



计算机工程与应用  
Computer Engineering and Applications  
ISSN 1002-8331, CN 11-2127/TP

## 《计算机工程与应用》网络首发论文

题目: 基于生成对抗网络的运动模糊图像复原  
作者: 桑亮, 高爽, 尹增山  
网络首发日期: 2018-06-25  
引用格式: 桑亮, 高爽, 尹增山. 基于生成对抗网络的运动模糊图像复原. 计算机工程与应用. <http://kns.cnki.net/kcms/detail/11.2127.TP.20180622.1618.006.html>



**网络首发:** 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式 (包括网络呈现版式) 排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

**出版确认:** 纸质期刊编辑部通过与《中国学术期刊 (光盘版)》电子杂志社有限公司签约, 在《中国学术期刊 (网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊 (网络版)》是国家新闻出版广电总局批准的网络连续型出版物 (ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

# 基于生成对抗网络的运动模糊图像复原

桑 亮<sup>1,2,3,4</sup>, 高 爽<sup>4</sup>, 尹增山<sup>1,2,4</sup>

SANG Liang<sup>1,2,3,4</sup>, GAO Shuang<sup>4</sup>, YIN Zengshan<sup>1,2,4</sup>

1. 中国科学院 上海微系统与信息技术研究所, 上海 200050

2. 中国科学院大学, 北京 100049

3. 上海科技大学 信息科学与技术学院, 上海 201210

4. 上海微小卫星工程中心, 上海 201203

1. Shanghai Institute of Microsystem And Information Technology, Chinese Academy of Sciences, Shanghai 200050, China

2. University of Chinese Academy of Sciences, Beijing 100049, China

3. School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China

4. Shanghai Engineering Center for Microsatellites, Shanghai 201203, China

**SANG Liang, GAO Shuang, YIN Zengshan. Motion deblurring based on generative adversarial networks. Computer Engineering and Applications**

**Abstract:** Image motion blur is a very challenging problem caused by camera shaking or object movements. In order to tackle this problem, the paper proposes a deep convolutional neural network based on generative adversarial networks method. The proposed method can restore a clear image in an end-to-end way without estimating blur kernel. By introducing adversarial loss based on generative adversarial networks and modifying the residual network structure, the proposed method can restore image details effectively. Then we train this deep convolutional neural network model on public datasets. Finally, it is proved that the proposed method achieves good results according to the test on blurry image benchmark datasets.

**Key words:** motion blur; image restoration; generative adversarial networks; deep learning

**摘 要:** 针对相机成像时相机抖动、物体运动等导致图像产生运动模糊这一十分具有挑战性的问题, 本文提出基于生成对抗网络的深度卷积神经网络来复原模糊图像的解决方案。该方案省略了模糊核估计的过程, 采用端对端的方式直接获取复原图像; 通过引入生成对抗网络思想的对抗损失和对残差网络进行改进, 有效地复原了图像的细节信息。最后通过训练此深度卷积神经网络模型并在相关模糊复原基准数据集上测试, 证明了该方案取得了较好的结果。

**关键词:** 运动模糊; 图像复原; 生成对抗网络; 深度学习

文献标志码: A 中图分类号: TP751.1 doi: 10.3778/j.issn.1002-8331.1712-0048

**基金项目:** 中科院先导培育项目(No. ZDBS16ZRJ1)。

**作者简介:** 桑亮(1992 - ), 男, 在读硕士, 研究领域为图像处理, E-mail: sangliang@shanghaitech.edu.cn; 高爽(1986 - ), 女, 博士, 副研究员, 研究领域为遥感卫星总体技术; 尹增山(1971 - ), 男, 博士, 研究员, 研究领域为图像与信息处理技术。

