

基于轻量级 GAN 的实时视频图像去模糊模型^①



贾 凡, 张英俊, 谢斌红

(太原科技大学 计算机科学与技术学院, 太原 030024)

通讯作者: 贾 凡, E-mail: 707278645@qq.com

摘 要: 由于手持拍摄设备抖动或目标运动等原因, 使得视频图像资料产生运动模糊问题, 降低了人类的感知质量. 针对该问题从原来如何处理得到清晰图像, 到现在如何高效快捷的获得清晰图像, 提出了一种基于轻量级 GAN (Generative Adversarial Network) 的实时视频图像去模糊新模型. 该模型通过定义 PatchGAN 作为判别网络, 并在其基础上设置了全局图像和局部特征的双尺度判别器; 生成网络以轻量级 MobileNetV3 为主干网并引入特征金字塔进行特征提取, 以解决判别网络中特征信息利用率低以及生成网络推理效率慢的问题. 该模型采用端到端的方式对视频图像进行快速高效去模糊. 经过在 GoPro 和 Kohler 数据集上进行实验, 结果表明该模型去模糊后的锐利图像具有较高的峰值信噪比和结构相似度, 同时比其他模型的推理速度提高了 1.7-127 倍.

关键词: 实时视频图像去模糊; 轻量级; GAN; 特征金字塔; MobileNetV3

引用格式: 贾凡, 张英俊, 谢斌红. 基于轻量级 GAN 的实时视频图像去模糊模型. 计算机系统应用, 2021, 30(10):31-39. <http://www.c-s-a.org.cn/1003-3254/8104.html>

Real-Time Video Image Deblurring Model Based on Lightweight GAN

JIA Fan, ZHANG Ying-Jun, XIE Bin-Hong

(College of Computer Science and Technology, Taiyuan University of Science and Technology, Taiyuan 030024, China)

Abstract: Due to the shaking of a handheld camera or the movement of targets, the video image data is subject to motion blur, which reduces the image quality of human perception. With regard to the problem, from how to obtain clear images from the original process to how to obtain clear images efficiently, a new model for real-time video image deblurring based on the lightweight Generative Adversarial Network (GAN) is proposed in this study. The model defines PatchGAN as a discriminant network and sets up a dual-scale discriminator for global images and local features on the basis of it; the generation network takes lightweight MobileNetV3 as the backbone network and introduces a feature pyramid for feature extraction to solve the problem of low utilization of feature information in the discrimination network and low inference efficiency of the generation network. This model uses an end-to-end approach to efficiently deblur the video image. After experiments on the GoPro and Kohler datasets, the results show that the sharp image deblurred by this model has a high peak signal-to-noise ratio and great structural similarity, and the inference speed reaches 1.7-127 times faster than that of other models.

Key words: real-time video image deblurring; lightweight; Generative Adversarial Network (GAN); feature pyramid; MobileNetV3

① 基金项目: 山西省重点研发计划 (重点) 高新领域项目 (201703D111027); 山西省重点研发计划 (201803D121048, 201803D121055)

Foundation item: Major High-Tech Field Program of Key R & D Plan of Shanxi Province (201703D111027); Key R & D Plan of Shanxi Province (201803D121048, 201803D121055)

收稿时间: 2020-12-24; 修改时间: 2021-01-25; 采用时间: 2021-02-02

近年来,随着科技时代的快速发展,手机、摄像机等智能设备已经成为人们日常拍摄的主要工具,但在实际应用时,运动模糊通常会出现在手持智能设备拍摄的照片图像或包含运动物体的低帧速率视频中.图像模糊会降低人类的感知质量,并对后续的计算机视觉分析提出挑战.为此从模糊图像中恢复潜在的锐利图像不仅可以提高图像的视觉效果,还有利于其他的视觉任务,如目标检测和识别^[1].

拍摄过程中造成视频图像模糊的原因可以分为运动模糊、失焦模糊和高斯模糊3大类.其中运动模糊是造成视频图像退化的主要原因,也是研究的热点之一.在进行科学研究和实际问题的处理中,模糊模型可以表示为:

$$I_B = K_{(M)} * I_S + N \quad (1)$$

其中, I_B 是模糊图像, I_S 是锐化图像, $K_{(M)}$ 是由运动场 M 确定的未知模糊核, $*$ 表示卷积, N 是噪声. 根据模糊核 $K_{(M)}$ 是否已知, 运动视频图像去模糊方法可划分为非盲去模糊和盲去模糊两大类. 早期的工作主要集中在非盲去模糊, 根据相机运动的信息, 往往会假定一个运动视频图像的模糊函数(核), 以此为基础对模糊图像 I_B 进行反卷积, 从而得到去模糊的图像 I_S . 近几年, 许多非盲去模糊的方法都被提出, 如 Schmidt U^[2]、Ren^[3]、Richardson-Lucy^[4] 方法都利用各种数学方法模拟出退化机理(模糊核), 再反卷积对模糊图像进行重新处理. 但该类模拟出退化机理的方法需要花费大量时间推理, 并且很难找到一个模糊核满足所有像素点.

盲运动去模糊方法主要目标是如何有效地抑制噪声对视频图像的影响, 从而提高对视频图像的恢复质量. 在实际处理中, 盲去模糊算法的关键在于能够精确估算模糊的核函数. 早期盲去模糊算法依赖于自然图像的先验去模糊的方法. 2006年 Fergus 等^[5] 用混合高斯模型来约束条件服从自然图像梯度分布的规律, 利用变分贝叶斯方法获得模糊核函数, 然后使用 RL 算法对图像进行去模糊处理. 2013年 Xu 等^[6] 利用一种参数化的先验计算模型对每次迭代后的模糊核和锐化图像的进行估计改进. 这类方法的缺点是运算时的成本高, 并且这类方法在用来处理具有一个更大图像尺度的模糊动态图像时, 稳定性也会更差.

