

课题组项目报告

伍瀚杰

华南理工大学

1. 论文概要翻译

1.1. 论文题目和作者

Assaf Shocher, Nadav Coheny, Michal Irani.
"Zero-Shot" Super-Resolution using Deep Internal Learning

1.1.1 问题提出

本文所要解决的问题是图像超分辨率重构的问题，即将低分辨率的图像通过一定的技术（本文通过基于 CNN 的神经网络模型）转变成高分辨率的图像。

1.1.2 研究动机

- 基于监督学习的 SR (super-resolution) 方法，需要特定的 LR-HR 图像数据对，其中 LR 图像数据是由 HR 图像数据通过例如双三次降尺度的方法来获得，训练的 LR 图像要求是没有干扰噪声的。但问题是在现实世界中，我们采集到的 LR 图像一般都是存在噪声的（手抖、光照影响等），这导致我们通过监督学习方法训练出来的最好的模型，应用在现实世界的图像中时，并不能表现出在训练时能达到的良好效果。以及用于训练的符合要求的 LR-HR 图像对是很难获得和数量少。
- 基于自然图像中存在局部信息循环的特性，产生了很多无监督学习的方法来训练模型（例如 unsupervised SR, Blind-SR）但是这些方法一般依赖于简单的欧氏距离和 K 最邻近搜索方法，这导致这些方法在不是重复出现的局部图像区域中不能产生好的效果，以及对于重复出现但并不是同样大小的局部图像区域也不能产生好的效果。
- 基于上面的两点认识，本文的作者决定采用无监

督学习的方法解决很难获得理想的训练数据对的情况，同时采用 CNN 神经网络来解决不能充分利用重复局部图像信息的问题。提出了 ZSSR 网络模型结构。

1.1.3 作者提出的解决方案

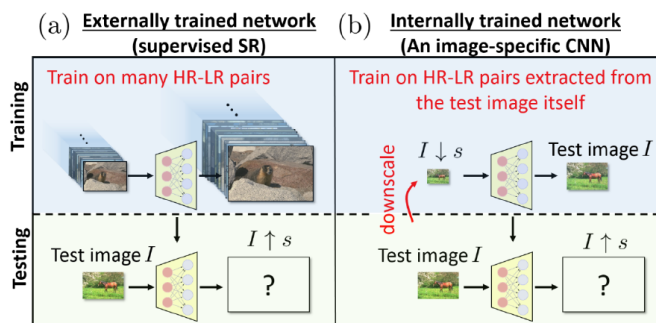


图 1. The Network Architecture of Supervised SR and ZSSR

- 网络结构：全卷积层的网络，有 8 层隐藏层，每层有 64 通道，每层卷积层后用 ReLU 激活。
- 训练数据：将我们需要进行 SR 操作的测试图像，降尺度获得不同尺度的图像，这些图像组成一个图像集合 ($I=I_0, I_1, I_2, \dots, I_n$)。将图像集合中的每张图片，都用 s 的分辨率比例 (s 即测试图片为期望的超分辨率比例) 下采样到对应的 LR 图像，每个这样生成的 LR 图像和原图构成训练的一组数据对。同时为了增加训练的数据量，我们将前面构成的训练数据对，进行四个方向的旋转和水平镜像、垂直镜像操作来进行数据增强从而获得很多的训练数据。
- 训练过程：将上面获得的训练数据对，对网络进行训练。为了能在小的 LR 图片上应用大的比例因子 s ，训练过程中使用多个比例因子，这些比例因子

从较小的比例因子 s_1 , 逐渐增加到我们期望的 s 。从小的比例因子开始训练, 训练产生的 HR 和原输入组成一个数据对, 加入训练数据中, 用于后面的训练。另外, 在训练时加入高斯噪声, 实验证明可以提高训练的效果。

- 训练细节: 网络的输入图像尺寸要插值成输出图像尺寸。采用 L_1 损失函数和 Adam 作为梯度下降的优化方法。初始的学习率定为 0.001。学习率会周期衰减, 当学习率小于 10^{-6} 时, 停止训练。
- 实验结果: 最后的实验结果证明了本文作者提出的 ZSSR 网络架构所训练出来的模型, 在理想图片输入的情况下, 它的评价指标和有监督学习算法相差不多 (甚至比 SRCNN 这种有监督算法的评价指标高), 同时它的评价指标比其他无监督算法都高。另外在图片输入是不理想的、存在噪声时 (例如历史老照片), 无论是对于监督学习算法还是其他无监督学习算法产生的模型, 本文的算法产生的模型达到了最好的评价指标。

1.2. 论文题目和作者

Muhammad Haris, Greg Shakhnarovich, and Norimichi Ukita, Deep Back-Projection Networks For Super-Resolution

1.2.1 问题提出

本文所要解决的问题是图像超分辨率重构的问题, 即将低分辨率的图像通过一定的技术 (本文通过一种不断上采样和下采样的神经网络模型) 转变成高分辨率的图像。

1.2.2 研究动机

- 目前的进行 SR 操作的神经网络架构, 都是利用从 LR 图像的提取的特征, 通过训练找到从 LR 图像空间非线性映射到 HR 图像空间的一个模型。这样只利用了从 LR 图像到 HR 图像的单方向的信息, 没有充分利用 LR 图像与 HR 图像相互转换的信息。
- 同时通过认知科学的研究表明, 人类的视觉系统也是存在这种类似于反馈的机制来引导视觉生成相关的结果。并且传统的迭代反投影算法在解决

