arXiv:2302.11338v1 [cs.CV] 7 Feb 2023

基于深度学习的视觉水印去除

Rongfeng WEI 香港城市
大学rfwei2@my.cityu.edu.hk

摘要

近年来，随着互联网时代不断发展，在社交媒体上分享图像已成为常见现象。在某些情况下，人们会使用水印来保护图像的所有权；然而，更多时候，人们可能希望去除图像中的水印，以便在不受遮挡的情况下获取原始图像。在这项工作中，我们提出了一种基于深度学习的方法，用于视觉水印去除。受*U*形结构出色的图像转换性能启发，我们提出了一种名为*AdvancedUnet*的端到端深度神经网络模型，可同时提取并去除视觉水印。此外，我们采用了一些有效的*RSU*模块，取代*UNet*中常用的残差块，在不会显著增加计算成本的情况下加深了整体架构。深度监督混合损失引导网络以多尺度、三级层次学习输入图像与真实值之间的转换。对比实验表明了我们方法的有效性。

1. 引言

水印去除是一项开放且具有挑战性的问题，旨在从带水印的图像中重建背景图像。水印可以叠加在背景图像的任意位置，而背景图像的尺寸、形状、颜色和透明度各不相同。此外，水印通常包含复杂图案，例如变形符号、细线和阴影效果。去除带水印的图像并恢复原始图像是一项艰巨的任务。这些水印的结构、位置和大小因图像而异；若无人工指导或对底层图像作出假设，就很难检测带水印的图像并重建原始图像。

近年来，随着 GPU 的持续发展，计算机的计算能力也显著提升，从而推动了深度学习方法的快速进步，并催生了一系列基于卷积神经网络的技术。通过连续的卷积和池化操作，这些技术能够提取具有丰富语义信息的多层次上下文高级特征，并学习图像潜在的普遍规律，因此远超传统方法。

总而言之，作为图像处理任务的一个分支，图像水印去除具有重要的应用价值和社会价值。在当今信息社会蓬勃发展的时代，水印去除技术的进步也持续推动着水印技术的对抗性发展，这对网络信息安全领域意义重大。与传统方法相比，深度学习方法具有更强的鲁棒性，且更加简洁高效。本文以 CVPR2019 提出的 BVMR [1], 为基准，改进基于深度学习的“盲”水印去除方法，并取得了显著优于基准结果的实验成果。

我们的主要贡献可总结如下：

- 我们提出了一种名为 Ad-vancedUnet 的单阶段神经网络，可同时提取和去除视觉水印。
- 在 U 形结构的每个阶段中，我们都嵌入了一个高效的 RSU 模块，以替代 UNet 中的原始模块，使网络能够学习更深层的特征，并在多尺度场景下表现得更加稳健。
- 我们提出了一种新颖的深度监督混合损失，在每个阶段融合二元交叉熵、结构相似性指数和交并比，从而更准确地预测水印掩码。
- 我们的方法大幅超越了基线。

2. 相关工作

图像水印去除可以视为一种两阶段的多任务技术，用于目标检测和图像重建；也可以视为一种单阶段的图像转换任务。

2.1. 显著性检测

人类视觉系统具有有效的注意力机制，能够从视觉场景中选择最重要的信息。

对于水印检测这类特定任务，最能吸引人类注意力的是水印，计算机也同样如此。此外，水印检测应达到像素级别，为背景重建提供输入。因此，我们将水印检测视为显著性检测任务。Qin 等人 [2] 提出了一种新颖的边界感知显著目标检测网络：BASNet。该网络由一个深度监督的编码器-解码器和一个残差细化模块组成。此外，他们提出了一种融合二元交叉熵、结构相似性指数和交并比的新型混合损失，用于在像素级、图像块级和映射级这三个层面监督训练，从而实现准确的显著目标预测。Qin 等人 [3] 还提出了一种新颖的双层嵌套 U 形结构，使网络能够加深并获得更高分辨率，同时不会显著增加内存和计算开销。

2.2. 图像重建

图像修复的主要目的是为受损图像的缺失区域生成视觉上合理的结构和纹理。许多著名研究将水印去除视为图像到图像转换任务，并直接将带水印图像映射为无水印图像。

Cun 和 Pun [4]基于一项经验观察展开研究：不协调的外观只会出现在局部不协调的对象中，而这些对象与背景区域共享相同的语义信息和外观。因此，本文提出了一种名为“空间分离注意力模块”的新型注意力模块，通过硬编码掩码分别学习前景和背景区域的特征。 [5]是一篇综述，全面概述了基于深度学习的图像修复领域的最新进展，并详细分析了不同图像修复算法的性能。Zamir 等人 [6]提出了一种多阶段架构，逐步为退化输入学习复原函数，从而将整体恢复过程拆分为更易处理的步骤。

