DOI: 10.3969/j.issn.1009-6868.2018.03.013
网络出版地址：http://kns.cnki.net/kcms/detail/34.1228.TN.20180622.1739.002.html

基于深度卷积神经网络的视觉SLAM去模糊系统

Deep Convolutional Neural Network for Visual SLAM Deblurring

中图分类号：TN929.5 文献标志码：A 文章编号：1009-6868 (2018) 05-0062-005

摘要： 提出了一种高效的、基于深度卷积神经网络（CNN）的图像去模糊算法。网络结构基于条件生成对抗网络，并使用堆叠的自编码器结构与跳跃相连接。相关的试验结果表明：该算法有良好的图像去模糊效果，并且能够大幅度地降低时间与内存开销。

关键词： 图像去模糊；卷积神经网络；对抗生成网络

Abstract: In this paper, an efficient deep convolutional neural network (CNN)-based image deblurring method is proposed. The network architecture is based on conditional generative adversarial network integrated with stacked encoder-decoder architecture and skip connections. Experiment results show that the proposed method achieves good image deburring performance and in the meanwhile reduce the testing time and required memory resource.

Key words: image deburring; CNN; generative adversarial network

缪弘/MIAO Hong
张文强/ZHANG Wenqiang
（复旦大学，上海 200433）
(Fudan University, Shanghai 200433, China)

1 模糊对视觉SLAM的影响及图像去模糊简介

同时定位与地图构建（SLAMS）的目的是让机器人利用各类传感器信息来得知自身的位置以及周围的环境。因此，SLAM是实现机器人自主移动的一项关键技术。视觉SLAM是指利用视觉传感器的信息的SLAM系统，其输入就是视觉传感器得到的图像。

在机器人运行过程中，因为相机抖动、景物移动等原因，都会造成图像模糊。无论是特征点法还是直接法，模糊的图像输入都会直接影响视觉SLAM系统，降低系统整体的运行效率。视觉SLAM系统需要将拍摄的前后两帧图像进行匹配，根据匹配结果对自身进行定位，这一过程称为跟踪。在跟踪过程中，模糊的输入图像会造成匹配失准或无法匹配，这被称为跟踪失败。当出现跟踪失败时，需要让整个机器人停止运动或者回退，重新拍摄清晰的图像，同时需要进行全局的地图搜索，定位当前机器人的位置，直至跟踪成功，机器人再重新开始运动。全局的地图搜索是一个相对耗时的操作，如果频繁地触发这一操作，会影响整个SLAM系统的运行效率。同时，每次机器人停止运动或者回退，都使得运行过程变得不连续，影响了流畅性。因此，模糊的输入图像是需要避免的。为了避免模糊的输入图像，我们可以使用去模糊算法对图像进行处理，恢复出清晰的图像。

相机抖动、相机与景物之间的相对运动造成模糊一般被称为运动模糊。图像中的运动模糊效果通常在空间上是不均匀的，这是由于不同对象的运动经常是彼此不同的。取模糊算法的目的就是恢复出一张没有模糊的清晰的图像。以前的大部分方法都是通过这个模型来建模图像上的模糊：

$$B=K\times S+n \tag{1}$$

其中B，K，S和n分别是模糊的图像、模糊核、潜在的清晰图像和噪声。在去模糊问题中，模糊核是未知的。因此，这些方法需要在只有给定的模糊图像B同时估计模糊核K和潜在清晰图像S，这其实可以看为一个病态的问题。

实际上，真实世界模糊图像的模糊核往往在空间上不均匀。估计空间非均匀的模糊核是一个难题，因为每个像素的模糊核都可能不同。因此，以前的一些方法[1-4]都对模糊来源做了一些简单的假设，以简化模糊核估计。然而，由于实际的模糊核通常比所假设的模糊核更加复杂，所以通