图像超分辨率重构的问题也利用了 LR 图像与 HR 图像相互转换的信息, 来进行超分辨率的图像重构。

- 基于以上的两点认识, 本文作者提出了一种不断迭代进行上采样和下采样的方式组织的神经网络架构来重构超分辨率图像。

1.2.3 作者提出的解决方案

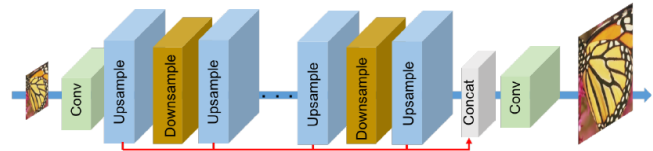


图 2. The Network Architecture of DBPN

- 首先我们用一个卷积层来提取出 LR 图像的特征, 在特征提取之后, 将提取后的特征经过交替的上采样层和下采样层来构建 LR 和 HR 图像的特征图, 经过一定数量交替的上下采样层后, 把每个上采样层构建的特征图连接起来, 再经过一个卷积层得到最后的 HR 图像。
- 其中上采样层和下采样层的具体网络结构如下图:

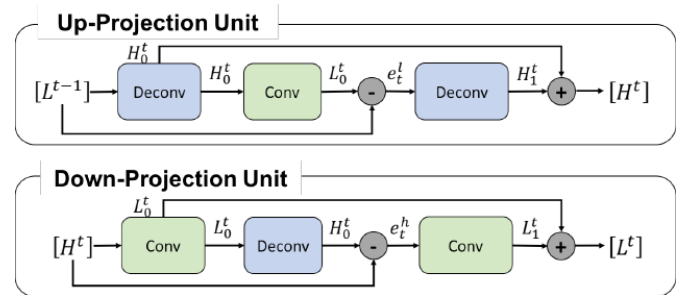


图 3. The Detail Architecture of Upsample Layer and Down-sample Layer

- 我们在这里分析上采样层的处理过程, 首先图像 $L^{(t-1)}$ 经过一个反卷积层并用比例因子 s 进行上采样处理得到中间结果 H_0^t 图, 接着将刚刚得到的图经过一个卷积层并用比例因子 s 进行下采样再得到 L_0^t , 将 L_0^t 与 $L^{(t-1)}$ 相减得到一个误差值, 后面将这个误差值通过比例因子上采样后, 再和前面的 H_0^t 相加得到输出特征图。

$$\text{scale up: } H_0^t = (L^{t-1} * p_t) \uparrow_s, \quad (1)$$

$$\text{scale down: } L_0^t = (H_0^t * g_t) \downarrow_s, \quad (2)$$

$$\text{residual: } e_t^l = L_0^t - L^{t-1}, \quad (3)$$

$$\text{scale residual up: } H_1^t = (e_t^l * q_t) \uparrow_s, \quad (4)$$

$$\text{output feature map: } H^t = H_0^t + H_1^t \quad (5)$$

图 4. The Representation of Upsample Layer

- 在上述网络结构的基础上，由于本文的网络比较深，在训练过程中容易出现梯度下降的情况，本文作者受残差网络的启发，将上述的网络结构进行改进，如下图所示，在交替的上采样和下采样层中，后面的每一个下采样层的输入都是前面所有上采样层的输出连接，而不只有前一层的上采样层的输出作为输入。同理，后面的每一个上采样层的输入都是前面所有下采样层的输出连接。

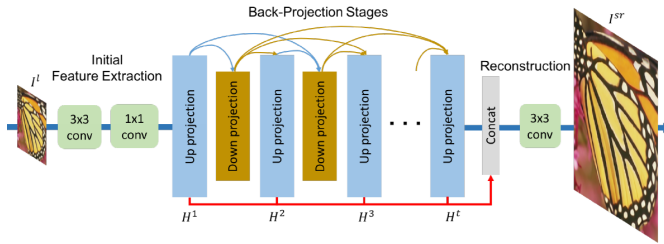


图 5. An implementation of D-DBPN for super-resolution

- 训练细节：DIV2K, Flickr 和 Imagenet 作为训练数据集，LR 图像的尺寸为 32*32，一批训练数量大小为 20。在不同比例因子的超分辨率任务上，设置不同的卷积核参数。

表 1. 参数设置

比例因子	卷积核大小	步长	填充大小
2	6*6	2	2
4	8*8	4	2
8	12*12	8	2

- 实验结果：本文提出的网络结构在大的比例因子的情况下，如比例因子为 8 时，达到了目前最好的效果。

1.3. 论文题目和作者

He Zhang, Vishal M. Patel, Density-aware Single Image De-raining using a Multi-stream Dense Network

1.3.1 问题提出

本文要解决的问题是将雨水图像除去雨水的问题。

1.3.2 研究动机

- 由于在雨天拍摄的图像中，雨水的密度分布不是均匀的、雨的大小和尺度不是一致的、雨的形状多样等多种情况，图像去雨是具有挑战性的，并且目前的达到最好效果的一些方法，不具有鲁棒性，只能在符合它的训练数据类型的雨水图像上得到好的结果，一旦雨水图像不是训练数据的类型，就产生不出好的结果。也有研究小组用多种雨水类型的图像作为训练数据去训练模型，但是他们是使用一个卷积神经网络去拟合这些数据，但由于存在多种类型，一个卷积神经网络不能很好的去拟合各种情况，这种方法产生的结果也不理想。
- 为了解决上面存在的问题，作者提出多流密集网络的网络架构去解决图像去雨的问题，同时也公布了一个更大的不同雨水密度的雨水图像数据集（第一个具有雨水密度标签的数据集）。