2.3. 水印去除

Hertz 等人 [1]提出了一种基于深度学习的水印盲去除技术。在盲设定下，水印的位置和精确几何形状均未知。该方法同时估计哪些像素包含视觉图案，并合成底层潜在图像。据我们所知，专门用于水印去除的深度学习方法为数不多，而这项工作是目前唯一一种不使用其他冗余模块、采用图像到图像转换方法的算法，因而相当简单且高效。但在后续研究中，研究人员通常认为细化模块对于水印去除任务不可或缺。Liu 等人 [7] 将其网络设计为 De-

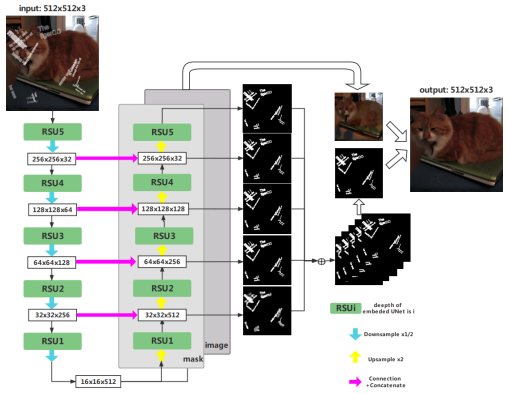


图 1. AdvancedUnet

compNet 和 RefineNet 的组合。 [8] 和 [9]还提出了一个两阶段框架，用以模拟检测、去除和细化的过程。

然而，两阶段模型必须以牺牲速度为代价来换取准确率，而且通常由许多复杂模块组成。本文延续 BVMR 简洁的设计理念，提出了一种名为 AdvancedUnet 的单阶段神经网络，可同时提取并去除视觉水印。实验表明，我们的方法大幅优于基线。

3. 数据

在图像水印去除技术中，我们需要使用水印的二值化掩码作为真实值。这类似于水印分割问题，需要进行非常精细的像素级人工提取，劳动强度很大。因此，本文使用自动水印图像合成方法来生成字符水印数据集。

在实验中，我们通过将视觉水印嵌入背景图像来合成水印图像，其中视觉水印简化为具有不同不透明度的随机字符串，背景图像则随机选自 Microsoft COCO val2014 数据集 [10]。将水印 (V_m) 抠图到图像 (I_m) 上可通过以下方式实现：

$$C_r = \alpha \circ V_m + (1 - \alpha) \circ I_m \tag{1}$$

其中， C_r 是合成受损图像， α 是空间变化透明度。

4. 方法

4.1. AdvancedUnet

图1展示了我们提出的图像水印去除算法模型 AdvancedUnet 的详细网络结构。

具体而言，该模型基于一种 Y 形网络结构，包含一个编码器和两个解码器，两个解码器分别输出预测的水印二值掩码和重建的非水印图像。输入为尺寸 512*512 的合成水印图像，特征图在 RSU模块中经过上采样、下采样、跳跃连接和卷积等操作，并且每经过一个 RSU模块，特征图的尺寸都保持不变。在编码器阶段，RSU模块输出后执行下采样操作，使特征图的长和宽减半，通道数加倍。在解码器阶段，RSU模块输出后，特征图与对应编码器阶段的编码输出在通道维度上连接，然后执行上采样操作。如图所示，对于二值水印掩码，我们将各阶段预测的二值掩码与真实值通过深度监督混合损失进行比较。

最终图像通过将原始受损图像 C_r 中估计的可视水印区域内的所有像素替换为$\widehat{\ }$重建图像 I_m中对应的像素来重建。最终生成的图像由重建图像 I_m 给出

$$Im_{final} = (1 - \widehat{Ma}) \circ C_r + \widehat{Ma} \circ \widehat{I_m} \tag{2}$$

，其中 ∘ 表示逐元素相乘 .

