

DOI: 10.19650/j.cnki.cjsi.J2514319

基于 CLIP 的无监督大尺寸 DR 图像增强算法研究*

陈明飞¹, 廖 望^{2,3}, 王广文^{2,3}, 吴义顺^{2,3}, 沈 宽^{2,3}

(1. 中国航发哈尔滨东安发动机有限公司 哈尔滨 150066; 2. 重庆大学 ICT 研究中心 重庆 400044;
3. 重庆大学光电工程学院 重庆 400044)

摘 要: X 射线数字成像技术 (DR) 已广泛应用于工业无损检测领域。然而在实际应用中, 存在大量的结构不规则、厚度变化范围大的工件。DR 检测容易在工件厚度厚的地方呈现曝光不足, 在工件厚度薄的地方出现曝光过度的现象, 导致 DR 扫描图像质量差, 结构信息严重缺失。伴随探测器像素矩阵增至 4 K×4 K 以上, 大多数算法难以在消费级设备上处理 DR 扫描产生的大尺寸图像。同时工业检测场景难以获取大量的成对配对标签。针对大尺寸 DR 推理与标签稀缺的问题, 将对比语言-图像预训练 (CLIP) 与限制对比度自适应直方图均衡化 (CLAHE) 先验耦合, 构建两阶段轻量级无监督增强框架, 全程无需配对或分块操作。该算法在第 1 阶段初步学习提示向量引导冻结的 CLIP 图像编码器, 通过 CLIP 增强损失、结构一致性损失, 以及 CLAHE 特征图感知损失指导训练; 第 2 阶段通过排序损失迭代对提示权重值迭代精炼, 交替更新增强网络直至视觉收敛。实验结果表明, 与同期的无监督算法相比, 峰值信噪比 (PSNR)、感知图像相似度 (LPIPS)、结构相似度 (SSIM) 等指标分别提高 1.0 dB、1.6% 和 2.0%, 在多个参考指标优于对比方法。在推理时只需加载 0.279 M 的参数, 处理 5 732×2 333 的大尺寸图像单张耗时约 1.5 s。仅用 380 张铸件图像训练的模型即可直接在未训练过的碳纤维线路板以及其他材料的物品上迁移, 并展现出良好的泛化能力, 为工业检测落地提供实时增强方案。

关键词: 大尺寸 DR 图像; CLIP; CLAHE; 损失函数; 图像增强

中图分类号: TH878.1 **文献标识码:** A **国家标准学科分类代码:** 510.40

Research on unsupervised large-size DR images enhancement algorithm based on CLIP

Chen Mingfei¹, Liao Wang^{2,3}, Wang Guangwen^{2,3}, Wu Yishun^{2,3}, Shen Kuan^{2,3}

(1. AECC Harbin Dongan Engine Co., Ltd., Harbin 150066, China; 2. Industrial Computed Tomography Research Center, Chongqing University, Chongqing 400044, China; 3. College of Optoelectronic Engineering, Chongqing University, Chongqing 400044, China)

Abstract: X-ray digital radiography (DR) has been widely used in the industrial nondestructive testing. However, there will be a large number of workpieces with irregular structures and large thickness variations in the practical applications, which cause the DR detection prone to the underexposure in thick parts and overexposure in thin parts. On one hand, for the detector pixel array of 4 K×4 K, most algorithms can hardly handle these large DR scan images with the customer-level devices. On the other hand it is difficult to obtain a large number of paired labels for the industrial inspection. To address the problem of large-size DR inference and label scarcity, a lightweight unsupervised enhancement framework is proposed by coupling the contrastive language-image pretraining (CLIP) vision-language model with contrast-limited adaptive histogram equalization (CLAHE) priors. The first stage learns prompt vectors to guide a frozen CLIP image encoder with the CLIP enhancement loss, structural consistency loss, and CLAHE feature-map perception loss. The second stage refines the prompts iteratively through the ranking loss and alternately updates the enhancement network until the visual convergence. Experimental results show that peak signal-to-noise ratio (PSNR), learned perceptual image patch similarity (LPIPS), and structural similarity (SSIM) are improved by 1.0 dB, 1.6%, and 2.0%, respectively, outperforming other unsupervised algorithms on multiple metrics. Additionally the inference needs only 0.279 M parameters and processes a 5 732×2 333 image in 1.5 s. Furthermore the model trained with merely 380 casting images generalizes directly to unseen carbon-fiber circuit boards and other materials,

收稿日期: 2025-07-25 Received Date: 2025-07-25

* 基金项目: 国家重点研发计划 (2022YFF0706400) 资助

demonstrating the strong potential for industrial deployments.

Keywords: large-size DR image; CLIP; CLAHE; loss function; image enhancement

0 引言

X 射线数字成像技术 (digital radiography, DR) 已广泛应用于工业无损检测领域^[1], DR 技术成像具有空间分辨率高、信息量大、动态范围大等优点^[2]。更为重要的是其摄影速度快, 能够快速有效地检测出铸件表面及内部缺陷, 实现实时成像检测。然而, 对于一些结构不规则、厚度变化范围大的复杂工件, 单一能量的 X 射线无法实现对整个工件同时曝光成像, 容易在有效厚度厚的地方呈现曝光不足, 在有效厚度薄的地方出现曝光过度的现象, 导致被测工件的 DR 图像质量差、空间分辨率低、结构信息严重缺失^[3]。

近年来深度学习技术在图像增强领域大放异彩, 其中, Uformer^[4] 构建了一个 U 形网络, 在图像增强任务中表现良好。在 Restormer^[5] 中, 修改了 Transformer 块, 以提供更好的图像恢复长距离依赖性特征。在 LLFormer^[6] 中, 改进了 Transformer 中的注意力机制, 以捕捉上级图像特征。然而, Transformer 由于其极高的算法复杂度和大量 GPU 内存占用导致在显存有限情况下只能采取分块推理图像的方法。这不仅导致推理速度的急剧下滑也容易出现分块效应。

然而, 基于有监督算法实现图像增强始终具有很大的局限性, 实际场景中配对数据集的获取较为困难, 促使无监督算法相继提出^[7-10]。EnlightenGAN^[11] 使用生成对抗网络无监督地训练光照增强模型, 可以将过曝光/弱光的图像增强为正常的光照效果。零参考深度曲线估计 (zero reference deep curve estimation, ZeroDCE)^[12] 提出了一种轻量无监督的网络, 该方法将图像增强表述为一个关于各像素的值的二次曲线, 网络预测参数图迭代地施加在每个颜色通道上, 从而实现增强。Xiong 等^[13] 提出两阶段的增强, 第 1 阶段实现对对比度增强, 第 2 阶段实现去噪。Zeng 等^[14] 提出利用 3D Look Up Table 来解决较大图像的实时高效增强, 一定程度上解决了资源消耗较大的问题。对比语言-图像预训练 (contrastive language-image pretraining, CLIP) 模型^[15] 凭借从大规模图像-文本数据中学习到的知识, 在零样本无监督分类任务中表现出了显著的性能, 其通用性在高级任务中得到了验证^[16]。目前研究表明 CLIP 中丰富视觉语言先验可以用于以零样本方式评估图像的质量和抽象感知, 并且可以利用 CLIP 进行背光图像增强。提示学习作为视觉与语言模型的核心, 是最近新兴的研究方向。CoOp^[17] 将提示学习引入到视觉-语言模型的适应中, 应用于下游视觉任

务。CoCoOp^[18] 通过允许每个输入实例条件化提示, 而不是固定一次学习, 进一步提高了模型的泛化能力。现有的提示学习方法主要集中于为高级视觉任务获取更好的提示信息。基于 CLIP 的无监督背光增强 (CLIP-LIT)^[19] 方法利用提示学习提取更准确的低级图像表示 (颜色、曝光和饱和度), 同时忽略 CLIP 中的高级语义信息, 进行无监督的图像增强。