收稿日期：2018-03-10
网络出版日期：2018-06-22

过这些方法估计的模糊核是不准确的。不准确的模糊核的估计直接会降低潜在的清晰图像的质量。因此，这些方法只适用于几种特定的模糊类型。

近年来，越来越多的方法使用卷积神经网络（CNN）来解决去模糊问题的方法[5-10]。由于缺乏真实场景下的模糊清晰图像对，文献[5-8]中的方法通过合成模糊核进行卷积来产生模糊图像进行训练。另外，这些方法不是以端到端的方式，并且仍然需要估计模糊核或逆模糊核。因此，这些方法仍然存在模糊核的估计不准确的问题，而且它们在真实模糊图像上的表现比人工生成的模糊图像要差。文献[9]提出了一个由高速摄像机拍摄的真实场景下的模糊清晰图像数据集，文献[9-10]中的模型在这个数据集上进行了训练。此外，两种方法都是以端对端的方式，直接生成清晰图像，没有进行模糊核的估计。因此，这两种方法在性能上都超越了以前的方法。然而，文献[9]中的方法运行缓慢，文献[10]中的方法相对较快，但仍需要大量内存资源，这使得人们很难在实践中应用这些方法。

基于上述的研究现状，我们提出了一种基于深度卷积神经网络的图像去模糊算法。算法整体基于条件对抗生成网络，在网络结构上使用堆叠的自编码器结构与跳跃连接。通过在基准数据集上的实验，算法表现出了良好的图像去模糊效果，并且能够大幅度地降低时间与内存开销。算法的高效性使其更容易与视觉SLAM系统相结合。

2 基于深度卷积神经网络的去模糊算法

2.1 网络结构

我们的网络是基于对抗生成网络设计的，包含有1个生成器和1个鉴别器。生成器的任务是从输入的模糊图像中提取特征，利用特征生成出一张足以“骗过”鉴别器的图像。鉴别器的任务是正确地判别出一张图像是真实的清晰图像，还是一张生成器生成出来的图像。通过让生成器和鉴别器互相对抗式地学习，生成器和鉴别器的能力都能得到提升，最终生成器能从一张输入的模糊图像中生成出一张真实的清晰图像。

生成器的网络结构如图1所示。生成器的结构包括3个部分：头部、中部与尾部。

头部只包含有一个5×5的卷积层。这个卷积层将3通道的输入图像映射为一个64通道的特征映射，作为生成器中部的基础。我们并没有在头部的卷积层后接一个激活层，这是因为我们在生成器中部的构建模块中使用了文献[11]中提到的前置激活层的方法，所以头部卷积层的激活层包含在了中部的构建模块中。

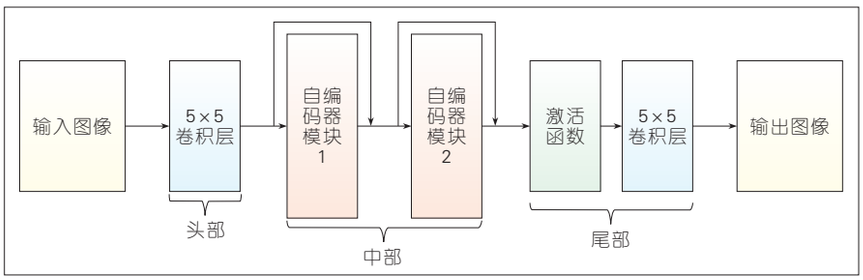
中部包含有连续N个构建模块，并且每一个构建模块都有一个残差连接。因为构建模块是基于自编码器结构的，所以我们把构建模块称为“自编码器模块”。因为输入的模糊图像与要输出的清晰图像在数值上很接近，所以网络所需要学习的函数比起零映射更接近与恒等映射，而带有残差连接的结构更容易学习到一个恒等映射[12]。我们选择将N个自编码器模块顺序地堆叠，因为这使得网络能够重复地从整张图像中提取特征。每一个自编码器模块只需要在输入的特征映射上做一点改进，最终就能得到一个足够好的特征映射。在实验中，我们选择N=2。

尾部包含有一个激活层和一个5×5的卷积层。尾部的任务是将中部产生的特征映射变换到最终的输出图像。在整个生成器中，我们都没有使用任何归一化层，因为我们发现添加归一化层反而会使得结果变差，同时会带来更大的时间与内存开销。