4.2. RSU模块

残差U形块（RSU） [3], 用于捕获阶段内的多尺度特征。 $RSU - L(C_{in}M, C_{out})$的结构如图2所示，其中 L 为编码器的层数， C_{in}， C_{out} 分别表示输入和输出通道，M 表示 RSU 内部层的通道数。在我们的模型架构中，RSU 的深度因阶段而异。这一设计使网络能够直接从每个残差块中提取多尺度特征；特征图分辨率越大，RSU 就越深。更值得注意的是，U形结构带来的计算开销很小，因为大多数操作都应用于下采样后的特征图。因此，我们采用 RSU模块来增强模型的特征提取能力，同时避免大幅增加计算开销。

4.3. 深度监督混合损失

训练过程中，我们采用了与 HED 类似的深度监督方法 [11], , 即为两个解码器的每一层都输出一个预测图，将预测图通过双线性插值调整至输入图的尺寸，并分别计算各阶段预测图以及各预测图融合图相对于真实图的损失值。

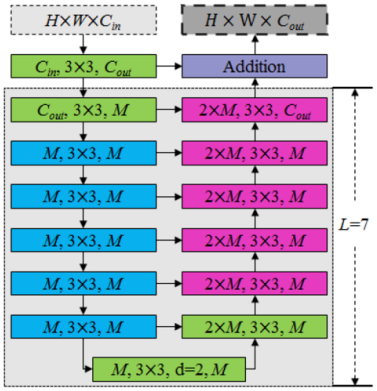


图 2. 残差U形块

与显著性目标检测和阴影检测问题类似，水印掩码检测也是一个二值划分问题，因此我们选择使用二元交叉熵损失，不仅计算各层的损失，还计算水印掩码解码器分支融合特征图的损失，以得到预测图的损失。为了增强模型在图像结构方面的鲁棒性，我们将结构相似性指数损失引入混合损失中。

混合损失结合二元交叉熵（BCE）、结构相似性指数（SSIM）和交并比损失，从像素级、图像块级和图级三个层次监督训练过程。

二值掩码的损失:

$$\begin{aligned} Loss_{mask}(M) = \sum_{l \in layers} & [L^{ssim}(M_l) + l^{bce}(M_l)] \\ & + L^{ssim}(M_{fuse}) + l^{bce}(M_{fuse}) \end{aligned} \tag{3}$$

重建图像的损失:

$$Loss_{image}(M) = \sum_{l \in layers} L^{ssim}(M) + L^{1}(M) \tag{4}$$

总损失:

$$Loss(M) = Loss_{mask}(M) + Loss_{image}(M) \tag{5}$$

5. 实验

5.1. 实现细节

我们使用 PyTorch [12] 和 CUDA v11.2 实现了我们的算法。在 Colab 平台上使用 Tesla P100 GPU 时，训练批次大小设为 8。为进行更公平的比较，我们采用与基线相同的优化器 [1], , 并且我们的项目基于 [1]。

在实验中，通过观察训练过程中各层解码器的输出损失折线图（如图 3 所示），我们发现 $loss_{mask1}$ 的波动最大，这与我们的算法流程一致：随着层级升高，损失波动逐渐减小。在较低层级，

方法/指标	MAE	mIoU	SSIM	PSNR
基线	0.023	0.748	0.966	28.041
Ours	0.010	0.868	0.976	39.762

表1。定量分析

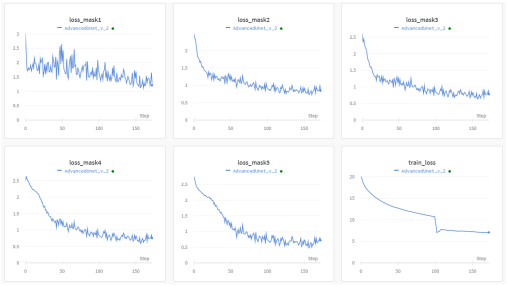


图3。各级输出掩码的损失示意图

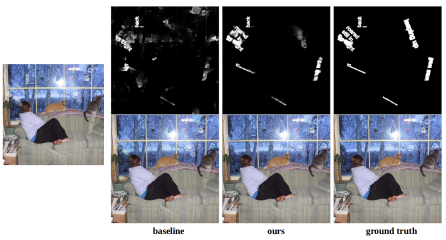


图4. 基线模型与我们的方法的预测掩码和重建图像对比

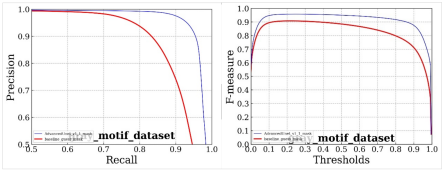


图5. 定量分析对比

特征图尺寸较小，其相对于真实图像的损失计算需要通过上采样，使尺寸与真实图像一致。这种上采样操作往往会造成巨大的精度损失，而且低层级获得的编码器跳跃连接输入补偿信息也较少，因此可靠性较低，所以在实际实验的损失函数计算中，我们会赋予 $loss_{mask1}$ 较小的权重。