然而目前深度学习网络在消费级设备 GPU 内存不足的场景, 对高分辨率的 DR 图像通常无法处理, 且对配对数据具有高度依赖性。限制对比度自适应直方图均衡化 (contrast limited adaptive histogram equalization, CLAHE) 等传统增强算法^[20-23] 无需配对数据即可增强, 在工业 DR 领域已表现出独特优势。着眼于这一问题, 提出基于 CLIP 与 CLAHE 的无监督增强框架。实验结果表明, 在推理时只需加载 0.279 M 的参数使得推理 4 K 图像非常简单, 为工业落地提供支持。

1 方 法

1.1 基于 CLIP 的无监督增强算法框架

无监督算法包括两个阶段。在第 1 阶段, 通过限制提示对和 CLIP 嵌入空间中相应图像之间的文本图像相似性来学习初始提示对 (负/正提示指的是曝光异常/曝光正常的图像)。对于初始提示对, 使用冻结的 CLIP 模型来计算提示和增强结果之间的文本图像相似性。换言之, 提示对 (Prompts) 设计为可学习权重向量, 通过 CLIP 嵌入空间中相应图像之间的文本-图像相似性做损失函数进行对比学习。此时 CLIP 的编码器参数是冻结的, 反向传播时让网络学习 Prompts (学习 Prompts 时无需配对图像)。Prompts 学习完成之后将冻结 Prompts 的参数, 此过程如图 1 所示。然后设计一个增强网络, 网络会接收一个曝光异常的图像, 使用 3 个损失函数让网络学习如何输出一个曝光正常的图像, 通过这 3 个损失函数反向传播更新这个增强网络的模型参数即可, 此过程如图 2 所示。

另外, 为了进一步提高增强网络的性能, 设计了如图 3 所示的第 2 阶段, 通过排序学习利用曝光异常、增强结果和曝光正常的图像来细化可学习的提示 (此时依然无需配对图像)。精炼的提示可用于微调增强网络以进一步提高性能。接着交替进行提示细化和增强网络的微调, 直到获得视觉上满意的结果。

值得说明的是, CLIP 模型在整个学习过程中保持不变, 并且除了提示词初始化和细化之外, 不会引入任何

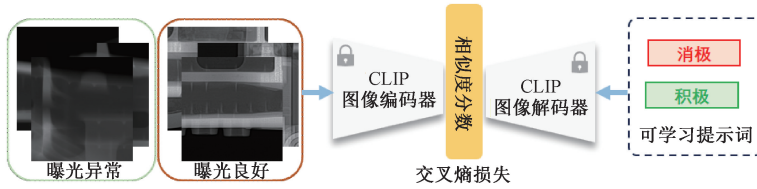


图 1 学习 Prompts 的过程

Fig. 1 The process of learning Prompts

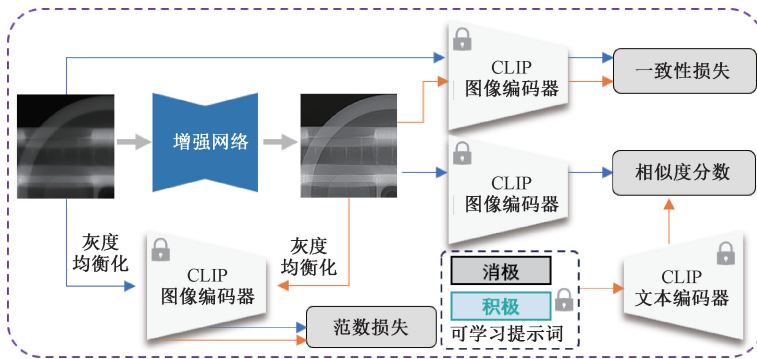


图 2 通过 3 个损失函数指导学习增强网络的过程

Fig. 2 The learning process of enhancing the network guided by three loss functions

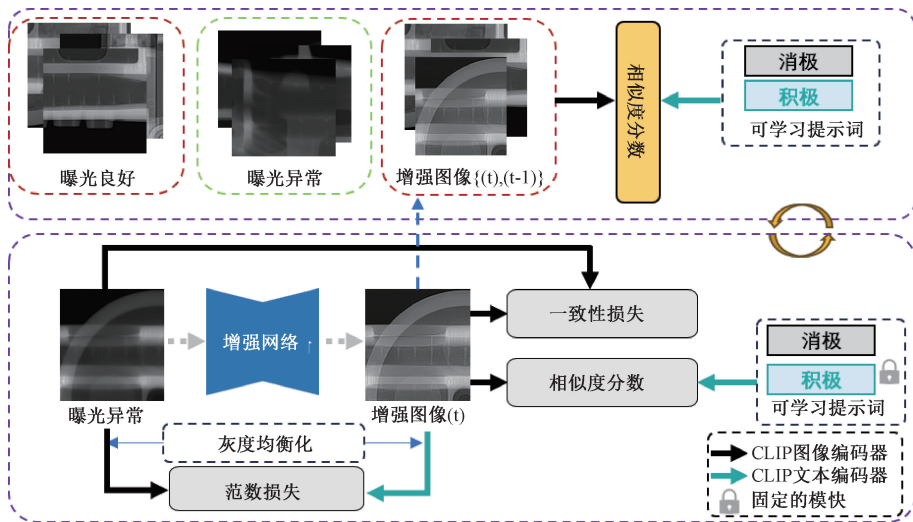


图 3 第 2 阶段: Prompts 微调与增强网络训练

Fig. 3 The second stage: Prompts fine-tuning and enhanced network training

额外的计算负担。在推理应用阶段,将无需 CLIP 模型,也不需要 Prompts 向量权重,仅仅需要加载增强网络即可成功推理,这将是能够推理 4 K 大尺寸图像至关重要的原因之一。

1) 基于 CLIP 的无监督增强算法框架

第 1 阶段为 Prompts 学习与增强网络训练,首先通过 Prompts 学习以粗略地表征曝光异常和曝光正常的图像。对于给定一个形状为的 $H \times W \times 3$ 曝光异常图像 I_b 与曝

光正常的图像 I_w (图像无需配对),随机初始化一个形状为 $N \times 512$ 的可学习积极正向提示向量 T_p 与消极反向提示向量 T_n 。其中 N 代表着每个提示词里嵌词 token 的数量,接着将曝光异常和曝光正常的图像送入已经预先训练好的 CLIP 图像编码器中获得图像的潜在编码表示,同时将积极正向提示向量和消极反向提示向量送入已经预先训练好的 CLIP 的文本编码器中获得文本的潜在编码表示。此时将 CLIP 的预训练权重全部冻结,基于 CLIP

潜在空间的文本-图像相似性,使用曝光异常和曝光正常的图像进行分类的二元交叉熵损失函数来对比学习,如式(1)所示,反向传播更新 Prompts 权重参数来学习到 Prompts 的正确表示。

$$\mathcal{L}_{\text{initial}} = - (y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})) \quad (1)$$

$$\hat{y} = \frac{e^{\cos(\Phi_{\text{image}}(\mathbf{I}), \Phi_{\text{text}}(T_p))}}{\sum_{i \in \{n, p\}} e^{\cos(\Phi_{\text{image}}(\mathbf{I}), \Phi_{\text{text}}(T_i))}} \quad (2)$$

其中, $\mathbf{I} \in \{\mathbf{I}_b, \mathbf{I}_w\}$, y 是当前图像的标签, 0 是负样本, 1 是正样本。

接下来训练初始的增强网络, 由于前面已经初步的学习到了 Prompts 的文本表示, Prompts 知晓了什么是对应的正常曝光图像, 什么是对应的曝光异常的图像, 受此启发可以将这个表示作为损失函数来监督增强网络学习。故此定义 CLIP 增强损失 $\mathcal{L}_{\text{CLIP}}$ 为, 增强网络输出的结果被 CLIP 图像编码器 (Image Encoder) 转换的图像语义表示, 学习到的 Prompts 通过 CLIP 文本编码器 (Text Encoder) 转换的文本语义表示, 将这两者计算相似性分数作为如式(3)所述的损失值反向传播, 这一操作旨在让增强网络的输出去逼近一个曝光良好的图像。

$$\mathcal{L}_{\text{CLIP}} = \frac{e^{\cos(\Phi_{\text{image}}(\mathbf{I}_l), \Phi_{\text{text}}(T_n))}}{\sum_{i \in \{n, p\}} e^{\cos(\Phi_{\text{image}}(\mathbf{I}_l), \Phi_{\text{text}}(T_i))}} \quad (3)$$