1.3.3 作者提出的解决方案

本文作者提出的网络架构主要分为两个部分：雨水密度估计分类、雨水去除网络。雨水密度估计分类器是为了确定一张给定图像的雨水密度等级（大雨、中雨、小雨）。获取到估计的雨水密度等级后，通过第二部分雨水去除网络，将估计的雨水密度融合到多流密集连接网络中（其中多流密集连接网络考虑了雨水不同尺度和形状的变化），通过这个网络得到最后去除雨水后的图像。

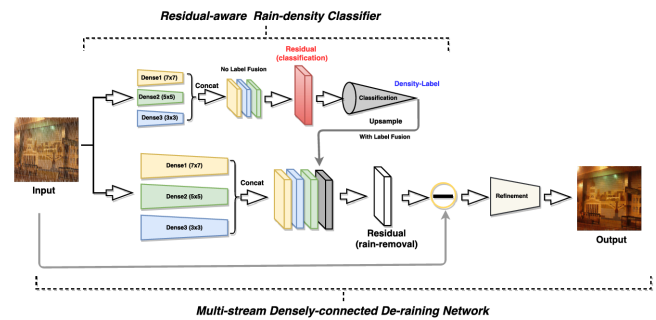


图 6. An overview of the proposed DID-MDN method

- 使用不同卷积核大小 (3*3、5*5、7*7) 的 DenseNet

作为特征提取网络，在第一部分的估计雨水密度的网络，通过多流 DenseNet 提取出特征后计算出残差值，作为分类网络的输入，输出分类的标签（大、中、小雨）。第二部分同样是先提取特征，之后结合第一部分估算的标签值上采样的标签图，形成一个雨水的残差值，用原图像减去这个残差值，再通过一个小的卷积网络去保证图像的细节最后生成去雨后的图像。

- 分类器网络和 DenseNet 网络细节如下图

Dense1: three transition-down layers, three transition-up layers and kernel size 7×7 .

Dense2: two transition-down layers, two no-sampling transition layers, two transition-up layers and kernel size 5×5 .

Dense3: one transition-down layer, four no-sampling transition layers, one transition-up layer and kernel size 3×3 .

图 7. Detail of DenseNet

Details of the classifier are as follows:

Conv(3,24)-Conv(24,64)-Conv(64,24)-AP-FC(127896,512)-FC(512,3),

图 8. Detail of Classifier

1.4. 论文题目和作者

Anat Levin, Yair Weiss, Fredo Durand, William T. Freeman, MIT CSAIL, Weizmann Institute of Science, Hebrew University, Adobe

1.4.1 问题提出

本文分析评估现有的算法在同样的数据集上的效果。

1.4.2 研究动机

- 盲反卷积进行图像复原是把一张模糊的照片（在模糊核未知的情况下）复原成一张锐化图像的过程。
- 作者通过研究盲反卷积的算法的原理和不足，对比不同算法或是某一算法在不同参数下（采取不同的先验知识）的实验结果差异，为未来进行相关研究的人员提供一些指导结果和假设。

1.4.3 作者的研究结果

- 本文中没有提出新的盲反卷积的算法，本文先证明了对 x (清晰图像), k (模糊核) 同时联合进行最大后验概率 (MAP) 估计存在的局限，会导致估计的模糊核有误差，导致实验图像复原结果不佳。接着提出了只对 k 进行最大后验概率估计， x 用非盲的其他图像复原方法恢复，最后研究者定量地去评价关于盲反卷积的一些算法（在同样的数据集上，对比不同算法的结果。）
- 盲去卷积的图像恢复模型为： $y = k \otimes x + n$ 。其中 y 是模糊的图像是输入，已知的。而右边的 k （模糊核）、 x （清晰图像）、 n （噪声）都是未知需要估计的。同时在这篇论文中，研究者假设待恢复的图像中的噪声都服从高斯分布，为高斯噪声。
- 作者指出自然图像的先验知识及自然图像的导数分布为稀疏的： $\log p(x) = -\sum_i |x'_i|^\alpha, \alpha < 1$ ，该稀疏先验知识将用于 $MAP_{x,k}$ 估计上。最原始的 $MAP_{x,k}$ 估计的公式为： $\min -\log p(x, k|y) = \frac{1}{\eta^2} |k \otimes x - y|^2 + \sum_i |x'_i|^\alpha, \alpha < 1$ ，复原的过程就是找到一对 (x, k) ，使这个估计公式尽可能地小，这个值越小， x 应该越清晰。但是作者在实验中发现，如果按照上式进行模糊核的估计和清晰图像的复原的话，由于未知的参数很多，这个问题属于一个 ill-pose 问题，符合要求的解可能有很多（例如图中的两个解都是符合的），但是这可能会出现不正确的情况，如图所示，图中的两组 (x, k) 都是上述公式估计的解，但是我们会发现 $(x = \text{模糊图像}, k = \text{delta})$ 会比 $(x = \text{清晰图像}, k = \text{对应正确的模糊核})$ 的联合概率更大，这说明在选择时，算法会输出 $(x = \text{模糊图像}, k = \text{delta})$ 为正确结果，而不是真正的清晰图像作为正确结果。文章中证明了这种现象会在 $MAP_{x,k}$ 估计中经常发生，文中通过实验指出在现实清晰的图片中，一些噪声和纹理可能会被作为脉冲，导致上式的稀疏先验变大，这也是导致这种估计方法经常失败的原因。作者通过实验展示了一些失败的案例。

$$P(\text{Latent image}, \text{kernel}) > P(\text{Latent image}, \text{kernel})$$