## 1 引言

成像系统的局限性、环境的复杂性、目标物体的动态性和非配合性等诸多因素都会导致获取的图像为有强噪声、低品质和失真等特性的退化图像。图像模糊是一种典型的图像退化形式，焦距、相机抖动、目标物体的运动等因素都会造成图像产生模糊<sup>[1]</sup>。研究模糊图像的复原问题，即图像去模糊，从人眼视觉的角度来将有助于提升图像的视觉质量；从实际应用的角度来说将有助于改善视觉应用的鲁棒性，因此具有重要的实际应用价值。

早期的模糊图像复原研究往往假设退化模型已知，主要集中在数学模型和算法方面。图像模糊复原的方法一般是基于如下的产生模糊退化模型：

$$B = I \otimes K + N \quad (1)$$

其中， $B$  和  $I$  分别表示模糊图像和清晰图像， $\otimes$  为卷积操作， $K$  为模糊核或点扩散函数(PSF)，即图像的模糊核， $N$  为噪声。模糊图像复原的常规方法是从模糊图像中估计出模糊核，常见方法有正则化方法、基于边缘预测的方法等，然后再利用模糊核和模糊图像做解卷积，从而恢复出清晰图像  $I$ <sup>[2]</sup>。

根据模糊核是否已知，可将模糊图像的复原分为两大类：非盲图像复原和盲图像复原。非盲图像复原是在准确知道模糊退化模型中的某些先验知识，即模糊核已知的前提条件下对模糊图像进行复原。当图像成像的过程中，相机运动轨迹等条件未知的情况下，模糊核未知的运动模糊图像复原的过程叫做盲图像复原。由于缺乏先验信息，相比非盲图像复原具有更高的病态性。在上面的图像模糊退化建模中，图像中所有的像素点与同一个模糊核进行卷积操作，Levin<sup>[3]</sup>等人提出全局一致模糊核的假设并不能够充分描述由于相机的抖动等原因导致图像运动模糊的这一过程，实际生活中非均匀模糊图像更为普遍。因此，本文主要讨论解决非均匀模糊图像的盲模糊图像复原问题。

对于每个像素值找出相对应的模糊核是一个

严重的不适定问题，因此，许多方法把图像统计信息作为先验信息，尝试着对模糊退化模型带有一定的假设。近几年来，卷积神经网络 (Convolutional Neural Networks, CNN) 的方法被广泛应用到各种计算机视觉问题并展现出可喜的成果，其中也包含模糊图像复原问题。Xu<sup>[4]</sup>等人提出了一个非盲图像解卷积的卷积神经网络来进行去模糊。他们根据模糊核可分解的特性组成这个神经网络，这就使得这个神经网络要针对不同的模糊核进行训练。另一方面，Schuler<sup>[5]</sup>等人提出了一个基于卷积神经网络的盲去模糊方法。这一神经网络模仿了传统的优化图像解卷积能量的方法，用迭代的方法以粗糙到精细 (Coarse-to-Fine) 的方式求解出估计的模糊核和清晰图像。他们把清晰图像和用高斯模型生成的模糊核进行卷积以合成模糊和清晰图像对来作为数据集。但是，该文章中提到该神经网络对用较大的模糊核合成的模糊图像恢复效果不佳。类似于Cousin-Devy<sup>[6]</sup>等人的工作，Sun<sup>[7]</sup>等人提出了一个按次序的去模糊方法。首先，他们合成了许多用 73 个不同方向不同大小的线性模糊核卷积所得的模糊图像块和对应的清晰图像块组成一对的数据集；接着，他们训练了一个分类的卷积神经网络来衡量模糊图像块相对应的运动模糊核概率分布，进一步通过优化马尔可夫随机场模型获得图像逐点不同的运动模糊核。最后，基于上述估计的运动模糊核，通过传统的解卷积算法获得去模糊图像。

