

基于无监督学习的金属表面高光去除算法研究

成懋¹, 李炳然², 陈炜铎², 洪海波³, 薛静³

(1.西北工业大学 软件学院, 陕西 西安 710072; 2.清华大学 机械工程系, 北京 100084;
3.上海航天精密机械研究所, 上海 201699)

摘要:金属材料表面的高光会严重破坏图像的连续性,产生一定的伪边缘,导致高光区域的纹理细节减弱甚至消失,干扰后续表面的区域分割和缺陷检测等操作。针对现有高光去除方法效率低、损失大、易失真、金属高光数据难标定等问题,提出了一种基于卷积神经网络(CNN)的无监督感知增强网络模型。使用生成对抗的方法生成大量拥有高光特征信息的金属图像,用于增加训练集中高光金属图像数据集的数量;在上下文聚合网络中引入细节增强模块(detail enhance model, DEM)和色彩增强模块(color enhance model, CEM)提高低分辨率权重图中的特征细节保留率;评估函数采用多尺度结构相似性函数代替原有结构相似性函数,解决在图像尺寸过大时细节不敏感的问题。实验表明,所提模型相比其他多曝光图像融合模型,在融合速度领先情况下,融合图像的互信息和平均梯度等评估指数提升 10% 左右,能够保留更多的纹理特征信息。

关键词:高光去除;无监督学习;MEF;图像融合;感知增强

中图分类号:TP391

文献标志码:A

文章编号:1000-2758(2024)04-0744-09

在智能制造推动下,机器视觉不断深入应用于制造行业的各个领域,使用计算机视觉完成工艺检测工作已成为智能装备的主流趋势。以视觉为主导的智能制造产业不断进步与创新,在实际工业复杂环境下的应用也越来越广泛。利用机器视觉集合工业大数据对在机加工件进行实时缺陷检测,可以及时发现加工过程中的问题,也可以在一定程度上解决人工检测带来的不确定性和低效率问题。

然而目前的机器视觉都需要基于高质量的图像来发挥作用,带有高光的金属图像不满足高质量的要求。由金属材料构成的航空发动机大部分组成部件在工业复杂光线条件下易产生高光区域,导致该区域的纹理细节以及特征信息缺失,并且金属特性容易导致其表面产生极端高光的情况。极端的高曝光使得金属曲面上信息丢失,干扰后续视觉检测过程,通过此类高光图像无法获得完整信息。因此,金

属表面的高光去除对于后续以机器视觉为主的智能检测系统具有重要意义。

基于输入图像的数量,高光去除方法可以分为 2 类:基于单图像的方法和基于多图像的方法。高光去除的早期研究始于 1985 年 Shafer^[1] 的双色反射模型,该模型首次将图像中的高光分解为镜面反射和漫反射 2 种分量,其中前者的光谱分布与光源的相似,而后的光谱分布在很大程度上取决于材料表面的属性,根据该模型使用图像的颜色分析和光谱分析便可以消除图像高光。后续学者们基于此模型,发展出了很多基于单图像的高光去除算法模型。然而基于单图像的方法大都会使用像素填充,在图像纹理或者光线环境较复杂情况下高光去除效果差^[2],单图像方法在高光区域信息完全被掩盖情况下通常只能降低区域亮度,不能保留其中的纹理信息和色度信息。

随着计算机运算能力的增长以及图像处理算法对高质量图像数量需求的增多,基于多图像的高光去除方法也逐渐发展起来。多图像方法中多曝光图像融合(multi-exposure image fusion, MEF)方法最具代表性,其主要思想是获得同一图像在不同曝光属

收稿日期:2023-07-20

基金项目:航天科技创新应用研究-慧眼行动(22BE5EEC)资助

作者简介:成懋(1999—),硕士研究生

通信作者:李炳然(1992—),助理研究员

e-mail:libingran1992@126.com

性下的状态,包括曝光角度、曝光强度等,然后对这些图像进行特征提取,通过原图与特征图的加权结合获得最终的融合图像^[3]。相比基于单图像的方法,MEF 方法能够获得高光区域中更多的纹理细节,但获得更多信息的代价是降低算法模型的实时性和泛化能力,因为大量固定的参数使其无法在多样化的数据上表现如一,并且传统的多图像方法在处理极端曝光情况下的图像时效果较差。

自 2012 年以来,深度学习在计算机视觉领域被广泛应用,学者们针对不同问题提出了各种针对性的网络框架,基于这些基础性框架的优化更是层出不穷。在深度学习多曝光图像融合领域,主要分为 2 类方法:基于无监督的方法和基于有监督的方法。前者属于先聚类再定性,可适应多样化的输入图像,如 MEF-Net^[4]、DeepFuse^[5]、MEF-GAN 等^[6],模型运行速度较快;后者则是按照已有标签进行分类,一般用于指定图像类型,如 ECNN^[7]、MFNet 等^[8],此类模型生成的图像质量较高。