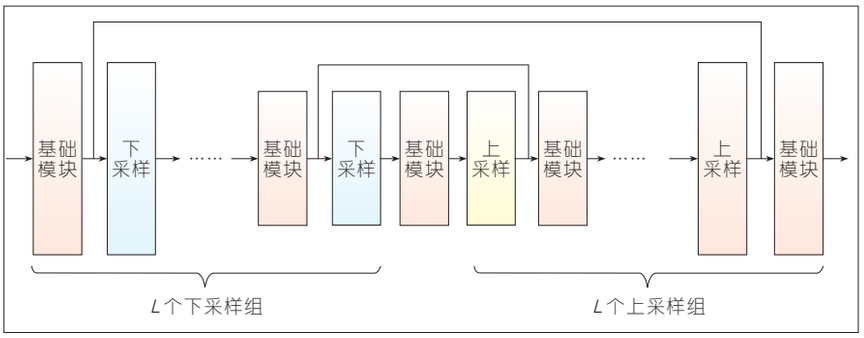
2.2 自编码器模块结构

自编码器模块的结构如图2所示。我们使用一种残差模块、最大池化层和最近邻插值层来构建自编码器模块。我们称这种残差模块为“基础模块”。基础模块能生成维度与其输入一样的特征映射。我们将一个基础模块和一个最大池化层定义为一组“上采样组”，将一个基础模块和一个最近邻插值层定义为一组“下采样组”。在自编码器模块中，输入的特征映射先经过L组下采样组不断下采样，直至到达瓶颈层（包含一个基础模块），然后再经过L组上采样组不断上采样。同时，我们在第i个最大池化层和第$L-i$个最近邻插值层中添加了跳跃连接，共L个跳跃连接。在实验中，我们选择L=4。

自编码模块的结构与Hourglass Network[13]和U-Net[14]的结构类似。类似自编码器的结构能够提取不同尺度的特征，而跳跃连接能够将它们组合起来。因为同一张图像，模糊的程度会随着尺度的降低而降低，所以不同尺度的特征可以用来处理不同程度的模糊。因为输入图像上各处的模糊程度都可能相同，所以提取不同尺度的特征是很重要的[15]。我们使用跳跃连接是因为跳跃连接能直接将网络的低层信息传递到网络的高



▲图1 生成器结构



▲图2 自编码器模块结构

层，这能让网络的输出共享低层信息。另外，跳跃连接还能直接将梯度信息从高层传递到低层，这会让网络的训练更加容易。

2.3 基础模块结构

基础模块的结构如图3所示。基础模块的输入与输出维度相同，我们将输入与输出的通道数定义为 C_{hr}。在一个基础模块中，共有 C 条路径。每一条路径包含两个卷积核大小为3×3的卷积层，并且在每个卷积层之前都有一个激活层。第1个卷积层的输出与第2的卷积层的输入通道数相同，都为 D。每条路径除了卷积层的膨胀系数都相同。C 条路径中，每个卷积层的膨胀系数从1增加到 C。基础模块也包含一个残差连接。我们将所有路径的输出与模块的输入相加，得到最后的输出。在实验中，我们选择 C_{hr}=64，C=4，D=16。

基础模块的结构设计受到了ResNeXt[16]中残差模块的启发。这2种模块都使用了残差连接，并且将多路操作聚合起来。但与ResNeXt中的残差模块不同的是：基础模块中每一路操作都不同，而ResNeXt中每一路操作都相同。每一路中使用不同的膨胀系数，可以在不增加参数量的情况下增大了感受域，同时还能提取到不同尺度的特征。

2.4 鉴别器结构

鉴别器是基于条件对抗生成网络设计的，需要两组图像对作为输入。一组图像对包含一张模糊图像与对应的清晰图像，另一对图像对包含模糊图像和对应的经生成器处理的图像。与传统的对抗生成网络相比，条件对抗生成网络的鉴别器需要一张额外的模糊图像作为输入。这样做的好处是在让生成器生成的图像“欺骗”鉴别器的同时，还能保持与输入的模糊图像的一致性。

鉴别器结构的设计参照PatchGAN[17]，只包含5个卷积层。鉴别器输出的是一个特征映射，特征映射中的每一个值都对应于输入图像中的一块。因此，比起整张图像，鉴别器更着重于局部的图像块，这会鼓励生成器去生成更清晰的局部边缘与结构。而且，浅层的鉴别器结构也能节约训练的时间。

