**An End-to-End Steel Surface Defect Detection Approach via Fusing Multiple Hierarchical Features**

**摘要**

完整的缺陷检测任务旨在实现图像中每个缺陷的特定类别和精确位置，这使其在实践中仍然具有挑战性。缺陷检测是分类和定位的复合任务，导致相关方法通常很难兼顾两者的准确性。 缺陷检测的实现取决于包含昂贵的手动注释的特殊检测数据集。 在本文中，我们提出了一种基于深度学习的新型缺陷检测系统，并针对实际的工业应用：钢板缺陷检测。 为了实现强大的分类能力，该系统采用基线卷积神经网络（CNN）在每个阶段生成特征图，然后提出的多级特征融合网络（MFN）将多个分层特征组合为一个特征，其中可以包含更多缺陷的位置详细信息。 基于这些多级特征，采用区域提议网络（RPN）来生成关注区域（ROI）。 对于每个ROI，由分类器和边界框回归器组成的检测器将产生最终的检测结果。 最后，我们建立了一个缺陷检测数据集NEU-DET来训练和评估我们的方法。 在NEU-DET上，我们的方法通过使用300个建议，使用基线网络ResNet34 / 50实现了74.8 / 82.3 mAP。 此外，仅使用50个建议，我们的方法就可以在单个GPU上以20 ft / s的速度进行检测，并达到上述性能的92％，因此具有进行实时检测的潜力。

**索引词**-自动缺陷检查（ADI），缺陷检测数据集（NEU-DET），缺陷检测网络（DDN），多级特征融合网络（MFN）。

**一、引言**

缺陷检查是保证工业生产质量的关键步骤，特别是对于钢板。 但是，此过程通常是手动执行的，这是不可靠且耗时的。 为了代替手工作业，期望允许机器使用计算机视觉技术来自动检查钢板的表面缺陷。

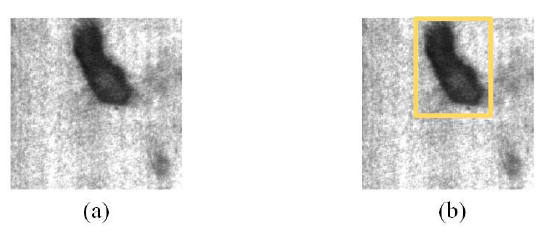


图1.缺陷分类和缺陷检测任务。 （a）缺陷分类任务的目标是“什么”，仅输出缺陷类别分数。 （b）缺陷检测任务的目标是“什么”和“在哪里”，输出带有缺陷类别分数的边界框。

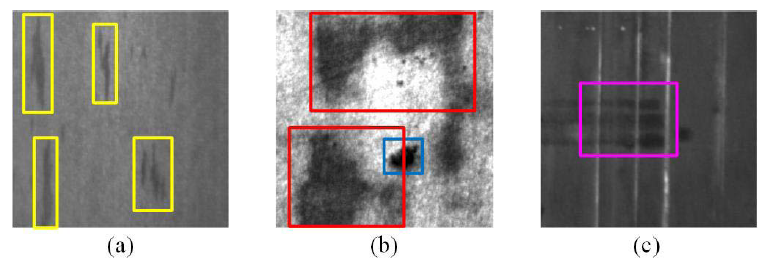


图2.复杂的缺陷 （a）多个缺陷。黄色框表示缺陷属于同一类别。 （b）多类缺陷。 红色和蓝色框表示不同类别的缺陷。 （c）重叠缺陷。 粉色框包围不同类别缺陷的重叠区域。

计算机视觉的创始人，英国神经生理学家Marr认为，视觉任务可以定义为“什么在哪里”，即发现图像中所含内容及其位置的过程[1]。 因此，对象分类和检测是计算机视觉研究领域中最基本的问题[2]。 同样，自动缺陷检查（ADI）也可以分为两种类型：缺陷分类和缺陷检测。 给定缺陷图像，缺陷分类任务是解决该图像是否包含某些类别的缺陷[图2. 1（a）]，缺陷检测任务是解决此图像中存在缺陷的地方，该缺陷由带有等级评分的边界框表示[图1（a）， 1（b）]。 因此，完整的缺陷检测任务包括两部分：缺陷分类，确定特定的缺陷类别以及缺陷定位，获取详细的缺陷区域。 对于钢板上的缺陷检查，检测任务比复杂的缺陷（例如，多个缺陷2（a）]，多类缺陷[图2（b）]，和重叠的缺陷[图2（c）]）具有优越的优势。分类任务只能找到图像中类别置信度最高的缺陷，并且不知道图2（a）所示的缺陷数量，图2（b）所示的缺陷类别以及图2 （C）所示的重叠缺陷。 但是，对于后续质量评估系统，缺陷的数量，类别和复杂性将作为评估钢板质量的主要指标。 显然，缺陷检测可以实现钢板表面更全面的信息反射。

以前的ADI方法存在两个常见问题：一个是对手工功能的不清楚使用[3] – [5]。 特征的确定过于主观，因此人类经验通常在其中起决定性作用。 另一个问题是缺陷定位不精确[图2。 3（a）]。 大多数方法仅执行缺陷分类[6] – [8]或不完整的缺陷检测。 例如，一些方法执行二进制分类以找到缺陷区域[9]，[10]或仅提供缺陷区域的粗略区域[11]，[12]。 最近开发的深度学习（DL）技术可以克服传统ADI方法的弊端，并在许多视觉任务上取得了显著成果。 DL可以通过深度网络[例如，卷积神经网络（CNN）]提取判别表示。 这些表示可以达到较高的抽象水平，因此具有很强的表示能力。 相比之下，手工功能只是低级功能的组合[16]。 此外，DL可以训练带有位置注释的样本以获得准确的位置信息。