目前已有学者将多曝光图像融合应用于自然图像处理领域并加以优化和改进,均在一定程度上取得了不错的效果。2018 年,Peng 等^[9]基于 CNN 设计了 MergeNet 融合网络,此方法将光流配准^[10]和深度学习网络相结合,相比之前多位学者提出的基于运动检测的方法,有效地解决了融合图像的重影问题,然而鲁棒性有限。后续 Yan 等^[11]提出了 CAHDRNet,一种基于多尺度的上下文注意的图像融合网络,这是一个端到端的网络,由特征提取网络和特征融合网络 2 个子网络构成,解决了速度慢的问题。在这些研究的基础上,Ma 等提出了一种快速多曝光图像融合方法 MEF-Net,此方法适用于任意空间分辨率和曝光次数的图像,使用了“下采样-执行-上采样”的方案,中间生成低分辨率权重图,上采样过程中利用了权重图的平滑性,在保证融合图像质量的同时,极大程度上提高了运行速度,在 GPU 上处理相同图像时,速度比其他深度学习方法快 10~100 倍。

当前的多曝光图像融合方法主要针对自然曝光图像,很少有在金属曝光图像上的研究,主要几点原因有:一是工业数据集的特殊性,导致金属高光图像的数据集稀少,很难找到专业的金属高光数据集用于研究;二是金属特性的原因,其表面极易产生极端高曝光或者低曝光的情况,难度较大。除此以外,还有一个重要的待解决问题是模型的运行速度,高光

去除作为一个图像预处理步骤,不应花费太多的时间。因此,本文以目前速度最优的 MEF-Net 为基础,引入细节增强模块(detail enhance model,DEM)和色彩增强模块(color enhance model,CEM)^[12]提高融合图像质量,并且使用 DCGAN^[13]生成金属高光图像扩充数据集,一定程度解决金属图像数据集缺少的问题,实现无监督下的金属表面高光去除。

1 算法原理

1.1 MEF-Net 模型介绍

MEF-Net 是一种端到端的多图像融合方法,它在保证融合图像质量的同时,极大提高了运行速度。MEF-Net 利用了一个全卷积网络使得用户可以输入任意数量的图像,为了提高运行速度,对输入图像直接进行线性下采样操作,难以避免会损失一些高频的细节信息。MEF-Net 的整个模型是基于 MEF 结构相似性(MEF-SSIM)^[14]指数进行优化的,这是整个模型质量保证。目前大量的实验表明,与多数最新 MEF 方法相比,MEF-Net 融合的图像在视觉感知方面有一定优势,且其运行速度远快于其他方法。

1.2 MEF-Net 网络结构

MEF-Net 由一个双线性下采样器、一个基于扩张卷积^[15]的特征预测网络(context aggregation net,CAN)、一个引导滤波器^[16]和一个加权融合模块组成。获取特征的结构如图 1 所示。

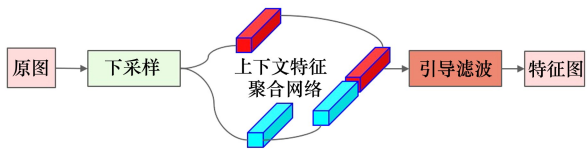


图 1 MEF-Net 获取特征结构图

MEF-Net 首先对输入序列进行降采样,并输入 CAN 网络低分辨率版本用来预测低分辨率权重图。以原图和特征图作为输入,使用引导滤波器获得高分辨率权重图,该操作也称为计算机视觉中的联合上采样。最后,计算特征图与原图的融合图像获得最终图像,如(1)式所示, W_k 表示原图像 X_k 的权重图, Y 为最终加权融合获得的无高光图像。

$$Y = \sum_{k=1}^K W_k \odot X_k \tag{1}$$

MEF-Net 核心模块是一个卷积网络,它将低分辨率的输入序列转换为相应的权值图。网络需要允

许任意空间大小和曝光数,并产生相应大小和数量的特征图。为此,MEF-Net 使用完全卷积网络来处理所有的曝光(即不同曝光的图像共享相同的权重

生成网络),这可以通过 batch 维度的分配来高效实现。MEF-Net 的核心网络有 7 个卷积层,其配置如表 1 所示,它的输出与输入具有相同的分辨率。

表 1 CAN 网络配置

网络层次	卷积	扩张	宽度	偏移	自适应归一化	非线性	感受野	卷积
1	3×3	1	24	×	✓	✓	3×3	3×3
2	3×3	2	24	×	✓	✓	7×7	3×3
3	3×3	4	24	×	✓	✓	15×15	3×3
4	3×3	8	24	×	✓	✓	31×31	3×3
5	3×3	16	24	×	✓	✓	63×63	3×3
6	3×3	1	24	×	✓	✓	63×65	3×3
7	1×1	1	1	✓	×	×	65×65	1×1

2 改进的 MEF 算法

2.1 DCGAN 图像生成

本文的方法主要应用在复杂工业环境下的高光金属表面,当前已有的高光图像数据集均为自然图像。实际加工现场收集到的图像数量很少,远远达不到深度学习模型训练要求,若采用现有数据集直接进行训练,容易导致实验结果出现过拟合情况。传统的数据集扩充方法主要采用旋转、缩放以及添加噪音等方式,这样的方式不仅繁琐,而且本身依赖于一定的数据数量^[17]。为基于目前有限的数据集获得更多金属高光图像,本文采用深度卷积生成对抗网络(deep convolutional generative adversarial networks, DCGAN)进行金属高光图像生成。