可以注意到，以上的方法都要估计出精确的模糊核来恢复潜在的清晰图像，但是，估计模糊核会带来一些问题。估计模糊核的方法对图像噪声和饱和很敏感，除非模糊的模型是非常仔细的设计过程<sup>[8]</sup>。此外，估计错误的模糊核会直接影响复原图像的质量，产生振铃现象；对每个像素找出对应的模糊核需要占用大量的计算资源和内存，计算复杂度较高<sup>[1]</sup>。因此，本文提出了基于生成对抗网络 (Generative Adversarial Networks, GAN) 的深度卷

神经网络模型采用端对端 ( end-to-end ) 的方式，由输入的一幅模糊图像直接获取复原图像，这可以有效避免估计模糊核带来的一些问题。深度学习的方法可以在给定模糊图像的情况下，有效的预测出模糊信息的概率分布并恢复潜在的清晰图像。

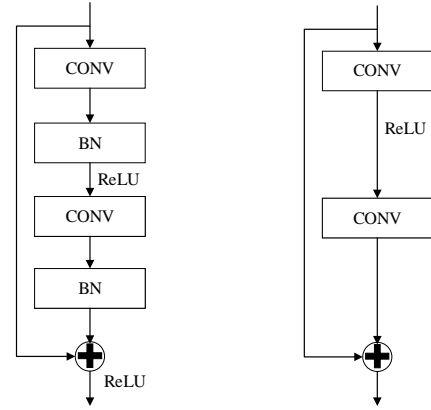
## 2 相关的思路及算法

### 2.1 模型建立

我们提出的网络结构由生成器 ( Generator ) 模型和判别器 ( Discriminator ) 模型组成。生成器的任务是学习图像数据样本的潜在分布并生成新的图像数据样本，并欺骗判别网络生成的图像数据样本来自真实的目标样本集。因此，我们可以在给定一幅模糊图像的条件下生成人眼视觉接受的模糊复原清晰图像。与此同时，判别器的任务是区分来自图像数据集中的图像样本和生成的图像样本。接下来我们会详细介绍我们的生成器和判别器模型。

#### 2.1.1 生成器模型

我们把模糊图像作为输入替换在原始的生成对抗网络中的随机噪声输入。与 He<sup>[9]</sup>等人不同，我们并不会降低输入图像信息的维度而是在整个模型网络中保持不变。虽然这一方法会导致对显卡显存有所要求，但可以避免模型中使用反卷积操作来生成图像所带来的棋盘效应 ( Checkerboard Artifacts )<sup>[10]</sup>。我们提出的生成器模型如表 1 所示，其中修改后残差块的结构如图 1 ( b ) 所示。其中 CONV、BN 和 ReLU 分别指卷积 ( Convolution ) 层、批规范化 ( Batch Normalization ) 层和修正线性单元 ( Rectified Linear Units ) 激活函数。



(a)基本的残差网络 (b)修改后的残差网络

图 1 基本的残差网络结构和修改后的版本

从表 1 中可以看出，我们提出的生成器模型主要采用了基本的残差网络结构<sup>[9]</sup>，并对此结构做了一些细微的调整。相比于其他的卷积神经网络模型，使用残差网络结构能够使训练的网络具有更深的结构。此外，模糊清晰图像对在像素值上非常接近，残差网络结构的能够有效的学习到两幅图像之间的映射关系。我们发现删去在的残差网络结构中快捷连接 ( Shortcut Connection ) 后的修正线性单元后，在训练时的收敛速度会加快。由于显卡显存的限制，训练时的批大小 ( Batch Size ) 设为 6，这要小于通常用到批规范化层的批大小，故删去了两个批规范化层。为了描述的方便性，我们把这个修改后的残差网络结构简称为残差块。

输入的图像块首先经过一个卷积层和修正线性单元会得到 64 个特征图 ( Feature Map )。接着，在这之后排列着 19 个残差块，最后接着一个卷积层和修正线性单元，把 64 个特征图转化为输入图像维度。通过设置每个卷积层的填充输入图像的边界 ( Padding ) 为 0，使得输入输出图像的分辨率保持不变。通过排列足够多数量的卷积层，此神经网络的感受野 ( Receptive Field ) 会逐渐增长到能学习到整个图像块 ( Patch ) 的信息。对所有的卷积层，卷积核的大小设为  $5 \times 5$ ，这虽然增加了参数量，却

