**EDRNet: Encoder-Decoder Residual Network for Salient Object Detection of Strip Steel Surface Defects**

**摘要**—由于带钢复杂的变化，包括可变的缺陷类型，杂乱的背景，低对比度和噪声干扰，检测带钢的表面缺陷仍然是一项艰巨的任务。 现有的检测方法不能有效地从复杂背景中分割出缺陷对象，实时性较差。 为了解决这些问题，我们提出了一种基于Encoder Decoder残差网络（EDRNet）的新的显着性检测方法。 在编码器阶段，我们使用全卷积神经网络提取丰富的多级缺陷特征，并融合注意力机制以加快模型的收敛速度。 然后在解码器阶段，我们分别采用信道加权块（CWB）和残差解码器块（RDB）来整合较浅层的空间特征和较深层的语义特征，并逐步恢复预测的空间显着性值。 最后，我们使用1D滤波器（RRS\_1D）设计残差细化结构，以进一步优化粗略显着性图。 与现有的显着性检测方法相比，深入监督的EDRNet可以准确地分割边界明确的完整缺陷对象，并有效滤除无关的背景噪声。 大量的实验结果证明，我们的方法始终优于最新技术，具有较大的裕度和强大的鲁棒性，并且在单个GPU上的检测效率超过27fps。

**索引词**-编码器-解码器，表面缺陷，显着物体检测，残差细化结构

**一、引言**

带钢是不可或缺的重要原料，已广泛应用于汽车，轮船和桥梁等。然而，受制造工艺和生产环境等复杂因素的影响，表面可能会出现各种缺陷。 带钢的表面，例如夹杂物，补片，划痕[1]。 这些缺陷不仅会影响产品的外观，还会对其性能和安全性产生长期不利影响。因此，带钢表面缺陷的检测是控制其质量的关键步骤。手动检查技术很难满足严格的在线检测的高精度和实时性要求。当前，基于机器视觉的自动检测方法可以解决人工检查中存在的问题（例如，效率低，劳动强度高和主观因素），因此越来越受到研究人员的关注。 此外，受人类视觉注意力机制的启发，我们引入了显着性检测[2]技术来快速检测缺陷对象。 显着性检测可以捕获图像的重要视觉信息的子集以进行进一步处理，并滤除大量多余的背景干扰。 作为一种快速的图像预处理操作，它已广泛用于各种计算机视觉任务，包括图像分割[3]，缺陷分类[46]，缺陷检查[7-12]和对象跟踪[13]。

当前，根据特征提取的不同策略，显着性检测方法主要分为两类，即基于传统模型的方法和基于深度学习的方法。 前者充分探索了手工视觉特征的内在特征，并根据不同的假设或启发式先验，开发了多种模型来检测显着物体，例如稀疏编码[14]，流形排序[15]，低秩矩阵恢复[16，17]和先验知识[18]。 但是，这些人工设计的特征描述符相对复杂，提取的特征大多是较浅的特征（例如，纹理，颜色，边缘，和对比），无法有效地表征背景复杂的复杂图像。 为了解决上述问题，基于深度学习的显着性检测方法可以在地面实况的监督下自动学习丰富而有区别的图像表示，显着提高显着性检测的性能。 尽管这些方法与常规方法相比取得了令人印象深刻的结果，但是它们的预测显着性图在对象完整性和边界保留方面仍然存在缺陷。

更具体地说，现有的显着性检测方法（例如，[17]，[18]，[19]，[20]）在准确的缺陷对象分割中仍然具有以下缺点。 首先，对于具有小的显着缺陷对象的具有挑战性的图像（如图1的第一行所示），很难从紧凑的背景中分割出完整的缺陷对象。 此外，当处理背景混乱或对比度低的情况时（图1的第二行），基于传统模型（BC [18]，SMD [17]）的方法往往会引入背景噪声，而无法均匀地突出显示整个缺陷地区。 主要原因是启发式先验或手工制作的功能几乎无法捕获有关缺陷的全局和高级语义信息。 然而，由于缺少部分对象细节，由DSS [19]和CPD [20]生成的显着图的质量下降。 对于具有精细结构的高对比度图像，现有方法易于输出边缘模糊的显着性图，如图1的最后一行所示。

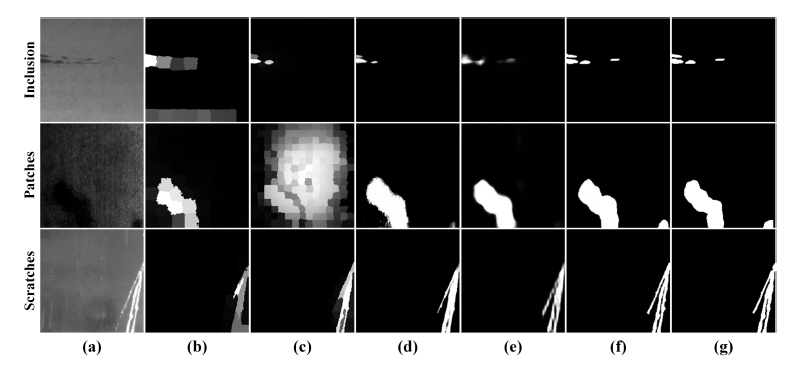


图1.我们的方法和基于传统模型（b-c）和深度学习（d-e）的方法的典型显着性检测结果。 从左到右：（a）源图像（b）BC [18]（c）SMD [17]（d）DSS [19]（e）CPD [20]（f）EDRNet（我们的）（g​​）真实情况 。

针对上述挑战，我们建立了一种基于编码器-解码器-残差网络的显着性检测方法，该方法具有较高的检测精度，较强的鲁棒性，较小的模型尺寸和较好的实时性。 在编码器阶段，通过充分利用全卷积网络（FCN）的优势来提取丰富的多尺度特征[21]。 之后，我们设计通道加权块（CWB），使模型更加关注重要的缺陷区域。 在解码器阶段，建议的残留解码器块（RDB）与CWB组合以恢复在FCN多层中编码的显着性信息。 为了进一步优化粗糙图中缺陷对象的边界细节，我们设计了一个细化网络以促进生成令人满意的结果。 另外，在深度监督机制[22]的启发下，在模型上施加了七个监督信号，以学习预测的边输出图和地面真实情况之间的残差。 这种策略不仅促进了模型训练，而且还帮助模型更好地定位了显着缺陷区域的位置。 所提出模型的整个框架如图2所示。