图 9. 采用 $MAP_{x,k}$ 估计时不正确的情况

- 作者也在文中表明 $MAP_{x,k}$ 估计方法失败的原因，不是因为有足够的图片数据也不是因为稀疏先验这个正则项选择是错误的，而是这个算法本身存在问题。
- 所以作者提出第二种估计方法， MAP_k 估计即只估计模糊核 k ，公式： $\argmax p(k|y) = \int p(x, k|y)$ 。与上面 $MAP_{x,k}$ 估计的不同是，我们在评价 $p(k|y)$ 时，会把所有可能的 x 考虑进去，而不是只针对单一的最好的 x 。实验证明这种方法会选择正确的模糊核。这种方法能成功的原因是在 x 和 k 之间有很强的维度不对称性。这体现在模糊核一般都是固定而比图像的大小小的，一般不会随图像的尺寸变大而变大的。同时可以采用拉普拉斯估计等方法去估计。
- 在一维实验中，作者采用了盒滤波器作为模糊核，并且假设在图像复原过程中，知道模糊核是盒滤波器，所以要求的是盒状滤波器的大小，由图我们可以看到虚线即为正确的盒大小的位置（Exact 就是上面提到的 MAP_k 估计方法），我们可以看到 MAP_k 估计方法虚线的位置正好是最小值的位置，这说明 MAP_k 估计方法能找到正确的参数位置，但是 $MAP_{x,k}$ 最小值的位置并不是虚线的位置。图中还有其他的一些算法的结果。

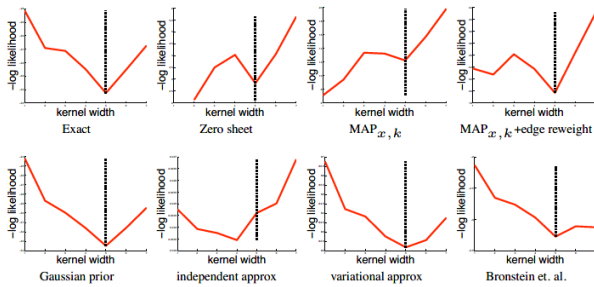


图 10. 一维实验结果

- 在二维图像上的实验结果，表明 Fergus 提出的算法的表现在当时来说是最好的。 MAP_k 估计方法的表现并不是很好，原因是采用了高斯先验而不

是稀疏先验。但即使是这样， MAP_k 估计方法也是比 $MAP_{x,k}$ 估计方法（使用稀疏先验）表现得要好，这也表明了对于估计方法的选择比先验知识的选择重要。

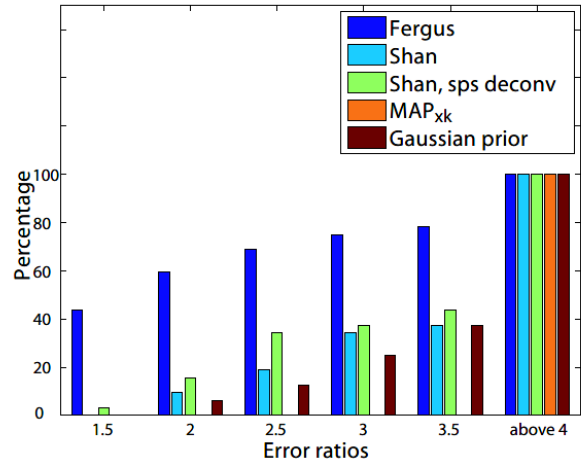


图 11. 二维实验结果

- 文章最后还指出，目前（论文发表的时候）盲反卷积去模糊的算法都假设图片在不同位置上的模糊程度是一样的，但是作者通过实验证明现实并不是这样的，如图，作者在一张图像的四个角取大小相同的局部区域，求导数，看到四个角的模糊是不一样的。指出未来可以在模糊程度分布的估计上进行研究。

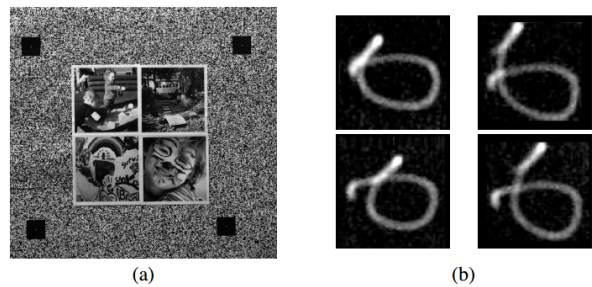


图 12. 模糊在空间上不是均匀分布的

2. 图像超分辨率、去模糊、去噪声、去雨、去雾研究现状以及对比

2.1. 超分辨率

将低分辨率的图像生成高分辨率的图像的过程。超分辨率可分为两类：从多张低分辨率图像重建出高分辨率图像和从单张低分辨率图像重建出高分辨率图像。图像超分辨率是一个不适定逆过程，因为对于任何 LR 输入都存在多种解决方案。为了解决这个逆问题，研究者们已经提出了大量的图像 SR 算法，包括基于插值、基于重建和基于学习的方法。

- 传统方法：稀疏编码方法：一幅图像能够在非常苛刻的条件下由它的一组稀疏表示系数在超完备字典上得到精确重建。通过对低分辨率图像块字典和高分辨率图像块字典的联合训练，我们可以强化低分辨率和高分辨率图像块与之对应真实字典稀疏表示的相似性，从而低分辨率图像块的稀疏表示和高分辨率超完备字典一起作用可以重建出高分辨率图像块，然后由高分辨率图像块连接得到最终完整的高分辨率图像。
- 深度学习方法：