DCGAN 是目前生成对抗网络在实际工程应用中被采用最多的衍生网络,它是生成对抗网络针对图像生成质量以及稳定性优化而产生的。DCGAN 模型由生成网络 G 与判别网络 D 组成,结构如图 2 所示,网络结构均为 5 层,生成网络接收一个随机输入的噪声,通过该噪声生成一张目标图像;判别网络接收一张图像,对该图像做出判断后,计算出一个概率值,若该图像来源为生成网络,则输出 0,若该图像来自真实的数据分布,则输出 1。两者根据反馈的结果,不断更新自己的参数,直至达到纳什均衡。

生成模型的网络结构一共有 5 层,其中采用反卷积代替上采样来获得指定大小的图片。它的主要结构如图 3 所示。

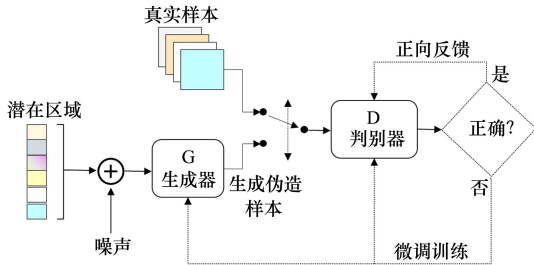


图 2 生成对抗模型

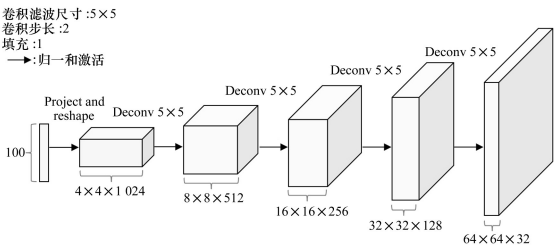


图 3 生成器 G 的网络结构

网络根据输入的向量信息,分步获取输入图像的特征信息,包括边缘、色度等,随后根据网络的深入,不断优化图片细节。输入一张图片,根据需要的特征图大小设置相应的卷积核,卷积核大小影响了生成网络的可学习空间特征值大小。在除了输出层之外的每一层,加上批归一化处理,缓解模型崩溃问题。最后,输出一个像素的三通道 RGB 图像。

判别模型的网络结构类似于反向的生成器,网络同样有 5 层,采用下采样方法。整个网络结构没有池化层,采用 LeakyReLU^[18] 作为激活函数,通过一个全连接层输出判别结果,表示输入图像属于真

实样本还是由生成器所生成,其结构如图4所示。

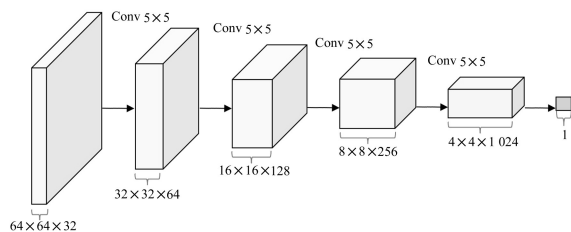


图4 判别器D的网络结构

生成网络和判别网络的输出层激活函数分别为 Tanh 函数和 Sigmoid 函数,两者都是 S 型函数,其函数特性会导致反向传播算法收敛速度降低,使用 BCELoss 函数后,解决了因 Sigmoid 函数导致的梯度消失问题。

用 DCGAN 模型生成的部分金属高光图像如图 5a) 所示。这些生成的图像带有金属高光的特征,并且可以通过调节参数和输入来改变生成图像的曝光程度,如图 5b) 所示。这些生成的图像分辨率为 256×256 ,并且带有一定的噪音点,可以作为训练集提高模型泛化能力。

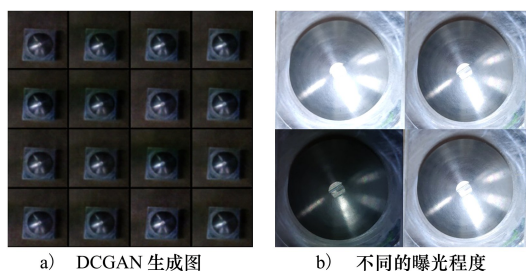


图5 DCGAN 生成图像结果

2.2 引入 DEM 模块和 CEM 模块

MEF-Net 为了满足速度要求,使用了下采样-处理-上采样的方法进行特征融合,其中下采样直接使用了线性下采样 (bilinear downsampling),这样会导致高光区域中高频的纹理信息丢失,下采样得到的图像序列输入 CAN 进行低分辨率权重预测。本文在这个过程中引入细节增强模块 DEM 和色彩增强模块 CEM 来增强网络的感知能力,减少因线性下采样导致的细节丢失和颜色失真问题,并且可以在几乎不影响原来速度的情况下实现。整体感知增强流程如图6所示,将色彩空间从 RGB 转换为 YCbCr 是很自然的,因为与 RGB 色彩空间相比,YCbCr 可以有效地将亮度和纹理(Y通道,亮度分量)与颜色(Cb 和 Cr 通道,色度分量)分开,这样细节增强和色