本文的主要贡献概述如下：

我们提出了一种新颖的端到端编码器-解码器残留网络（EDRNet），用于缺陷对象的显着性检测。EDRNet也可以使用设计的通道加权块和残差解码器块逐步从编码的多级语义特征中恢复预测的空间显着性值，从而促进检测出完整的缺陷对象并抑制非显着背景。

我们设计了一个优化网络，即带有1D过滤器（RRS\_1D）的残差优化结构。 它可用于进一步优化粗略的预测显着性图，并在缺陷对象区域上以准确的边界和紧凑的显着性输出高质量的检测结果。 此外，它采用一维滤波器具有很高的计算效率，并且由于其高度模块化的结构，可用于提高现有深度模型的性能。

在真实的具有挑战性的带钢基准数据集（SD-saliency-900 [23]）上进行的实验测试证明，我们提出的模型显着优于其他最新的显着性检测方法，并且具有强大的鲁棒性和更好的检测效率（运行速度超过27fps） 。

**二、相关工作**

在过去的二十年中，研究人员开发了巨大的代表性显着性检测方法。 在本节中，我们主要讨论两类显着对象检测，即基于传统模型的方法和基于最新深度学习的方法。

**A.基于传统模型的方法**

作为早期的杰出工作，Itti等人[24]首先提出了视觉注意的生物学上合理的架构。 该显着性模型利用具有不同特征（包括颜色，强度和方向）的中心-周围差异机制来输出注视预测图。 后来，Achanta等人 [2]重新定义了显着性，并引入了频率调谐方法来生成具有清晰定义的显着对象边界的全分辨率显着性图。 考虑到先前的边界先验[25，26]很难使突出物体接触其边界来处理这些具有挑战性的图像。 为了解决这个问题，Zhu等人[18]提出了一个鲁棒的边界连通性（BC），并将其集成到显着性优化框架中，产生了更加自信的结果。 最近，Peng等。 [17]根据低秩矩阵恢复理论构造了一个快速结构化矩阵分解（SMD）模型。该模型假定可以将图像表示为两个部分的组合：代表视觉一致背景的低秩矩阵和代表独特前景对象区域的稀疏矩阵。 黄等 [27]利用对象建议来估计在多实例学习（MIL）框架下每个超像素的显着性得分。 这种学习策略显然提高了显着性检测的准确性，但是具有很高的计算成本。 此外，周等 [28]将显着性种子散布在构造的2层稀疏图上，由于去除了多余的节点信息，提高了检测效率。 读者应参考Bori等人的调查论文 [29]有关基于传统模型的方法的更多详细信息。

本质上，传统的显着性检测方法通过使用手工制作的功能来评估显着性值，并且基于各种假设[14-17]或启发式先验[18、26]。 但是，这些浅层特征对噪声干扰很敏感。 当图像遭受照明不均匀或图像缺陷对象和背景之间存在大量相似部分时，这些假设或先验将被破坏，从而限制了它们的应用。

**B.基于深度学习的方法**

近年来，流行的全卷积网络（FCN）已成功地用于显着性检测中，并且刷新了以前所有的最新记录。 主要原因是FCN提取的多级特征具有强大的表示能力。 来自浅层的低层特征保留用于重构对象边界的空间细节，而来自深层的高层特征对语义信息进行编码，以获得对象的抽象描述。因此，罗等 [30]构建了一个具有多分辨率4×5网格结构的非局部深度特征（NLDF）网络，以集成局部对比度特征和全局信息。 受Mumford-Shah函数[31]的启发，采用包括交叉熵和边界IOU的融合损失来指导NLDF网络，以保留显着物体区域的边界细节。侯等人 [19]通过引入与整体嵌套边缘检测器（HED [22]）体系结构的跳过层的短连接，对显着对象进行了分段。 邓等 [32]开发了一个递归残差细化网络（R3Net），通过交替结合浅层和深层的特征来不断细化中间预测图。Chen等 [33]设计了反向注意模块，以自顶向下的方式指导侧输出残余显着性学习。 刘等 [34]提出了两个基于池的模块来逐步完善高级语义特征，产生细节丰富的显着图。 后来，Zhao和Wu [35]建立了金字塔特征注意网络（PFANet），以提取有效特征来定义图像的显着区域。秦等 [36]提出了一种边界感知显着性检测网络（BASNet），该网络经过混合损失训练，可以生成具有清晰边界的高质量结果。 Wu等[20]提出了一种级联的部分解码器（CPD）框架，该框架丢弃了较浅层的高分辨率特征以降低深度聚合模型的复杂性。Wang等人对深度显着性检测网络进行了更全面的文献调查[37]。

尽管基于FCN的显着性检测方法取得了显着进展，但仍有很大的改进空间，尤其是对于精细的结构片段质量和对象检测的完整性。

**三、编码器-解码器残留网络**

在本节中，我们首先详细介绍拟议的编码器-解码器残留网络（EDRNet）中的三个主要子模块。 整个网络体系结构以单个图像作为输入，并以端到端的方式输出缺陷对象的相应预测显着性图。 我们进一步阐述了构造的融合损失，用于指导网络均匀地突出显示具有整个结构和精美边界的显着物体区域。

