

用于图像恢复的非局部循环网络

Ding Liu¹, Bihan Wen¹, Yuchen Fan¹, Chen Change Loy², Thomas S. Huang¹

¹University of Illinois at Urbana-Champaign ²Nanyang Technological University
{dingliu2, bwen3, yuchenf4, t-huang1}@illinois.edu ccloy@ntu.edu.sg

Abstract

许多经典方法已经表明自然图像中的非局部自相似性是图像恢复的有效先验。然而，通过深度网络利用这种内在属性仍然不清楚且具有挑战性。在本文中，我们提出了一种非局部循环网络（NLRN），这是将非局部操作合并到循环神经网络（RNN）中以进行图像恢复的首次尝试。这项工作的主要贡献是：（1）与现有的以孤立方式测量自相似性的方法不同，所提出的非局部模块可以灵活地集成到现有的深度网络中进行端到端训练以捕获深度特征相关性在每个位置及其邻域之间。（2）我们充分利用 RNN 结构的参数效率，并允许深度特征相关性沿着相邻的循环状态传播。这种新设计增强了对由于严重退化的图像导致的不准确相关估计的鲁棒性。（3）我们表明，在给定退化图像的情况下，必须保持一个受限邻域来计算深度特征相关性。这与部署整个图像的现有做法 [41] 形成对比。对图像去噪和超分辨率任务进行了广泛的实验。由于循环非局部操作和相关传播，所提出的 NLRN 以更少的参数取得了优于最先进方法的结果。代码可在 <https://github.com/Ding-Liu/NLRN> 获得。

1 简介

图像恢复是一个不稳定的逆问题，旨在从其退化的测量中估计底层图像。根据退化的类型，图像恢复可以分为不同的子问题，例如图像去噪和图像超分辨率（SR）。成功恢复的关键通常依赖于基于图像先验的有效正则化器的设计。本地和非本地图像先验在过去都被广泛使用。以图像去噪为例，局部图像属性，如高斯滤波和基于总变差的方法[31]，在早期研究中被广泛使用。后来，自然图像中自相似性的概念引起了更多的关注，并被基于非局部的方法所利用，例如非局部均值 [2]、协同过滤 [8]、联合稀疏性 [27]、和低秩建模 [15]。这些非局部方法被证明可有效捕获非局部补丁之间的相关性以提高恢复质量。

虽然非局部自相似性已在文献中得到广泛研究，但很少探索用深度网络捕捉这种内在属性的方法。最近用于图像恢复的卷积神经网络（CNN）[10, 20, 28, 49] 比传统方法取得了令人印象深刻的性能，但没有明确地在图像中使用自相似属性。为了纠正这个弱点，一些研究 [23, 30] 在将补丁输入 CNN 之前将块匹配应用于补丁。然而，块匹配步骤是孤立的，因此不能与图像恢复网络联合训练。

在本文中，我们首次尝试在 CNN 中结合非局部操作进行图像恢复，并提出了一种非局部循环网络（NLRN）作为一种高效且有效的网络

带有非本地模块。首先，我们设计了一个非局部模块来为给定严重退化图像的自相似性测量产生可靠的特征相关性，它可以灵活地集成到现有的深度网络中，同时享受端到端学习的好处。为了在不影响恢复质量的情况下实现高参数效率，我们部署了一个类似于 [21、35、36] 的递归神经网络 (RNN) 框架，以便递归地应用具有共享权重的操作。其次，我们仔细研究了深度特征空间中非局部操作的行为，发现限制相关计算的邻域可以提高其对退化图像的鲁棒性。受限邻域有助于将计算集中在空间附近的相关特征上，而忽略噪声特征，这符合传统的图像恢复方法 [8, 15]。此外，我们允许在 RNN 的相邻循环状态之间传递非本地操作的消息。这种特征相关性的状态间流动有助于更稳健的相关性估计。通过将非局部操作与典型卷积相结合，我们的 NLRN 可以有效地捕获和利用局部和非局部图像属性来进行图像恢复。

值得注意的是，最近的工作在视频分类 [41] 上采用了类似的想法。然而，我们的方法在以下方面与它有显著不同。对于每个位置，我们仅在其邻域内测量每个位置的特征相关性，而不是像 [41] 中那样在整个图像中测量。在我们的实验中，我们表明对计算非局部先验有用的深层特征更有可能位于邻近区域。较大的邻域（整个图像作为一个极端）可能导致对降级测量的相关性估计不准确。此外，我们的方法充分利用了 RNN 架构的优势——相关信息在相邻的循环状态之间传播，以增加相关估计对不同程度退化的鲁棒性。此外，我们的非本地模块可以灵活地处理各种大小的输入，而 [41] 中的模块仅处理固定大小的输入。

