优化器，首先是 SGD 采用了随机梯度下降策略，在神经网络的课程部分，我们已经学习了这种优化策略，不多赘述。

其次是 MBGD，小批量梯度下降策略，与 SGD 的区别在于，随机梯度下降每次更新只考虑前一步对这一步的影响，而小批量梯度下降利用加权平均的方法在更新步骤时，通过计算前面多步骤的平均值来替代一次的梯度值。

$$\frac{\Delta J(\theta_0, \theta_1)}{\Delta \theta_j} = \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

就其本质而言，梯度下降的过程中有两个参数值可以进行优化，第一个是步长，而第二个是梯度值，MBGD 就是在计算梯度的过程中综合考虑全面众多步。

Momentum 方法对梯度 MBGD 进行了改进，采取制数加权平均的方式综合考虑前面多步梯度的影响。

$$v_t = \beta * v_{t-1} + (1 - \beta) \theta_t$$

该方法的优势在于每步仅需要带入迭代公式计算一次即可求得指数加权的梯度值，在计算量上远远小于 MBGD。

同样还有针对每一步迭代的补偿做出修正的 NAG 算法以及将指数加权平均应用于步长修正策略中的 Adagrad 策略，最后 Pytorch 还提出了一种集大成的优化函数 Adams，该函数较为复杂，优化效果同样较好。

4. 工业线缆缺陷检测图像的预处理

在开始训练以及执行识别任务之前，良好的数据预处理和参数初始画能够达到事半功倍的效果。除常规的预处理方式之外，我还针对高空架设的线缆在识别任务之前拟采取视频去模糊的操作。

4.1 数据预处理

数据预处理是机器学习中提出的概念，在图像识别领域同样使用，首先分析如图所示的图像。

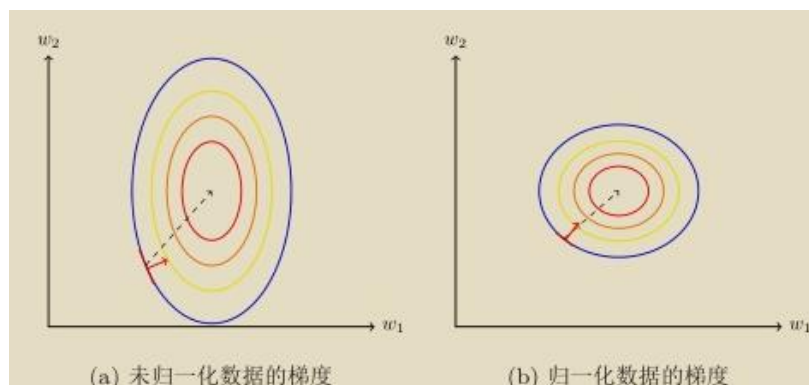


图6. 数据归一化对梯度的影响

简单的数据方式是对数据点进行减均值除方差，这样处理之后能够将数据点集中在(b)图的分布范围之内，这样处理之后可以壁面在训练过程中的波动，增加训练过程的稳定性，若不进行数据归一化，严重还会导致训练过程不收敛。

4.2 运动模糊去除

图像模糊还原属于图像复原技术的一种，图像模糊有许多情况，常见的有运动模糊、高斯模糊、散焦模糊等，本文只考虑在图片采集过程中高空设备不稳定所造成的全局运动模糊展开去模糊处理。

与传统的卷积神经网络不同，在这一部分我初步拟采取基于GAN网络(对抗生成网络)提出的去模糊网络DebluredGAN展开叙述。

生成式对抗网络是一种深度学习模型，是近年来复杂分布上无监督学习最具前景的方法之一。模型通过框架中(至少)两个模块：生成模型和判别模型)的互相博弈学习产生相当好的输出。原始 GAN 理论中，并不要求 G 和 D 都是神经网络，只需要是能拟合相应生成和判别的函数即可。但实用中一般均使用深度神经网络作为 G 和 D。一个优秀的 GAN 应用需要有良好的训练方法，否则可能由于神经网络模型的自由性而导致输出不理想。GAN 最常使用的地方就是图像生成，根据输入图像自动生成虚拟图像。