**A.编码器**

网络受U-Net [38]的网络体系结构的启发，我们以编码器-解码器样式开发了显着性检测方法。 这种网络样式可以有效地集成低级富裕空间细节和高级上下文信息，从而有助于精确预测缺陷对象的位置。 对于编码器网络，与大多数以前使用VGG-16 [42]模型提取不同分辨率的特征图的方法[19、20、30、35、39-41]不同，我们选择ResNet-34 [43]作为主干特征提取器。做出此选择的原因有两个。 一方面，通过在VGG网络采用的普通网络（简单地逐一堆叠卷积层）上使用跳过层连接（即通过快捷方式进行身份映射），可以更轻松地优化残差学习框架。 另一方面，残余结构易于实现更深的网络，并且仍然具有较低的复杂度。 以这种方式，由于接收场扩大，即覆盖更多的上下文信息，模型可以从深度显着增加的深度中获得准确性增益。如图2所示，整个编码器部分包含一个输入卷积层，来自ResNet-34的四个残差学习块（分别为conv2-3，conv3-4，conv4-6和conv53）和一个桥接模块。 而且，与原始的ResNet-34不同，我们的输入层具有64个通道，内核大小为3×3，步幅为1，而不是内核大小为7×7，步幅为2。然后，在步幅的尾部添加步幅2的最大池化操作。 输入层进一步扩大了接收场的大小。 我们执行这种手术来编码更精细的空间细节，并分别在合并操作之前和之后捕获更好的过滤器响应。 类似于[44]，每个卷积输出被馈入一个批处理归一化[45]层，以平衡特征的尺度，然后是ReLU [46]激活函数，以增强非线性表示能力。

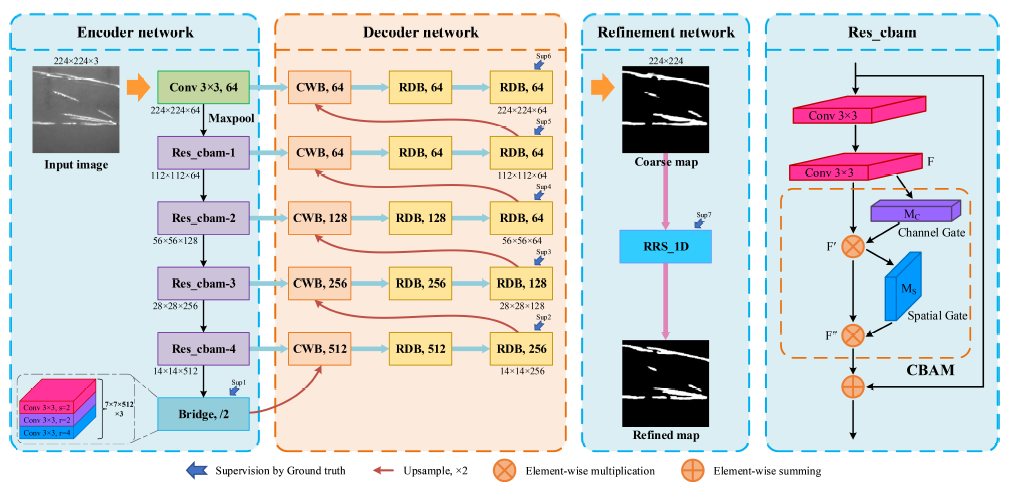


图2.我们提议的EDRNet的示意图。 在编码器网络中，我们使用ResNet-34 [43]骨干排除了输入层和最终分类层，以提取丰富的多层功能。 此外，我们将卷积块注意模块（CBAM）融合到ResNet的每个残差基本块中，以加速EDRNet的收敛，表示为Res\_cbam-i（i∈{1，2,3，4}）。 Res\_cbam的详细结构显示在最右侧。 特征图的大小定义为高度×宽度×通道（例如224×224×64）。 在解码器网络中，CWB表示信道加权块，RDB表示残留解码器块。 在优化网络中，RRS\_1D表示带有一维滤波器的残余优化结构。

形式上，给定输入图像，其中H，W，C分别表示其高度，宽度和通道数。我们可以在分辨率为的六个级别上抽象为的多尺度特征。考虑到注意力机制具有强大的学习准确和紧凑特征的能力，并且由于其有效性和效率而被广泛用于各种计算机视觉任务中。然后由Chen等人的工作激发 [47]，我们将轻量级的卷积块注意模块（CBAM [48]）嵌入到ResNet-34的每个残差基本块中，以进行快速收敛，表示为Res\_cbam-i（i∈{1,2,3,4}）。Res\_cbam的详细结构显示在图2的最右边。图3直观地显示了编码器网络中带有/不带有CBAM的EDRNet的训练损失和预测损失。 通过引入CBAM的关注机制，我们可以清楚地观察到EDRNet更易于优化，并以更快的收敛速度有效地减少了训练错误。 此外，这两种加工方式的定量比较示于表Ⅰ的Ⅳ-B部分。

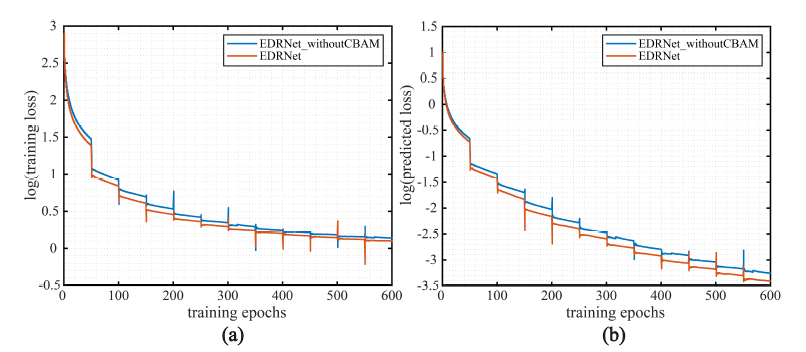


图3. SD-saliency-900数据集上的训练曲线。 （a）训练损失代表总损失，其计算方法是将七个基本输出的所有损失相加，并由地面事实监督。 （b）预测的损失代表了改进的显着性图和地面真实情况之间的误差。

由于显着物体检测需要对均匀区域进行完全分割，这比仅需要简单梯度信息的边缘检测要困难得多。为此，我们在编码器网络的末端设计了一个附加的桥接模块，以进一步捕获全局上下文感知信息，这些信息适合准确定位缺陷对象的区域。该模块由三个512通道3×3卷积层组成带有扩张[49]。其中r表示膨胀率。 为了保持与原始ResNet-34相同的特征图分辨率，第一个卷积层的步长为2降采样。然后在这些卷积层中的每一个之后进行批归一化[45]和ReLU [46]激活功能。

**B.解码器网络**

在解码器网络中，我们可以选择使用建议的信道加权块（CWB）和残差解码器块（RDB）逐渐恢复以先前的多尺度特征编码的显着性信息。 以下小节专门阐述如何构建这两个模块。