然而仅有这个损失是远远不够的, 考虑到训练 Prompts 时采取的是未配对的数据集, 因此增强网络的输出虽然会变成曝光正常的图像, 但很有可能导致失真或者输出结构与输入图像的结构出现偏差, 这并不是本研究所期待的结果, 故而考虑添加一致性损失, 即将原始的曝光异常图像和增强网络输出的图像分别送入 CLIP 的 Image Encoder 的各个层去计算二范数损失, 这一操作旨在鼓励增强网络的输出在结构方面与曝光异常的图像相似, 具体描述为:

$$\mathcal{L}_{\text{identity}} = \sum_{l=0}^4 \alpha_l \|\Phi_{\text{image}}^l(\mathbf{I}_b) - \Phi_{\text{image}}^l(\mathbf{I}_l)\|_2 \quad (4)$$

式中: α_l 是 CLIP 图像编码器模型中 ResNet101 的第 l 层。

2) Prompts 细化与增强网络微调

在第2阶段, 迭代地进行 Prompts 细化和增强网络调整。Prompts 细化和增强网络调整以交替的方式进行。目标是提高学习 Prompts 的准确性, 以区分曝光异常和曝光正常的图像, 以及感知不同亮度的异质区域。本研究观察到, 在某些情况下, 仅使用从曝光异常和曝光正常的非配对数据集获得的初始 Prompts 不足以增强对比度和细节。这是因为初始 Prompts 可能无法捕获曝光异常和曝光正常图像之间的细粒度差异。为了解决这个问题, 本文引入了对可学习的正向积极 Prompts 和反向消极 Prompts 的进一步改进。给定由当前增强网络增强的结果 \mathbf{I}_l , 通过排序损失来更新 Prompts。

一般 Prompts 对和图像之间的负相似对得分定义为式(5), 即:

$$S(\mathbf{I}) = \frac{e^{\cos(\Phi_{\text{image}}(\mathbf{I}), \Phi_{\text{text}}(T_n))}}{\sum_{i \in \{n, p\}} e^{\cos(\Phi_{\text{image}}(\mathbf{I}), \Phi_{\text{text}}(T_i))}} \quad (5)$$

那么, 排序损失可以定义为式(6), 即:

$$\mathcal{L}_{\text{prompt1}} = \max(0, S(\mathbf{I}_w) - S(\mathbf{I}_b) + m_0) + \max(0, S(\mathbf{I}_l) - S(\mathbf{I}_b) + m_0) + \max(0, S(\mathbf{I}_w) - S(\mathbf{I}_l) + m_1) \quad (6)$$

其中, $m_0 \in [0, 1]$ 表示 CLIP 嵌入空间中曝光良好/增强结果的得分与曝光异常图像之间的边距。将 m_0 设置为 0.9, 以尽可能扩大曝光异常图像和曝光良好图像之间的距离。同时, m_1 表示 CLIP 嵌入空间中增强结果的得分与光照良好图像之间的边距。将 m_1 设置为 0.2, 以确保增强结果与曝光良好图像相似。这些超参数是根据算法在验证集上的性能经验选择的。

为了保证迭代学习在每一轮迭代中都能提高性能, 本研究保留了前一次增强网络在排序过程中得到的增强结果 \mathbf{I}_{l-1} , 并将 \mathbf{I}_{l-1} 和 \mathbf{I}_l 两组增强结果加入约束中, 使新学习的提示更加关注图像的细节和对比度分布, 而不是图像中的高级内容。损失函数修改为式(7), 即:

$$\mathcal{L}_{\text{prompt2}} = \max(0, S(\mathbf{I}_w) - S(\mathbf{I}_b) + m_0) + \max(0, S(\mathbf{I}_{l-1}) - S(\mathbf{I}_b) + m_0) + \max(0, S(\mathbf{I}_w) - S(\mathbf{I}_l) + m_1) + \max(0, S(\mathbf{I}_l) - S(\mathbf{I}_{l-1}) + m_2) \quad (7)$$

其中, m_2 表示新的增强结果和先前增强结果之间的边缘差距。设置 $m_2 = m_1$, 因为边缘 m_1 和 m_2 具有相同的目标, 保持两个图像组相似。

增强网络的微调遵循第1阶段相同的过程, 除了使用改进的 Prompts 来计算 CLIP 增强损失和限制对比度直方图均衡化先验损失, 并从更新的网络生成增强的训练数据以进一步改进 Prompts。

1.2 损失函数设计

1) CLAHE 的特征图先验损失

直方图均衡化先验是基于直方图均衡化增强后的图像, 传统的限制对比度的自适应直方图均衡化方法通过对直方图的处理来扩展暗图像的动态范围。CLAHE 能够有效地增强图像中不同区域的局部对比度, 而不仅仅是全局对比度, 尤其适合处理那些具有局部亮度变化的图像。CLAHE 通过限制对比度的增强 (即限制直方图的最大灰度值), 可以减少在图像噪声较多区域的过度增强, 避免噪声放大。并且相较于全局直方图均衡化, CLAHE 能够根据不同区域的局部信息进行自适应处理, 使得图像的细节得到更好展示。此外, CLAHE 通过对对比度进行限制 (即限制局部区域内的最大灰度范围), 避免了图像中某些区域对比度过度增强, 导致细节丢失或产生伪影。遗憾的是, CLAHE 算法也有它的不足之处,

CLAHE 的性能依赖于两个主要参数:局部窗口大小和对比度限制值。如果参数选择不当,可能会导致图像过度增强或不足,影响最终结果的质量。因此它的泛化性较差,面对不同的图像可能需要调整超参数。在某些情况下,CLAHE 可能会在图像的边缘或均匀区域产生不自然的伪影,尤其是在局部对比度变化剧烈的区域。尽管 CLAHE 在一定程度上可以抑制噪声,但在高噪声环境下,它可能仍然会放大图像中的噪声,尤其是在噪声较为均匀或随机分布的区域。

因此直接让网络预测的输出逼近 CLAHE 算法的处理结果是不理想的,受文献[24]的启发,可以采用 CLIP 的 Image Encoder 特征图来约束低光图像及其直方图均

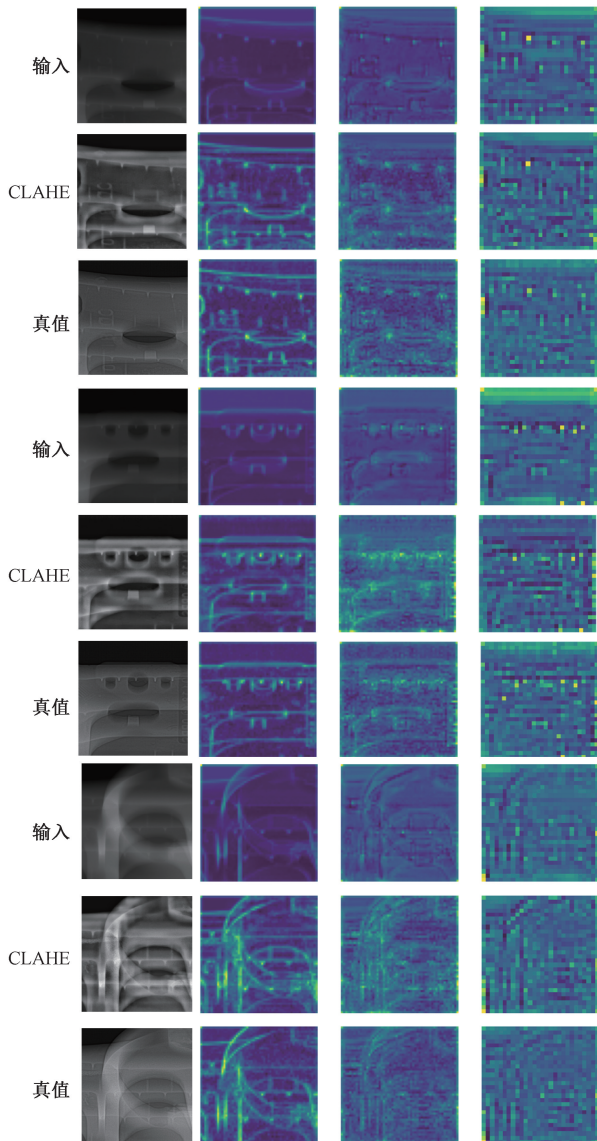


图 4 输入图像、CLAHE 及标签在 CLIP 的 Image Encoder 上使用预训练权重的不同层特征图

Fig. 4 Input image, different layered feature maps of CLAHE and labels for the CLIP Image Encoder with the pre-trained weights