2.5 损失函数

生成器的损失函数包含了 ℓ_1 损失函数和对抗损失函数。ℓ_1 损失函数常常被用于图像恢复任务，它可以让生成图像与目标图像的像素值更接近。然而，只使用 ℓ_1 损失函数会导致结果过于平滑。为了防止过于平滑，我们将对抗损失函数与 ℓ_1 相结合。我们没有使用文献[18]中使用的对抗损失函数形式，而是使用了最小二乘生成对抗网络（LS-GAN）[19]中的形式。鉴别器的对抗损失函数定义如式（2）：

$$\mathcal{L}_{adv}^{D}=\frac{1}{2}E_{x\in B,y\in S}\left[\left(D(x,y)-b\right)^2\right]+\frac{1}{2}E_{x\in B}\left[\left(D(x,G(x))-a\right)^2\right] \quad (2)$$

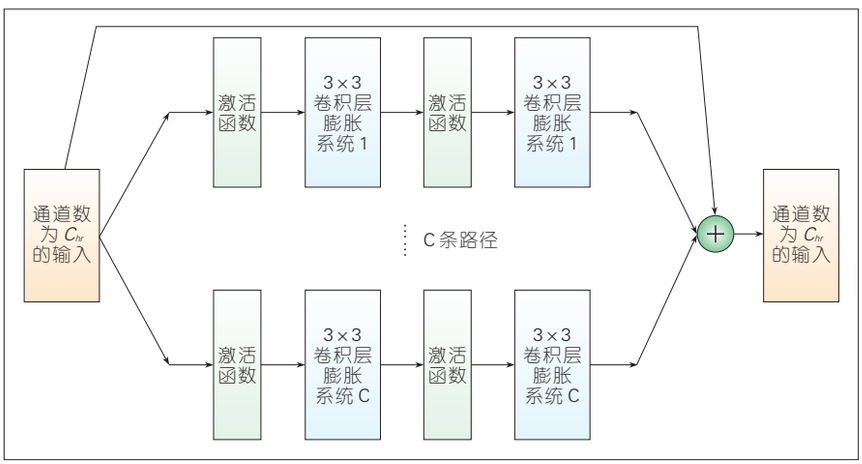
生成器的对抗损失函数定义如式（3）：

$$\mathcal{L}_{adv}^{G}=\frac{1}{2}E_{x\in B}\left[\left(D(x,G(x))-c\right)^2\right] \quad (3)$$

其中，B 和 S 分别表示模糊图像集合和清晰图像集合，a 和 b 分别表示生成图像和真实图像的标签，c 表示生成图片想要达到的标签。根据文献[19]中的设置，我们选择 a=0，b=1，c=1。与文献[18]中的对抗损失函数相比，LS-GAN中的形式在训练中更加稳定，更容易训练。最后，整体的损失函数如公式（4）所示：

$$\mathcal{L}_{total}=\mathcal{L}_{\ell_1}+\lambda\mathcal{L}_{adv}^{G} \quad (4)$$