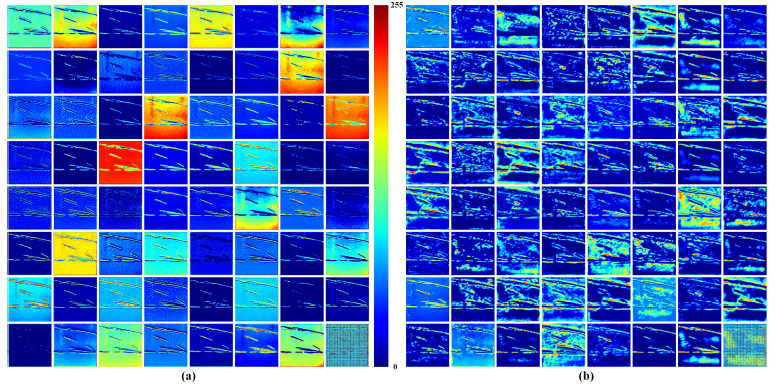
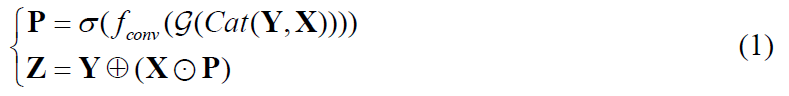


图4.信道加权块（CWB）的影响示意图。 （a）和（b）分别表示在不使用和使用CWB处理的情况下从Res\_cbam-1提取的特征热图。

**1）通道加权块**

如图4（a）所示，直接从编码器网络生成的特征图更多地集中在非突出背景区域（即发热量较高的区域）上。 主要原因是未充分考虑全局上下文信息，从而导致不正确的显着性预测结果。 为了解决这个问题，我们设计了通道加权块（CWB）来捕获更有效的特征区域并滤除背景噪声干扰因素。 在对提出的CWB进行处理之后，该模型将更多地关注缺陷对象区域或其边缘，这可以从图4（b）直观地观察到。更具体地说，CWB的输入是当前编码特征X和来自下一解码器级的输出特征Y的级联特征图。 此外，我们将特征Y上采样2倍，以保持与X相同的分辨率。此后，将全局平均池应用于上采样连接特征，以获取信道方向向量，目的是学习全局上下文感知信息并增强其一致性。之后，我们构造了一个具有两个1×1卷积层和一个嵌入式非线性PReLU [50]激活函数的瓶颈结构，以限制模型的复杂性并有助于推广。 然后，我们对先前的输出采用S型函数，以得到权重向量P，该向量归一化为[0,1]。最后，我们根据残差学习的优势获得CWB的最终输出。 它是通过将修改后的特征X和初始深度特征Y相加而得出的。CWB的详细结构如图5（a）所示。 形式上，信道加权块Z定义如下：



其中Cat表示上采样串联操作。表示全球平均池。被称为特征融合的瓶颈结构。σ是SIGMOD激活函数。表示逐元素乘法。 ⊕是逐元素求和运算。

**2）残留解码器块**

为了逐渐恢复以前的多级特征中编码的显着性信息，我们构造了残留解码器块（RDB），如图5（b）所示。 RDB的结构基本上类似于ResNet-34 [43]的原始残差基本块（RBB），以实现较低的计算复杂度并促进优化。 与RBB不同，我们在两个3×3卷积层之间实现了信道混洗[51]操作。 这样的操作可以增强泛化能力并挖掘出更多潜在的显着性信息。 此外，值得注意的是，信道混洗不包含任何其他复杂的操作，例如ShuffleNet [51]中使用的组卷积或深度卷积，可以实现更高的检测效率并降低模型复杂度。 像大多数现代方法一样，两个3×3卷积层中的每一个都紧跟着批量归一化[45]和PReLU [50]激活函数。 另外，出于两个原因，我们添加了一个额外的1×1卷积层。 一方面，它可以提高跨渠道信息交互的能力。 另一方面，它用于减小尺寸。 此外，为了证明RDB的合理性，我们将其与RBB和普通卷积块（PCB）进行了比较。 具体分析见Ⅳ-B节表Ⅰ。