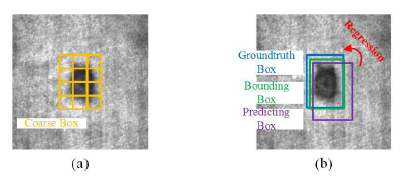


图3.获得缺陷区域的不同样式。 （a）许多以前基于手工特征的探测器通过各种特殊方法将相关的空间像元直接组合成一个块。 该块被视为检测区域，它是未经精炼的粗框。（b）基于DL的检测器主要使用回归方法来完善预测框。 通过大量的迭代学习，预测框逐渐接近地面真实框。 最后，将改进后的框视为缺陷的边界框，可以表示缺陷的精确位置信息。

目前，一些研究已经将DL用于ADI。然而，由于缺乏特殊的数据集，大多数方法只能进行缺陷分类[18]-[21]。 缺陷分类似乎过于简化，无法提供位置信息。 其他方法使用DL和传统图像处理的组合来执行缺陷检测或分割[17]。 这些方法始终将DL分类器与基于传统图像处理的检测器或分割器并行使用。 这种方式可以消除对特殊训练数据集的需求，但会破坏DL系统的端到端特性，并在某种程度上失去智能和概括性。 与上述方法不同，我们尝试建立ADI的端到端缺陷检测系统，该系统可以为包围盒提供分类评分，以精确地分类和定位缺陷[Fig。 3（b）]。像Mask R-CNN [13]这样的基于DL的分割器似乎更适合显示缺陷的形状。 但是，这种分割器会消耗大量的计算资源，无法满足工业检测的实时需求。 此外，对于行业来说，构建大型实例级缺陷分割数据集是非常不切实际的，因此这种分割器几乎是不可能应用的。 因此，目前对ADI进行缺陷检测是最好的折衷方案。

本文主要针对三个挑战。 首先，检测系统需要强大的分类能力。 常见的分类问题，例如类间相似性，类内差异和背景干扰，也存在于ADI [9]，[11]中。 因此，我们在系统中装备了一个深层网络ResNet作为骨干[23]。 作为当前在转移学习中的研究[15]，驱动大型网络的关键是在ImageNet上进行预培训[22]。 通过在足够的数据上训练ResNet，检测系统可以获得强大的分类能力。

其次，在基于DL的方法中使用CNN功能执行缺陷定位的挑战仍然存在。 众所周知，CNN的卷积层可以视为滤镜，这导致当图像在CNN中流动时某些位置细节将逐渐丢失。 通常，基于DL的方法基于最后的卷积特征图执行定位[14]，[28]，[34]。 我们的方法是融合多个特征图。 因为特征图在CNN的每个阶段都表现出不同的特征：浅层特征具有丰富的信息，但没有足够的区分性；深层特征在语义上具有鲁棒性，但是丢失了太多细节。在其他领域[34]，Hypernet也使用更多功能，但是它们主要是从网络的后半部分选择的。拟议的多级特征融合网络（MFN）结合了涵盖所有阶段的多个功能。 我们从工业角度解决检测问题。 由于灰度图像比彩色图像具有更少的信息，因此MFN必须包含HyperNet丢弃的较低级功能。此外，MFN在融合之前统一多个特征的大小，这不仅可以保存更多的图像细节，而且可以使用更少的模型参数。

第三，在缺陷检测中，数据注释非常昂贵，因为必须绘制一个缺陷的边界框并为其分配类别标签。 该领域的最新进展可归因于两个因素：1）ImageNet预训练模型； 2）大型基线CNN，这在基于DL的缺陷分类中取得了很大进展[18] – [20]。 但是，有限的数据和昂贵的注释仍然限制了缺陷检测的发展。在本文中，我们为微调模型打开了一个缺陷检测数据集NEU-DET。 当DL模型完成了对特殊数据集的训练后，便可以将其用于执行缺陷检测任务。

本文旨在建立一种称为缺陷检测网络（DDN）的端到端ADI系统，以克服上述挑战。 DDN 1）在缺陷分类中采用了强大的ResNet； 2）提出MFN，以集合更多的地点细节； 3）建立缺陷检测数据集以进行微调并报告改进情况。 更详细地说，首先，我们在ImageNet上对ResNet进行预训练，然后在NEU-DET上微调所有模型。 MFN可以将选定的功能融合到一个多级功能中，该功能具有涵盖ResNet所有阶段的特性。 接下来，在基于多级特征的提案生成中采用区域提案网络（RPN），然后DDN可以输出类别得分和边界框的坐标。 最后，我们评估了在NEU-DET上提出的方法，结果可以证明明显优于其他ADI方法。

总而言之，本文的主要贡献如下。

1）引入了端到端缺陷检测管道DDN，该管道将ResNet和RPN集成在一起，以进行精确的缺陷分类和定位。

2）提出的用于融合多级特征的MFN。 与其他融合方法相比，MFN可以将低级和高级功能组合在一起，从而使多级功能具有更全面的特性。

3）用于微调网络的缺陷检测数据集NEU-DET，并证明所提出的DDN在该数据集上具有非常有竞争力的性能。

**二、相关工作**

**A.缺陷检查**

通常，​​缺陷分类方法包括两部分：特征提取器和分类器。 经典的特征提取器是要获取手工特征，例如HOG和LBP，并且始终跟随有分类器，例如SVM。 因此，不同特征提取器和分类器的组合产生了多种缺陷分类方法。例如，Song和Yan [3]改进了LBP以抵抗噪声，并采用NNC和SVM对缺陷进行分类。Ghorai等。 文献[9]基于少量的小波特征，并使用SVM进行缺陷分类。 与上述两种方法不同，Chu等。 [8]采用通用特征提取器并增强SVM。 从计算机视觉的角度来看，缺陷分类任务本质上是缺陷图像分类，它在复杂的缺陷图像中难以解决。 为了解决这个问题，简单直接的方法是在缺陷分类之前执行缺陷定位，使检查任务在缺陷区域上进行分类，而不是对整个缺陷图像进行分类，这就是缺陷检测任务。 例如，[11]和[12]中的缺陷检测器首先执行0-1分类以判断特征是属于缺陷类别还是非缺陷类别，然后根据缺陷类别特征的边界找到缺陷区域， 最后执行不同的分类方法以确定缺陷的特定类别。 另外，对于快速检测的要求，还有另一种简化的检测器，该检测器仅关注缺陷区域，而与缺陷属于不同类别无关[10]。