由于传统去模糊方法存在着计算量大、受噪声影响大、对模糊核估计要求较高等缺点, 因此基于深度学习的去模糊方法得到了迅速发展. 目前, 卷积神经网络

(Convolutional Neural Networks, CNN)^[7] 在图像的去模糊领域中已被广泛应用. 2015年 Sun 等^[8] 提出了利用 CNN 预测运动图像模糊概率分布的深度学习方法, 输入为已知模糊概率分布的模糊图像, 使得训练输出的模糊概率和已知先验信息的模糊概率同分布, 然后再反卷积得到清晰图像. 2017年 Nah 等^[9] 提出一种端到端的方式直接去除动态场景模糊的深度多尺度卷积网络; Ramakrishnan 等^[10] 利用条件生成对抗网络框架与卷积神经网络相结合的方法, 进行盲无核图像去模糊处理; Kupyn 等^[1] 用条件生成对抗网络和内容损失函数相结合的方法, 直接去除运动所产生的图像模糊. 这些方法避免了模糊核的估计问题, 直接端到端的生成锐化图像, 并且取得较好效果, 但缺点是推理速度慢, 难以达到对模糊视频图像的实时处理.

针对上述研究, 本文提出了一种基于轻量级生成对抗网络的实时视频图像去模糊新模型, 避免了对模糊核的估计, 以端到端的方式直接实现快速高效的视频图像复原.

1 理论基础

2014年, 由 Goodfellow 等^[11] 提出的生成对抗网络 (Generative Adversarial Network, GAN) 可以使用生成器生成逼真的高质量图像, 并为计算机视觉应用提供了新的技术和手段. GAN 提出的初衷是为了判断生成人工数据样本的质量, 目前已满足了许多领域的研究和应用需求^[12], 如语音和语言处理^[13]、图标设计^[14]、电脑病毒检测^[15] 等.

GAN 模型启发于博弈论中的纳什均衡^[16], 生成对抗网络模型如图 1 所示, 它包括生成网络 (Generative net, G) 和判别网络 (Discriminative net, D) 两个部分. 两个网络对抗训练交替进行, 生成网络接受噪声作为输入, 输出生成数据, 判别网络接受生成数据和真实数据进行更新参数进行优化时固定生成网络参数, 目的是将它们区分判断真假. 更新参数优化生成网络时固定判别网络参数, 目的是提高生成数据的可信度.

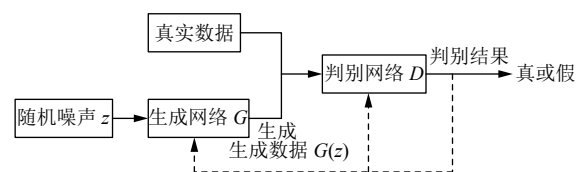


图 1 生成对抗网络模型

GAN 目标函数定义为:

$$\min_G \max_D V(D, G) = E_{x \sim P_{\text{data}}(x)} [\log(D(x))] + E_{z \sim P_z(z)} [\log(1 - D(G(z)))] \quad (2)$$

其中, D 表示判别器, G 表示生成器, E 表示期望, x 为真实数据, P_{data} 为生成器的数据分布, z 为随机噪声, P_z 为输入噪声的先验概率, \log 没有具体底数, 在运算过程中求导, 一般情况下要大于1, 保证结果与原数据正相关. 对于判别器, 当输入为真实数据时希望 $D(x)$ 值越接近1越好, 当输入为生成数据时希望 $D(G(z))$ 值越接近0越好, 两项合在一起希望最大化判别器的目标函数. 对于生成器, 期望生成的数据 $G(z)$ 输入到判别器时 $D(G(z))$ 值越大越好, 即 $1 - D(G(z))$ 越小越好, 意味着生成数据越逼近真实数据, 即最小化生成器目标函数.

2 轻量级高效模型

2.1 生成网络

本文引用 GAN 网络的思想, 为了达到实时高效的视频图像去模糊效果, 在生成网络部分选用轻量级 MobileNet, 该网络具有计算量少, 体积小, 精度高等优势. MobileNetV1^[17]中使用了大量的深度可分离卷积, 它是由一层深度卷积 (DepthWise convolution, DW) 和一层逐点卷积 (PointWise convolution, PW) 组合而成; MobileNetV2^[18]中引入了 Bottleneck 块, 该块第1层使用逐点卷积升维扩充通道数提取更多特征, 第2层使用深度卷积进行特征融合, 其激活函数均使用 ReLU6, 第3层使用逐点卷积降维, 降维后使用 Linear 激活函数; MobileNetV3^[19]分为 Large 和 Small 两个版本, 它们分别针对高资源和低资源的使用情况, 在 MobileNetV2 的基础上将 5×5 大小的深度卷积代替了部分 3×3 的深度卷积. 引入并改进 Squeeze-and-Excitation (SE) 模块^[20]使得有效的权重大, 无效或效果小的权重小, 可以训练出更好的模型, 并且减少了参数量和计算量, 还有部分 ReLU 激活函数被 H-Swish (HS) 激活函数替代以提高模型精度, MobileNetV3 (Small) 网络结构如表1, 其中, NL (NonLinear) 表示使用的非线性激活函数有 HS 和 RE 两种, NBN (No BatchNormalization) 表示不使用批量标准化, k 类表示 1×1 卷积抽取用于训练最后分类器的特征, 最后划分到 k 类.