在实验中，我们将权重系数设为



▲图3基础模块结构

$\lambda = 0.01$。

3 相关实验

所有的实验都是在同一台使用Titan XP显卡的工作站上进行的。我们的模型使用pytorch库来实现。

3.1 GOPRO数据集上的实验

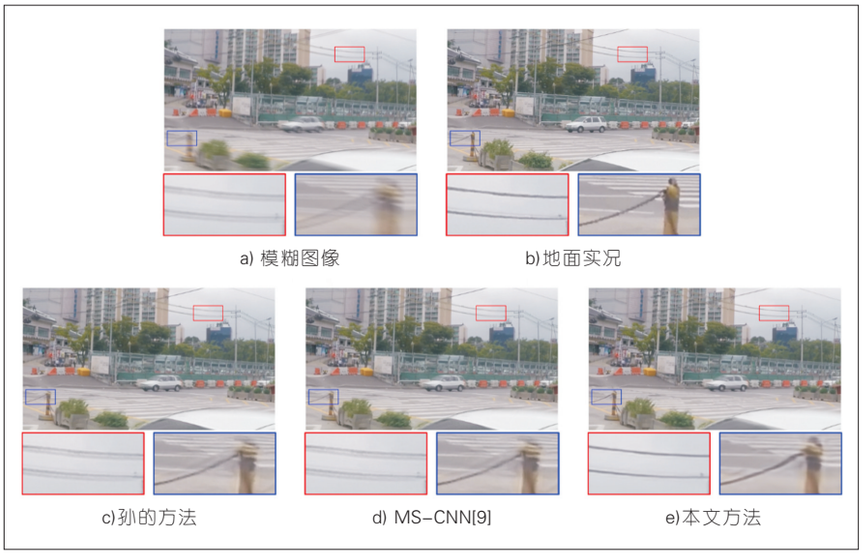
GOPRO数据集包含了3 214对高速摄像机拍摄的模糊清晰图像对，其中训练集包含2 103对，测试集包含1 111对。我们与其他先进的去模糊算法进行了对比实验，并进行了定性与定量的分析。图4展示了一些去模型效果图，从中我们能看出多尺度卷积神经网络（MS-CNN）方法[9]与孙的方法[7]都出现了振铃效应，而我们的方法则避免了这一情况。表1展示了定量分析的结果，我们的方法在峰值信噪比（PSNR）和结构相似性（SSIM）的指标上都远远超越了其他的一些方法。

3.2 Köhler数据集上的实验

Köhler数据集[20]包含4张清晰图像，每张清晰图片有12张对应的模糊图像。作者记录了12条不同的相机轨迹来生成12张不同的模糊图像。我们在Köhler数据集上进行了对比实验，并做了定量分析，如表2所示。

3.3 时间与内存开销

我们在时间与内存开销上与其他方法做了对比。为了公平起见，我们用pytorch库重新实现了MS-CNN[9]与深度对抗滤波（DGF）[10]。对于每一个方法，我们分别测试了1 000张1 280×720的图片，计算平均的时间与内存开销。对于时间测试，我们只计算正向传播的时间，不考虑反向传播的时间。对于内存测试，我们只计算生成器的内存开销，不考虑鉴别器的内存开销。表3展示了时间与内存开销的对比实验。我们的方法比DGF快3.4倍，比MS-CNN快8.4倍，同时消耗的内存是DGF的25.59%，是MS-CNN的27.65%。这显示出我们的方法更加高效，更容易应用于实际场景中。



▲图4 在GOPRO数据集上的对比实验效果图

▼表1 在GOPRO数据集上的对比实验结果

方法	PSNR	SSIM
孙的方法[7]	24.6980	0.8561
MS-CNN[9]	28.4498	0.9008
本文方法	29.2168	0.9208

MS-CNN：多尺度卷积神经网络
PSNR：峰值信噪比
SSIM：结构相似性

▼表2 在Köhler数据集上的对比实验结果

方法	PSNR	MSSIM
孙的方法[7]	25.12	0.7748
MS-CNN[9]	26.51	0.8083
本文方法	25.91	0.8115

MS-CNN：多尺度卷积神经网络
MSSIM：结构相似度均值
PSNR：峰值信噪比

▼表3 平均时间与内存开销表

方法	时间/s	内存/MB
MS-CNN[9]	2.9285	8 279
DGF[10]	1.1823	7 665
本文方法	0.3478	2 119

DGF：深度对抗滤波
MS-CNN：多尺度卷积神经网络

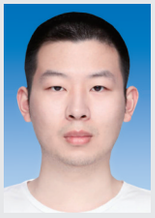
4 结束语

本文中，我们提出了一种基于深度CNN的图像去模糊方法。与现有方法相比，该方法更加高效。通过在不同数据集上的测试，该方法与目前最先进的方法效果相当，同时速度更快，所需内存空间更少。运行速度快与所需内存少的特性，使这种方法更容易应用于包含视觉SLAM系统在内的实际应用中。