5.2 与基线的比较

5.2.1 定性分析

实验结果可视化如图3所示，并从水印掩码和重建图像两个方面进行分析。

（1）基准模型 [1] 在水印掩码预测中能够大致预测水印的位置。但对于透明度较高的水印，它无法进行像素级预测，而且会产生大量伪影，如图4所示。相比之下，我们的模型能够更准确地预测水印的各个部分，边缘处理效果也更理想；但对于上图，在反射光严重的位置存在一处水印区域，导致模型预测受到干扰。

（2）基准模型能够去除部分水印，但仍残留大量水印；而我们的模型效果十分出色，几乎无需与带水印图像进行比较，就能让人无法分辨我们的算法重建图像与原始无水印图像。

5.2.2 定量分析

从具体的定量指标来看（如表 1 所示），我们的模型预测二值水印图像与真实二值水印图像之间的平均绝对误差为 0.010，较原始模型的 0.023 降低了 0.013；按百分比计算，优化幅度达 56%。同样，在水印二值掩码预测方面，从平均交并比指标可以明显看出，我们的模型取得了出色的结果。我们还计算了预测二值水印掩码与真实水印掩码之间的结构相似性指数（SSIM）。从上述数据可以看出，我们的模型输出的二值掩码伪影更少、边缘也更加平滑。我们还计算了重建图像与无水印原始图像之间的结构相似性指数和峰值信噪比，结果显示，相较于原始模型，两项指标分别提升了 1% 和 42%。

6. 结论

我们通过一系列实验验证了所提方法的有效性。本文提出了一种名为AdvancedUnet的单阶段神经网络，可同时提取并去除视觉水印，且性能远超基线。采用RSU的AdvancedUnet具备强大的多尺度逐像素特征提取能力，而新颖的损失函数也有助于生成更平滑的重建图像。除了盲水印去除，我们的方法未来还可应用于其他相关任务，例如阴影去除和反射去除。

参考文献

[1] Amir Hertz、Sharon Fogel、Rana Hanocka、Raja Giryes和Daniel Cohen-Or。单张图像中的盲视觉图案去除，2019年。1、2、3、4[2] Xuebin Qin、Zichen Zhang、Chenyang Huang、Chao Gao、Masood Dehghan和Martin Jagersand。BASNet：具备边界感知能力的显著目标检测。载于2019年IEEE/CVF计算机视觉与模式识别会议（CVPR），第7471–7481页，2019年。2[3] Xuebin Qin、Zichen Zhang、Chenyang Huang、Masood Dehghan、Osmar R. Zaiane和Martin Jagersand。U2-net：采用嵌套U形结构进行显著目标检测。《模式识别》，106：107404，2020年10月。2、3[4] Xiaodong Cun和Chi-Man Pun。通过空间分离注意力模块提升合成图像的协调性。《IEEE图像处理汇刊》，29：4759–4771，2020年。2[5] Hanyu Xiang、Qin Zou、Muhammad Ali Nawaz、Xianfeng Huang、Fan Zhang和Hongkai Yu。深度学习图像修复综述。《模式识别》，134：109046，2023年。2：。[6] Syed Waqas Zamir、Aditya Arora、Salman Khan、Munawar Hayat、Fahad Shahbaz Khan、Ming-Hsuan Yang和Ling Shao。多阶段渐进式图像复原。载于2021年IEEE/CVF计算机视觉与模式识别会议（CVPR），第14816–14826页，2021年。2[7] Yang Liu、Zhen Zhu和Xiang Bai。Wdnet：用于可见水印去除的水印分解网络，2020年。2[8] Xiaodong Cun和Chi-Man Pun。先拆分后细化：用于盲单图像可见水印去除的堆叠式注意力引导ResUNet，2020年。2[9] Jing Liang、Li Niu、Fengjun Guo、Teng Long和Liqing Zhang。通过自校准定位和背景细化去除可见水印。载于第29届ACM国际多媒体会议论文集。ACM，2021年10月。2[10] Tsung-Yi Lin、Michael Maire、Serge Belongie、Lubomir Bourdev、Ross Girshick、James Hays、Pietro Perona、Deva Ramanan、C. Lawrence Zitnick和Piotr Doll´ar。Microsoft COCO：上下文中的常见物体，2014年。2[11] Saining Xie和Zhuowen Tu。整体嵌套边缘检测，2015年。3[12] Adam Paszke、Sam Gross、Francisco Massa、Adam Lerer、James Bradbury、Gregory Chanan、Trevor Killeen、Zeming Lin、Natalia Gimelshein、Luca Antiga等。PyTorch：一种命令式、高性能深度学习库。《神经信息处理系统进展》，32，2019年。3