另外, 在图像的去模糊领域里, 感受野的选取对于去模糊问题有着重要作用, 为了使得去模糊后的视频

图像纹理细节丰富和美观, 就需要通过选取多个感受野进行特征提取, 从而获得更多的图像上下文语义信息. 因此, 本文方法又加入了多尺度感受野模型特征金字塔网络 (Feature Pyramid Net, FPN). FPN 模型最初设计用于目标检测^[21], 它产生多个特征映射层, 这些特征映射层编码不同的语义, 并且包含更好的质量信息. FPN 有自下而上和自上而下的两条路径, 其中, 自下而上的路径是使用卷积网络提取不同尺度的特征, 在该过程中空间分辨率被下采样, 可以提取和压缩更多的语义信息. 另一方面, 通过自上往下的路径, FPN 从语义丰富的层重建出更高的空间分辨率. 自下而上和自上而下通道之间的横向连接补充了高分辨率细节, 有助于定位物体.

表1 MobileNetV3 (Small) 网络结构

Input	Operator	Output	SE	NL	s
$256^2 \times 3$	conv2d, 3×3	16	—	HS	2
$128^2 \times 16$	Bottleneck, 3×3	16	√	RE	2
$64^2 \times 16$	Bottleneck, 3×3	24	—	RE	2
$32^2 \times 24$	Bottleneck, 3×3	24	—	RE	1
$32^2 \times 24$	Bottleneck, 5×5	40	√	HS	2
$16^2 \times 40$	Bottleneck, 5×5	40	√	HS	1
$16^2 \times 40$	Bottleneck, 5×5	40	√	HS	1
$16^2 \times 40$	Bottleneck, 5×5	48	√	HS	1
$16^2 \times 48$	Bottleneck, 5×5	48	√	HS	1
$16^2 \times 48$	Bottleneck, 5×5	96	√	HS	2
$8^2 \times 96$	Bottleneck, 5×5	96	√	HS	1
$8^2 \times 96$	Bottleneck, 5×5	96	√	HS	1
$8^2 \times 96$	conv2d, 1×1	576	√	HS	1
$8^2 \times 576$	pool, 7×7	—	—	—	1
$1^2 \times 576$	conv2d, 1×1 , NBN	1024	—	HS	1
$1^2 \times 1024$	conv2d, 1×1 , NBN	k类	—	—	1

生成网络在 DeblurGAN 模型的基础上修改添加了多尺度特征 FPN 框架和轻量级 MobileNetV3 (Small) 主干网两部分, 主干网络一共有 16 层, 本文选取表1中的 5 层 (1、2、4、9、12) 提取图像特征, 并作为多尺度 FPN 模型不同尺度的特征信息来进行合并, 由此生成更清晰的去模糊图像. 本文模型利用 MobileNetV3 的轻量快捷和 FPN 重建的高分辨率丰富语义信息特点, 使生成网络在速度和精度之间取得了平衡, 生成网络模型如图2所示.

生成网络具体实现过程如下:

1) 如表1, 首先对输入的模糊图片 $256 \times 256 \times 3$ 进行 3×3 的有效卷积下采样, 得到第一个 $128 \times 128 \times 16$

的特征图 lateral0.

2) 随后再通过主干网提取 4 个不同尺度特征: $64 \times 64 \times 16$ (lateral1), $32 \times 32 \times 24$ (lateral2), $16 \times 16 \times 48$ (lateral3), $8 \times 8 \times 96$ (lateral4).

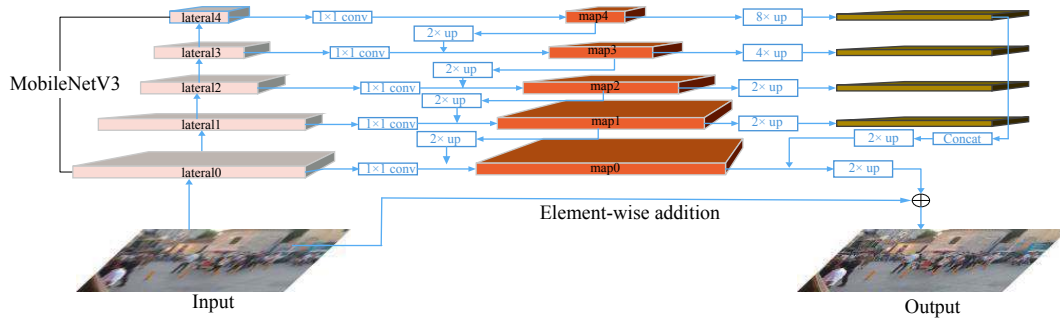


图2 生成网络模型

4) 将 map1–map4 进行 3×3 相同卷积改变通道数为 64 并上采样到原输入尺寸的 1/4 大小并连接成一个张量 ($64 \times 64 \times 256$), 其中包含不同级别的语义信息. 将该张量如同上述操作相同卷积后通道数变为 64.

5) 将该张量 2 倍上采样与 map0 进行特征融合后再次二倍上采样并进行 3×3 下相同卷积通道变为 3, 恢复原始图像大小并减少伪影.

6) 最后引入了一个从输入到输出的直接跳过连接, 以便学习集中在剩余部分.