彩增强就可以分别在亮度和色度组件上进行。

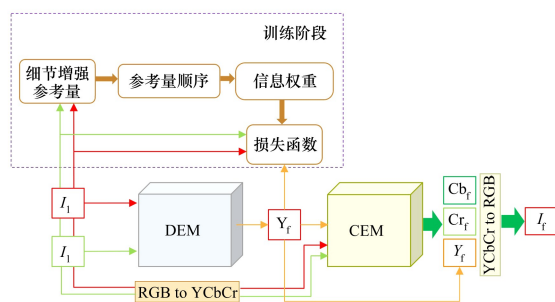


图6 感知增强流程

在细节增强模块中,给定图像 I ,在图像中可以轻松得到全局收益 $I^a = \alpha \times I$, α 为带曝光调整的比例系数。图像曝光升高时 $\alpha > 1$,降低时 $\alpha < 1$ 。图像中可能同时存在过度曝光、适度曝光和曝光不足的区域,通过全局调优 α 时并不能完全解决所有区域存在的曝光问题,因此需要找到一个自适应规则。引入一个标准差变量 σ ,在半径为 r 的窗口内,图像的每个位置为 I_{ij} ,这样标准差就可以看作是细节丰富度的衡量标准,其公式表示为

$$\sigma_{ij} = \frac{1}{2r+1} \sqrt{\sum_{-r \leq p, q \leq r} (I_{i+q} - \mu_{ij})^2} \quad (2)$$

在 $\alpha > 1$ 的预设集合中逐渐增加其标准差,将相应得到 $\sigma_{ij}^a = \alpha \times \sigma_{ij}$,到达合适的曝光。对于曝光不足的区域,上述操作增强其对比度,但无法处理过度曝光的区域。因此通过反转原图像获得 $I_{inv} = 1 - I$,最初过度曝光的区域看起来就像曝光不足的区域。将完全相同的规则应用于 I_{inv} ,便可获得双向的细节增强。

在色彩增强模块中,源图像被转换为 YCbCr 色彩空间,然后仅在 Y (亮度) 通道上执行融合策略,而 Cb 和 Cr (色度) 通道的融合规则仍然以直接的形式设计,如(3)式所示。

$$C_f = \frac{C_1(|C_1 - \tau|) + C_2(|C_2 - \tau|)}{|C_1 - \tau| + |C_2 - \tau|} \quad (3)$$

式中, C_1 和 C_2 表示输入图像的 Cb 和 Cr 通道, C_f 是相应的融合色度通道, τ 值一般设定为 128。当原图像曝光严重时,颜色信息会受到干扰甚至破坏,并且不同照明条件下的颜色也会不一致,因此 CEM 中通过获取 2 个源图像 (亮度和色度分量) 的全部信息以及目标亮度作为输入来推断最适合融合亮度 (由 DEM 生成) 的色度,如(4)式所示

$$[C_{1,f}, C_{2,f}] = N_{CEM}(I_1, I_2, Y_f, \theta_{CEM}) \quad (4)$$

式中: N_{CEM} 表示带有参数的 CEM 网络; θ_{CEM} 是需要学习的参数。CEM 设置为编码器-解码器联合结构, 每层有 4 个分支, 以探索输入图像之间的颜色映射关系。

2.3 评估损失函数的改进

MEF-SSIM 是 MEF 领域普遍适用的评估函数, 它的构建遵循 SSIM (结构相似性) 指数^[19] 的定义, 评价指标主要基于人眼查看图像时提取的结构化信息。SSIM 方法是一种全参考的评价方法, 对于图像 x 和 y , 其 SSIM 定义如(5)式所示。

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + N_1)(2\sigma_{xy} + N_2)}{(\mu_x^2 + \mu_y^2 + N_1)(\sigma_x^2 + \sigma_y^2 + N_2)} \quad (5)$$

式中: μ_x 和 μ_y 分别表示输入图像的平均强度; σ_x, σ_y 和 σ_{xy} 表示局部方差和协方差; N_1 和 N_2 是 2 个较小的常数, 以防止其不稳定。在本文方法中由于可以输入不同空间分辨率的图像, 而 SSIM 需要在特定配置下才能表现良好, 在灵活性方面没有较好的结果, 所以采用多尺度结构相似度 (multi-scale SSIM, MSSSIM)^[20], 作为评估函数进行优化替代。

MSSSIM 采用了多尺度分块的方法, 对于不同分辨率都能保持性能稳定。将对比图像进行输入后, 依次使用低通滤波器进行 1/2 下采样。假设原始图像为 Scale 1, 最高尺度为 Scale M 经过 $M-1$ 次迭代得到。对于第 j 尺度, 只计算对比度 $C(x, y)$ 和结构相似度 $S(x, y)$ 。通过 Scale M , 计算亮度相似度 $I(x, y)$, 最终的 MSSSIM 是将各个尺度的结果连接起来。

$$\text{MSSSIM}(x, y) = [I_M(x, y)]^{\alpha_M} * \prod_{j=1}^M [C_j(x, y)]^{\beta_j} * [S_j(x, y)]^{\gamma_j} \quad (6)$$