带来更大的感受野，图像模糊复原的效果更好。

表 1 生成器网络结构

输出大小	网络层
$w \times h \times c$	输入图像
$w \times h \times 64$	$5 \times 5$ CONV, ReLU
$w \times h \times 64$	残差块
...	...
$w \times h \times 64$	残差块
$w \times h \times c$	$5 \times 5$ CONV, ReLU
$w \times h \times c$	输出图像

表 2 判别器网络结构

输出大小	网络层
$128 \times 128 \times 3$	输入图像
$128 \times 128 \times 32$	$3 \times 3$ CONV, LReLU
$64 \times 64 \times 32$	$3 \times 3$ CONV stride 2, LReLU
$64 \times 64 \times 64$	$3 \times 3$ CONV, LReLU
$32 \times 32 \times 64$	$3 \times 3$ CONV stride 2, LReLU
$32 \times 32 \times 128$	$3 \times 3$ CONV, LReLU
$16 \times 16 \times 128$	$3 \times 3$ CONV stride 2, LReLU
$16 \times 16 \times 256$	$3 \times 3$ CONV, LReLU
$8 \times 8 \times 256$	$3 \times 3$ CONV stride 2, LReLU
$8 \times 8 \times 512$	$3 \times 3$ CONV, LReLU
$4 \times 4 \times 512$	$3 \times 3$ CONV stride 2, LReLU
8192	Flatten
1024	FC, LReLU
1	FC, sigmoid
1	输出分类标签

### 2.1.2 判别器模型

判别器模型是一个二分类器，估计一个样本来自于训练数据（而非生成数据）的概率，如果样本

来自于真实的训练数据，判别器模型 D 输出大概率，否则，输出小概率。

判别器网络结构如表 2 所示，其中 CONV、FC 和 LReLU 分别指卷积（Convolution）层、全连接（Fully Connected）层和带泄露修正线性单元（Leaky Rectified Linear Units）激活函数。我们遵循 Radford<sup>[11]</sup>等人的设计样式，全部使用步长为 2 和滤波器大小为  $3 \times 3$  的卷积层和斜率为 0.2 的带泄露修正线性单元和来减小图像的空间维度直到特征图大小为  $4 \times 4$ ，接着连接两个全连接层和一个 sigmoid 激活函数得到一个 0 或 1 的分类标签。

## 2.2 损失方程

为了使最优化这个神经网络模型，我们使用的损失函数由内容损失和残差损失组成。

### 2.2.1 内容损失

从根本上来说，我们想要模型的输出恰好就是对应的清晰图像，因此，损失方程选用  $L_2$  损失或均方差表示：

$$L_{cont} = \|I_{est} - I_{gt}\|_2^2, \quad (2)$$





图 2 实验结果对比图

其中，

$$\|I\|_2^2 = \frac{1}{whc} \sum_{w,h,c} (I_{w,h,c})^2, \quad (3)$$

其中， $I_{est}$ 、 $I_{gt}$  分别指此模型输出的复原图像和对应的真实清晰图像， $w, h, c$  分别指输入模型图像的宽度、高度和通道数。内容损失通过除以图像所有像素数量值归一化。

### 2.2.2 残差损失

传统上来说，图像复原任务往往采用复原图像和真实图像的  $L_1$  或  $L_2$  损失作为目标方程。但是，在深度卷积神经网络只用  $L_1$  或  $L_2$  损失会导致复原图像过于平滑，这是因为像素级别的误差方程趋向于在所有可能的解的平均值处收敛<sup>[13]</sup>。这会使图像丢失锐利的边缘，而且边缘处的模糊依然大部分存在。因此，我们采用把内容损失和对抗损失结合起来共同作为损失方程。