2.2 判别网络

判别器网络的架构与 PatchGAN^[22] 相同. 为了使得判别网络在辅助生成网络生成更加逼真的去模糊图像, 本文在 PatchGAN 上设置了双尺度判别器, 分别为尺寸 70×70 的局部特征判别器和 256×256 的全局图像判别器来产生两个对抗性损失, 以便捕获图像纹理的局部连续性和图像的全局特征, 从而为生成网络提供更加丰富的信息, 判别网络如图 3 所示.

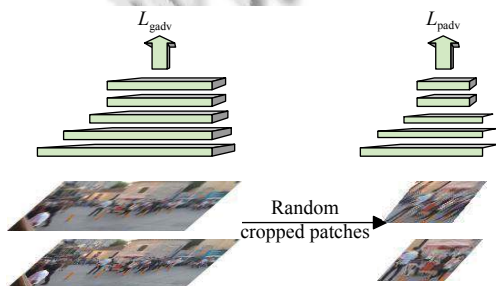


图3 判别网络模型

2.3 损失函数

本文将像素空间损失、内容损失和对抗损失相结

3) 将 lateral1–lateral4 进行 1×1 的卷积改变特征通道数为 128, 并进行 2 倍上采样, 与上一层的特征 lateral 融合分别得到特征图 map1–map4, 将 lateral0 进行 1×1 的卷积改变特征通道数为 64 map0.

合训练网络. 其中 L_{adv} 包含了局部判别器损失 (L_{padv}) 和全局判别器损失 (L_{gadv}). a 、 b 、 c 为权重因子, 用于平衡各项损失函数, 总体损失函数定义为:

$$L = a \times L_P + b \times L_C + c \times L_{adv} \quad (3)$$

(1) 对抗损失

GAN 早期的优化对抗损失是基于 JS (Jensen Shannon divergence) 和 KL (Kullback Leibler divergence), 但是这种优化方式会使训练过程变得不稳定, 并且很容易导致梯度消失和模型崩溃. 针对上述问题, 2017 年 Gulrajani 等^[23] 提出 WGAN-GP, 使网络训练收敛速度加快, 产生更多的纹理细节信息和视觉更好的锐利图片. 2018 年 Jolicœur-Martineau 认为对抗损失中真实数据的概率应该是一个逐步下降的过程, 假数据的概率应该是一个逐步上升的过程, 这是以往标准 GAN 所欠缺的. 故此提出了相对论判别器 (relativistic discriminator)^[24], 本论文采用当中具有梯度惩罚的 RaGAN-LS 对抗损失来训练:

$$L_{adv}^{RaGAN-LS} = E_{S \sim P_{sharp}(S)} \left[(D(S) - E_{B \sim P_{blurred}(B)} D(G(B)) - 1)^2 \right] + E_{B \sim P_{blurred}(B)} \left[(D(G(B)) - E_{S \sim P_{sharp}(S)} D(S) + 1)^2 \right] \quad (4)$$

其中, S 表示清晰图像, P 表示概率, $G(B)$ 表示生成器生成的图像, $S \sim P_{sharp}(S)$ 表示在真实清晰图像集中取样, $B \sim P_{blurred}(B)$ 表示在生成图像集中取样, E 为期望值, 相比于其他 GAN 相比, RaGAN-LS 可以显著提高网络训练速度和稳定性, 所产生的结果具有更高的感知质量和更清晰的输出.

(2) 像素空间损失

像素空间损失用于保证生成图像与真实图像之间的相似性, 选用 L_2 损失 (MSE) 有助于纠正颜色和纹理失真. 像素空间损失函数定义如下:

$$L_p = \frac{1}{WH} \sum_{x=1}^W \sum_{y=1}^H (S_{x,y} - G(B)_{x,y})^2 \quad (5)$$

(3) 内容损失

像素空间损失往往会产生过于平滑的像素空间输出, 本文使用感知距离作为内容损失 L_C 的一种形式. 它将生成的图像和真实图像送入训练好的 VGG19 网络中, 与 L_2 相比通过计算第 3 个池化层前的第 3 个卷积层的特征差异来计算内容损失, 使得产生的图像与目标清晰图像在语义上更为相似. 内容损失函数定义为:

$$L_{\text{content}} = \frac{1}{W_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\varphi_{i,j}(S)_{x,y} - \varphi_{i,j}(G(B))_{x,y})^2 \quad (6)$$

其中, S 表示清晰图像, B 表示模糊图像, $W_{i,j}$ 、 $H_{i,j}$ 是特征的维度, 表示特征映射图的宽和高, 是第 i 个池化层之前的第 j 个卷积所获得的特征映射.

3 实验与结果分析

3.1 数据集

GoPro 数据集^[9] 使用 GoPro Hero 4 相机每秒捕获 240 帧 (fps) 视频序列, 随后通过平均连续的短曝光帧生成模糊图像. 包含 3214 个模糊、清晰图像对, 使用 2103 对进行训练, 其余 1111 对进行评估测试.

Kohler 数据集^[25] 由 4 幅图像组成, 每一幅图像都使用 12 种不同的模糊核进行模糊处理. 它是评估盲去模糊方法的标准数据集. 该数据集是通过记录和分析实际的相机运动生成的, 并在机器人平台上进行回放, 从而记录了一系列清晰的图像, 以对 6D 相机的运动轨迹进行采样.

DVD 数据集^[26] 包含手机、运动相机, 以及单反相机以 240 fps 的速度拍摄的 71 段视频, 然后通过平均连续短曝光帧来近似长曝光, 生成了 6708 个模糊和清晰对.