其训练模式也区别与传统的前馈

神经网络，是机器学习中的对抗式训练，其目的是使生成器 G 与判别器之间达到一种理想的平衡状态。

如图六所示，是将 GAN 网络应用到图像去运动模糊，简称 DeblurGAN，其基本原理为:其中清晰图像和模糊图像是训练数据集中的一对数据，当模糊图像送入生成器中，生成器会重建一幅生成的图像，每一次训练会计算数据集中清晰图像和生成器生成的图像之间的对抗损失和内容损失,对抗损失用来约束生成的图像清晰度高、真实性好;内容损失用于制约生成图像和数据集中的清晰图像在内容上一致。判别器输出重建的清晰图像为真实图像的概率值，根据 Wassertein 距离来决定继续训练生成器或判别器，Wassertein 距离是度量两个概率分布，之间的距离

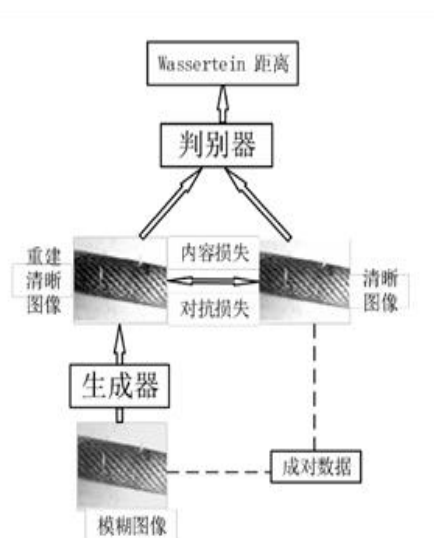


图7. 图像去模糊算法流程图

直观上来说就是两个分布在最优路径下的最小消耗，这个数值越小就代表网络的泛化性能更好，由生成器输出的图像质量就更加真实，训练完成后，生成网络 G 就可以生成逼真的图片。在实际的处理过程中，DebluredGAN 基于全局的内容损失以

及对抗损失来进行训练优化，因此对于局部点的特征可能不过于关注甚至可能将局部小缺陷当作模糊处理掉，而作为缺陷检测的预处理网络，这无疑是不容许的，因此如何在去模糊的过程中增强局部缺陷的特征是我接下来的研究关键。



图8. 边缘提取后图像的比较

实际应用过程中，并不是所有的传回的视频均会产生运动模糊，如何设定一个合理的阈值来选择是否对输入的视频进行去模糊处理也是至关重要的。由于进行检测的场景较为固定，这也对阈值的设定提供了可能性。

模糊图像会导致图像边缘信息的损失，因此利用 Opencv 提供的边缘检测函数提出边缘之后，清晰图片与模糊图片会有明显的区别(如图 8 所示)

针对这一特性，提出假设的解决方案：采集一系列图片后进行边缘特征处理，利用 k-means 聚类算法自动将采集图片分类，分析比较分类后两类图片之间的边缘特征值的均值，选取合适的值设定为分类均值。

5. 工业线缆缺陷检测

图像识别的任务可以划分为三类：分类、检测和分割。分类要解决的问题是给定一张图片或一段视频判断里面包含什么类别的目标。对于一幅图像来说，分类就是聚类，即分割；对于一组图像样本来说，分类是检测出样本中有相同目标的图像。检测解决“是什么？在哪里？”的问题，即定位出这个目标的位置并且知道目

标物是什么。分割分为实例的分割（Instance-level）和场景分割（Scene-level），解决“每一个像素属于哪个目标物或场景”的问题。

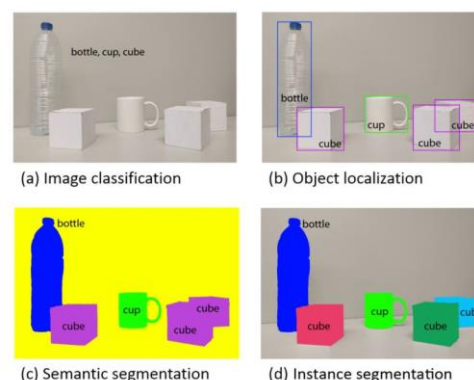


图9. 图像识别的三类任务

缺陷检测的主要任务就是检测缺陷是哪种缺陷以及在哪里的问题。

主流的目标检测网络分为两种，一种是以 Faster-RCNN 为代表的 Two-stage 目标检测器；另一种是以 YOLO 为代表的 One-stage 目标检测器，我将分别简要的概述。

5.1 Two-stage 目标检测器

两阶段的目标检测器，也称为基于候选区域的目标检测，顾名思义即

是在训练过程中神经网络划分为两部分训练，其中一个部分训练候选框的生成，而另一部分对训练对于候选框的回归以及分类任务。

两阶段的目标检测网络的发展可以由下图所概述。

RCNN->SPP-net->Fast-RCNN->Faster-RCNN->Mask R-CNN:

