Abstract

一句话点题+介绍网络结构

自动缺陷检测技术得到了极大地提升，随着深度学习的逐步渗透。然而很多领域中缺陷检测任务依然有挑战性，由于其产品的多样性。在这篇文章中，我们关注于轮胎工业中的缺陷检测并提出了一个...网络，（which）它包含两个并行的子网络（network或branch）用来感知（挖掘）多尺度的缺陷特征。具体的（Specifically），蕴藏在深网络层中的（需要深层网络获取的）缺陷形状和位置信息被挖掘通过结构感知网络，（which）以传统的（已存在的）全卷积网络为框架。与此同时，为了弥补深层网络造成的细节丢失，一个新奇的纹理感知网络被用来（is developed for）尽可能的挖掘cover小尺寸的缺陷。之后，通过融合含有结构和细节信息的特征得到最后的密集预测结果。实验结果表明，我们的网络可以得到令人满意的效果，特别是小缺陷的检测。

Though automatic detection method has been tremendous improved, with the gradual penetration of deep learning. Defect detection in many industrial processes is one of the remaining challenging tasks due to the diversity of its products. In this work, we focus on detection tasks in the tire industry and develop a *Multi-scale Defect Detection Network (MDDN)*, which contains two parallel sub-networks to capture multi-scale defect features. Specifically, high-abstracted semantic features containing defect shapes and locations are mined via a *Semantic-aware Sub-network*, simplified by an off-the-shelf fully convolutional network. Furthermore, to complement the details filtered by the deep network, a novel *Texture-aware Sub-network* is used to exploit the small size of the cover edge features and small defects as much as possible. Finally, the pixel-wised detection results are obtained by fusing features with semantic and texture information. Extensive experiments demonstrate that *MDDN* can produce comparable results and achieve significantly performance improvement in small defects detection.

不像之前的网络，它可以同时定位并分割缺陷。在...数据集上的实验结果表明...

Introduction

自动检测技术，被用来提高质量和加速生产，已经成为不可或缺的一部分在许多工业领域，例如\*\*\*，\*\*\*。特别是在轮胎产品中，缺陷检测任务一直受到广泛关注并且已经有很多研究被提出[\*\*\*]（在过去的几十年中）。most real-world 检测通常首先从获得的X-ray 图像中挖掘出缺陷区域，which（由\*\*\*造成的），然后有缺陷的产品根据缺陷的位置、形状等信息被分等级处理。由于已存在方法的局限性和图像的特殊性[\*，\*]，大多依然采用人工的形式，造成了\*\*（人工的缺点）。因此轮胎缺陷检测依然是最有挑战的工业检测任务之一。

Automatic defect detection, used to improve quality and accelerate production, has become an indispensable part in industrial processes, such as fabrics[1,2,3], steel[4], semiconductors[5], and solar wafers[6]. Especially in tire manufacturing, numerous detection algorithms have been proposed [7,8,9,10,11] and aroused extensive attention over the past two decades. In most real-world applications, tire defect detection is first carried out by deriving the defective region from tire X-ray images, which contains various types of defects caused by unclean raw materials and undesired manufacturing facilities[12]. Then, the defective product is hierarchical processed according to the location and size of defects. Due to unique properties of the tire image, for instance complexity and low-quality, illustrated in previous study[13, 14], most inspection processes are performed by human observers, which increases the risk and reduces the efficiency. Therefore, tire defect detection remains one of the most challenging inspection tasks.