3.2 实验环境

本实验使用的软件环境为 Ubuntu 16.04.6 操作系统、PyTorch 1.5 深度学习框架、Python 3.7 语言版本;

硬件环境主要有 NVIDIA GeForce RTX 2080Ti 显卡, 11 GB 显存, Intel(R)Xeon(R)Silver4210CPU@2.20 GHz 处理器, 10 GB 内存.

3.3 训练细节

网络训练采用的数据集通过从 GoPro 数据集里 2103 对和 DVD 数据集 6708 对混合组成训练集, 来减少对任何特定数据集的过度拟合. 用于测试的数据集都为通用数据集, 分别为 GoPro 数据集 (1111 对), Kohler 数据集 (48 对).

模型 epoch 设置为 300, 式 (3) 中的权重因子 a 、 b 和 c 分别设为 0.5、0.006 和 0.01. 采用 Adam 优化器^[27] 来训练模型, batchsize 设置为 1, 学习率初始化为 10^{-4} , 在训练 150 个 epoch 后学习率线性衰减到 10^{-7} . 整个模型训练近 112 h. 采用客观量化的质量评价指标, 即峰值信噪比 (Peak Signal to Noise Ratio, PSNR) 和结构相似度 (Structural SIMilarity, SSIM)^[28], 其中, PSNR 衡量的是锐化 (生成) 图像与模糊图像在像素值上的差异, 值越高代表图像失真越小, 图像质量越高; 而 SSIM 从亮度、对比度、结构信息 3 方面来衡量两幅图之间的相似度, 其值越趋近于 1, 表明它们之间越相似.

训练过程的损失函数及 PSNR、SSIM 指标变化折线图如图 4 所示.

3.4 对比实验

本文将提出模型与近年来几种先进的深度学习算法模型进行比较, 包括文献 [8] 的模型、DG-wild^[1]、DeepDeblur^[9]、DG^[1] 和 IRv2^[29].

3.4.1 GoPro 数据集

本文所提出的模型在标准性能指标 (PSNR、SSIM)、推理速度 (在单个 GPU 上测量每个图像的平均运行时间) 以及模型大小与其他模型进行了比较, 结果如表 2 所示, 实验对比图如图 5 所示.

从表 2 可以看出, 本文去模糊模型在 PSNR 和 SSIM 值上都保持较高水平, 并且在推理速度有了显著提升. 表 2 中, 文献 [8] 的模型训练时所用的数据集是由模糊核和清晰图像卷积合成的, 这与真实场景中的模糊有很大不同, 因此应用到真实视频图像去模糊不能够有效恢复图像. 相比于使用 GoPro 数据集进行训练和测试的模型 DG-wild^[1], 本文模型的 PSNR 提升了 1.46 dB, SSIM 也有所提高. 文献 [9,29] 的模型、DG^[1] 与本文模型 PSNR 都达到了 28 以上, SSIM 达到了 0.9 以上, 说明本文模型也可以生成纹理信息丰富、与原始图像

相近的图像,与模型 DeepDeblur^[9] 相比推理速度提高了近 127 倍;与模型 IRv2^[29] 相比,本文所提出来的模型尺寸要小的多 (233 MB vs. 10.5 MB),同时推理速度提高了近 10 倍;与现有的轻量级网络 MobileNet 模型^[29] 相比,PSNR 和 SSIM 都有所提升,模型大小也占有优势,推理速度提升了近 1.7 倍.这表明本文模型可以对 29 fps 的视频进行实时帧去模糊处理.本文模型是能够同时实现速度和精度需求的去模糊模型.

3.4.2 Kohler 数据集

各模型在 Kohler 数据集上的测试结果如表 3 所示,实验对比图如图 6 所示.与 GoPro 结果类似,相比于文献 [8] 的传统去模糊模型,文献 [1,9,29] 的模型和本文

模型采用端到端的网络模型得到的去模糊图像 SSIM 较高,说明模型对图像去模糊效果依然显著. PSNR 提升并不明显、普遍较低,这是因为基于深度学习训练的端到端去模糊模型依赖于训练时的数据集,当测试数据集与训练数据集相关程度不高时,模型没有学习此类数据特征,导致重建的去模糊图像在纹理细节上较差.表 2 和表 3 中 DG-wild^[1] 和 DG^[1] 两个模型的测试结果相反是因为训练时数据集的分配比例不同所导致的,所以在端到端的去模糊网络模型训练中,训练数据是关键问题,在未来具有大量、准确训练对的数据情况下,所训练的网络模型泛化能力应有更大的提升.