衡增强版本之间的感知相似性。如图 4 所示,可以观察到原始输入的 DR 图像的特征图具有较少的语义信息。相比之下,经过限制对比度的自适应直方图均衡化增强后的图像的特征图具有更丰富的语义信息,至关重要的是,其与标签真值图像的特征图具有显著的相似性。这一发现启示本研究可以采用这个先验而不是真实图像来指导训练过程。

图 5 显示了在 CLIP 的 Image Encoder 上使用 CLAHE 增强图像预训练的网络不同层。靠近输入层的特征图更关注纹理信息的具体细节,一些特征图还可以显示工件的整体形状。远离输入层的特征图更关注语义和抽象信息,最深层的特征图变得更加模糊,不能提供足够的信息,而每组特征图之间的特征是相似的,基于这些信息,将最后一层的权重设置为 0。

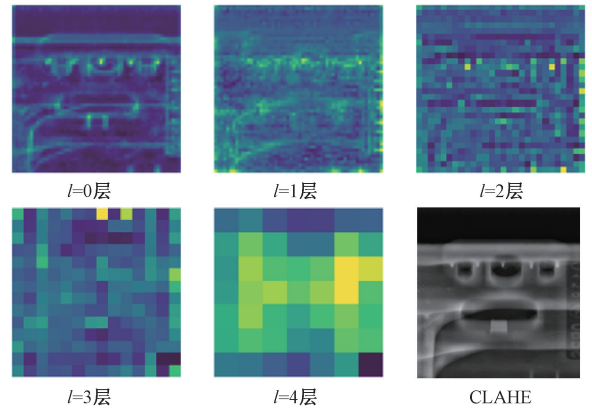


图 5 使用 CLAHE 算法的图像在预训练的 CLIP 图像编码器网络不同层上的特征图和输入原始图

Fig. 5 Input image and feature maps for different layers of CLIP Image Encoder networks pre-trained with a CLAHE enhanced images

故而这个先验损失可以表示为式(8),即:

$$\mathcal{L}_{\text{CLAHE}} = \sum_{l=0}^4 \gamma_l | \Phi_{\text{image}}^l(H(l_b)) - \Phi_{\text{image}}^l(H(l_t)) |_2 \quad (8)$$

其中, $H(\cdot)$ 表示通过限制对比度直方图均衡化算法。 γ_l 是 CLIP 图像编码器模型中 ResNet101 的第 l 层。

2) 无监督模型总损失

第 1 阶段的模型训练将由三大项损失构成—CLIP 增强损失、结构一致性损失、限制对比度直方图均衡化先验损失。公式描述如式(9)所述,即:

$$\mathcal{L}_{\text{enhance}} = \mathcal{L}_{\text{CLIP}} + \beta \mathcal{L}_{\text{identity}} + \omega \mathcal{L}_{\text{CLAHE}} \quad (9)$$

其中, β, ω 是平衡不同损失项的大小的权重,根据经验分别设置为 0.9、0.8。训练步骤分为两部分。首先,使用一致性损失和直方图先验损失来实现自我重建,因为它鼓励增强结果在像素空间与结构中曝光异常图

像相似。然后,通过使用一致性损失、CLIP 增强损失、直方图先验损失来训练网络。对于一致性损失和直方图先验损失,在自重建阶段,设置 $\alpha_{l=0,1,2,3} = 1, \alpha_{l=4} = 0.5$ 。在增强网络的训练过程中,将参数数值设置为 $\alpha_{l=0,1,2,3} = 1, \alpha_{l=4} = 0.5, \gamma_{l=0,3} = 1, \gamma_{l=1,2} = 1.5, \gamma_{l=4} = 0$ 。这是因为最后一层的特征与本文希望增强的结果所背离,第2和第3层的特征与图像的对比度更相关,这是想要调整的。

2 实验与结果

2.1 实验设置与环境配置

实验使用的数据集是来自中车齐齐哈尔车辆有限公司生产的铁路货运列车转向架铸钢摇枕铸件,对这些金属铸件的 DR 扫描数据集进行了实验。在训练阶段,不同于完全监督算法,无需使用上千张图像进行训练,为了减弱训练开销同时防止网络的过拟合泛化性降低,仅使用 380 张曝光异常对比度低下的图像,390 张非配对曝光良好的正常对比度图像作为参考图像进行训练。

实验在 NVIDIA RTX 4090 GPU 卡上使用 PyTorch 实现算法,使用 Adam 优化器,并且设置 $\beta_1 = 0.9, \beta_2 = 0.99$ 。为保持与 CLIP 预训练模型的兼容性,实验遵循先前工作的设置,将每个可学习 Prompts 的 token 数量 N 设为 16。将总训练迭代次数设置为 50 K,其中,自重构迭代次数设置为 1 K, Prompts 对初始化学学习迭代的次数设置为 10 K。提示初始化/细化和增强网络训练的学习速率设置为 5×10^{-6} 和 2×10^{-5} 。提示初始化/细化的批量大小为细化和增强网络训练设置为 8 和 16。在训练期间,将输入图像调整为 512×512 ,并使用翻转,缩放和旋转作为图像增强策略。参数设置如表 1 所示。

2.2 对比实验结果

为说明本文方法的性能,选取了近期的一些具有代表性的增强方法进行对比,包括灰度均衡化先验

表 1 模型的训练参数

Table 1 Training parameters of model

参数名称	参数值
图像大小	(512,512)
优化器	Adam
迭代次数	5×10^4
学习率	5×10^{-6} 和 2×10^{-5}
批次大小	8/16
β_1	0.90
β_2	0.99

(histogram equalization prior, HEP)^[24]、结合架构搜索的 Retinex 启发展开 (Retinex-inspired unrolling with architecture search, RUAS)^[25]、自校准照明 (self-calibrated illumination, SCI)^[26]、配对低光照图像增强 (PairLIE)^[27]、零样本照明引导的联合去噪与自适应增强 (ZERO-shot illumination-guided joint denoising and adaptive enhancement, ZeroIG)^[9] 和 CLIP-LIT 方法^[19]。

考虑常见的 DR 图像质量指标,客观评价结果如表 2 所示,包括 3 个有参考图像的指标:结构相似度 (structural similarity, SSIM),感知图像相似度 (learned perceptual image patch similarity, LPIPS),峰值信噪比 (peak signal-to-noise ratio, PSNR)。5 个无参考图像的指标平均梯度 (average gradient, AG),信息熵 (information entropy, IE),灰度标准差 (standard deviation, SD),空间频率 (spatial frequency, SF),无参考空间锐度 (no-reference spatial sharpness, NRSS)。评价指标表格中的上箭头“↑”表示数值越大指标越好,下箭头“↓”表示数值越小指标越好。

3 个有参考图像的评价指标在 SSIM 上提升 2%, LPIPS 提升 1.6%, PSNR 提升 1 dB,结果可以看出提出的算法的增强结果在结构上与标签最吻合,即失真最少,增