图10. Two-stage 目标检测器的发展

以 Faster-RCNN 对网络结构简要介绍。主要从以下三个关键点进行讲解

- (1) 如何生成候选框？
- (2) 如何针对候选框生成对应的特征图？
- (3) 候选框回归如何实现？

首先，与传统的 Two-stage 目标检测器不同，Faster-RCNN 网络首次提出了利用 RPN 网络来替代 selective search (selective searches 是指在图片上利用滑窗的思想依次选取若干的候选框进行回归)算法，用 RPN 先生成一堆 Anchor box，对其进行裁剪过滤后通过 softmax 判断 anchors 属于前景 (foreground) 或者后景 (background)，即

是物体 or 不是物体，所以这是一个二分类；同时，另一分支 bounding box regression 修正 anchor box，形成较精确的 proposal（注：这里的较精确是相对于后面全连接层的再一次 box regression 而言）

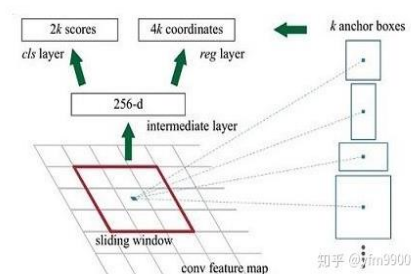


图11. RPN（区域生成网络）运行机制

其实 RPN 最终就是在原图尺度上，设置了密密麻麻的候选 Anchor。然后用 cnn 去判断哪些 Anchor 是里面有目标的 positive anchor，哪些是没目标的 negative anchor。所以，仅仅是个二分类而已。对于 positive anchor 做 bounding box regression 使其接近 Ground truth，至此即完成第一阶段定位。

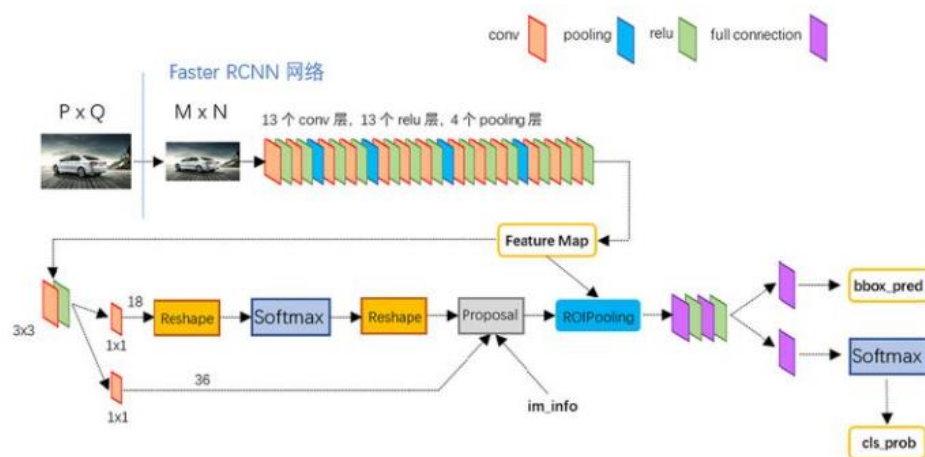


图12. Faster-RCNN 的网络结构

初代的 RCNN 网络对于已经得到的候选框是将其 Reshape 成固定尺寸输入到特征提取网络后得到特征图，Faster-RCNN 利用 RoI Pooling 模式，首先将整张图片送入到特征提取网络，随后将候选框映射到特征图上，得到候选框的特征图进行分类任务，完成第二阶段的分类。

图 12 中上半部分为特征提取网络生成 Feature Map，随后利用 RPN 生成候选框结合 RoI Pooling 输出候选区域的特征图进行分类任务。

在这个过程中还有一些与 One-stage 目标检测器通用的技术，如 NMS（非极大值抑制）、IOU 计算损失函数放在下一章节讲述

限于篇幅不对 Two-stage 目标检测器其他细节进行展开叙述。

5.2 One-stage 目标检测器

前面我们已经介绍了 Faster-RCNN，其对于候选区域的选择是利用 RPN 对每个点生成锚点框，对锚点框进行训练回归分类，从而产生优秀的 RPN 网络可以对输入图像产生

一个较为准确的候选区域，这个过程中定位任务与分类任务分别执行。One-stage 目标检测器省略了对于 RPN 定位网络的训练过程，直接对输入图片进行定位以及分类，一次完成。这样和 two-stage 相比，可以预见到检测速度会快很多。