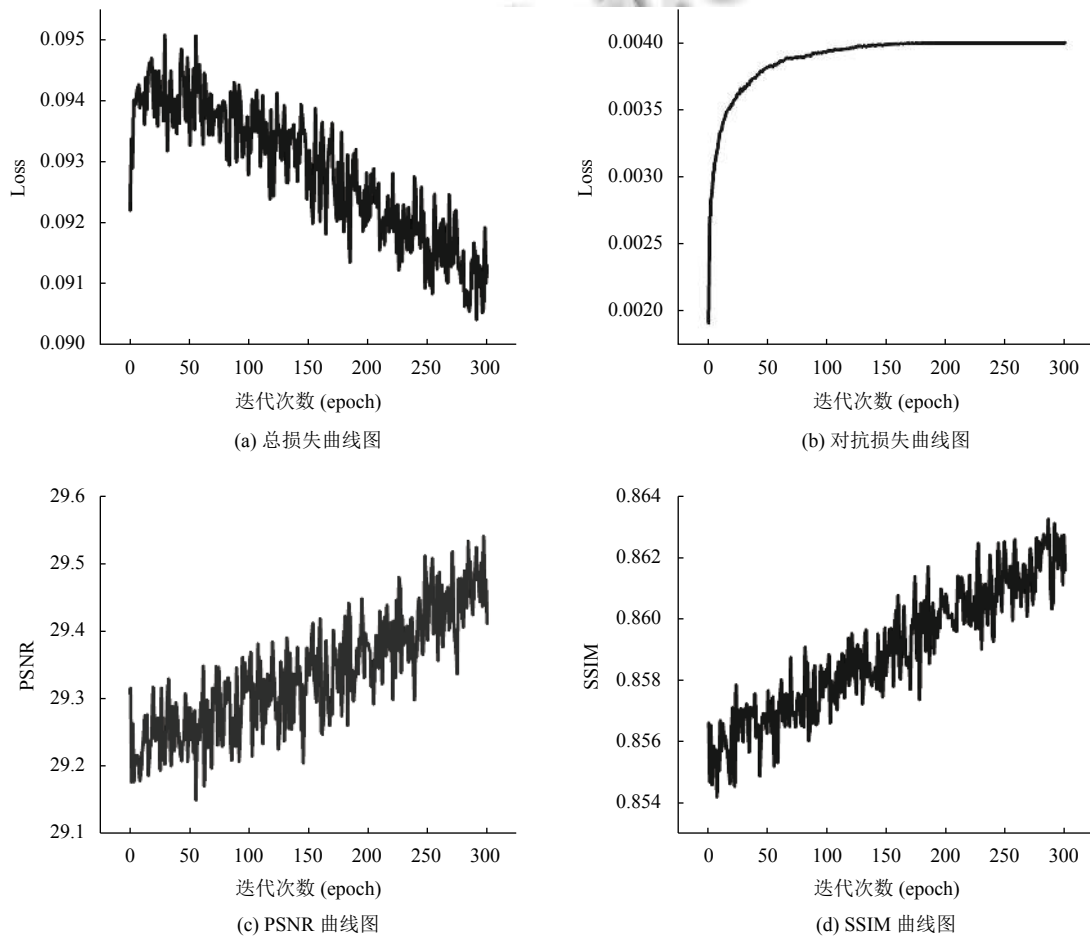


图 4 变化折线图

表 2 GoPro 数据集的实验对比结果

评价指标	文献[8]	DG-wild ^[1]	DeepDeblur ^[9]	DG ^[1]	IRv2 ^[29]	MobileNet ^[29]	本文方法
PSNR	24.64	26.82	29.23	28.70	29.55	28.17	28.28
SSIM	0.842	0.894	0.916	0.927	0.934	0.925	0.929
Time	20 min	0.85 s	4.33 s	0.85 s	0.35 s	0.06 s	0.034 s
Size (MB)	—	—	457	—	233	12.8	10.5