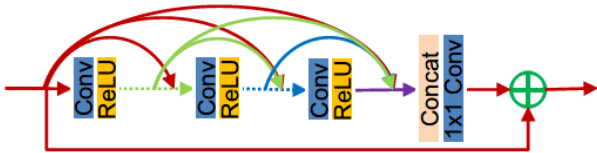


图 13. 提出的新的残差密集块结构

研究者提出了残差密集网络 (RDN) 如图，通过残差密集块 (RDB) 如图来充分利用原始 LR 图像的所有分层特征。对于一个很深的网络来说，直接提取 LR 空间中的每个卷积层的输出很难，可以说是不切实际的。这种方法的研究者将残差密集块 (RDB) 作为 RDN 的构建模块。RDB 包含密集连通层和带有局部残差学习 (LRL) 的局部特征融合 (LFF)。我们的残差密集块还支持 RDB 间的连续记忆。一个 RDB 的输出可以直接访问下一个 RDB 各层，从而使状态连续传递。RDB 每个卷积层都可以访问所有的后续层，传递需要保留的信息。将前面的 RDB 与当前 RDB 的所有前面层的

状态连接，LFF 通过自适应地保存信息来提取局部密集特征。此外，LFF 通过稳定更大网络的训练来实现极高的增长率。在提取多层局部密集特征后，研究者进一步进行全局特征融合 (GFF) 以全局方式自适应地保留分层特征。

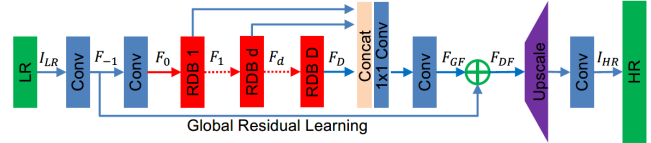


图 14. RDN 结构

RDN 主要包含四部分：浅层特征提取网络 (SFENet)、残差密集块 (RDBs)、密集特征融合 (DFE) 以及上采样网络 (UPNet)。研究者将 ILR 和 ISR 表示为 RDN 的输入和输出，具体来说，研究者使用两个 Conv 层来提取浅层特征。

2.2. 去模糊

将模糊的图像复原成清晰的图像的过程。实际生活与操作中，造成图像模糊的原因有很多，如设备抖动或设备与被摄物间的位移或空间变化引起的运动模糊，聚焦不准导致的离焦模糊等。本质是原始图像与模糊核 (或称点扩散函数 PSF) 间相互卷积形成的，则去模糊就是解卷积的过程。

- 传统方法：

1. 在模糊核已知的情况下，可采用 Richardson-Lucy 算法，该算法是一种非线性迭代图像复原技术。RL 算法是在假定图像像素服从泊松分布的条件下，运用最大似然法、基于贝叶斯分析来求解原始清晰图像的。RL 算法在图像复原过程中由于存在振铃效应和噪声放大的问题，会影响图像恢复质量。当迭代次数增加时，能使图像恢复更多的细节，但相应地，振铃效应增多，模糊图像中的噪声会被放大。其迭代方程为：
$$f(x,y)^{n+1} = f(x,y)^n \left[\frac{g(x,y)}{h(x,y) * f(x,y)^n} \right] \otimes h(x,y)$$
2. 在模糊核未知的情况下，采用盲解卷积算法，盲解卷积的方法是以最大似然估计为基础的，即一种用被随机噪声所干扰的量进行估计的最优化策略。似然函数用 $g(x,y)$ 、 $f(x,y)$ 和

$h(x,y)$ 来加以表达, 然后问题就变成了寻求最大似然函数。在盲解卷积中, 最优化问题用规定的约束条件并假定收敛时通过迭代来求解, 得到的最大 $f(x,y)$ 和 $h(x,y)$ 就是还原的图像和模糊核。

- 深度学习的方法: 由于现实世界中模糊照片的模糊核一般都是未知的, 根据模糊图像的建模公式: $B = I \otimes k + n$ 可知, 为了解决问题, 好模糊核和图像的额外约束和先验知识都是必需的, 传统的先验知识都是手工设计, 这些知识基于对特定图像的有限次观察统计, 具有局限性。目前最新的研究尝试通过深度学习的 CNN 网络去学习图像的先验知识, 并用于去模糊的任务。

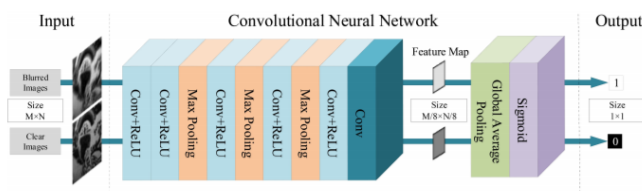


图 15. CNN 网络学习先验

研究者将图像先验表示为能够区分清晰图像和模糊图像的二值分类器。具体来说, 他们训练深度卷积神经网络来分类模糊图像 (标记为 1) 和清晰图像 (标记为 0)。由于基于 MAP(最大后验) 的去模糊方法通常使用 coarse-to-fine(由粗到精) 策略, 因此在 MAP 框架中插入具有全连接层的 CNN 无法处理不同大小的输入图像。为了解决这个问题, 他们在 CNN 中采用了全局平均池化层, 以允许学习的分类器处理不同大小的输入。此外, 为了使分类器对不同输入图像尺寸具有更强的鲁棒性, 他们还采用多尺度训练策略。然后将学习到的 CNN 分类器作为 MAP(最大后验) 框架中潜在图像对应的正则项。