在此同样选择一个代表网络 YOLOv3 就其特点讲解。

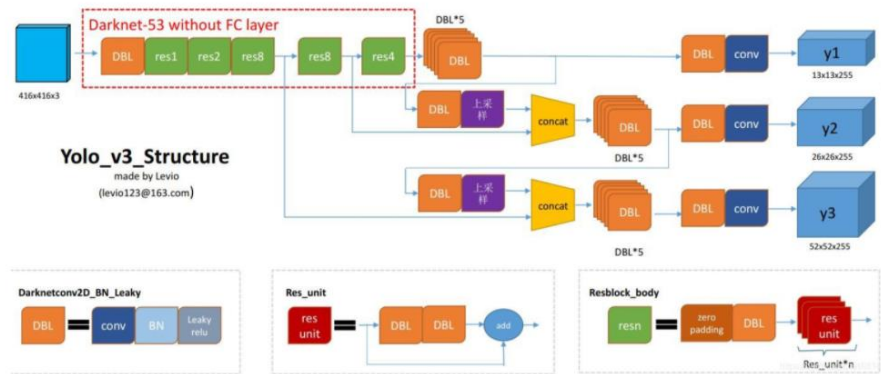


图13. yolov3 网络结构

首先 yolov3 的主体网络结构为 Darknetnet-53 去除线性层作为特征提取网络，乍一看神经网络的层数由 53 层，自然而然就想到了梯度消失的问题，这里就不得不提到一大经典网络 Resnet 残差网络。

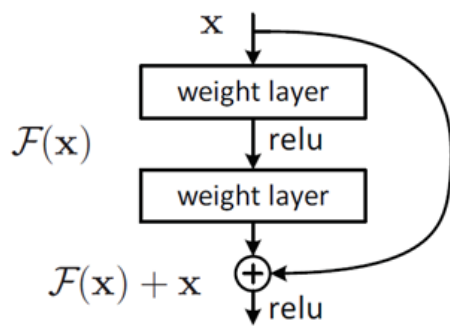


图14. 残差网络单元

提出了残差学习来解决退化问题。对于一个堆积层结构（几层堆积而成）当输入为 x 时其学习到的特征记为 $H(x)$ ，现在我们希望其可以学习到残差 $F(x)=H(x)-x$ ，这样其实原始的学习特征是 $F(x)+x$ 。之所以这样是因为残差学习相比原始特征直接学习更容易。当残差为 0 时，此时堆积层仅仅做了恒等映射，至少网络性能不会

下降，实际上残差不会为 0，这也会使得堆积层在输入特征基础上学习到新的特征，从而拥有更好的性能。残差学习的结构如图 14 所示。

而 Resnet 就是由许多个这样的残差块堆积而成的网络，这大大加深了网络层数在 resnet 之后，多种网络残差结构层出不穷，前不久大火的黄高博士设计的 DenseNet 也是依托于此思想，每一层都与下一层跳层连结。

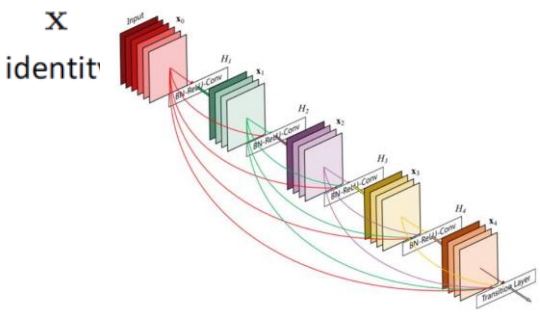


图15. DenseNet 网络结构

YOLOv3 并没有简单的利用 Darknetnet-53 进行特征提取，还利用的 FPN（特征金字塔）来对深层的特征和浅层的特征进行特征融合输出三个尺度下的特征图，除此之外 FPN 还起到了“分而治之”的思想。

至此还没有还没有看出 One-stage

目标检测器如何实现速度上的提升，其实速度提升的关键在于此处，YOLOv3 会针对训练数据集利用 K-mean 聚类算法获取训练数据集中的锚点框的尺寸，在训练过程中，将图片划分为 $n \times n$ 的 grid，每个 Grid 预测一定数量的锚框，仅认定目标中心点所在的 grid 为 positive，对每一个锚点框进行二分类，对于 positive 的框进行回归预测，在网络结构中全连接层输出特征向量各维度对应内容如下：