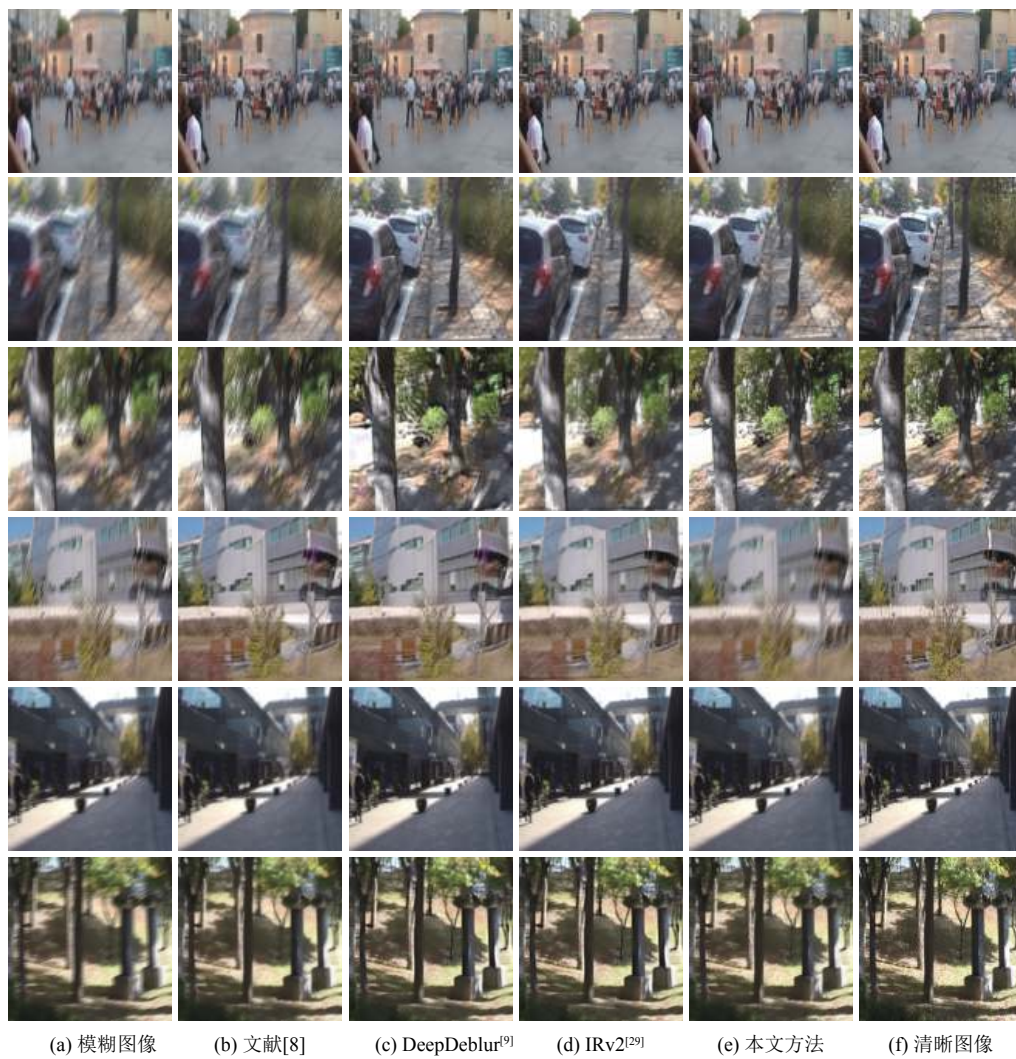


图5 实验对比图

表3 Kohler 数据集的实验对比结果

评价指标	文献[8]	DG-wild ^[1]	DeepDeblur ^[9]	DG ^[1]	IRv2 ^[29]	MobileNet ^[29]	本文方法
PSNR	25.22	26.10	26.48	25.86	26.72	26.36	26.05
SSIM	0.773	0.816	0.807	0.802	0.836	0.820	0.819



图6 实验对比图

3.4.3 消融实验

为验证本文模型各组成部分的有效性,在 GoPro 数据集进行了消融实验,起点为生成对抗网络(生成器部分 MobileNetV3、感知损失和 WGAN-GP,全局判别器)端到端直接去模糊,结果如表 4 所示。

表 4 本文方法的消融实验 (GoPro)

条件	PSNR	SSIM
DeblurGAN (MobileNetV3)	26.89	0.917
+FPN	27.92	0.922
+FPN+局部判别器	27.98	0.924
+FPN+局部判别器+RaGAN-LS	28.15	0.927
+FPN+局部判别器+RaGAN-LS+MES	28.28	0.929
移除感知损失 ($\alpha=0$)	27.13	0.914

从表 4 中可以看出本文提出的方法各部分都在稳步提升 PSNR 和 SSIM, FPN 模块的贡献最大,另外加入均方误差或感知损失有利于训练的稳定性 and 最终结果。

3.4.4 自行拍摄图像去模糊

许多经过训练好的模型应用到实际场景中,不能很好地获得去模糊效果,为了验证本文模型的实用性,在训练好的模型上对手机拍摄的校园模糊图像进行去模糊处理,结果如图 7 所示。

4 结论

本文主要研究单幅图像盲运动去模糊问题,针对的主要是由相机抖动或低帧率视频中运动物体引起的非均匀运动模糊。在 DeblurGAN 模型的基础上提出了一种基于轻量级生成对抗网络实时视频图像去模糊新方法,解决了早期去模糊算法计算复杂、推理效率慢以及恢复图像效果差等问题。该方法在 GAN 的生成器中引入轻量级网络 MobileNetV3 进行特征提取,并加入多尺度 FPN 框架进行特征融合,这两者结合使得生成网络在推理效率上有了显著提升。双判别网络可以为生成网络提供更加丰富的局部纹理细节和全局特征,更加准确的区分生成的去模糊图像和真实的清晰图像。实验结果表明,本文方法在 PSNR 和 SSIM 保持着较高的结果,同时速度比其它方法快 1.7–127 倍,并且网络模型仅 10.5 MB。训练测试时使用的图片类型为 jpg 或 png,本文的网络模型是完全卷积的,因此可以应用于任意尺寸大小的图像,另外还可以将本文方法用于车辆、人脸的去模糊任务中。未来工作一是进一步改进网络模型,在保持推理效率的同时,进一步提

升图像复原的质量;二是扩展该方法以实现实时视频增强。

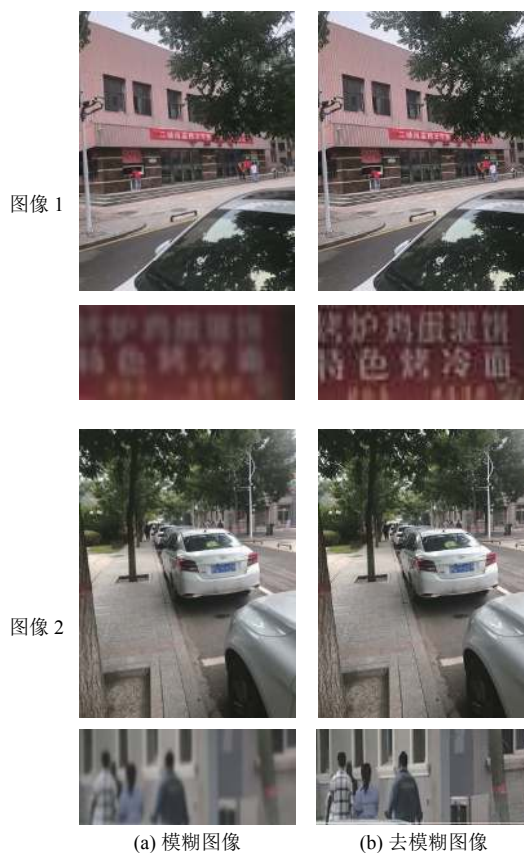


图 7 自行拍摄校园图像去模糊

参考文献

- 1 Kupyn O, Budzan V, Mykhailych M, *et al.* DeblurGAN: Blind motion deblurring using conditional adversarial networks. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 8183–8192.
- 2 Schmidt U, Rother C, Nowozin S, *et al.* Discriminative non-blind deblurring. Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland: IEEE, 2013. 604–611.
- 3 Ren WQ, Zhang JW, Ma L, *et al.* Deep non-blind deconvolution via generalized low-rank approximation. Proceedings of 32nd Conference on Neural Information Processing Systems. Red Hook: Curran Associates, Inc., 2018. 295–305.
- 4 Tai YW, Tan P, Brown MS. Richardson-Lucy deblurring for scenes under a projective motion path. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(8):