表 2 提出的方法与近期代表的方法在 DR 数据集的定量比较

Table 2 Quantitative comparisons between proposed model and state-of-the-art methods on DR datasets

方法	PSNR ↑	SSIM ↑	LPIPS ↓	AG ↑	IE ↑	SD ↑	SF ↑	NRSS ↑
HEP ^[24]	12.955 2	0.736 5	0.381 0	1.427 5	6.055 2	37.476 7	4.614 4	0.408 7
RUAS ^[25]	6.471 4	0.462 6	0.693 1	0.515 6	2.970 2	79.980 2	3.891 8	0.061 8
SCI ^[26]	12.530 0	0.736 5	0.381 0	1.574 6	5.827 5	68.547 3	5.475 7	0.502 2
PairLIE ^[27]	8.864 7	0.482 1	0.568 3	0.708 9	5.528 3	33.622 0	2.309 3	0.232 1
ZeroIG ^[9]	14.492 8	0.707 4	0.344 8	1.663 8	6.835 2	52.884 5	5.978 9	0.503 2
CLIP-LIT ^[19]	24.315 9	0.822 7	0.239 9	1.602 5	6.156 6	37.259 3	5.395 9	0.451 8
本文方法	25.316 5	0.842 2	0.223 2	1.952 5	6.461 6	51.839 3	6.638 1	0.534 4

强结果较好,并且更加接近真实标签。从 5 个无参考图像的评价指标可以看出提出的算法的增强结果有着更丰

富的信息。

增强结果局部放大对比如图 6 所示。

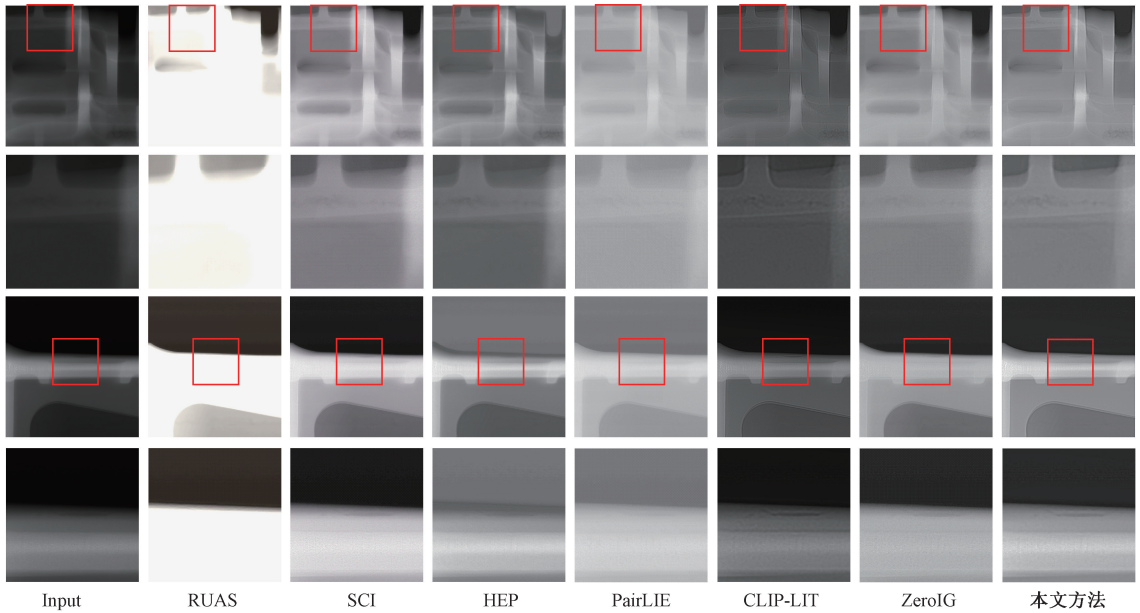


图 6 在 DR 数据集上与近期代表方法进行定性比较的图及其局部放大图

Fig. 6 Plots of qualitative comparisons with state-of-the-art methods on the DR dataset and the local zoom-in picture

从整体的结果来看, RUAS 算法^[25]增强质量最为低下,整体已经呈现出曝光过度的情况,细节信息完全消失,增强结果过于亮。而对于 SCI^[26]、PairLIE^[27]、ZeroIG^[9]算法增强相对较亮,整体对比度仍然较低,信息熵很低,细节边缘和铸件缺陷难以看清,缺乏局部增强。相比之下 HEP 算法^[24]在亮度对比度方面展示出不错的效果,遗憾的是 HEP 算法存在失真的情况,且容易放大噪声信息。较于其他算法, CLIP-LIT 算法和提出的算法在局部对比度增强方面的表现最佳,基本能够实现图像的自适应增强,也不存在失真情况而且不会放大噪声,在详细对比之下, CLIP-LIT 算法相较于提出的算法增强结果稍暗了一些,整体的灰度值呈现较低的数值,存在部分边缘细节信息被弱化的现象,故而细节缺陷边缘的增强仍有待提高。提出的算法在增强局部缺陷细节方面做得最好,经过提出的无监督增强算法,铸件所带缺陷可以较清晰地看清楚并且不会放大噪声。增强结果得到缺陷清晰、轮廓明显的图像。

如图 7 所示的增强对比结果可以看出,实际上在图像正中央位置其实是存在铸件的裂纹缺陷的,原始图像很难从肉眼观察出这一缺陷。RUAS 算法增强结果呈现出过亮的风格,细节信息完全被淹没,增强质量很差。对于 SCI、PairLIE、ZeroIG 算法增强结果,由于图像依然较亮导致对比度不高,因此图像中央的缺陷也很难明显地看出来,细节信息不清楚,边缘被弱化。相比之下 HEP 算法对比度稍微提高,可以大约看出一些缺陷的痕迹,遗

憾的是也未能看得较为清楚,且在图像的上端部分存在一些失真错误增强的情况。相比之下 CLIP-LIT 与提出的算法由于局部对比度增强方面表现最佳,可以较为明显地看出铸件缺陷纹理细节, CLIP-LIT 整体稍微偏暗,提出的算法相对看得更清楚。

在增强网络模型的设计中,为了能够推理 4 K 大尺寸图像,将网络模型设计得十分简单轻量,推理阶段需要加载的权重参数仅为 0.278 8 M,故而无需使用分块推理的操作,不会造成块状效应,对分辨率为 5 732×2 333 的 DR 铸件图像进行推理,结果如图 8 所示。可以看到,即便面临大尺寸图像,网络依然能够有效捕捉其细节特征,在局部缺陷位置的增强展现出了强大的对比度,可以非常直观地看出缺陷。

无监督方法的时效性在该领域通常不是主要指标。作为参考,在当前实验环境下,本方法处理一张 3 008×2 496 分辨率的 DR 图像约需 1.5 s,已能满足工业抽检(通常要求节拍在 2 s 以内)的时效性需求。

2.3 消融实验

为了验证限制对比度直方图均衡化损失的有效性,为此做了消融实验,与此同时对比了是将增强结果进入 CLIP 的 Image 编码器提取的特征图有效还是将增强结果送入在 ImageNet 数据集进行预训练的 VGG-19backbone 提取的特征图更有效进行了对比实验,实验客观评价指标结果如表 3 所示,结果表明使用 CLAHE 作为损失函