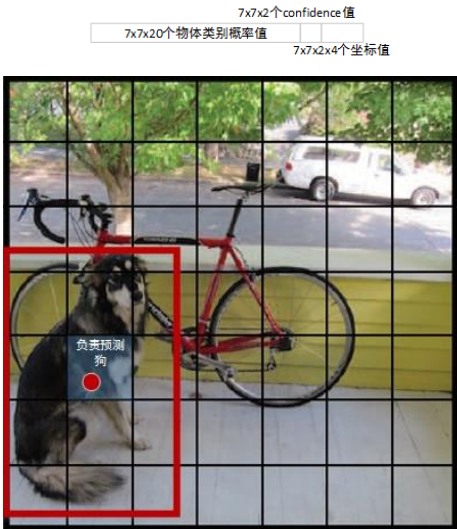


图16. 网格划分规则

在训练过程中利用 NMS 来对同意目标的多个预测框进行处理，仅保留与真实框 IOU（并交比）最大的预测框。

5.3 小目标检测

拟定的拍摄距离是间距为 50 厘米，且缺陷相对于整个图片的篇幅而言站相对较小的一部分，因此将工业线缆的缺陷检测部分归结于小目标检测，为了增加所选用的网络在小目标检测上的普遍适用性，我拟采取一些方法。

首先由于在缺陷采集过程中，同一篇幅中有较大的空白区域，拟利用图片处理的方法将缺陷复制粘贴于各部分。如图是我制作的由电流击穿的线缆图片，通过抠图的技术将缺陷复制粘贴可以增加同一张图片的训练价值。



图17. 模拟的线缆表面缺陷图

其次，由于图像识别的神经网络训练的主要方式采取的是梯度下降法，因此损失函数的选择对于训练过程是尤为重要的，Pytorch 为损失函数的选择提供了很多可选，交叉熵损失函数一直是分类任务重的首选，器表达形式为

$$L = \frac{1}{N} \sum_i L_i = \frac{1}{N} \sum_i -[y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i)]$$

如上图可见，该损失函数在设计的过程中并未考虑缺陷尺寸对损失函数的影响，可将预测框的尺寸面积作为比例因子引入该方程，增大网络对于尺寸的关注度将会有效提高该神经网络对于缺陷检测的训练速度以及泛化效果。

综上完成了对于目标检测的主要方法的介绍，拟采取 YOLO 系列用于工业线缆目标检测，在实际的使用过程中将结合最新的技术，近些年对于图像识别无 anchor 的方法也日益增多，并且前面提到 FPN（特征金字塔）有分而治之的作用，将这一思想应用于 YOLOv3 中可以进一步提升其训练效果，在完成了自己的数据集时我也会努力设计出自己的网络使之更加适配我所要完成的工业线缆检测的任务。但未完成自己的工业线缆缺陷数据集的制作，因此在这里不做展示。

六. 总结

通过本次课程报告完成对过往学习知识的总结，由于仍然在理论研究部分，围在线缆数据集上对网络进行

验证，是本文的不足，在后续的过程中，会补充完成实验部分，并且还希望可以做图像拼接方向的相关研究，通过特征点匹配完成多张图片拼接成全景图片的相关技术。

七. 参考文献

- [1]. Bharath A, Petrou M. Next Generation Artificial Vision Systems: Reverse Engineering the Human Visual System [M]. London , UK: Artech House Publishers, 2008
- [2]. o papermaking for printing [J]. Pattern Recognition and ImageAnalysis, 2011, 21 (3) : 486-490 . [DOI: 10 . 1134 /S1054661811020441].
- [3]. rjenaki O O, Moghaddam P A, Motlagh A M. Online tomatosorting based on shape, maturity, size, and surface defects using machine vision [J]. Turkish Journal of Agriculture and Forestry, 2013, 37(1) : 62-68. [DOI: 10. 3906 /tar-1201-10]
- [4]. Abdi M and Nahavandi S. 2017. Multi-residual networks: improving the speed and accuracy of residual networks [EB/ OL]. [2020-05-22]. <https://arxiv.org/pdf/1609.05672.pdf>
- [5]. Gao S H, Chen M M, Zhao K, Zhang X Y, Yang M H and Torr P.2021. Res2Net: a new multi-scale backbone architecture. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(2):652-662 [DOI: 10. 1109 / TPAMI. 2019. 2938758]
- [6]. He K M, Zhang X Y, Ren S Q and Sun J. 2015. Delving deep into rectifiers: Surpassing human level performance on imagenet classification // Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE: 1026-1034 [DOI:10. 1109 / ICCV. 2015. 123]