最近，生成对抗网络在生成逼真的图像领域取得了十足的进展<sup>[11]</sup>，目前已经成为人工智能学界一个热门的研究方向。生成对抗网络的基本思想源自博弈论的二人零和博弈，由一个生成器和一个判别器构成，通过对抗学习的方式来训练。在原始的设定中，生成器网络  $G$  被训练来学习，从服从某种分布的随机向量  $z$  到图像数据空间  $x$  的映射，即估计图像数据样本的潜在分布并生成新的图像数据样本，与此同时，一个判别器网络  $D$  被训练来区分来自图像数据集中的图像样本和生成的图像样本。训练生成器网络是为了最小化：

$$L_A = -\log_2(D(G(z))), \quad (4)$$

而判别器网络是为了最小化：

$$L_D = -\log_2(D(x)) - \log_2(1 - D(G(z))), \quad (5)$$

这导致了一个极小极大化的问题，将两者结合起来，得到残差损失定义为：

$$L_{adv} = E_{S \sim p_{gr}(S)} (\log_2 D(S)) + E_{B \sim p_{blurry}(B)} (\log_2 (1 - D(G(B)))) \quad (6)$$

其中， $G$  和  $D$  分别指 2.1 节中的生成器模型和判别器模型， $S$  和  $B$  分别指对应的清晰和模糊图像对。

### 3 实验分析

#### 3.1 实验设置

我们通过 Torch7 框架来实现我们的模型。所有的实验在一台配有 E5-1620V4 处理器和 NVIDIA GTX Titan X 显卡的工作站上实现。

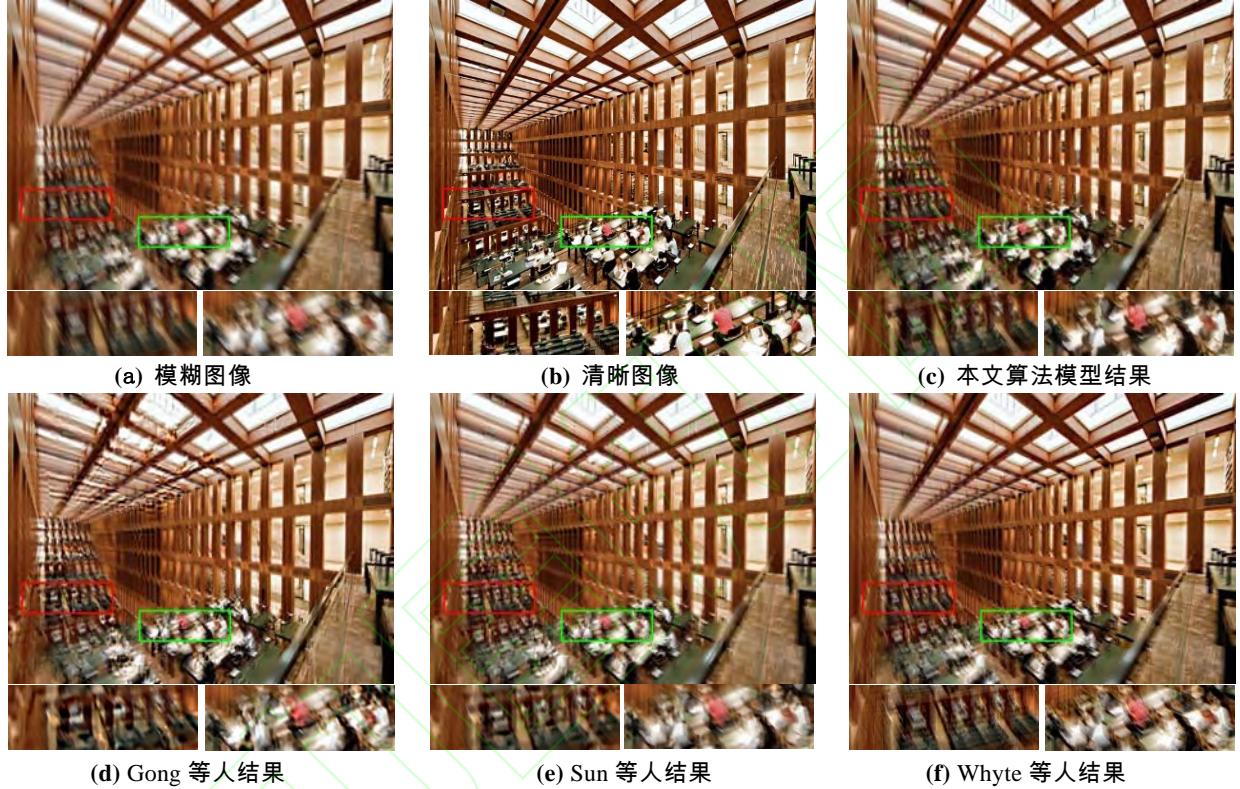


图 3 实验结果对比图

表 3 图像复原质量评价结果

测试图像	评价标准	Whyte	Sun	Gong	本文算法模型
GoPro 数据集	PSNR ( dB )	23.6169	25.8298	26.4322	29.9762
	SSIM	0.8017	0.8894	0.8994	0.9182
Lai 等人数据集	PSNR ( dB )	13.6199	13.5328	13.9127	13.9519
	SSIM	0.5833	0.6072	0.6262	0.6325