式中, α, β 和 γ 用于调整各分量的权重。图像细节的可感知性取决于图像信号的采样密度、图像平面到观察者的距离以及观察者视觉系统的感知能力。在实验中, 当这些因素发生变化时, 对给定图像的主观评价就会发生变化。因此 MSSSIM 的评估结果可以更贴近主观质量评估结果。

2.4 改进后的网络结构

本文使用了 DCGAN 的方法扩充了曝光金属图像的数量, 提升了网络的训练效果; 在 CAN 网络之前引入 DEM 和 CEM 模块, 使低分辨率图像进行特征预测时保留更多的纹理细节和色彩细节; 将评估函数 SSIM 替换为 MSSSIM, 更加注重图像整体的质

量, 提高网络的感知能力。整体改进后的多曝光图像融合网络结构如图 7 所示。

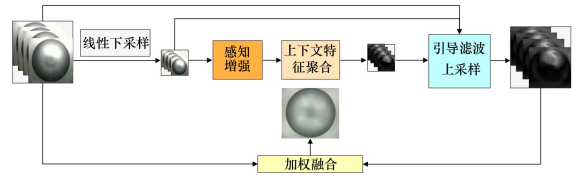


图 7 改进后的网络结构图

3 实验方案与结果

3.1 实验环境

本文的生成对抗网络以及多曝光图像融合网络均基于 Pytorch 深度学习框架而搭建, Python 版本为 3.9。网络训练采用的计算机操作系统为 Windows 10, CPU 为 6 核 12 线程的 Intel (R) Core (TM) i5-12400F, 图形处理器 (GPU) 采用 NVIDIA RTX 3070, 显存 12 GB。

3.2 实验参数设定

本文在训练中使用了自适应动量估计算法 (Adam) 作为其求解器, 学习率为 0.000 1, 动量参数 β_1 为 0.5, Adam 中的其他参数是默认情况下设置的。Batch-size 等于当前序列中的曝光次数, 即图像数量。epoch 设置为 200, 前 150 个 epoch 保持初始学习率 0.001 不变, 后 50 个 epoch 采取线性衰减方式至 0。对于损失权重的权值, 经过多次实验对比后设置 λ_1 为 15, λ_2 为 0.5。

作为一个无监督的模型, 本文中实验所用到的数据集均不需要进行标注, 同一图像的不同曝光数据放在相同目录下, 通过 Pytorch 中的 Dataset 工具包自动转化成数据流作为训练以及测试的输入。

3.3 评估指标

在图像融合领域, 评估方法分为 2 类: 一是定性分析, 通过主观的视觉质量来进行对比; 二是定量客观分析。目前在大多数 MEF 方法中用到的客观评估指标包括多曝光图像融合结构相似度 (multi-exposure fusion structural similarity index measure, MEF-SSIM)、信息熵 (entropy, EN)、互信息 (mutual information, MI) 等^[21], 详细如表 2 所示。

本文方法针对融合金属图像的特征信息和色彩信息进行了优化, 因此主要对应 MEF-SSIM、AG 以及 MI 3 个指标进行评估, 用于反映方法对细节和色彩的保留程度。

表 2 图像融合评估指标	
指标	作用
MEF-SSIM	图像感知质量评估
EN	衡量图像中信息的丰富程度
AG	表征图像的清晰程度
STD	表述像素灰度值的离散程度
SF	反映图像灰度变化率
MI	衡量融合图像中包含源图像序列的信息量

MEF-SSIM 是 MEF 领域被广泛应用的评估指标,它基于结构相似性方法原理和补丁结构一致性来计算度量指标,主要用于 MEF 结果的感知质量评估。该指标反映了融合图像中保留输入图像局部细节的程度,其数值越大代表融合结果越好。感知质量评价 S 与整体结构相似度定义为

$$S(\{x_k\},y)=\frac{2\sigma_{xy}+C}{\sigma_x^2+\sigma_y^2+C}\tag{7}$$

$$Q(Y)=\frac{1}{M}\sum_{j=1}^MS(\{x_k\}(j),y(j))\tag{8}$$

式中: σ_x^2, σ_y^2 分别代表 x, y 向量的局部方差; σ_{xy} 为 x 与 y 的局部协方差; M 为分割的图像块总数; $Q(Y)$ 是最终计算的整体结构相似度指标。

除上述对融合图像质量的评估指标外,本文引入了计算复杂度(computational complexity, CC)和融合时间(running time, RT),用于反映模型的规模以及速度。计算复杂度中假设输入通道的数量为 K , 每个图像都包含 M 像素,用于计算本地统计信息的窗口大小为 N^2 ,若模型只对整张输入图像进行 n 次处理,则计算复杂度为 $nKMN^2$ 。融合时间则为在输入相同数量及空间分辨率的曝光图像后,模型输出图像的运行时间。

3.4 结果对比分析

为了评估本文所改进的 MEF-Net 模型的高光去除效果,选取了部分带有曲面特征的金属高光图像进行实验,这部分图像与本文实验实际针对对象有相似的一些特征,其中包含字符信息、细小纹理以及大曲面等。将本文方法和原 MEF-Net 以及当前一些主流 MEF 方法,包括深度学习 MEF-GAN、U2Fusion 和传统 MGFF 方法进行对比。实验结果如图 8~11 所示,输入高光图像为 4 张不同曝光角度的金属高光图像。