参考文献
[1] GUPTA A, JOSHI N, ZITNICK C L, et al. Single Image Deblurring Using Motion Density Functions[C]// European Conference on Computer Vision. German: Springer, 2010: 171–184
[2] KIM T H, AHN B, LEE K M. Dynamic Scene Deblurring[C]//International Conference on Computer Vision. USA:IEEE, 2013:3160–3167. DOI: 10.1109/ICCV.2013.392
[3] KIM T H, LEE K M. Segmentationfree Dynamic Scene Deblurring[C]//Computer Vision and Pattern Recognition. USA:IEEE, 2014:2766–2773. DOI: 10.1109/CVPR.2014.348
[4] WHYTE O, SIVIC J, ZISSERMAN A,et al. Non–Uniform Deblurring for Shaken Images [J]. International Journal of Computer Vision, 2012, 98(2): 168–186
[5] CHAKRABARTI A. A neural Approach to Blind Motion Deblurring[C]//European Conference on Computer Vision. German: Springer, 2016: 221–235

[6] SCHULER C J, HIRSCH M, HARMELING S, et al. Learning to Deblur[J]. Transactions on Pattern Analysis and Machine Intelligence, USA: IEEE, 2016, 38(7): 1439–1451. DOI: 10.1109/TPAMI.2015.2481418
[7] SUN J, CAO W, XU Z, et al. Learning a Convolutional Neural Network for Nonuniform Motion Blur Removal[C]// Computer Vision and Pattern Recognition. USA: IEEE, 2015:769–777. DOI: 10.1109/CVPR.2015.7298677
[8] XU L, REN J S J, LIU C L, et al. Deep Convolutional Neural Network for Image Deconvolution[C]//Advances in Neural Information Processing Systems. USA: MIT Press, 2014: 1790–1798
[9] NAH S, KIM T H, LEE K M. Deep Multi–Scale Convolutional Neural Network for Dynamic Scene Deblurring[C]//Computer Vision and Pattern Recognition. USA: IEEE, 2017. DOI: 10.1109/CVPR.2017.35
[10] RAMAKRISHNAN S, PACHORI S, RAMAN S. Deep Generative Filter for Motion Deblurring[C]// International Conference on Computer Vision. USA: IEEE, 2017. DOI: 10.1109/ICCVW.2017.353
[11] HE K, ZHANG X, REN S, et al. Identity Mappings in Deep Residual Networks[C]// European Conference on Computer Vision. German: Springer, 2016: 630–645
[12] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[C]// Computer Vision and Pattern Recognition. USA: IEEE, 2016: 770–778. DOI: 10.1109/CVPR.2016.90
[13] NEWELL A, YANG K, DENG J. Stacked Hourglass Networks for Human Pose Estimation[C]// European Conference on Computer Vision. German: Springer, 2016: 483–499
[14] RONNEBERGER O, FISCHER P, BROX T. U–net: Convolutional Networks for Biomedical Image Segmentation[C]// International Conference on Medical Image Computing and Computer–Assisted Intervention. German: Springer, 2015: 234–241
[15] MICHAELI T, IRANI M. Blind Deblurring Using Internal Patch Recurrence[C]// European Conference on Computer Vision. German: Springer, 2014: 783–798
[16] XIE S, GIRSHICK R, DOLL Doll' AR P, et al. Aggregated Residual Transformations for Deep Neural Networks[C]//Computer Vision and Pattern Recognition. USA: IEEE, 2017: 5987–5995. DOI: 10.1109/CVPR.2017.634
[17] ISOLA P, ZHU J Y, ZHOU T H, et al. Image–to–Image Translation with Conditional Ad–Versarial Networks[C]//Computer Vision and Pattern Recognition. USA: IEEE, 2017. DOI: 10.1109/CVPR.2017.632
[18] GOODFELLOW I, POUGET–ABADIE J, MIRZA M, et al. Generative Adversarial Nets [C]//Advances in Neural Information Processing Systems. USA: MIT Press, 2014: 2672–2680
[19] MAO X, LI Q, XIE H, LAU R YK, et al. Least Squares Generative Adversarial Networks [C]//International Conference on Computer Vision. USA: IEEE, 2017. DOI: 10.1109/ICCV.2017.304
[20] KOHLER R, HIRSCH M, MOHLER B, et al. Recording and Playback of Camera Shake: Benchmarking Blind Deconvolution with A Real–world Database[C]//European Conference on Computer Vision. German: Springer, 2012: 27–40

作者简介



缪弘，复旦大学计算机科学技术学院在读研究生；研究方向主要包括计算机视觉和人工智能。

张文强，复旦大学研究员，智能机器人研究院副院长；研究方向主要包括机器人、人工智能、机器智能。