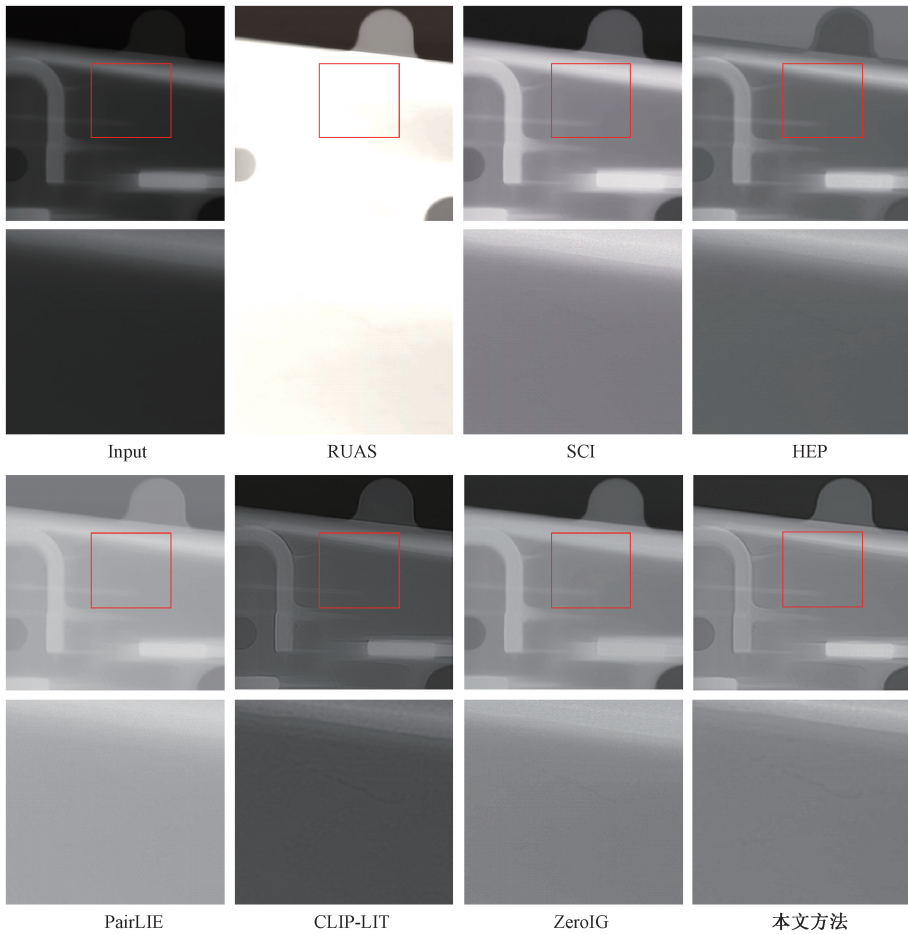


图 7 在 DR 数据集上与近期的代表方法进行定性比较

Fig. 7 Qualitative comparisons with state-of-the-art methods on the DR dataset

表 3 是否添加限制对比度直方图均衡化损失及提取特征网络的消融实验结果

Table 3 The ablation test results for the cases with and without the CLAHE loss and extract feature network

方法	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
不采用 $\mathcal{L}_{\text{CLAHE}}$	24.315 9	0.822 7	0.239 9
采用 $\mathcal{L}_{\text{CLAHE}}$ & VGG-19	24.745 0	0.833 9	0.230 1
采用 $\mathcal{L}_{\text{CLAHE}}$ & CLIP	25.316 5	0.842 2	0.223 2

数指导训练是可以增强图像质量的,并且将 CLAHE 的结果送入 CLIP 的 Image 编码器效果最佳。

同时比较了直接通过传统算法 CLAHE 增强后的结果,其定性结果如图 9 所示,如果直接进行 CLAHE 效果较差,很容易导致每一块之间的差异过大,局部过增强,放大噪声得到失真的结果,并且细节部分也难以看清,相较于不加入 CLAHE 损失的无监督网络,其对比度较低,整体呈现较暗风格。加入 CLAHE 损失的无监督网络,整体对比度较为均匀,缺陷细节部分也较为明显,不会出现

局部过增强情况。

2.4 泛化性实验

为了证明所提模型的稳健性和通用性,进行泛化性实验。将训练好的模型直接应用在其他网络未曾训练过的工件中,包括 ACCC 碳纤维复合芯导线和线路板等工件,其定性结果分别如图 10 所示。

可以看到原始碳纤维复合导线 DR 图的对比度非常之低下,几乎看不到有用信息,而其他无监督算法虽有一定的效果,但是其细节的对比度还是不足,灰度值较为单一,而对于有监督算法而言,其亮度的增强也较弱,整体画面依然较暗,相比之下,本文的无监督算法拥有更高的对比度,并且细节也能够看得更加清楚,相较于有监督算法,本文的无监督算法体现出更好的泛化性。实验的定量结果如表 4 所示,通过比对一些无监督算法和有监督算法,结果表明本文的无监督算法取得了最好的结果,在多数无监督评价指标中达到最好。图 11 展示了本文方法其他材料的无监督增强结果作为参考。

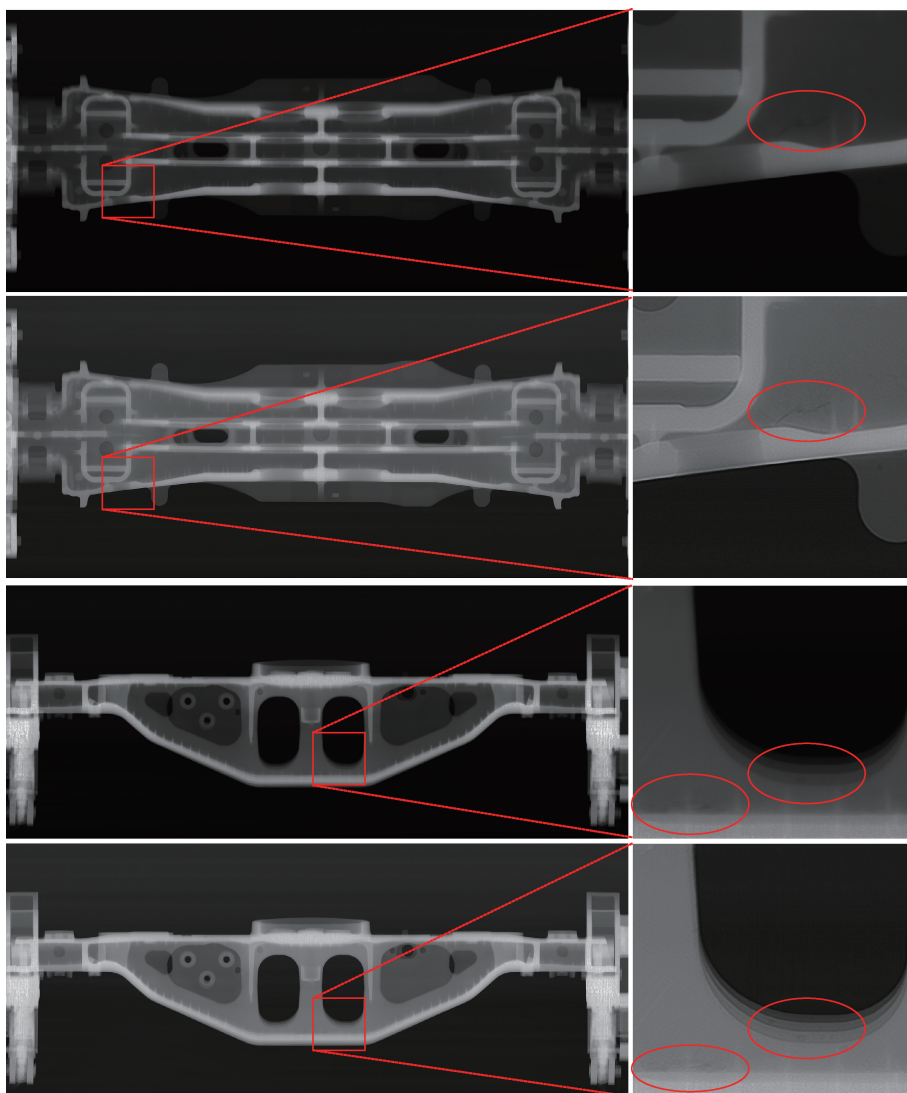
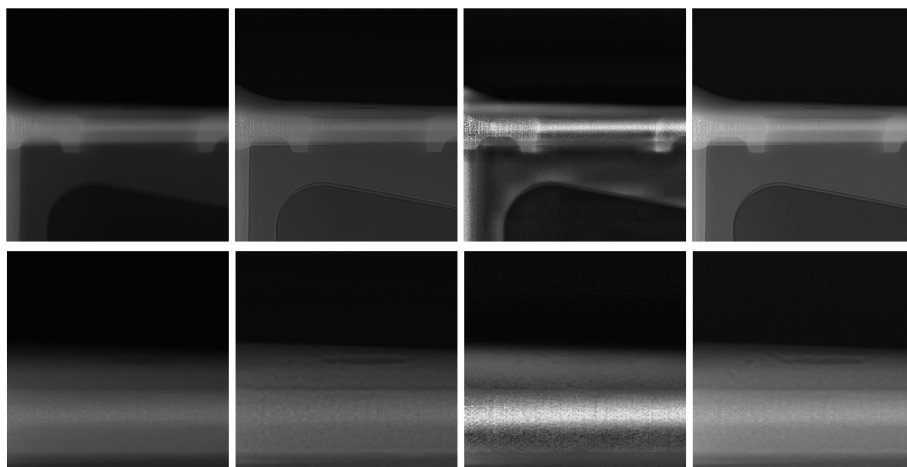