从图中可以看出,传统 MGFF 方法在实际处理金属高光时表现很差,丢失了大量纹理信息,在复杂

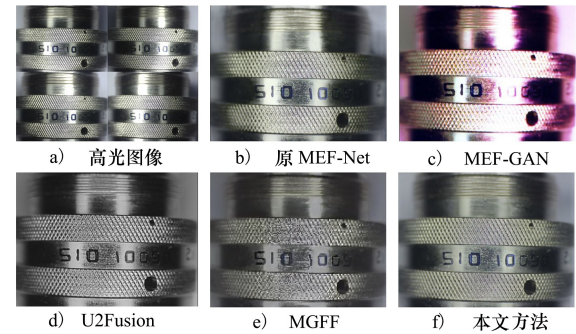


图 8 带字符金属高光图像融合结果

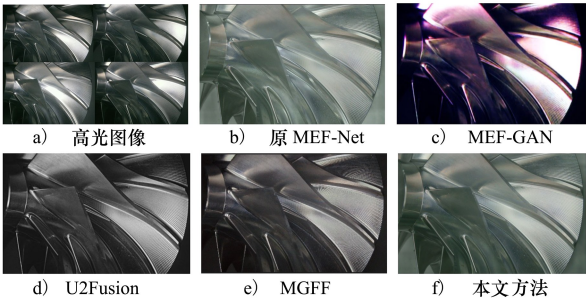


图 9 带复杂纹理金属高光图像融合结果

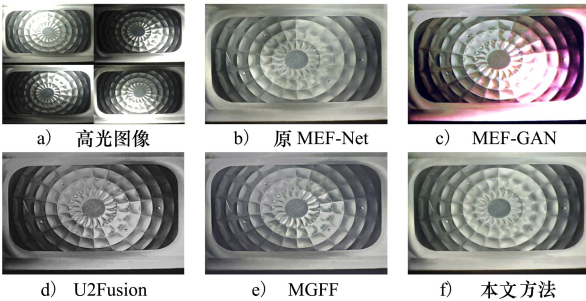


图 10 复杂曲面类金属高光图像融合结果

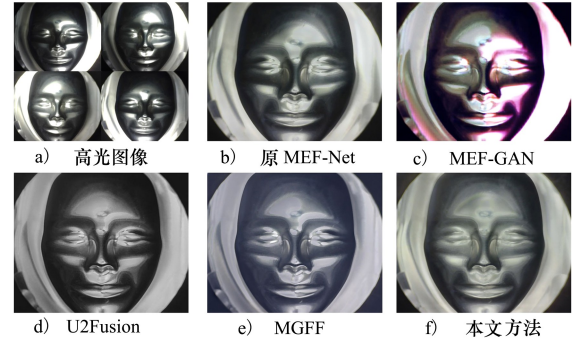


图 11 人脸曲面类金属高光图像融合结果

纹理的融合图像中产生了异常像素,在人脸这类大曲面金属工件上还产生了严重的色彩失真。MEF-GAN 方法在金属高光图像中的融合结果不仅在色彩上严重的失真,高光区域还存在特征信息丢失的情况,高光去除效果差。

U2Fusion 在融合图像的细节保留上表现良好,但由于其网络特性,只能产生单通道的灰度图,如果使用 OpenCV 对结果图像进行色彩转换也会出现一定程度失真。相比于原 MEF-Net,本文改进后的方法保留了高光区域更多纹理细节的同时,避免了失真,整体高光去除效果优于其他方法,无异常像素产生,一定程度上还原了图像无高光的场景。

表 3 中选取了部分评估指数在高光金属数据集下的表现,SSIM、MI 和 AG 参数指数主要反映融合图像与原图像之间信息的关系,时间和 CC(计算复杂度)参数用于评估模型的实时性和规模大小。如表中数据所示,MEF-GAN 方法在高光金属数据集中的表现普遍较差且模型规模最大,传统 MGFF 方法速度最快,因为其只对像素做了多尺度的算法运算,融合质量一般。相较于原 MEF-Net 方法,本文方法在基本不牺牲速度的情况下,整体评估质量提升了 10%左右,进一步表明了本文方法在金属高光区域有着较强的细节保留能力。

表 3 高光金属数据集中不同方法评估

方法	指标				
	时间/s	SSIM	MI	AG	CC
原 MEF-Net	0.82	0.620 9	2.571 6	3.531 5	KMN^2
MEF-GAN	1.5	0.344 5	1.284 5	2.632 0	$4KMN^2$
U2Fusion	1.3	0.637 2	2.712 2	3.947 2	$2KMN^2$
MGFF	0.7	0.626 1	2.536 5	3.184 6	KMN^2
本文方法	0.87	0.723 2	2.771 2	4.087 5	KMN^2

为探究本文方法在实时性方面所占优势的具体情况,选取了几种主流多曝光图像融合方法在不同分辨率图像下进行速度对比,实验结果如图 12 所示。图中横坐标代表不同分辨率,如 512s 表示 512×512 大小的图像,纵坐标表示融合所花费的时间,各种方法输入图像数量均为 2 张。