模型参数：我们通过 ADAM 算法优化损失方程，并通过随机梯度下降 (SGD) 算法使模型收敛。在实验中，我们设定全局的学习率 (learning rate) 为，动量 (momentum) 为 0.9。经过  $6.0 \times 10^5$  次迭代后此卷积神经网络模型收敛。

#### 3.2 数据集

为了训练我们的模型，我们采用了 Nah<sup>[8]</sup>等人公开的 GoPro 数据集。输入神经网络模型的一对模糊清晰图像块的分辨率大小为  $128 \times 128 \times 3$ ，是从训练数据集对中随机截取的。



为了防止训练的过程中过拟合,我们采取了下列的方法。为了防止图像的几何变换,截取的图像块被旋转随机的角度或水平、垂直翻转。图像的RGB通道也被随机打乱。考虑到图像退化的影响,HSV颜色模型中的饱和度被乘上一个在0.5和1.5之间的随机数。此外,随机高斯噪声也被加到模糊图像。最终,经过上述方法的图像的像素值被调整到0与1之间。

我们和其他具有代表性的非均匀模糊复原方法在一些公开的基准(benchmark)数据集上的结果进行比较。我们选取了比较流行的盲非均匀模糊图像复原算法,分别是Whyte<sup>[13]</sup>等人、Sun<sup>[7]</sup>等人和Gong<sup>[14]</sup>等人公开的代码。基准数据集如下:

(1) GoPro数据集:Nah等人公开的GoPro数据集通过累加GoPro拍摄的高速视频中的视频帧来模仿真实模糊图像的生成过程。整个数据集包含了3214对模糊清晰图像对,其中训练集包含2103对而剩下的作为测试集。我们使用此训练集训练我们的模型并选取剩下的测试集展示此方案的结果。

(2) Lai<sup>[15]</sup>等人生成的数据集:包含均匀和非均匀模糊核的真实和合成的图像并采用了一些常见的图像退化方法。而用于合成模糊图像的非均匀模糊核是按照Köhler<sup>[16]</sup>等人的文章中的方法生成的,他们是用6D相机拍摄时记录的运动轨迹回放生成模糊核。

实验结果如图2、3所示,其中图2为GoPro数据集的结果,图3为Lai<sup>[15]</sup>数据集的结果。我们的结果在主观视觉效果中均取得了较好的效果。可以注意到,Sun<sup>[7]</sup>等人和Gong<sup>[14]</sup>等人的方法在锐利边缘处或模糊是非线性的区域复原效果不是很成功,而Whyte<sup>[13]</sup>等人的结果包含了很严重的振铃现象。