目前存在基于计算机视觉的方法大多致力于通过区分缺陷区域与正常区域检测缺陷，一个关键的问题是。。。a key issue for such methods is texture feature extraction. 从特征表示的角度，这些方法可以被大体分为三类：固定特征，固定变换核学习系数，学习特征。\*\*\*\*；\*\*\*\*；\*\*\*\*；然而，大量实验结果表明，这种方法不够精确，对边缘细节和小缺陷不敏感。这个问题同样出现在使用深度网络分割自然图像的任务中。（这与深度网络分割自然图像时所遇到的问题一致。）

At present, existing computer vision based detection methods are mostly devoted to distinguish difference between defective regions and background (defective-free regions). Hence a key issue for such methods is feature extraction. Guo et al. [15] exploited a local kernel regression descriptor to derive feature vectors. By comparing the dissimilarity of the corresponding feature between one pixel and its neighbors, anomaly pixels can be located and segmented, even in the tread image. Nevertheless, this method is not suitable for real-time detection tasks because of the high computational complexity. A component decomposition based method was proposed in [15] , which separated the background from the image by means of two designed filters. Then through an adaptive thresholding processing, defects were derived from the residual image. Besides, Independent component analysis(ICA) was also used for defect detection tasks[8,9]. A major disadvantage of these fundamental methods is the limitation of the information contained in low-level clues and domain features. ~~A major disadvantage of these fundamental~~ ~~methods that rely on low-level clues is the limitation of representation capabilities.~~ To address the limitation, Zhang et al. [] and Zhang et al. [][] introduced radon transform and mulit-scale transform, for instance curvelet and wavelet transform, in detection tasks respectively. Furthermore, optimized edge detection and total variation algorithm are used to achieve more accurate results[]. Zhao et al.[] proposed a multiple kernel learning method, which combined various transform kernels to get more differentiated information. However, the representation capability of fixed kernels is not comprehensive enough. In addition, transform process is computationally expensive. Recently, Cui et al. [] attempted to classify tire defects by means of convolutional neural networks(CNN), which has outstanding performance in the recognition and segmentation tasks of natural images. With the excellent feature extraction capability of deep network, Wang et al. [] further implemented the detection and segmentation in tire images by a fully convolutional network(FCN). However, FCN is not sensitive to small defects and edge details, which is similar to that in dealing with natural image tasks.

为了解决这个问题，已经有很多多尺度的方法被提出在公共数据集中。（成功的开始介绍之前想过的多尺度方法）。这些方法虽然在一定程度上缓解了小目标的检测，但是两个关键的问题：1，没有考虑过比例的问题。2，细节纹理的提取只在浅层，而浅层过多会导致参数爆炸，过少又提取不到关键特征。

To overcome the shortcomings of insensitivity to small defects and detailed textures, many methods have been proposed in benchmark datasets. Most of them are based on multi-scale strategies and can be roughly classifified into image pyramids and in-network feature hierarchies. Image pyramids illustrated in Figure 1a[] were extensively used in the era of hand-crafted features[]. Even if the crafted features have largely been replaced by self-learning features, multi-scale testings on the image pyramid are still used to verify the adaptability and robustness. Nevertheless, image pyramid based methods is impractical for real applications due to the considerable increase in inference time. In-network feature hierarchies are formed by the forward propagation within deep convolutional networks(ConvNets). Due to a series of sub-sampling layers, in-network hierarchies produce feature maps of different spatial resolutions, with an multi-scale and pyramid shape[]. By fusing these multi-scale feature maps, the texture feature in shallow layers and the semantic information contained in deep layers can be perceived. The Single Shot Detector (SSD) [] is one of the first attempts at combining predictions from these features maps to detect objects of various sizes. Generally, shallow features are used to predict small objects, and deep features with large receptive fields are used detect large objects. However, the lack of semantic information is harmful to the detection of small targets in shallow layers. Another fusing way can effectively address this problem by concatenating multi-scale features and detecting on top of the expanded feature maps, as shown by the red line in Figure 1b[]. For example, FCN-8s and FNC-16s defined a skip architecture to produce more accurate segmentation. Similar top-down skip architectures are popular in recent research[]. There exists a basic problem that it is still not enough to mine the detail texture in these structures[]. Moreover, small targets will become smaller or even disappear as the increases of sub-sampling layers, even if they can be captured in shallow layers.