从图中的数据可以看出,在不同分辨率的图像融合中,本文当前的方法速度领先于其他主流图像

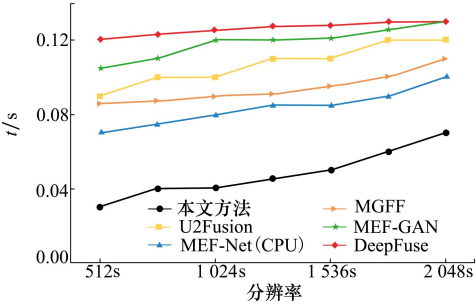


图 12 主多曝光图像融合方法速度对比

融合高光去除方法,并且 MEF-Net 得益于其端到端的特性,其在 CPU 模式下的运行速度也快于其他方法。

综上所述,本文改进后的 MEF 方法在金属表面的高光去除效果相比于当前的一些主流方法,无论是在视觉上还是指标上都有一定的优势,说明引入的 DEM 和 CEM 模块很好地解决了由于线性下采样所导致的纹理丢失问题,为 CAN 特征预测网络提供了良好的输入。此外,本文方法在速度上继承了 MEF-Net 的优势,远远快于其他方法,这对后续的实时性应用提供了很好的基础。

3.5 消融实验

为了验证 DPM(DEM 和 CEM)模块以及多尺度评估损失函数 MSSSIM 对融合图像质量的影响,本文将网络模型按照不同的组合在金属高光图像上进行了消融对比,分 3 组对照:第一组为原 MEF-Net 默认参数下的配置;第二组为引入 DPM 模块后,使用 SSIM 为评估函数;第三组为引入 DPM 模块,并且使用 MSSSIM 替换 SSIM 作为评估函数,实验结果如图 13 所示。

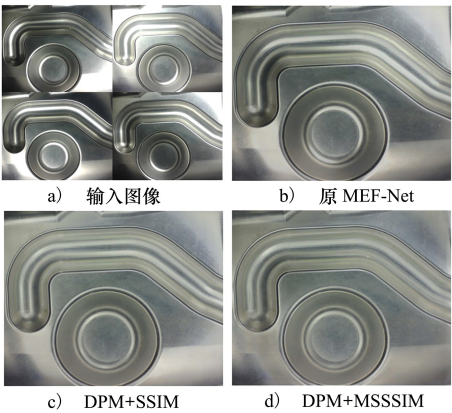


图 13 消融实验结果

从图中可以看出,引入 DPM 模块后高光区域局部亮度更加平和,保留了极端高光区纹理特征。使用 MSSSIM 评估函数后,高光区域整体更加平滑,图像边缘更加清晰,极大提升高光去除效果的同时,对图像中划痕、细小凹坑等特征信息进行了有效还原。

表 4 展示了消融实验各组的客观评估数据,由表中数据可知,DPM 模块的引入提高了融合图像的信息保留程度,MSSSIM 的使用提升了图像的整体结构相似度。上述实验结果表明,本文改进的 MEF 方法从特征信息细节和整体结构相似度上进行了优

化,能够对金属表面高光去除,获得更高质量的融合图像。

表 4 消融实验客观评估比较

方法	指标		
	SSIM	MI	AG
原 MEF-Net	0.620 9	2.571 6	3.531 5
DPM+SSIM	0.626 1	2.715 0	3.184 6
DPM+MSSSIM	0.723 2	2.771 2	4.087 5

4 结 论

本文针对金属表面高光问题,提出了一种以 MEF-Net 为基础,改进细节增强和色彩增强的 MEF 高光去除方法,使用 DCGAN 生成对抗的方法扩充

了金属高光数据集。在 CAN 特征提取网络前引入了 DEM 模块 CEM 模块,使其能保留高光区域更多的特征信息。最后使用多尺度的 MSSSIM 评估函数代替单尺度的 SSIM 损失,提高了融合图像的整体结构相似性。实验结果表明,在金属表面高光数据集上,本文改进的模型相较原 MEF-Net 模型,图像融合的视觉质量和客观评估指数均有所提升,其中反映图像中信息保留程度的评估指数互信息(MI)和平均梯度(AG)提升了 10%左右,并且保留了原模型的高运行速度。对比当前主流的 MEF 方法在高光去除效果以及速度方面均有明显优势。后续工作是继续优化网络结构,在满足实时性要求的同时,进一步提高图像的细节还原能力。

参考文献:

[1] SHAFER A S. Using color to separate reflection components[J]. Color Research & Application, 1985, 10(4): 210-218

[2] 祝新力, 张雅声, 方宇强, 等. 多曝光图像融合方法综述[J]. 激光与光电子学进展, 2023, 60(22): 2200003

ZHU Xinli, ZHANG Yasheng, FANG Yuqiang, et al. Review of multi-exposure image fusion methods[J]. Laser & Optoelectronics Progress, 2023, 60(22): 2200003 (in Chinese)

[3] FERIS S R, RASKAR R, TAN K, et al. Specular highlights detection and reduction with multi-flash photography[J]. Journal of the Brazilian Computer Society, 2006, 12(1): 35-42

[4] MA K, DUANMU Z, ZHU H, et al. Deep guided learning for fast multi-exposure image fusion[J]. IEEE Trans on Image Processing, 2020, 29(1): 2808-2819