2.3. 去雨

将雨天情景下, 拍摄的带有雨点雨水的图像, 去除照片中的雨水雨滴的过程。传统的关于图像去雨的图像恢复方法在某些特定的情况下会失效。所以目前图像去雨的比较好的研究都是基于深度学习的方法。

- 第一种方法是基于单图像深度联合雨水检测和去

除网络: 在该方法中定义的一个广义雨水模型为:

$$O = B + SR \quad (1)$$

其中 B 代表背景层, 即我们想要获取的没有雨痕的图像。 S 代表雨痕层。 O 代表有雨水的输入图像。 R 为一个区域二值图, 在雨水区域和非雨水区域进行不同处理, 其中值 0 表示对应的像素没有雨痕, 值 1 表示对应的位置存在雨痕。基于雨痕带具有不同的形状和不同的方向且雨痕之间会相互重叠, 以及暴雨情况下的雨水积聚产生的雾气效果导致远处场景的能见度降低这两个主要的问题。进一步定义了一个细致的雨水模型:

$$O = \alpha(B + \sum_{t=1}^s s_t R) + (1 - \alpha)A \quad (2)$$

S 表示一个雨痕层, 这层的雨痕的方向是一致的, 其中下标 t 是表示第几个雨痕层。 A 是全局光照, 是用于对雨水产生的雾气建模。 α 是全局光照的透射系数。这个更加细致的雨水模型, 同时对雨水和由于雨水产生的雾气进行建模, 更加符合真实情况。

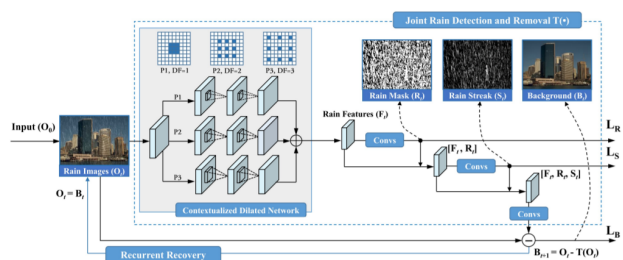


图 16. 联合雨水检测和去除的深度卷积神经网络

该深度架构包含一个新颖的网络结构, 即基于情境信息的上下文扩张网络, 此结构用于提取雨水图像的可识别特征, 为后续的检测和去除做基础。使用情境化的上下文扩张网络可以聚合多尺度的上下文信息来学习雨水图像的特征信息。扩张卷积与普通卷积相比, 在卷积核的大小是一样的情况下, 神经网络中即参数数量相同, 区别在于扩张卷积具有更大的感受野。在图中可以看到该网络包含三条卷积路径, 每条卷积路径均使用 3×3 的卷积核, 第一个卷积路径使用普通的卷积核, 其余两条路径则使用扩张卷积技术, 所以具有不同的

- 最近的第二种方法是使用 Attentive GAN 对单图像去雨。该方法定义的雨水模型为：

$$I = (1 - M) \cdot B + R \quad (3)$$

这里的 I 表示输入图像, M 是基于整张图像的每个像素二进制的掩模 (对于像素 x , 如果被雨滴覆盖, 则 $M(x) = 1$, 否则 $M(x) = 0$), B 是图像的背景 (即想要得到的目标图像), R 是雨滴带来的影响综合效应 (图像背景信息、环境反射光和附着在挡风玻璃或者镜头的雨滴的折射光的一种复杂混合, 因为雨水是透明的, 由于雨滴形状和折射率的问题, 图像雨水区域内的一个像素会受到周边像素的影响, 是一个综合的效应表现), 操作符表示逐像素乘法。基于这个模型, 目标是从输入降质图像 I 获取目标图像 B 。使用 M 作引导生成注意力映射图, 并应用到 GAN 来实现目标图像的生成。

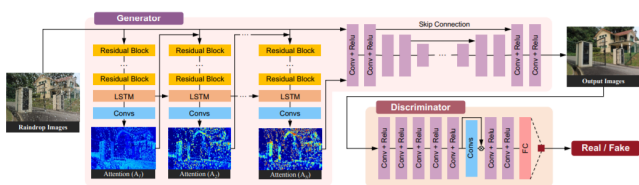


图 17. Attentive GAN 网络

使用一个注意力判别器，即注意力循环网络生成的注意力映射图应用到判别网络。使用注意力映射图来引导判别器聚焦相应的区域，更好的来判断图像的真实性。

2.4. 去噪声

将图像在拍摄和传输过程中，产生在图像上的脉冲噪声或电噪声等噪声去除的过程。从噪声的来源可分为加性噪声、乘法噪声、量化噪声。

- 传统方法去噪:

1. 均值滤波：均值滤波即线性滤波，主要采用几何领域平均法。均值滤波的基本原理是用当前待处理像素点 (x,y) 以及其若干领域像素组成一个模板，并求出模板中所有像素的均值，用以取代原图像中当前像素的值。这种方法适合清除由于扫描而产生的颗粒噪声。
2. 中值滤波：是一种常见的非线性滤波器，其基本原理是把图像中一点的像素值用一个领域中各点值的中值替换，它能让周围像素值差比较大的像素改取与周围的像素值接近的值。这种方法适合清除椒盐噪声。
3. 自适应维纳滤波：这种方法能根据图像的局部方差来调整滤波器的输出，当局部方差越大时，滤波器的平滑作用越强。它的目的是让恢复图像 $f(x,y)$ 与原始图像 $g(x,y)$ 的均方误差 $\theta^2 = E(f(x,y) - g(x,y))^2$ 。这种方法适合清除白噪声的图像。
4. 小波变换，是一种窗口大小固定不变但其形状可改变的时频局部化分析方法。由于噪声大多分布在高频区域，图像通过小波变换后，噪声信息大多集中在次低频、次高频及高频子块中，高频子块几乎以噪声信息为主。因此，将高频子块置为零，并对低频、高频进行一定抑制，便可达到去噪的目的。