但是，基于DL的方法与上述方法有根本不同。 手工特征提取器在本地分析单个图像并提取特征。 但是，CNN将通过大量学习来构造所有输入数据的表示。 CNN具有良好的概括性和可移植性，因此存在一些基于CNN的缺陷检查方法。 例如，Chen和Ho [21]证明可以通过某种方式将像Overfeat [24]这样的对象检测器转换为缺陷检测器。 与[18]和[19]相似，他们证明了使用顺序CNN提取特征可以提高缺陷检查的分类精度。 同样，基于连续的CNN，Ren等人 [17]对分类结果执行额外的缺陷分割任务，以定义缺陷的边界。 此外，纳塔拉延等。 [20]使用更深的神经网络VGG19进行缺陷分类。 随着CNN的深入，缺陷分类的准确性得到了进一步提高。

**B.基准网络**

当前有三种流行的CNN架构，它们被用作预训练的基准网络。 早期成功的网络基于顺序流水线体系结构[25]，它建立了CNN的基本结构并证明了网络深度的重要性。 随后，初始网络采用模块化单元，在不增加计算成本的情况下增加了网络的深度和宽度[26]。 第三类是ResNet，它使用残差块使网络更深而不会过度拟合[23]。ResNet广泛应用于各种视觉任务，只需几个参数即可获得具有竞争力的结果。

选择合适的基线网络是获得DL方法的良好结果的关键。 大型网络对输入数据具有很强的表示能力，因此在高抽象级别提取了特征，但是对训练数据有很大的需求。

**C. CNN检测器**

CNN检测器旨在用边界框对每个目标进行分类和定位。 它们主要分为两种方法：一种是基于区域的方法，另一种是直接回归方法。 基于区域的最著名的探测器是“ R-CNN家族” [27]，[28]，[14]。 在此框架中，采用了数千个与类无关的区域建议进行检测。 基于区域的方法精度较高，但需要更多的计算量。 代表性的直接回归方法是YOLO [29]和SSD [30]。 他们将图像直接分成小网格，然后为每个网格预测边界框，然后将边界回归到ground-truth框。 直接回归法检测速度快，但在小情况下难以解决。

**三， 缺陷检测网络**

在本节中，将详细描述DDN（请参见图4）。CNN处理任意大小的单尺度图像，并在ConvNet的每个阶段生成卷积特征图（ConvNet表示CNN的卷积部分）。 我们提取了多个特征图，然后通过使用轻量级MFN以相同的维度聚合它们。 这样，MFN功能具有ConvNet的多个层次级别的特征。接下来，使用RPN [14]生成区域建议MFN特征上的[关注区域（ROI）]。最后，通过ROI池[28]和全局平均池（GAP）层，将与每个ROI对应的MFN特征转换为固定长度的特征。 要素被馈入两个完全连接的（fc）层。 一个是（C +1）缺陷分类层之一（cls），另一个是边界框回归层（loc）。

本节的其余部分介绍了DDN的详细信息，并说明了为什么我们需要将MFN设计到网络中以进行缺陷检测任务。

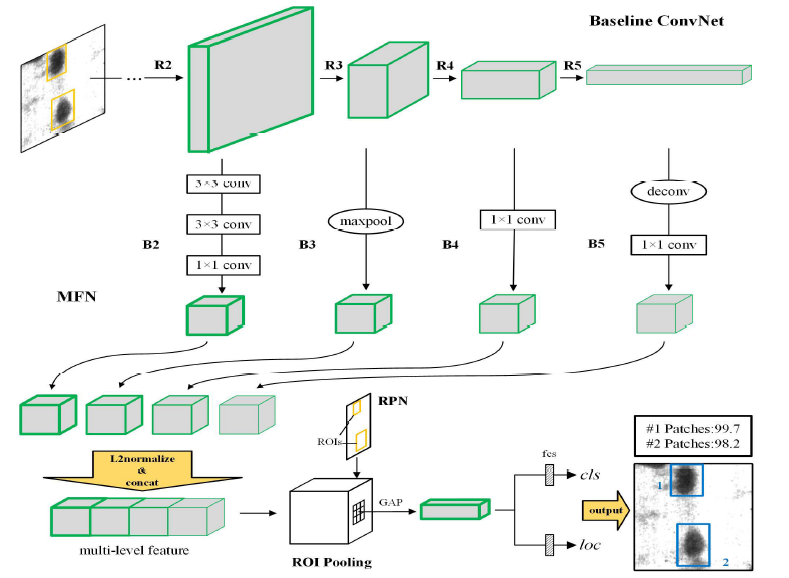


图4. DDN。 一次即可从Baseline ConvNet的每个阶段提取要素，然后通过MFN将其融合为多级要素。 采用RPN基于多级功能生成ROI。 对于每个ROI，通过ROI池和GAP层，将相应的多级特征转换为定长特征。 两个fc层处理每个固定长度特征，并馈入输出层，产生两个结果：（C +1）个缺陷类别预测（cls）和精确的边界框坐标（loc）。

**A.基线ConvNet体系结构**

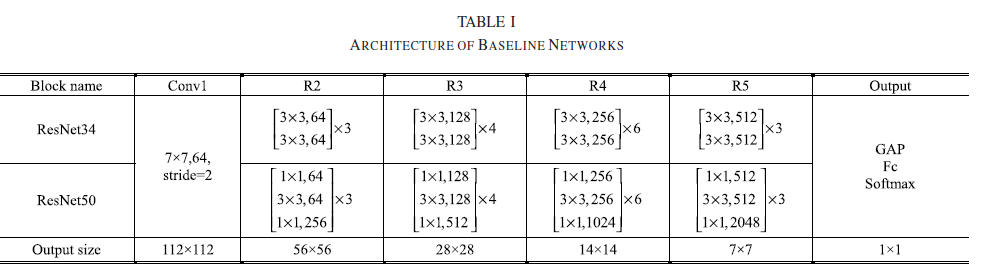
众所周知，对ImageNet数据集进行预训练对于获得竞争性能很重要，然后可以在相对较小的缺陷数据集上微调此预训练模型。 在本文中，我们选择最近成功的基准网络ResNet作为主干。 ResNet具有以下几个吸引人的优点。