[5] PRABHAKAR K R, SRIKAR V S, BABU R V. DeepFuse: a deep unsupervised approach for exposure fusion with extreme exposure image pairs[C]//IEEE International Conference on Computer Vision, 2017: 4724-4732

[6] XU H, MA J, ZHANG X P. MEF-GAN: multi-exposure image fusion via generative adversarial networks[J]. IEEE Transa on Image Processing, 2020, 29(5): 7203-7216

[7] AMIN-NAJI M, AGHAGOLZADEH A, EZOJI M. Ensemble of CNN for multi-focus image fusion[J]. Information Fusion, 2019 (51): 201-214

[8] XIANG Y, ZULQARNAIN S G, HANLIN Q, et al. Structural similarity loss for learning to fuse multi-focus images[J]. Sensors, 2020, 20(22): 6647-6664

[9] PENG F, ZHANG M, LAI S, et al. Deep HDR reconstruction of dynamic scenes[C]//IEEE International Conference on Image, Vision and Computing, 2018: 347-351

[10] ILG E, MAYER N, SAIKIA T, et al. FlowNet 2.0: evolution of optical flow estimation with deep networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2462-2470

[11] YAN Q, GONG D, SHI Q, et al. Attention-guided network for ghost-free high dynamic range imaging[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 1751-1760

[12] HAN D, LI L, GUO X, et al. Multi-exposure image fusion via deep perceptual enhancement[J]. Information Fusion, 2022, 79: 248-262

[13] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks[J/OL]. (2016-01-07) [2023-05-22]. <http://arxiv.org/abs/1511.06434v1>

[14] MA K, ZENG K, WANG Z. Perceptual quality assessment for multi-exposure image fusion[J]. IEEE Trans on Image Processing, 2015, 24(11): 3345-3356

[15] FU J, LI W, DU J, et al. DSAGAN: a generative adversarial network based on dual-stream attention mechanism for anatomical and functional image fusion[J]. Information Sciences, 2021, 576: 484-506

[16] HE K, SUN J, TANG X. Guided image filtering[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2012, 35(6): 1397-1409

[17] 罗凌杰. 基于深度学习的图像去高光技术研究[D]. 杭州: 杭州电子科技大学, 2020
LUO Lingjie. Research on image highlight removal based on deep learning[D]. Hangzhou: Hangzhou Dianzi University, 2020 (in Chinese)

[18] XU J, LI Z, DU B, et al. Reluplex made more practical: Leaky ReLU[C]//2020 IEEE Symposium on Computers and Communications, 2020: 1-7

[19] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE Trans on Image Processing, 2004, 13(4): 600-612

[20] HUANG H, LIN L, TONG R, et al. Unet 3+: a full-scale connected unet for medical image segmentation[C]//2020 IEEE International Conference on Acoustics, Speech and Signal Processing, 2020: 1055-1059

[21] 张小利, 李雄飞, 李军. 融合图像质量评价指标的相关性分析及性能评估[J]. 自动化学报, 2014, 40(2): 306-315
ZHANG Xiaoli, LI Xiongfei, LI Jun. Validation and correlation analysis of metrics for evaluating performance of image fusion [J]. Acta Automatica Sinica, 2014, 40(2): 306-315 (in Chinese)

Research on metal surface specular removal algorithm based on unsupervised learning

CHENG Mao¹, LI Bingran², CHEN Weihua², HONG Haibo³, XUE Jing³

1.School of Software, Northwestern Polytechnical University, Xi'an 710072, China;
2.Department of Mechanical Engineering, Tsinghua University, Beijing 100084, China;
3.Shanghai Aerospace Precision Machinery Research Institute, Shanghai 201699, China

Abstract: The highlights on the surface of metal materials can seriously destroy the continuity of the image, produce certain false edges, and cause the texture details in the highlights area to weaken or even disappear, which interferes with the subsequent operations such as surface region segmentation and defect detection. Aiming at the low efficiency, high loss, easy distortion and difficult calibration of metal highlight data, an unsupervised perceptual enhancement network model based on the convolutional neural network (CNN) is proposed. Firstly, the method of generating antagonism is used to generate a large number of metal images with high-light feature information, which is used to increase the number of high-light metal image data sets in the training set. Secondly, a detail enhance model (DEM) and a color enhance model (CEM) are introduced into the context aggregation network to improve the feature detail retention rate in the low resolution weight graphs. Finally, the multi-scale structural similarity function is used to replace the original structural similarity function to solve the insensitive detail when the image size is too large. Experiments show that comparing with other multi-exposure image fusion models, the present model can improve the evaluation index of mutual information and average gradient of fused images by about 10%, and can retain more texture feature information.

Keywords: highlight removal; unsupervised learning; MEF; image fusion

引用格式: 成懋, 李炳燃, 陈炜铎, 等. 基于无监督学习的金属表面高光去除算法研究[J]. 西北工业大学学报, 2024, 42(4): 744-752

CHENG Mao, LI Bingran, CHEN Weihua, et al. Research on metal surface specular removal algorithm based on unsupervised learning[J]. Journal of Northwestern Polytechnical University, 2024, 42(4): 744-752 (in Chinese)