- 采用深度学习的方法，神经网络模型可以避免在概率学习和推断过程中的计算困难。下图是采用深度学习的一种方法 DnCNN 的网络结构，这种方法提出一个端到端的可训练的 CNN 网络，采用残差学习策略，在网络的隐层隐式地移除干净图片。即输入为带噪的观测图片，输出是移除了干净图片的残差（噪声）图片。这样做的动机是，用残差学习刻画恒等映射或近似的恒等映射，效果要比直接学习干净图片要相对好。采用残差学习和批归一化加速训练并提升性能。

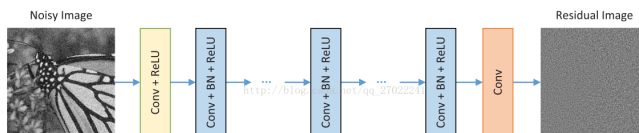


Fig. 1. The architecture of the proposed DnCNN network

图 18. DnCNN 网络结构

2.5. 去雾

将雾天的情况下，拍摄的带有雾气的图像，去除照片中的雾气的过程。现有的图像去雾技术都基于一个简单的物理模型。大气散射模型，这个模型说明了在有雾的情况下，物体是如何在相机上成像的。阳光在物体表面形成反射光 $J(x)$ ，反射光在穿过雾进入相机成像的过程中发生散射，导致只有部分的光 $J(x)t(x)$ 能到达摄像头。另一方面，太阳光也在雾中散射形成大气光 α 被摄像头接收。因此摄像头最后的成像由两部分组成：

$$I(x) = J(x)t(x) + \alpha(1 - t(x)) \quad (4)$$

其中 $t(x)$ 是媒介透射率，即光能透过雾到达摄像机的比率。可以知道媒介透射率和物体与摄像机的距离成反比，即离摄像机越远的物体受雾的影响越大， $t(x)$ 越大。在图像去雾的任务中，我们最核心的任务是如何更加精确地估计出一张图像的媒介透射率。

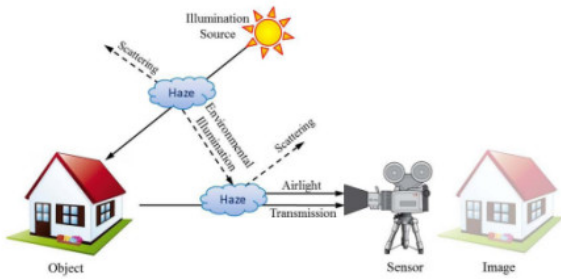


图 19. 大气散射模型

解决图像去雾问题，可以利用传统方法和深度学习的方法。

- 何凯明提出的暗通道先验方法，本方法基于一个先验知识：在绝大多数图像局部的非天空的清晰区域中，某些像素总会有至少一个颜色通道具有很低的值。基于这个先验知识，可以把原始的大气散射模型进行变换，用下式来恢复图像：

$$J(x) = \frac{I(x) - A}{\max(t(x), t_0)} + A \quad (5)$$

其中 $I(x)$ 是待去雾的图像， A 是大气光强度

- DehazeNet 是深度学习的方法，智能地学习雾的特征。DehazeNet 采用“卷积 + Maxout”的结构作为网络第一层去提取特征，同时借鉴于 GoogLeNet

中的 inception 结构，采用 3 组不同尺度（ 3×3 ， 5×5 ， 7×7 ）的滤波器实现 DehazeNet 的尺度鲁棒性。激活函数采用 BReLU，BReLU 的非线性回归对应于传统去雾方法中的边缘抑制操作（如 DCP 和 CAP）。双边约束引入先验信息缩小参数搜索空间，使得网络更加容易训练。局部线性避免 Sigmoid 函数梯度不一致带来的收敛困难。

2.6. 异同点

- 相同点：图像去模糊是指去除或减轻在获取数字图像的过程中发生的图像质量下降的问题，从而使图像更加清晰，更加接近真情实景。去模糊的图片模型为， $B = I \otimes k + n$ ，其中 n 是噪声，所以其实去模糊的过程中也包含的去噪声的过程。所以图像去噪、图像去雨、图像去雾都是属于图像去模糊的领域，只是在不同的造成图像退化的情况下，进行图像去模糊。图像去噪是在图像采集、传输、处理过程中受到噪声干扰。图像去雨和图像去雾是由于天气因素，但因为雨和雾的天气情况和对摄像机成像的影响是具有不同的特点的，所以分别研究它们是有意义的。图像超分辨率本质是将较大的图像放大其空间尺度，并且增加其分辨率的能力，图像超分辨率其实也是从一些由于各种原因（噪声、雾、雨）退化的图像，恢复成清晰（超分辨率率）的图像的过程，与上述的其他领域其实本质上都是图像复原。
- 不同点：图像超分辨率在把图像恢复清晰的同时，会放大图像的初度空间，增加像素的数量。但是其他例如图像去模糊、图像去雾、图像去雨、图像去噪的一般来说都不会改变原图像的尺度空间。