1）与相同大小的顺序管线架构的CNN（ResNet50与VGG16、0.85 M与138 M参数）相比，ResNet可以用极少的参数实现最先进的精度。这意味着ResNet具有较低的计算成本，并且过拟合的可能性较小。

2）ResNet使用GAP来处理最终的卷积特征图，而不是双重堆叠的fc层，从而可以保留图像中缺陷的更全面的位置信息。

3）ResNet具有模块化的ConvNet，易于集成。

在本文中，我们选择ResNet34和ResNet50作为基准网络。 这两个网络的详细结构如表I所示，残差块表示为{R2，R3，R4，R5}。



**B.生成多级特征**

以前的优秀方法仅利用高级特征来提取区域提议（例如在最后的卷积特征图上更快的R-CNN提取提议）。 为了获得优质的区域建议，使用单层特征应该扩展到多级特征。 显然，最简单的方法是从多个图层组合特征图[31]。因此，现在出现了一个问题，应该合并哪些层？ 有两个基本条件：不相邻，因为相邻层具有高度的局部相关性[32]；以及覆盖范围，包括从低层到高层的特征。 对于ResNet，最直观的方法是在每个残差块中合并最后一层。

为了融合不同级别的功能，将建议的网络MFN附加到预训练模型上。 MFN有四个分支，分别表示为{B2，B3，B4，B5}，每个分支都是一个小型网络。 B2，B3，B4和B5依次连接到R2，R3，R4和R5的最后一层。 当图像流过基线ConvNet时，将按顺序生成Ri特征。 Ri特征是指从残差块Ri的最后一层输出的特征图，i = 2，...5。 类似地，Bi特征是从MFN批次Bi的最后一层产生的特征图，i = 2，... 5.然后，Ri特征中的每一个都被引到MFN中产生Bi特征的相应分支。 最后，通过合并来自CNN不同阶段的B2，B3，B4和B5特征获得多级特征。

最后一点，MFN是有效的计算和强大的概括。 MFN可以通过修改1×1转换的滤波器数量来减少所需的参数。 此操作可能会损害准确性，但在训练数据不足的情况下会防止过拟合。

**C.提取区域建议**

RPN用于通过在多级特征图上滑动来提取区域建议。RPN将任意大小的图像用作输入并输出锚定框（候选框），每个锚定框都有一个分数来表示其是否为缺陷。 RPN的独创性是“锚定”方案，可以使锚定框具有多种比例和纵横比。 然后，将锚框分层映射到输入图像，以便生成具有多个比例和纵横比的区域建议。 由于MFN功能的分辨率大小，可以将RPN视为在R4特征上滑动。 按照[14]，我们设置三个长宽比{1：1，1：2，2：1}。考虑到缺陷的多种尺寸，我们设置了四个等级。 因此，RPN在每个滑动位置产生12个锚定盒。

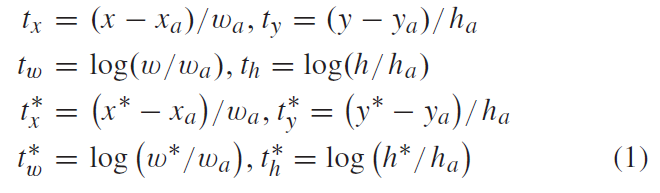
区域提议提取器始终以ROI池层结束。 该层对每个ROI内的特征图执行最大池化操作，以将其转换为固定大小为W×H的小特征向量（ResNet34使用512-d，ResNet50使用2048-d）（本文中为7×7）。 最后，基于这些小立方体，计算每个区域建议与相邻的真实框之间的偏移量以及是否存在缺陷的概率。

对于单个图像，RPN可以提取数千个区域建议。 为了处理冗余信息，通常采用贪婪非最大抑制（NMS）来消除高重叠区域建议。 我们将NMS的联合交叉（IOU）阈值设置为0.7，这可以丢弃大多数区域建议。 在NMS之后，从其余区域中选择排名前K的区域提案。 在下文中，由于提取的质量区域建议，我们使用前300个区域建议对DDN进行了微调，但是减少了此数量以加快检测速度而又不损害测试时的准确性。

**IV、培训**

**A.多任务丢失特征**

缺陷检测任务可以分为两个子任务，因此DDN具有两个输出层。 cls层针对C + 1类（C缺陷类加一个背景类）上的每个ROI输出离散的概率分布k =（k1，...，kC）。通常，k由softmax函数计算。 cls损失Lcls是两类（有缺陷或无缺陷）的对数损失。 ，其中k ∗是地面真理类。 loc层为C个缺陷类别中的每个缺陷类别输出边界框回归偏移。 如[28]中所示，位置损失Lloc是平滑的L1损失函数。 Lloc =其中，t ∗是与正样本关联的底线框。 对于边界框回归，我们采用[27]中给出的t和t ∗的参数化