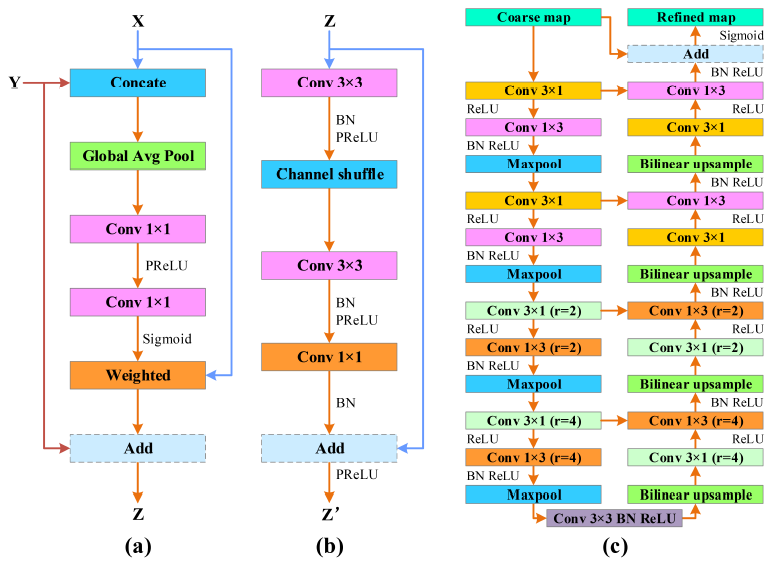


图5. EDRNet中提出的基本组件的详细结构。（a）CWB（信道加权块）（b）RDB（残余解码器块）（c）RRS\_1D（具有一维滤波器的残余细化结构）

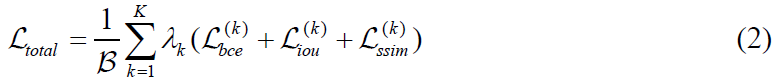
**C.残留优化网络**

在训练过程中，我们采用深度监督机制[22]在EDRNet中每个阶段的最后卷积层上的特征上施加监督信号（地面真相）（详细信息见图2）。以便于有用信息传播到缺陷对象区域。 因此，给定一个输入图像，我们在Encoder-Decoder网络中获得六个侧面输出显着性图。 每个侧面输出显着度图都是通过一个内核大小为3×3的单通道卷积层生成的，其后是一个双线性上采样层（用于将相同分辨率与输入图像进行匹配）和一个Sigmoid激活函数，用于将预测值映射到[0 ，1]。考虑到最后的侧面输出显着性图（如图2的粗图所示）具有更高的检测精度和更丰富的显着性信息，我们将其作为细化网络的输入。 我们设计残差细化网络的原因是为了进一步优化粗糙地图中缺少的边界和区域细节。如图5（c）所示，我们提出了带有一维滤波器[52]（RRS\_1D）的残差细化结构，该结构也遵循编码器-解码器样式。 编码器和解码器部分都由四个级组成，每个级都包含两个卷积层，后跟一个最大池化或双线性上采样单元。 桥接部分具有64通道3×3卷积层，然后进行批量归一化[45]和ReLU [46]激活函数。最大池用于下采样的地方，可以实现更深的网络并有助于减少计算量。 双线性插值用于上采样，以匹配特征尺寸。 为避免占大多数计算复杂度的逐点卷积，我们采用了两个专门的1D滤波器（即3×1和1×3卷积），而不是使用3×3卷积，从而有效地平衡了细化性能和计算效率。此外，使用扩张卷积[49]（扩张率r = 2,4）使我们的网络可以获得更大的接收场，从而提高了准确性。 与使用更大的内核大小相比，已证明该技术在计算成本和参数方面更有效。 为了便于训练，堆叠的输出通过身份映射的分支与输入的粗略映射相加。 然后，将经过S型映射后的精炼图作为我们EDRNet的最终结果显着图。

**D.损失函数**

由于显着物体检测从本质上也可以视为密集的二进制分类问题，因此其输出表示每个像素成为前景物体的概率得分。 因此，大多数先前的方法[19、20、32、35、47、53]始终使用交叉熵（通常应用于分类任务）作为训练损失。 但是，这种简单的策略很难指导网络捕获显着对象的全局结构信息，从而导致边界模糊或检测结果不完整。 为了克服这个问题，受到秦等人的启发。 [36]，我们构造了一种融合损失来监督网络的训练过程，以学习边界位置和结构捕获方面的更多详细信息。

更具体地说，聚变损失包括三个部分，包括二进制交叉熵（BCE [54]）损失，联合边界交集（boundary IOU [55]）损失和结构相似性（SSIM [56]）损失。 这样，我们网络的总损耗被定义为每边输出和地面真相之间所有融合损耗的加权总和。 因此，总损失合计公式为：



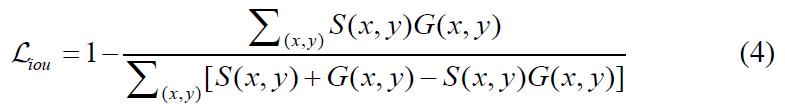
其中表示批次大小。K是深度监督的显着性图的总数，在本文中设置为7。λk表示第k个边输出融合损耗的权重，根据经验将其设置为1。分别表示BCE [54]丢失，边界IOU [55]丢失和SSIM [56]丢失。

如上所述，BCE [54]损失广泛用于二进制分类或分段任务中，以反映预测值和实际标签之间的概率分布差异。 它定义为：



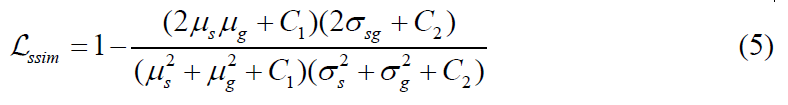
其中，G（x，y）∈{0,1}（0表示背景，1表示缺陷对象）表示地面真实情况中位置（x，y）的像素的标签，而S（x，y）∈[0,1]表示在相应像素处成为显着物体的预测概率。在此，，。

由于BCE [54]的损失是逐像素的，因此会不加选择地估算所有像素的总体分类精度，即在前景像素和背景像素上分配相等的权重。 难以解决缺陷对象检测中常见的类不平衡问题。 考虑到这个问题，我们进一步采用边界IOU [55]的损失来惩罚不准确的分类，并增强区域一致性和边界响应。 它定义为：



是可微的，损失计算过程是端到端可训练的。

SSIM [56]最初用于图像质量评估，用于测量预测的显着性图和地面实况之间的结构相似性。 因此，它被整合到我们的训练损失中，以指导网络从地面真相中学习缺陷对象的结构信息。令和分别是分别从侧面输出显着性图S和地面实况G裁剪而来的两个对应图像块（大小：N×N）的离散像素值。 s和g的SSIM定义为：



其中和分别是s和g的均值和方差。σsg是相应的协方差。 C1和C2为了避免被零除，它们是小的常数，并根据经验分别设置为。

**IV实验**

**A.实验设置**

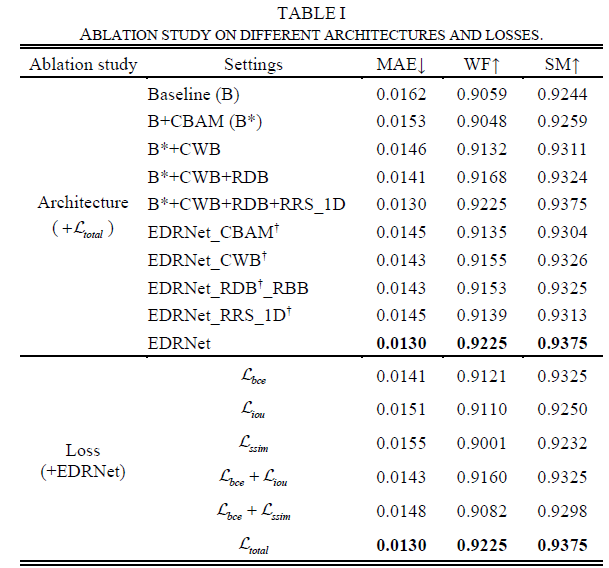
为了验证我们提出的模型的有效性和鲁棒性，我们对具有挑战性的公共带钢基准数据集SD-saliency-900 [23]进行了各种实验。该数据集包含总共900张裁剪后的图像，其中包含三种类型的缺陷，包括“包含”，“补丁”和“划痕”。此外，SD-saliency-900还具有Song等人提供的相应的像素级地面实况注释 [23]。每个图像的分辨率为200×200像素。