图 8 对分辨率为 $5\ 732 \times 2\ 333$ 的 DR 图像的推理结果
Fig. 8 Inference results of DR image with the resolution of $5\ 732 \times 2\ 333$



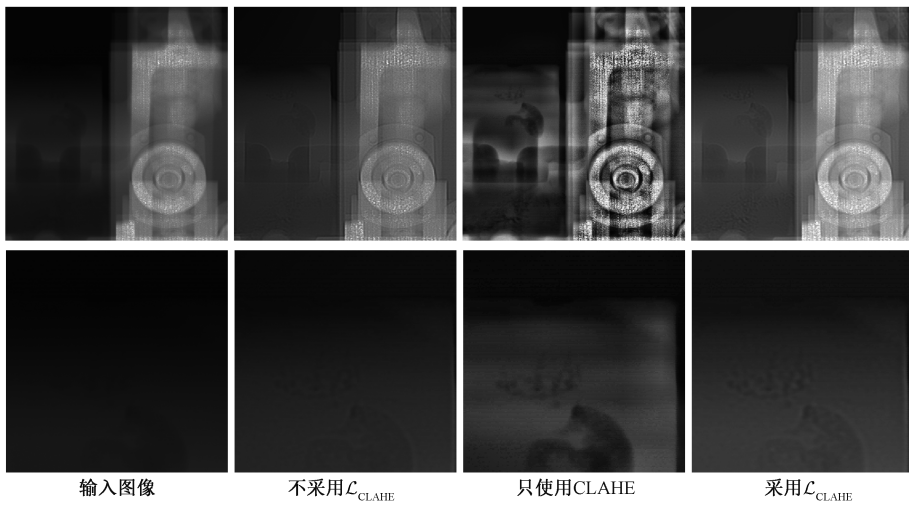
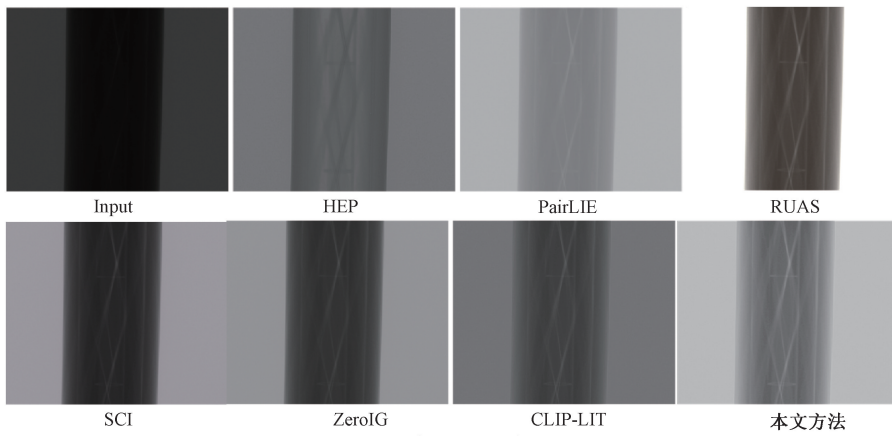
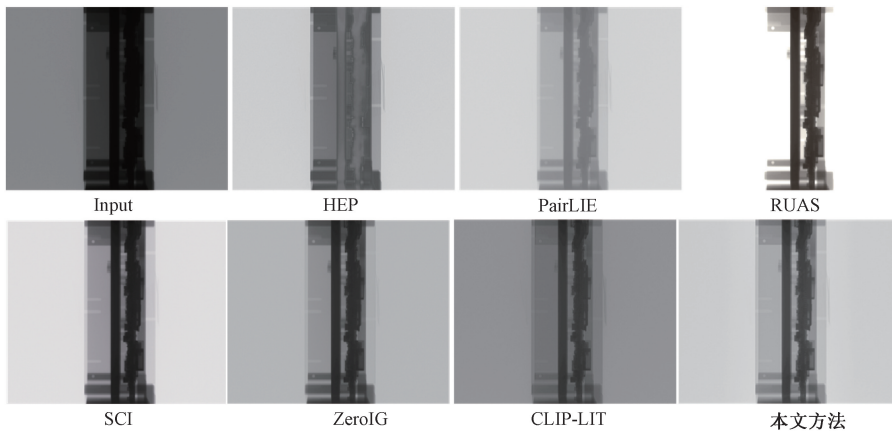


图 9 对只做 CLAHE 算法和是否使用 CLAHE 损失函数的消融实验定性结果

Fig. 9 Qualitative results of ablation experiments with only CLAHE algorithm and that with or without CLAHE loss function



(a) 碳纤维导线增强结果
(a) Reinforced structure of carbon fiber composite core conductor



(b) 线路板增强结果
(b) Reinforced structure of circuit board

图 10 提出的模型与有监督算法及其他的无监督算法泛化性实验的定性比较

Fig. 10 Quantitative comparison of generalization performances for the proposed model, supervised algorithm and other unsupervised algorithms

表 4 提出的模型与本文的有监督算法及其他的无监督算法泛化性实验的定量比较

Table 4 Qualitative comparison of generalization performance for the proposed model supervised algorithm in this thesis and other unsupervised algorithms

方法	AG \uparrow	IE \uparrow	SD \uparrow	SF \uparrow	NRSS \uparrow
HEP ^[24]	0.870 7	4.937 1	40.884 5	2.259 9	0.132 9
RUAS ^[25]	0.312 3	2.942 6	72.203 0	1.692 2	0.024 8
SCI ^[26]	0.959 1	4.726 2	66.000 8	2.785 9	0.140 9
PairLIE ^[27]	0.451 7	4.502 6	31.264 5	1.202 8	0.050 0
ZeroIG ^[9]	0.961 3	4.950 0	48.858 2	2.805 1	0.156 3
CLIP-LIT ^[19]	0.854 3	5.024 1	28.615 4	2.635 9	0.210 9
本文方法	1.307 4	5.024 1	44.217 2	3.993 9	0.339 9

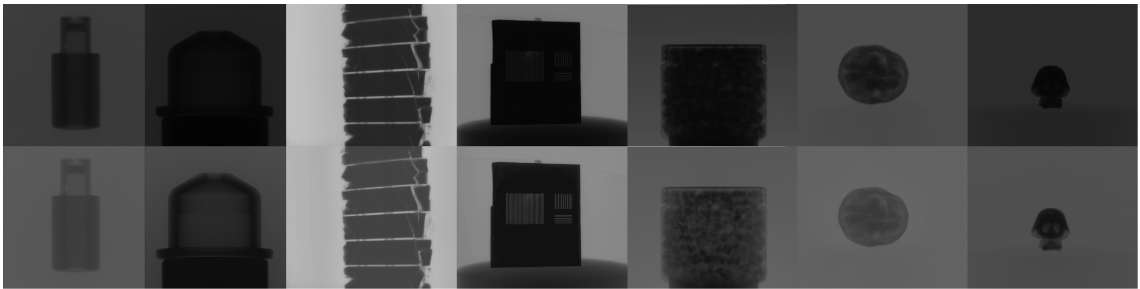


图 11 提出的模型在其他材质物件的无监督增强结果

Fig. 11 Unsupervised improvement results of proposed model implemented for the items made of other materials