表3为两组实验的图像质量评价结果,从表中的量化指标中可以看出,本文所提出的深度卷积神经网络模型不论是峰值信噪比(PSNR)或结构相似度(SSIM)都要优于其他三个算法。因此,本文所提

出的模型在主观视觉效果和图像质量评价结果中均取得了较好的效果。

## 4 结束语

针对相机成像时相机抖动、物体运动、景深等会导致运动模糊的问题,本文提出了一种基于生成对抗网络的端对端的盲图像复原方法。与传统的方法不同,我们的模型省略了模糊核估计的过程,这可以避免模糊核估计以及估计错误带来的问题。按照生成对抗网络的思想,整个模型由生成网络和判别网络组成。生成网络由经过修改的残差网络块组成,而判别网络参考Radford<sup>[11]</sup>等人经验设定。通过实验分析对比,本文所提出的模型可以有效地复原出了一幅较为清晰的高质量图像。

## 参考文献:

- [1] 程俊廷,左旺孟.快速非均匀模糊图像的盲复原模型[J].黑龙江科技大学学报,2017,27(2):196-199.
- [2] 张广明,高爽,尹增山等.基于模糊图像和噪声图像的遥感图像运动模糊复原方法[J].电子设计工程,2017,25(18):82-86.
- [3] Levin A, Weiss Y, Durand Y, Freeman W. Understanding and evaluating blind deconvolution algorithms[C]//IEEE Conference on Computer Vision and Pattern Recognition,2009:1964-1971.
- [4] Xu L, Ren J, Liu C, Jia J. Deep convolutional neural network for image deconvolution[C]//Advances in Neural Information Processing Systems,2014:1790-1798.
- [5] Schuler C, Hirsch M, Harmeling S, Schölkopf B. Learning to deblur [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2016,38(7):1439-1451.
- [6] Couzinie-Devy F, Sun J, Alahari K, Ponce J. Learning to estimate and remove non-uniform image blur[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2013:1075-1082.
- [7] Sun J, Cao W, Xu Z, Ponce J. Learning a convolutional neural network for non-uniform motion blur removal[C]// IEEE Conference on Computer Vision and Pattern Recognition,2015:769-777.
- [8] Nah S, Kim T, Lee K. Deep multi-scale convolutional neural network for dynamic scene deblurring[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2017:3883-3891.



- 
- 
- [9] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition,2016:770-778.
- [10] Sajjadi M, Schölkopf B, Hirsch M. EnhanceNet: single image super-resolution through automated texture sythesis [C]//IEEE International Conference on Computer Vision, 2017:4491-4500.
- [11] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[C]//International Conference on Learning Representations,2016:780-789.
- [12] Ramakrishnan1 S, Pachori S, Gangopadhyay A, Raman S. Deep generative filter for motion deblurring[C]//IEEE International Conference on Computer Vision,2017:2993-3000.
- [13] Whyte O, Sivic J, Zisserman A, Ponce J. Non-uniform deblurring for shaken images [J]. International Journal of Computer Vision,2012,98(2):168–186.
- [14] Gong D, Yang J, Liu L, Zhang Y et al. From motion blur to motion flow: a deep learning solution for removing heterogeneous motion blur[C]//IEEE Conference on Computer Vision and Pattern Recognition,2017:3806-3815.
- [15] Lai W, Huang J, Hu Z, Ahuja N, Yang M. A comparative study for single image blind deblurring//IEEE Conference on Computer Vision and Pattern Recognition, 2016:1701-1709.
- [16] Köhler R, Hirsch M, Mohler M, Schölkopf B, Harmeling S. Recording and playback of camerashake: benchmarking blind deconvolution with a real-world database[C]// European Conference on Computer Vision, 2012:27-40.