**1）评估指标**

我们使用几种普遍认可的指标评估了各种显着性检测方法的性能，这些指标包括精确召回（PR）曲线，F度量曲线，平均绝对误差（MAE），加权F度量（WF）得分，重叠率（OR）和结构量度（SM）。 给定一个显着图，我们通过比较二进制掩码序列和相应的地面真相来计算精度和召回率。 通过将阈值从0更改为255的显着性图进行二值化获得这些二进制掩码。F-measure [2]定义为精确度和召回率的加权谐波平均值，用于全面评估显着性图的质量。MAE [57]测量归一化显着图S与地面真实性G之间的差异，定义为 WF [58]是传统的广泛使用的F量度的加权版本。通过考虑邻域信息，它为不同位置的不同错误分配了不同的权重（w）。 因此，WF修正了当前使用的度量的内插，依赖和等重要性缺陷。定义为，其中β2默认设置为1。OR是分割后的对象蒙版S'与地面真实情况G之间的重叠率，定义为OR = |S'∩G| / |S'∪G|，其中S'是使用自适应阈值通过二值化S获得的，即 ，是[17]中S平均值的两倍。为了更好地评估检测结果，我们引入了最近提出的SM [59]指标，该指标同时考虑了S和G之间的对象感知和区域感知结构相似性。考虑到上述指标无法直观地表明分割蒙版的边界质量， 因此，我们介绍了普拉特的品质因数（PFOM [60]）来解决这个问题。PROM被广泛用于边缘检测评估中，其定义为PROM = 。其中NG和NS分别是从地面真值和二进制显着性掩码中提取的理想和实际边缘点的数量。 α是缩放常数，设置为0.1或1/9。 dk是第k个理想边缘点和相应的检测到的边缘点之间的欧几里得距离。

**2）参数设置和实现细节**

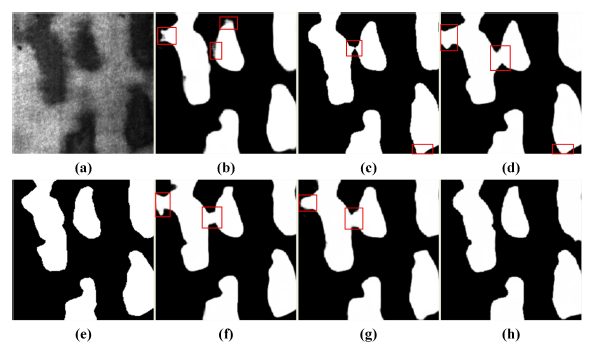
我们基于带有CuDNN后端的Pytorch框架和RTX 2080 GPU用于加速来实现建议的模型。为了训练各种深度模型，我们构建了一个标准的统一训练集，其中包含810张图像。该训练集包括从SD-saliency900数据集中随机选择的540个源图像（每种缺陷类型为180张图像）和盐和胡椒噪声（ρ= 20％）干扰的270个图像（每种缺陷类型为90张图像）。 这些噪声图像的源图像是从之前的540张图像中随机收集的，这有助于增强网络的鲁棒性。在训练期间，首先将每个图像I调整为256×256大小，并随机裁剪为224×224，然后通过（I-μ）/σ进行归一化。 其中，平均μ和标准偏差σ分别设置为0.4669和0.2437。此外，在训练损失收敛之前，我们不会使用任何验证集并训练模型。 我们采用He初始化[50]策略来初始化特征提取网络Res\_cbam的参数，而不是使用ResNet-34的预训练模型。而其他层是使用Pytorch的默认设置初始化的。 整个网络以端到端的方式训练。 我们利用RMSprop [61]优化器训练我们的网络，其超参数设置如下：学习率lr = 1e-3和alpha = 0.9，而其他参数固定为默认值。 训练过程大约需要7.5小时，并且在大约60K迭代之后以8的批量大小收敛。在测试时，每个图像都简单地调整为256×256的大小并馈入网络以获得其显着性图。之后，通过使用双线性插值来调整显着图的大小，以保持与输入图像相同的分辨率（200×200）。 源代码和实验结果将在https://github.com/950216/EDRNet上发布，以备将来比较。

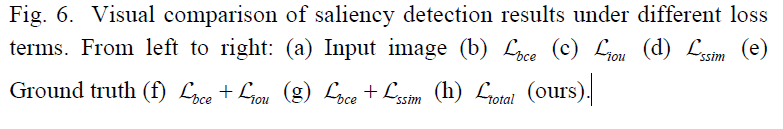


上标“†”表示没有此模块的网络体系结构。 EDRNet\_CBAM†意味着使用EDRNet作为训练模型，但删除了CBAM。 “ \_RDB†\_RBB”表示用剩余的基本块（RBB）替换RDB。 “ \_RDB†\_PCB”表示用普通卷积块（PCB，Conv + BN + ReLU）替换RDB。 基线模型仅由基本的Encoder-Decoder网络构建而成，不包含任何建议的模块。其中B \* + CWB + RDB + RRS\_1D等于EDRNet。

**B.消融研究**

在本小节中，我们进行了消融实验，以验证提议的EDRNet模型中使用的每个关键组件的有效性。 该消融研究包括架构分析和损耗分析。 所有实验均在SD-saliency-900数据集上进行。





**1）架构分析**

为了证明EDRNet模型中关键组件的有效性，我们使用MAE，WF和SM指标报告了模型对相关架构的定量评估结果，如表Ⅰ所示。 我们可以发现，每次递增添加关键组件时，模型的性能也会相应提高。 包含所有组件（即卷积块注意模块（CBAM），通道加权块（CWB），残差解码器块（RDB）和带有1D滤波器的残差细化结构（RRS\_1D））的模型可实现最佳性能。 与基准模型相比，EDRNet的MAE大幅降低了19.8％，WF和SM度量分别提高了1.8％和1.4％。 它客观地证明了所有关键组件对于所提出的模型以获得最佳缺陷对象检测结果都是有用和必要的。