3 结 论

本文着重解决在实际生产应用过程中标签图像难以获取导致有监督算法难以应用的问题,即针对无标签数据集进行图像增强任务。由于 CLIP 是从互联网收集了共 4 亿的文本-图像对数据集且涵盖各个方面各个领域,并且在 Zero-Shot 推理领域展示出了惊人的能力,故而本文基于 CLIP 的思想进行无监督图像增强,提出双阶段从粗增强到细粒化增强,通过对 Prompts 的初步学习并将其设置为损失函数指导训练,并且提出限制对比度的自适应直方图均衡化的特征图先验的思想作为损失函数指导训练,消融实验结果表明这个损失函数的设置是可以提高图像增强质量的。最后,本文对比几个较好的无监督算法,结果表明本文算法在缺陷细节的增强更明显,同时在 4 K 大尺寸图像进行了推理,结果表明此算法依然能够很好地增强 4 K 大尺寸图像,在泛化性实验上表明提出的算法具有更好的泛化性,这为实际落地工业应用奠定了坚实的基础。

参考文献

- [1] SEIBERT J A. Digital radiography: The bottom line comparison of CR and DR technology [J]. Applied Radiology, 2009, 38(5): 21-28.
- [2] VAN DER STELT P F. Better imaging: The advantages of digital radiography [J]. The Journal of the American Dental Association, 2008, 139: S7-S13.
- [3] DEMAIO D N, HERRMANN T, NOBLE L B, et al. Best practices in digital radiography [J]. Radiologic Technology, 2019, 91(2): 198-201.
- [4] WANG ZH D, CUN X D, BAO J M, et al. Uformer: A general U-shaped transformer for image restoration [C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 17662-17672.
- [5] ZAMIR S W, ARORA A, KHAN S, et al. Restormer: Efficient transformer for high-resolution image restoration [C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 5718-5729.
- [6] WANG T, ZHANG K H, SHEN T R, et al. Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method [C]. 37th AAAI Con-

- ference on Artificial Intelligence, 2023, 37(3): 2654-2662.
- [7] LI J L, LI B L, TU ZH ZH, et al. Light the night: A multi-condition diffusion framework for unpaired low-light enhancement in autonomous driving [C]. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 15205-15215.
- [8] LYU X Q, ZHANG SH P, WANG CH Y, et al. Fourier priors-guided diffusion for zero-shot joint low-light enhancement and deblurring [C]. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 25378-25388.
- [9] SHI Y Q, LIU D, ZHANG L G, et al. ZERO-IG: Zero-shot illumination-guided joint denoising and adaptive enhancement for low-light images [C]. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 3015-3024.
- [10] WANG W, YANG H, FU J L, et al. Zero-reference low-light enhancement via physical quadruple priors [C]. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 26057-26066.
- [11] JIANG Y F, GONG X Y, LIU D, et al. EnlightenGAN: Deep light enhancement without paired supervision [J]. IEEE Transactions on Image Processing, 2021, 30: 2340-2349.
- [12] GUO C L, LI C Y, GUO J CH, et al. Zero-reference deep curve estimation for low-light image enhancement [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1777-1786.
- [13] XIONG W, LIU D, SHEN X H, et al. Unsupervised low-light image enhancement with decoupled networks [C]. 2022 26th International Conference on Pattern Recognition, 2022: 457-463.
- [14] ZENG H, CAI J R, LI L D, et al. Learning image-adaptive 3D lookup tables for high performance photo enhancement in real-time [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 44(4): 2058-2073.
- [15] RADFORD A, KIM J W, HALLACY C, et al. Learning transferable visual models from natural language supervision [C]. International Conference on Machine Learning, 2021: 8738-8753.
- [16] ZHOU CH, LOY C C, DAI B. Extract free dense labels from clip [C]. Computer Vision-ECCV 2022, 2022: 696-712.
- [17] ZHOU K Y, YANG J K, LOY C, et al. Learning to prompt for vision-language models [J]. International Journal of Computer Vision, 2022, 130(9): 2337-2348.
- [18] ZHOU K Y, YANG J K, LOY C C, et al. Conditional prompt learning for vision-language models [C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 16795-16804.
- [19] LIANG ZH X, LI CH Y, ZHOU SH CH, et al. Iterative prompt learning for unsupervised backlit image enhancement [C]. 2023 of the IEEE/CVF International Conference on Computer Vision, 2023: 8060-8069.
- [20] 林立芃,杨朝阳,伍明诚,等. 改进型 CLAHE 图像增强算法及其 FPGA 实现 [J]. 电子测量技术, 2024, 47(10): 126-133.
- LIN L P, YANG ZH Y, WU M CH, et al. Research on the improved CLAHE image enhancement algorithm and its FPGA-based implementation [J]. Electronic Measurement Technology, 2024, 47(10): 126-133.
- [21] 羊肇俊,曾理. 基于加权最小二乘滤波和引导滤波的铸件 DR 图像融合 [J]. 仪器仪表学报, 2021, 42(6): 211-220.
- YANG ZH J, ZENG L. Casting DR image fusion based on weighted least squares filter and guided filter [J]. Chinese Journal of Scientific Instrument, 2021, 42(6): 211-220.
- [22] 袁文浩,屈庆洋,梁春燕,等. 基于感知条件网络的可控语音增强模型 [J]. 仪器仪表学报, 2023, 44(5): 53-60.
- YUAN W H, QU Q Y, LIANG CH Y, et al. Controllable speech enhancement model based on perceptual conditional network [J]. Chinese Journal of Scientific Instrument, 2023, 44(5): 53-60.
- [23] 王斌,刘祎,王祥. 基于密集连接的双能 DR 图像融合网络 [J]. 国外电子测量技术, 2023, 42(11): 33-41.
- WANG B, LIU Y, WANG X. Dual-function DR image fusion network based on dense connection [J]. Foreign Electronic Measurement Technology, 2023, 42(11): 33-41.
- [24] LIU Y B, LI SH Q, ZHENG X Y, et al. A low-light image enhancement method with histogram equalization prior [C]. 2023 9th International Conference on Computer and Communications, 2023: 2024-2029.
- [25] LIU R SH, MA L, ZHANG J AO, et al. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement [C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 10556-10565.

- [26] MA L, MA T Y, LIU R SH, et al. Toward fast, flexible, and robust low-light image enhancement [C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 5627-5636.
- [27] FU ZH Q, YANG Y, TU X T, et al. Learning a simple low-light image enhancer from paired low-light instances [C]. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 22252-22261.

作者简介



陈明飞, 2012 年于南昌航空大学获得学士学位, 现为中国航发哈尔滨东安发动机有限公司高级工程师, 主要研究方向为航空发动机无损检测。

E-mail: snack27@163.com

Chen Mingfei received his B.Sc. degree from Nanchang Hangkong University in 2012. He is currently a

senior engineer in AECC Harbin Dongan Engine Co., Ltd. His main research interest includes non-destructive testing of aero-engines.



沈宽 (通信作者), 1999 年、2002 年、2010 年于重庆大学分别获得学士、硕士、博士学位, 现为重庆大学工业 CT 无损检测教育部工程研究中心副研究员, 主要研究方向为 CT 理论设计、图像处理、深度学习和缺陷识别。

E-mail: shenk@cqu.edu.cn

Shen Kuan (Corresponding author) received his B.Sc., M.Sc., and Ph.D. degrees all from Chongqing University in 1999, 2002, and 2010, respectively. He is currently an associate researcher in Engineering Industrial Computed Tomography Nondestructive Testing of the Ministry of Education, Chongqing University. His main research interests include CT theory and design, image processing, deep learning, and defect recognition.