3. 求解 x 最大后验概率形式和不同正则项的选择

3.1. 求解 x 最大后验概率形式

$$\hat{x} = \operatorname{argmax}_x \ln p(y|x, A) + \ln p(x) + \ln p(A)$$

其中 $p(x)$ 为图像先验概率、 $p(k)$ 为模糊核的先验概率。

3.2. 正则项的选择

- 由于自然清晰图像具有导数稀疏的特点，故可以选择稀疏先验作为图像先验概率正则项： $\sum_i |x'_i|^2$

- 由于模糊核用来描述相机的运动轨迹，具有稀疏性，即核中大部分像素值接近于 0，这个分布可以建模为 D 维指数分布： A^2

4. 传统方法与深度学习

4.1. 深度学习的优点

- 深度学习利用了非线性函数的特性，可以增强模型的表达能力，从而更好的拟合数据的曲线。因为人类的认知能力有限和抽象问题的能力也有限，对比传统方法，人为地描述一个模型。深度学习能超越人类的认知，学习一个表达能力更强的模型。

4.2. 深度学习的缺点

深度学习的本质缺点就是严谨的理论解释去让研究者知道是如何得到结果的。这个本质的缺点将导致下面两个缺点。

- 不能更好地利用已有的网络架构去产生更好的结果。超参数的选择会对深度学习最后的结果产生很大的影响（初始值。每层的节点数、层数、学习率等），而目前超参数一般都是通过经验或者尝试得出一个好的结果，有很大的随机运气的成分在里面以及也很难找到好的理论依据支持我们为什么选择这样数值的超参数。这种随机性不解决的话或者说我们如果不能掌握着这些随机的规律，这将会变成深度学习继续前进的瓶颈。所以这也导致目前的研究者都采用更深的网络架构或者充分利用数据（如残差网络，通过跳跃连接避免数据特征在深层网络中丢失）去产生更好的结果。
- 不能发现或预测已有的训练好的模型的错误，因为模型的参数都是通过数据进行训练得到的，而且我们也不能解释每个具体参数的微小变动会直接影响到结果的哪个部分，这导致我们不能主动地发现整个训练好的模型可能出现的错误，只能在测试过程中，犯错之后才能发现。

4.3. 传统方法的优缺点

- 优点：传统方法的一些思想可以应用到深度学习上，传统方法对问题的解决流程有个整体的认识以及传统方法中的一些的数据处理方法、特征处理的方法也可以应用到深度学习。如 DehazeNet

的第一层是用传统的方法去提取去雾特征，后面再用网络学习这些特征从而得到一个好于传统方法的模型。

- 缺点：由于是人为地对提取特征的过程进行定义和特征的表示进行定义，人的认知高度决定了这种算法的上限，比如在人脸关键点检测时，我们要自己定义鼻子、嘴巴、眼睛等这些特征的数学表示（hog 特征等），很难定义到一个鲁棒性和准确性都比较好的一个检测模型。

总结来说，我认为在一些领域，例如图像超分辨率、图像分类、图像分割这些领域，在这些领域，就算我们在理论上不能严谨地解释这个网络结构为什么好，但是只要它能产生一个很好的结果。这也不妨碍我们将它应用到实际，毕竟它能产生传统方法都产生不了的很好的结果。但像自动驾驶这些领域，这些训练出来的模型可能对人造成不可挽回的影响的（自动驾驶识别出错，撞人或者车），在这些领域，深度学习如果不可解释，就像一个定时炸弹一样，我觉得是不能成为这些领域的解决方案的。

5. 对自己的认识

5.1. 自我优势和劣势

- 优势：数学基础良好（体现在数学类学科成绩优秀），编程基础扎实（体现在曾经的项目经历上），自己想做的事为做到最好（体现在曾获得国奖）。
- 劣势：有小小的拖延症，也尝过这带来的后果，在之后的研究生学习生活中会努力消除这个缺点。

5.2. 研究生计划

- 目标一：建立计算机视觉领域的合理的知识体系结构，选择自己感兴趣的领域中的细分方向深入研究。
- 实现过程：首先找计算机视觉领域最早的论述专著或教材，比如 David Marr 《Vision: A Computational Investigation into the Human Representation and Processing of Visual Information》。把这个领域的基本概念的内涵以及相互之间的关系梳理清楚，同时也要广泛涉猎这个领域的基本知识，尽量了解这个领域的方方面面都在进行什么工作和研究，例如计算机视觉中语义分割、图像处

理（图像超分辨率、图像上色等）、图像生成（图像风格转化、图像翻译等）、VR/AR 等。我所希望建立的体系结构是类似星型结构，中心是我理解的这个领域的核心知识和各个细分方向的共性的地方，每个分支就是不同的细分方向的针对该类问题特殊的解决思想方法。这个建立过程是贯穿我研究生的整个生涯。

- 目标二：学习如何研究，掌握研究的方法论。
- 实现过程：多和自己身边优秀的导师和同门师兄师姐交流请教，总结不同优秀的人对于研究一个问题时的思考方式的异同。重点思考是从哪些角度思考问题，以及如果有幸能参与到师兄师姐的研究中，学习师兄师姐研究过程的阶段特点和他们如何安排时间，如何解决问题。先找到可能的研究的方法和步骤，加以实践，在实践中结合自身特点形成自己的研究方法论。
- 目标三：在国际期刊上发表高水平的论文。
- 实现过程：上面的建立知识体系和掌握研究方法论，最终还是要转化为具有现实意义的论文。首先要通过大量阅读论文，锻炼自己形成一种如何将一个问题有逻辑地论述清楚的能力。于此同时，也要加强英语能力的提高，英语是作为这个学科研究的基本能力了。在目标一和目标二的实现过程中，论文的写作也是一个水到渠成的过程。