其中下标x，y，w和h表示每个框的中心坐标及其宽度和高度。 变量x，xa和x ∗分别代表预测框，锚定框和地面真实框（y，w和h的规则相同）。

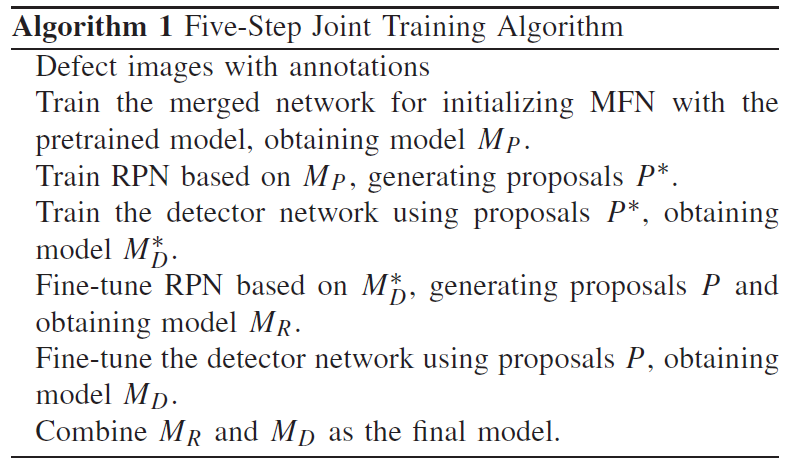
通过这些定义，我们最小化了多任务损失函数，该函数定义为



其中λ是权重参数，用于平衡cls和loc项。 在训练期间，我们设置λ= 2表示DDN致力于获得更好的缺陷位置。 p ∗是loc项的激活参数。 本地化损耗仅对正样本（p ∗ = 1）参与后续计算，否则将被禁用（p ∗ = 0）。我们遵循[14]中的“ IOU”策略来确定锚点的正样本和负样本。

**B.联合训练**

对于预训练网络，MFN和RPN是新的层。因此，我们需要通过训练使这三个网络共享共同的卷积特征。 预训练模型本质上是分类网络，从MFN生成的多级特征可以直接输入到cls层。 因此，可以将预训练的网络和MFN合并为一个网络，然后进行端到端训练。 没有RPN，其余的DDN是检测器网络。为了与RPN共享功能，采用[14]中的四步交替训练策略，在训练RPN和训练检测器网络之间交替进行。 结合这两种策略，我们开发了一种可行的五步联合训练算法，如算法1所示。



在步骤2和步骤3之后，依次使用ImageNet预训练模型初始化RPN和检测器网络。但是，这两个网络目前尚未共享卷积特征。 他们得到它，直到完成步骤3和步骤4的微调过程为止。 具体来说，我们冻结共享的卷积层，仅微调未共享的层。 最后，我们将两个网络合并为一个联合网络。

**C.实施**

对于DDN，我们采用以图像为中心的训练策略。 调整图像大小，使其短边为600像素。 我们使用随机梯度descen进行训练，权重衰减为0.0001，动量为0.9。 我们在每个小批量迭代中获取一张图像。 最小批量大小对于探测器网络训练（包括MFN训练）是64，对于RPN训练是128。 对于200k个小批量迭代，我们使用0.001的学习率对其他100k个小批量迭代使用0.0001的学习率对模型进行微调。 我们使用“ Xavier”初始化所有新层[33]。为了避免过度拟合，我们还使用了几种数据增强方法，例如旋转，反射和移位，但删除了辍学模块。

**V.实验**

DDN的性能根据我们的缺陷数据集：NEU-CLS和NEU-DET进行评估。 我们证明DDN达到了合理的设计和令人鼓舞的结果。

A. NEU-DET数据集

NEU表面缺陷是我们七年前打开的缺陷分类数据集[3]。 热轧钢板有六种类型的缺陷，包括开裂，夹杂物，斑块，表面凹陷，氧化皮和划痕。每个类别有300张图像，但这并不意味着一幅图像包含一个缺陷。 缺陷图像的示例如图5所示。

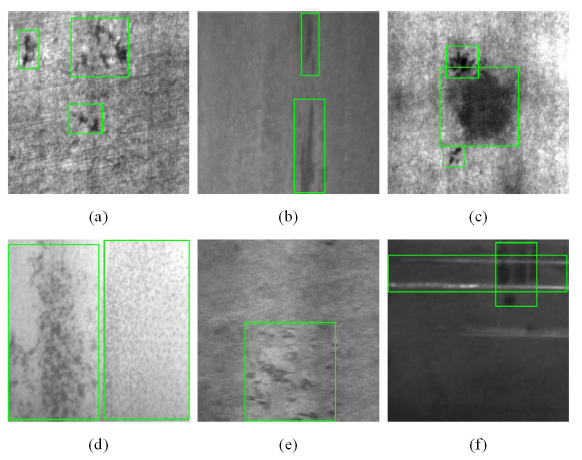


图5. NEU中带有注释的缺陷图像示例。 绿色框是一个ground-truth框，它具有一个类标签和该框的两个角坐标（左上和右下）。 图像所属的类别：（a）裂纹，（b）夹杂物，（c）色块，（d）有凹痕的表面，（e）卷起的水垢和（f）划痕。

为了执行缺陷检测任务，我们提供了另存为XML文件的注释。 使用它们，分类数据集将升级为检测数据集。 注释标记图像中出现的每个缺陷的类别和边界框。每个边界框都被视为地面真实框，由其左上角和右下角坐标表示。 总共有将近5000个地面真理盒子。 为简单起见，我们将原始数据集称为NEU-CLS，并将补充数据集称为NEU-DET。 注释的示例也显示在图5中。

**B. NEU-CLS上的缺陷分类**

如上所述，MFN可以合并到基线CNN中以进行缺陷分类任务。 因此，我们首先报告缺陷分类的结果，表明我们的方法可以比其他相关方法获得更高的竞争准确性，并且合并MFN待遇不会显着影响分类能力。 图6显示了与其他方法相比的缺陷分类结果。 根据图6，我们可以得出以下结论。

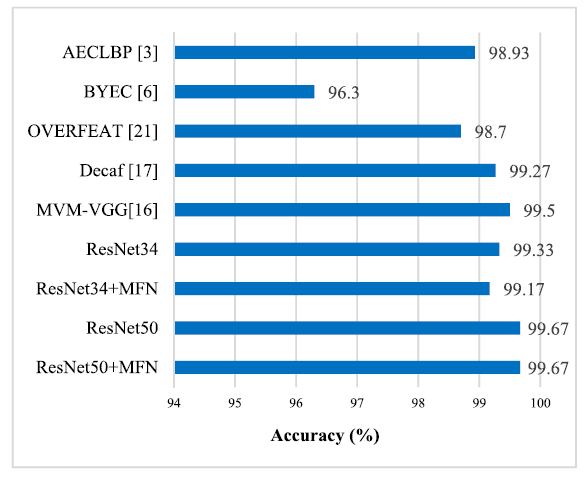


图6. NEU-CLS数据集的分类结果。

1）具有MFN的网络可以很好地进行缺陷分类，因此多级特征仍然具有很强的语义能力。

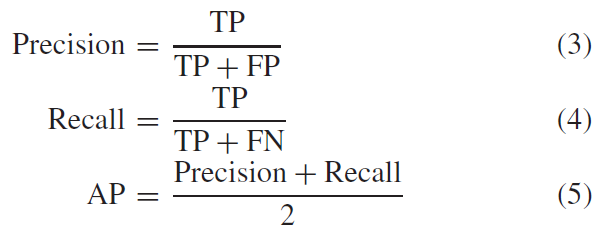
2）对于ResNet34，MFN会稍微损害分类结果。 但是，对于更深的网络ResNet50，这种影响消失了。 这表明从更深的网络中提取的特征更具特色，因此整个网络变得更加健壮。