Inspired by MTGAN, we construct a end-to-end network named Multi-scale Defect Detection Network (MDDN) consisting of a semantic-aware sub-network and a texture-aware sub-network. Based on tire X-ray images, a image patch strategy is adopted in the texture-aware sub-network. Unlike natural image patches, defects (objects) are still significant and discernible in the tire image patches, as shown in Figure 1(c). On the one hand, The proportion of the area of defective regions in images increases, which is advantageous for better capturing of detailed information. As shown in Figure 1(b), for a 256\*256 tire image, the proportion of defects is increased from \*\*\* to \*\*\*. On the other hand, image patches as input data can be used without reducing the number of parameters in the case of reducing the pooling layer. On the other hand, sub-sampling layers can be discarded to retain more shallow features without increasing the parameters.

Details can be better captured by the network.

texture-aware sub-networks use image blocking strategies to solve the above two problems.

We construct a end-to-end network named Multi-scale Defect Detection Network (MDDN) to solve these two issues based on tire X-ray images.

我们提出了\*\*\*\*\*\*\*\*去解决这两个问题，

Introduction v4

第一段：1. 检测很重要，2.在很多领域都有研究，3. 轮胎也有很多研究，4.

Introduction v3

1. 一句话介绍，自动检测的重要性，\*\*\*，\*\*\*，\*\*\*领域均有研究，其中轮胎检测一直广泛受到关注，尽管\*\*\*\*（把一些不好融合到related work中的轮胎检测一并说了）都做出了大量的贡献（就大概这个意思吧），但依然存在很大的挑战。

在实际的工业应用中的轮胎检测环节，根据X光的射线图像（如图所示）定位并检测缺陷（类型、形状、位置、面积等用来解释挑战），之后根据缺陷的位置和面积决定对轮胎进行对应的处理。多数工厂普遍采用人工检测，效率低，不安全，浪费资源......

1. 为了解决这个问题（任务或现状），近几年有一些研究被提出和使用，可以分为\*\*\*和\*\*\*和\*\*\*。分别简单介绍一下。（包含机器学习的方法，只介绍机器学习在轮胎上的应用即可，其他的不用说）。\*\*\*证明了深度学习（分割方法）在轮胎检测的可行性，但是结果表明这种pixel-to-pixel的网络对细节信息和小目标检测效果差，这同样也是自然图像上遇到的问题（和自然图像中的表现一致）（这里重新组织一下，重点是要把话题转移到自然图像、多尺度上）。
2. 为了解决这个问题，已经有很多多尺度的方法被提出在公共数据集中。（成功的开始介绍之前想过的多尺度方法）。这些方法虽然在一定程度上缓解了小目标的检测，但是忽略了一个本质的原因...\*\*\*首先引入超像素的方法解决小缺陷的问题，为占比问题提供了一个新的思路并证实了占比的重要性。\*\*\*在小人脸检测任务中采用\*\*\*策略，在原有的金字塔模型上固定目标框，缩放图片（这样做主要是因为人脸检测中的图像允许他这样做，通过这一条再说对其他图像并不适用）。针对轮胎图像的特点（可以简单的列举一下特点，或者直接引用某篇文献），我们提出了一种......方法并新建了一个网络，经常被用在传统方法的预处理中，\*\*\*、\*\*\*、\*\*\*也使用了这种方法。（大概列几个文献，证明可行性）。