- 1603–1618. [doi: [10.1109/TPAMI.2010.222](https://doi.org/10.1109/TPAMI.2010.222)]
- 5 Fergus R, Singh B, Hertzmann A, *et al.* Removing camera shake from a single photograph. *ACM Transactions on Graphics*, 2006, 25(3): 787–794. [doi: [10.1145/1141911.1141956](https://doi.org/10.1145/1141911.1141956)]
 - 6 Xu L, Zheng SC, Jia JY. Unnatural L0 sparse representation for natural image deblurring. *Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition*. Portland: IEEE, 2013. 1107–1114.
 - 7 Razavian AS, Azizpour H, Sullivan J, *et al.* CNN features off-the-shelf: An astounding baseline for recognition. *Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*. Columbus: IEEE, 2014. 512–519.
 - 8 Sun J, Cao WF, Xu ZB, *et al.* Learning a convolutional neural network for non-uniform motion blur removal. *Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition*. Boston: IEEE, 2015. 769–777.
 - 9 Nah S, Kim TH, Lee KM. Deep multi-scale convolutional neural network for dynamic scene deblurring. *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017. 257–265.
 - 10 Ramakrishnan S, Pachori S, Gangopadhyay A, *et al.* Deep generative filter for motion deblurring. *Proceedings of 2017 IEEE International Conference on Computer Vision Workshops*. Venice: IEEE, 2017. 2993–3000.
 - 11 Goodfellow IJ, Pouget-Abadie J, Mirza M, *et al.* Generative adversarial nets. *Proceedings of the 27th International Conference on Neural Information Processing Systems*. Cambridge: MIT Press, 2014. 2672–2680.
 - 12 杜立平, 宋燕红, 李晓东, 等. GAN 在 AI 领域的应用研究. *北京电子科技学院学报*, 2018, 26(3): 85–93. [doi: [10.3969/j.issn.1672-464X.2018.03.013](https://doi.org/10.3969/j.issn.1672-464X.2018.03.013)]
 - 13 田艳梅. 基于 GAN 的音频关键词识别技术研究与应用 [硕士学位论文]. 成都: 电子科技大学, 2020.
 - 14 甘翔. 基于深度学习和生成对抗网络的图标设计研究 [硕士学位论文]. 广州: 华南理工大学, 2019.
 - 15 曹启云. 基于生成对抗网络的恶意代码识别研究与应用 [硕士学位论文]. 北京: 北京化工大学, 2018.
 - 16 Ratliff LJ, Burden SA, Sastry SS. Characterization and computation of local Nash equilibria in continuous games. *Proceedings of 2013 51st Annual Allerton Conference on Communication, Control, and Computing*. Monticello: IEEE, 2013. 917–924.
 - 17 Howard AG, Zhu M, Chen B, *et al.* MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv: 1704.04861v1, 2017.
 - 18 Sandler M, Howard A, Zhu ML, *et al.* MobileNetV2: Inverted residuals and linear bottlenecks. *Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 4510–4520.
 - 19 Howard A, Sandler M, Chen B, *et al.* Searching for MobileNetV3. *Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*. Seoul: IEEE, 2019. 1314–1324.
 - 20 Hu J, Shen L, Albanie S, *et al.* Squeeze-and-excitation networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(8): 2011–2023. [doi: [10.1109/TPAMI.2019.2913372](https://doi.org/10.1109/TPAMI.2019.2913372)]
 - 21 Lin TY, Dollár P, Girshick R, *et al.* Feature pyramid networks for object detection. *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017. 2117–2125.
 - 22 Isola P, Zhu JY, Zhou TH, *et al.* Image-to-image translation with conditional adversarial networks. *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017. 5967–5976.
 - 23 Gulrajani I, Ahmed F, Arjovsky M, *et al.* Improved Training of Wasserstein GANs. *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Red Hook: Curran Associates Inc., 2017. 5767–5777.
 - 24 Jolicoeur-Martineau A. The relativistic discriminator: A key element missing from standard GAN. arXiv: 1807.00734v3, 2018.
 - 25 Köhler R, Hirsch M, Mohler B, *et al.* Recording and playback of camera shake: Benchmarking blind deconvolution with a real-world database. *Proceedings of the 12th European Conference on Computer Vision*. Berlin Heidelberg: Springer, 2012. 27–40.
 - 26 Su SS, Delbracio M, Wang J, *et al.* Deep video deblurring for hand-held cameras. *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017. 237–246.
 - 27 Kingma D, Ba J. Adam: A method for stochastic optimization. arXiv: 1412.6980v9, 2017.
 - 28 Horé A, Ziou D. Image quality metrics: PSNR vs. SSIM. *Proceedings of 2010 20th International Conference on Pattern Recognition*. Istanbul: IEEE, 2010. 2366–2369.
 - 29 Kupyn O, Martyniuk T, Wu J, *et al.* DeblurGAN-v2: Deblurring (orders-of-magnitude) faster and better. *Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*. Seoul: IEEE, 2019. 8877–8886.