3）使用MFN，ResNet34可获得ResNet50准确性的99％，这表明，实际上，缺陷分类任务实际上并不需要非常深的网络。

众所周知，更强的缺陷分类性能应与更强的缺陷检测性能正相关。 良好的分类结果是后续缺陷检测实验的前提。

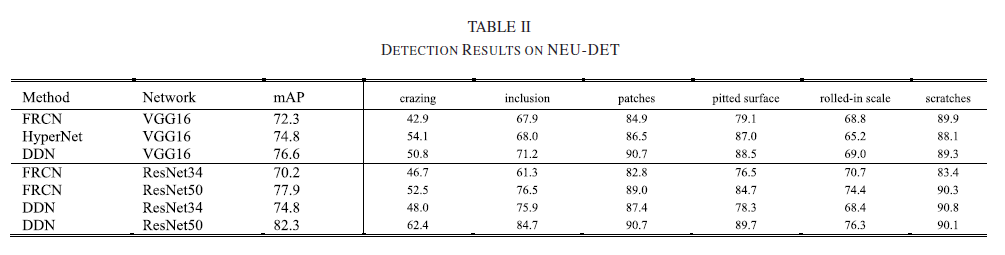
**C. NEU-DET上的缺陷检测**

我们对NEU-DET数据集进行缺陷检测实验。 按照惯例，我们将NEU-DET分为训练集和测试集，并固定训练/测试集。 包含1260张图像的训练集用于微调第IV-B节中介绍的网络，而包含540张图像的测试集。 我们在测试集上将DDN与更快的R-CNN和HyperNet [34]进行了比较，两种方法都使用了他们论文中提到的相同基准网络（VGG16 [40]）。此外，由于类似的提案生成器，在ResNet34 / 50上也对DDN和更快的R-CNN进行了实验。 与缺陷分类不同，在缺陷检测的情况下，仅准确性不是适当的性能指标。 因此，我们通过平均精度（AP）评估了检测实验的结果，这是两个重要检测指标：Precision和Recall之间的良好折衷。 这些索引定义如下：



其中TP，FP和FN分别代表真阳性，假阳性和假阴性的数量。 还计算平均AP（mAP）以评估整体性能，这是所有类别的AP的平均值。

表II示出了缺陷检测实验的结果。在基线ResNet34 / 50下，DNN的mAP为74.8 / 82.3，比更快的R-CNN高4.6 / 4.4。 该结果表明，从多级特征中提取的建议优于从单级特征中提取的建议。 在相同的基准网络（VGG16）下，更快的R-CNN的mAP为72.3，而HyperNet的mAP为74.8。 DNN的mAP达到76.6，比更快的R-CNN高4.3点，比HyperNet高1.8点。HyperNet还是基于多种功能的检测器，但是我们的方法可以提取更高质量的区域建议，这将在第六节中详细讨论。 NEU-DET上的检测结果示例如图7所示。



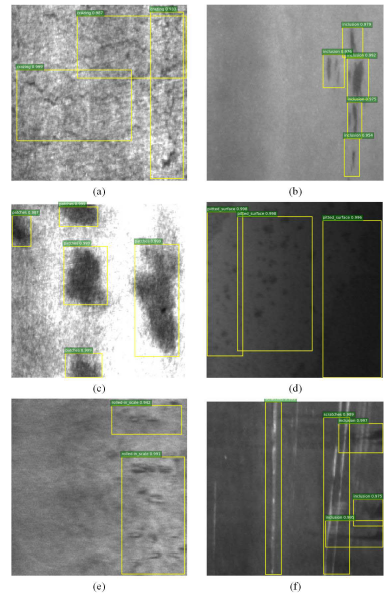


图7. NEU-DET上的检测结果示例。 对于每个缺陷，黄色框是指示其位置的边界框，绿色标签是类分数。 图像所属的子集（a）裂纹，（b）夹杂物，（c）色块，（d）带凹痕的表面，（e）滚入的水垢和（f）划痕。

通过先前的缺陷分类实验，证明了MFN对分类精度的影响很小。因此，mAP的改进得益于从多级特征中提取的质量区域建议。这意味着MFN有助于提高定位精度。 我们在第V-D节中专门评估了MFN的性能。

**D.对MFN的分析**

为了验证MFN能够提高定位精度，我们与几个区域提议提取器，滑动窗口，边框[35]和选择性搜索[36]进行了比较。 除这些方法外，还将RPN + MFN与朴素的RPN（基于单层功能的建议提取）进行比较。如果提议的质量得到改善，检测器可以使用更少的提议和更严格的IOU阈值，而不会损害召回率。 因此，我们用不同数量的提案和IOU阈值评估NEU-DET测试集的召回率。 提案数是通过这些方法选择的排名前K位的区域提案。 IOU表示预测框与地面真实框的交集和并集之间的比率。