**2）损失分析**

为了验证构造的融合损失的合理性，我们基于EDRNet架构进行了一系列具有不同损失项的实验。 如表Ⅰ的最后一行所示，具有融合损耗Ltotal的EDRNet获得了卓越的性能。与采用广泛使用的交叉熵损失Lbce相比，WF和SM分别增加了1.14％和0.54％，而MAE减少了7.8％。 为了进一步直观地说明不同损耗的影响，我们给出了视觉比较结果，如图6所示。我们发现使用Ltotal可以很好地保留缺陷对象的边界，并且可以有效地抑制背景噪声的干扰。 。 因此，这些结果也表明融合损失有助于指导我们的模型学习边界定位中更详细和正确的信息。

**C.与最新技术的比较**

将提议的EDRNet模型与基于传统模型或深度学习的十四种最新显着性检测方法进行比较，包括RCRR [62]，2LSG [28]，BC [18]，SMD [17]，MIL [27]，PFANet [35]，NLDF [30]，DSS [19]，R3Net [32]，BMPM [63]，PoolNet [34]，PiCANet [41]，CPD [20]和BASNet [36]。 为了公平起见，我们使用默认参数运行作者发布的源代码或可执行文件。 值得一提的是，所有比较的深度学习模型都是在与我们相同的训练集上进行训练的。

1）视觉比较

图7提供了我们的方法与竞争对手在不同输入图像上的视觉比较。 我们可以观察到，所提出的模型实现了最佳的显着性检测结果，在各种挑战性案例中，这些结果都非常接近相应的基本事实。 具体来说，对于对比度低且前景混乱的图像（例如第2行和第3行），大多数以前的方法错误地将某些背景区域视为缺陷对象，而我们的模型可以有效消除背景噪声的干扰而不会出现任何错误检测 。 即使对于带有明显缺陷对象的简单图像（例如，第1和8行），大多数现有方法都无法捕获整个显着对象，但是我们的模型统一地突出显示了它。 当图像具有复杂的背景时（例如第4、5、6行），大多数竞争方法在轮廓模糊或凸显色块分散的情况下产生的效果很差。 相比之下，我们的模型可以有效地从复杂背景中识别出缺陷对象，并生成高对比度显着性图。最后，对于具有复杂对象边界和精细结构的图像（例如，第7行和第9行），借助所提出的RRS\_1D，只有我们的模型才能以明确定义的边界准确地分割整个缺陷对象。 以上所有结果证明了所提出的EDRNet模型的合理性和优越性。

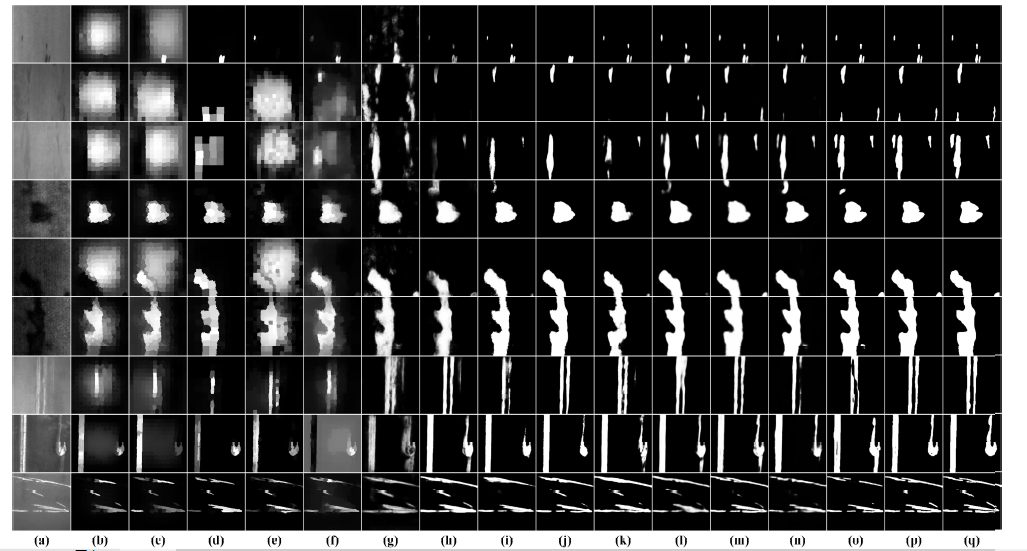
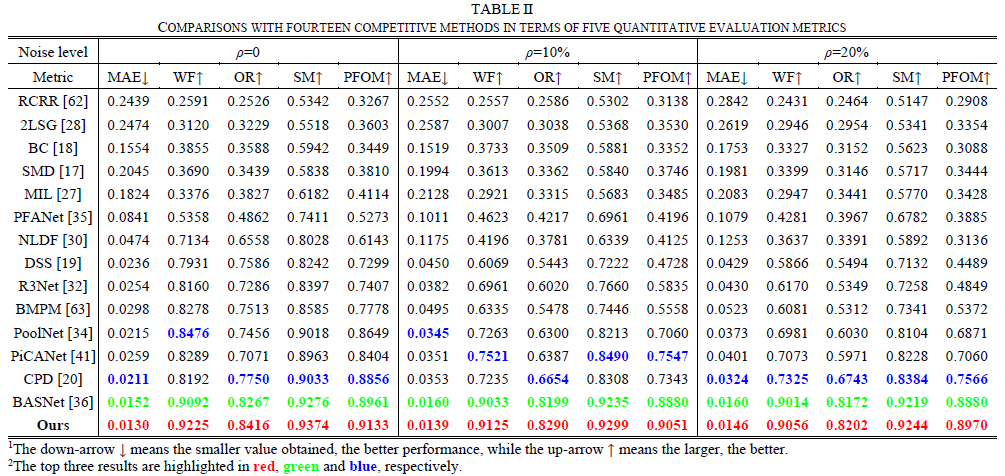