Introduction v2

1. 一句话，自动检测的重要性，\*\*\*，\*\*\*，\*\*\*领域均有研究，轮胎检测依然持续受到关注，其中\*\*\*\*（把一些不好融合到related work中的轮胎检测一并说了）都做出了大量的贡献（就大概这个意思吧）。
2. 在实际的工业应用中，轮胎检测的基本流程：a.拿到x光射线的图像（如图所示）；b.检测；c.根据检测结果决定对轮胎进行怎么样的处理（写流程太傻了，中心思想是“需要根据检测结果对轮胎进行下一步处理”，突出检测到缺陷大小的重要性）。人工检测，效率低，不安全，浪费资源......
3. 目前已存在的检测算法大体可以分为\*\*\*和\*\*\*和\*\*\*。分别简单介绍一下。（包含机器学习的方法，只介绍机器学习在轮胎上的应用即可，其他的不用说）。\*\*\*证明了深度学习（分割方法）在轮胎检测的可行性，但是结果表明这种pixel-to-pixel的网络对细节信息和小目标检测效果差，这同样也是自然图像上遇到的问题（和自然图像中的表现一致）（这里重新组织一下，重点是要把话题转移到自然图像、多尺度上）。
4. 为了解决这个问题，已经有很多多尺度的方法被提出在公共数据集中。（成功的开始介绍之前想过的多尺度方法）。这些方法虽然在一定程度上缓解了小目标的检测，但是忽略了一个本质的原因...\*\*\*首先引入超像素的方法解决小缺陷的问题，为占比问题提供了一个新的思路并证实了占比的重要性。\*\*\*在小人脸检测任务中采用\*\*\*策略，在原有的金字塔模型上固定目标框，缩放图片（这样做主要是因为人脸检测中的图像允许他这样做，通过这一条再说对其他图像并不适用）。针对轮胎图像的特点（可以简单的列举一下特点，或者直接引用某篇文献），我们提出了一种......方法并新建了一个网络，经常被用在传统方法的预处理中，\*\*\*、\*\*\*、\*\*\*也使用了这种方法。（大概列几个文献，证明可行性）。

Introduction v1

（1）轮胎缺陷检测+轮胎特征分析

1、展示轮胎图像，介绍其特征（**图和第二部分——预处理，分块处理共用一张图**）

2、讲一下为什么会属于分割任务（因为不同的缺陷面积会有不同的分级和处理方法）；引到自然图像的分割上。

（2）随着深度学习的流行，分割技术越来越多，大概 举例。

1、归归类，说说它们的做法有哪些不同（角度可以从感知多尺度方法入手。可以最后写）。

（3）这些方法针对自然图像有着很好的效果，但是由于轮胎数据的特殊性以及关注程度，出现的文章并不多，其中有\*\*方法。这些方法存在\*\*问题，我提出了一种\*\*，解决了\*\*\*。

The proposed method

（1）大体描述框架。（**图2，网络框架图**）

（2）预处理，分块处理（解释为什么轮胎图像可以分块，分块对比图）

（3）细的分割

1、解释一下为什么要用3层卷积，或者不用3层卷积用什么（可以最后写）。

2、分块之后怎么得到的

（4）粗的分割

借鉴了FCN的想法（暂时的，如果后面可以改成别的网络，但是重点我们只是借助于FCN的稠密预测能力，因为有细节提取网络，所以精细问题可以不用考虑）

实现细节

Experiments

数据集和评价指标

消融实验结果比较

与现有方法比较

实验结果比较

和存在的轮胎图像的方法作对比

和其他的分割领域（用轮胎数据集fine-turn过）的方法对比

指标以及指标结果的表格

Conclusion

我们提出了\*\*\*，有效的解决了\*\*\*，实验结果证明我们方法的有效性。同时，哪些地方依然存在这问题。

.

0.0

需要阐明的问题：

1. 轮胎图像分块的可行性。一般图像处理没有人用分块，大多数采用推荐框的形式，缺点是一旦推荐框检测不出来，后续的网络就没法检测。我们用分块不是为了精确地找到缺陷，而是为了不丢失每一处细节，尽管它有可能不是缺陷。
2. 分块的好处。最大的好处是感知细节；间接增加了数据集，增大了有缺陷图像的比例，解决了检测没有缺陷的问题。
3. 可以加一个简单的实验，看看没有缺陷的小图像块有多少是真的检测出是没有图像的，说明对增强鲁棒性有一定的作用。
4. 一个网络感知细节信息，一个网络感知全局信息，两只眼睛看世界，具体最后的结果是谁做主导谁做辅助，也是靠训练来的。