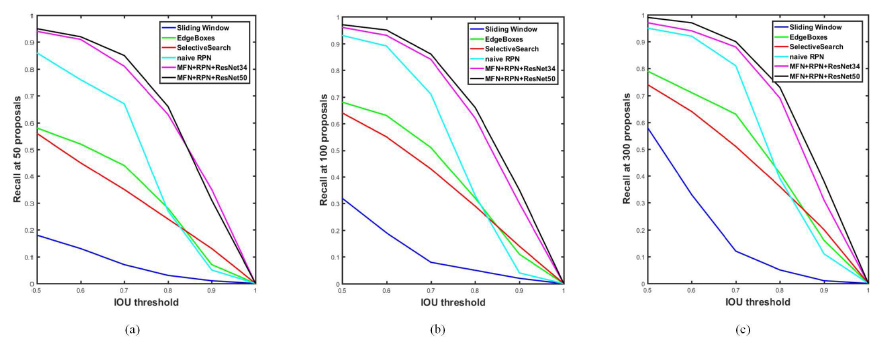


图8：在不同数量的区域提案中，NEU-DET的召回率与IOU阈值之间的关系。 （a）50个区域提案。 （b）100个区域提案。 （c）300个区域提案。

图8显示了在三个不同数量的区域提议中具有各种IOU阈值的缺陷召回。 IOU阈值越大，选择建议的质量越高。毫不奇怪，基于卷积特征的方法的性能要比没有CNN的方法高得多[37]。 当IOU> 0.7时，与RPN + MFN相比，天真RPN的召回率急剧下降。 天真的RPN仅从高级特征中提取建议，并且某些位置信息已由前一个过滤层使得提议质量下降。 随着提案数量的增加，当IOU> 0.7时，幼稚的RPN会急剧下降。 这是因为RPN提取了太多低质量的提案，并且随着提案的增加而变得更加明显。 幼稚的RPN在严格的IOU阈值（例如IOU> 0.7）下无法正常工作。 MFN可以帮助RPN从低级和中级功能获取位置信息，这使得RPN在严格的IOU阈值下具有较高的容忍度。

提议的数量增加可以使召回率提高，但是这将大大增加检测的运行时间[38]，更糟糕的是，低质量的提议会参与检测过程，从而导致缺陷检测失败。 某些情况下。 因此，一个好的检测器应该选择尽可能少的建议，同时要有相对严格的IOU阈值。 图9显示了在三个不同的IOU阈值下具有各种建议数量的缺陷召回情况。 幼稚的RPN可以通过前300个提案获得理想的召回率，但是RPN + MFN只需要前100个提案即可获得类似的性能。

如图10所示，对于使用ResNet34的RPN + MFN，通过仅选择50个提案，我们实现了选择300个提案的92％的性能，这将运行时间减少了一半。 我们认为选择前50个建议是在实际缺陷检测任务中的一个很好的权衡。

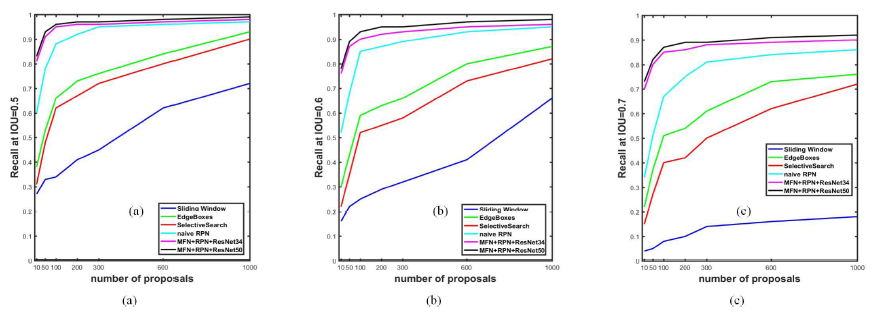


图9.在不同的IOU阈值下，NEU-DET上的提案数量与召回数的关系。 （a）IOU阈值为0.5。 （b）IOU阈值为0.6。 （c）IOU阈值为0.7。

**VI、讨论**

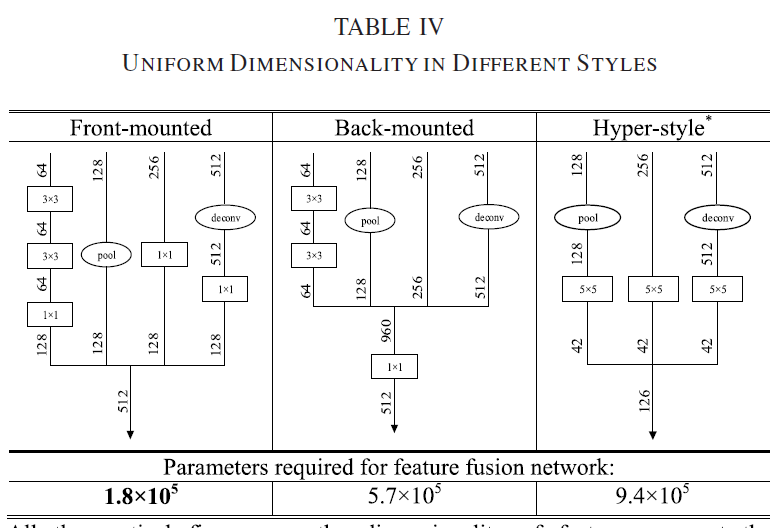
在本节中，为了证明我们的设计是合理的和先进的，我们讨论了一些可能影响缺陷检测的隐式因素。

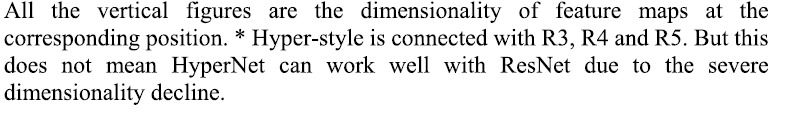
**A.合并哪一层用于MFN？**

MFN将各个级别的功能组合到一个多级特征中，这对于提高检测效率是有效的。在第III-B节中，简要讨论了应该组合哪种层。 在DDN中，我们选择四个层作为R1，R2，R3和R4的最后一层。 因此，这四个层的其他组合方式是否可以导致更好的性能。 因此，我们在NEU-DET数据集上以五种不同的组合方式训练DDN + ResNet34。 如表III所示，将所有四个层组合在一起的性能优于其他方式。 这表明多级特征对于提高检测精度是有效的。此外，在缺陷检测中，与高级特征（例如，R5特征）相比，低级特征（例如，R1特征）应引起更多关注，因为R2特征比R5特征具有更丰富的位置信息。