图7.显着图的视觉比较。 （a）源图像（b）RCRR [62]（c）2LSG [28]（d）BC [18]（e）SMD [17]（f）MIL [27]（g）PFANet [35]（h） NLDF [30]（i）DSS [19]（j）R3Net [32]（k）BMPM [63]（l）PoolNet [34]（m）PiCANet [41]（n）CPD [20]（o）BASNet [36]（p）EDRNet（我们的）（m）基本事实。



**2）定量比较**

为了定量评估分割缺陷对象的质量，我们在图8（a）中针对14种最新方法展示了模型的性能。 我们可以看到，在PR和F测度曲线方面，所提出的模型始终优于所有竞争方法。特别是，我们的模型的精度在较大的阈值范围内仍保持在90％以上。 我们进一步分析了各种显着性检测方法在严重的盐和胡椒噪声干扰下的鲁棒性，如图8（b-c）所示。 显然，我们的模型在所有评估曲线上都大大优于现有方法。 即使对于目前最好的BASNet [36]方法，我们提出的模型仍然具有轻微的优势。 因此，这些结果令人信服地验证了所提出模型的有效性和鲁棒性。 另外，表Ⅱ详细总结了全部十五种方法的五个定量评价指标。 它客观地表明，在所有情况下，我们的模型均在检测精度方面获得最佳性能。 特别是，与现有的CPD [20]方法相比，该模型在WF，OR，SM和PFOM指标上分别平均提高了12.6％，8.6％，3.8％，3.1％，并且大幅下降了38.4％ 在MAE指标上。 此外，与最新的BASNet方法相比，我们的模型还使MAE大幅降低了14.5％。 对于边界质量，相对于BASNet方法，我们的模型在PROM指标方面获得了1.9％的收益。事实证明，对于真实场景中的应用，我们的模型是更好的选择。

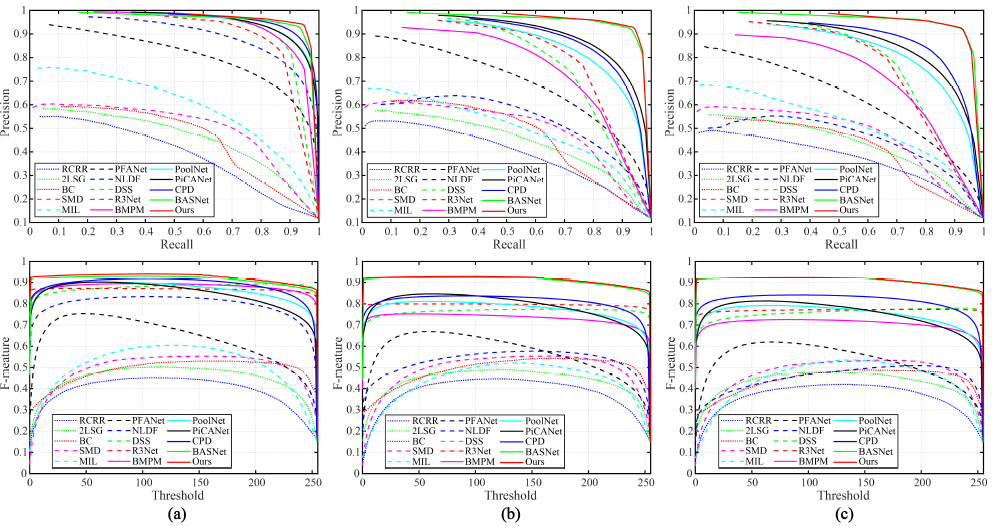
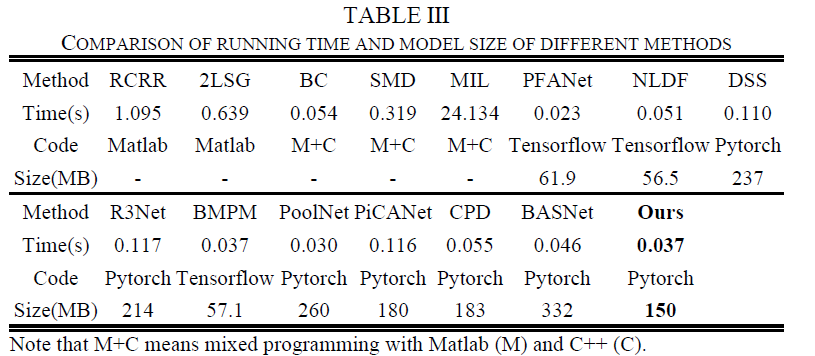


图8.定量评估曲线以及抗盐和胡椒噪声的鲁棒性。 从（a）到（c）的噪声级别分别为0、10％和20％。 从上到下的相应曲线分别是PR曲线和F测度曲线。

**3）运行效率**

我们总结了在SDsaliency900数据集上执行的竞争显着性检测方法的平均运行时间和模型大小，如表Ⅲ所示。 以上所有实验均在配备Intel Core i7-9700K 3.6GHz CPU，32GB RAM和RTX 2080 GPU的PC上进行了测试。 对于200×200的测试图像，我们的方法仅需0.037秒即可生成最终的显着图，而无需任何其他后处理。 它比大多数竞争方法都快得多，仅比PFANet [35]和PoolNet [34]慢。 特别是，BASNet [36]和我们的模型均获得了明显优越的检测结果，但BASNet具有最大的模型大小。 相比之下，我们的模型尺寸小于其尺寸的一半。 以上结果表明，我们的EDRNet模型可以更好地满足当前对实时性和高准确度的要求，并且仍有进一步提高效率的空间，可用于将来的工作。



**五、结论**

在本文中，我们提出了一种新型的端到端编码器-解码器残差网络（EDRNet），以有效地检测带钢的表面缺陷。 通过结合深入的监管机制和融合损失，EDRNet模型能够捕获精细的细节，并使网络更易于优化。 所提出的模型可以直接输出高质量的显着图，几乎与相应的地面真实情况相近。 生成的显着性图以清晰的边界均匀地突出显示缺陷对象，同时有效滤除背景噪声。 大量实验证明，与SD-saliency-900数据集上的其他十四种最新方法相比，我们的EDRNet模型在七个评估指标上均达到了最佳性能，并且具有很强的鲁棒性。 此外，我们的EDRNet不需要任何后处理，并且可以在单个GPU上以27 fps的实时速度运行。