B.简单设计对MFN更有效吗？

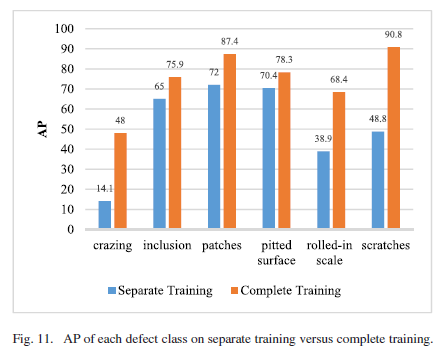
MFN的主要作用是在分辨率和尺寸上统一不同级别的特征。 为了保持尺寸一致，一种简单的方法是使用1×1转换来减小/增加尺寸。 1×1转换有两种放置方式：正面安装和背面安装。前置模式意味着在连接多级特征之前放置1×1转换。 我们在本文中使用的是正面安装的模式，即在MFN的每个分支的末尾放置一个1×1的转换，而背面安装的模式意味着在多级连接后放置一个1×1的特征转换。 这种模式看似简单，但实际上需要更多参数。 类似于[34]，我们使用多个5×5转换来同时使分辨率和维数均匀。 但是，5×5 conv是昂贵的操作，其效果与双层3×3 conv相同，但需要附加参数。 表IV详细显示了三种模式之间的可比结果。前置样式使用的参数比后置样式少三倍，比超样式少五倍。 因此，前装式的MFN不太可能过度装配。 此外，在相同分辨率大小的情况下，由于MFN功能的尺寸比Hyper功能的尺寸更大（512与126），因此可以保留更完整的信息。





**C.我们需要更多的缺陷数据吗？**

众所周知，物体检测器可以通过更多的训练数据来提高性能[39]。 因此，该规则对于工业缺陷数据是否也有效？ 为了弄清楚这个问题，我们不仅在完整的NEU-DET数据集上而且在每个子集上分别训练DDN。 如图所示在图11中，对于每个缺陷类别的AP，单独训练的性能通常比完整训练差。具体而言，开裂，轧制氧化皮和划痕急剧下降，而夹杂物，斑块和麻点表面呈现中等程度的下降。 这可能是因为前者需要比后者更多的数据来学习。 尽管训练数据的总量相同，但是结果却出现了巨大的差异。 我们认为，更多的训练数据可以提高CNN在特殊情况下的表示能力。也就是说，如果可以在更多的检测数据上训练DDN，则还可以改善AP。 最后，需要强调的是，其他类型的训练数据可能是无用的（例如，公共对象），因为DDN在ImageNet预训练模型上进行了微调。



**D.失败案例分析**

尽管我们的方法总体上取得了令人满意的结果，但在某些情况下，缺陷类别（如“裂纹”，“包含”，“补丁”和“累积规模”）的性能较差。 结合图7所示的成功案例，如图12所示，对一些故障案例进行分析以进行分析，并尝试探究检测结果不令人满意的原因。我们可以观察到DDN对连续的线性“裂纹”缺陷具有鲁棒性，但是在图12（a）的右下方找不到不连续的缺陷。 这意味着难以正确识别过分明显的缺陷，由于提供的缺陷数据，该缺陷并不全面。 如图12（b）所示，很难定义混乱的缺陷，甚至人眼也无法准确地将其与背景区分开。 如图12（c）所示，两种缺陷（“包含”和“斑块”）重叠，并且“包含”得到较低的分数。 毫无疑问，DDN具有处理重叠缺陷的能力和成功案例如图7（f）所示。 我们猜想原因是图中的“包含”和“补丁”相似，并且当它们非常接近时它们会相互影响。 对于“滚动比例尺”，由于这种缺陷太分散而无法定义其范围，因此边界框可能会忽略图12（d）所示的某些边缘缺陷。 由于仍存在改进的空间，因此仍需要一种更理想的缺陷检测器。

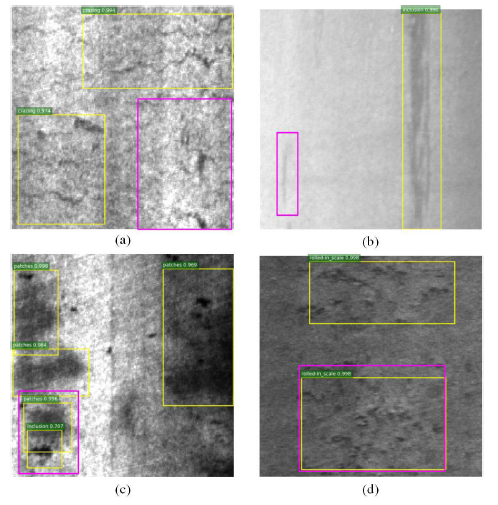


图12.失败案例的例子。 黄色框表示DDN产生的检测结果，粉红色框表示故障检测。 （a）过于明显的缺陷。 （b）令人困惑的缺陷。 （c）相似缺陷之间的干扰。 （d）范围不确定。

**七、结论**

本文提出了DDN，一种钢板缺陷检测系统。 该系统是DL网络，可以通过融合多级功能来获取缺陷的特定类别和详细位置。 对于缺陷检测任务，我们的系统可以为质量评估系统提供详细且有价值的指标，例如缺陷的数量，类别，复杂性和面积。 此外，我们建立了宝贵的缺陷检测数据集-NEU-DET。 实验表明，DDN缺陷分类任务的准确率达到99.67％，缺陷检测任务的准确率达到82.3 mAP。 此外，该系统可以以20 ft / s的检测速度运行，同时将mAP保持在70。在此功能中，我们将重点关注以下两个方向：一个是数据增强技术，这是由于检测数据集中的昂贵的手动注释所致。 另一种是利用DL技术执行缺陷分割任务，从而获得更精确的缺陷边界。