我们通过首先将我们提出的模型与统一框架中的其他经典和现有非局部图像恢复方法相关联来介绍 NLRN。我们通过广泛的消融研究彻底分析了 NLRN 中的非本地模块和循环架构。我们与最近的竞争对手进行了全面比较，其中我们的 NLRN 在图像去噪和 SR 方面在多个基准数据集上实现了最先进的性能，证明了非局部操作和递归架构在图像恢复方面的优越性。

2 相关工作

图像自相似性作为一种重要的图像特征已经被用于许多基于非局部的图像恢复方法中。早期的工作包括用于图像去噪的双边滤波 [38] 和非局部方法 [2]。最近的方法通过施加稀疏性来利用图像自相似性 [27, 44]。或者，相似的图像块使用低秩 [15] 或通过协作维纳过滤 [8, 47] 建模。邻域嵌入是图像 SR [5, 37] 的常用方法，其中每个图像块由流形中的多个相似块近似。基于自样本的图像 SR 方法 [14, 12] 利用局部自相似性假设，仅从不同尺度的低分辨率图像中提取 LR-HR 样本对以预测高分辨率图像。图像去模糊采用了类似的想法 [9]。

深度神经网络已普遍用于图像恢复。开创性的工作包括用于图像去噪的多层感知器 [3] 和用于图像 SR [10] 的三层 CNN。采用反卷积来节省计算成本并加快推理速度 [34, 11]。非常深的 CNN 旨在提高 [20, 22, 24] 中的 SR 准确性。各种残差块之间的密集连接包含在 [39] 中。类似地，在 [28, 49, 50, 26] 中开发了基于 CNN 的图像去噪方法。作为预处理步骤的块匹配与 CNN 级联以进行图像去噪 [23, 30]。除了 CNN 之外，RNN 也被应用于图像恢复，同时享有高参数效率 [21, 35, 36]。

除了图像恢复，特征相关性与神经网络一起在许多其他领域得到广泛利用，包括图形模型 [51, 4, 17]、关系推理 [32]、机器翻译 [13, 40] 等。限于篇幅，我们在此不一一详述。

3 图像恢复的非局部操作

在本节中，我们首先提出用于图像恢复方法的非局部操作的统一框架，例如协同过滤 [8]、非局部均值 [2] 和低秩建模 [15]，并讨论他们之间的关系。然后，我们提出了建议的非本地操作模块。

3.1 总体框架

通常，非局部操作将多通道输入 $X \in \mathbb{R}^{N \times m}$ 作为图像特征，并生成输出特征 $Z \in \mathbb{R}^{N \times k}$ 。这里 N 和 m 分别表示图像像素数和数据通道数。我们提出了一个具有以下公式的通用框架：

$$Z = \text{diag}\{\delta(X)\}^{-1} \Phi(X) G(X) \quad (1)$$

这里， $\Phi(X) \in \mathbb{R}^{N \times N}$ 是非局部相关矩阵， $G(X) \in \mathbb{R}^{N \times k}$ 是多通道非局部变换。每个行向量 X_i 表示位置 i 中的局部特征。 $\Phi(X)_j$ 表示 X_i 和 X_j 之间的关系，每个行向量 $G(X)_j$ 是 X_j 的嵌入。 δ 对角矩阵 $\text{diag}\{\delta(X)\} \in \mathbb{R}^{N \times N}$ 使用归一化因子 $\delta_i(X)$ 对每个第 i 个像素的输出进行归一化。

3.2 经典方法

所提出的框架适用于各种经典的非局部图像恢复方法，包括基于低秩[15]、协同过滤[8]、联合稀疏性[27]以及非局部均值过滤[2]的方法。

块匹配(BM)是在传统方法[15, 8, 27]中利用非局部图像结构的常用方法。一个 $q \times q$ 空间邻域被设置为以每个位置 i 为中心，并且 X_i 减少到以 i 为中心的图像块。BM 从该邻域中选择 K_i 个最相似的块 ($K_i \leq q^2$)，它们共同用于恢复 X_i 。在提议的非本地框架下，这些方法可以表示为

$$Z_i = \frac{1}{\delta_i(X)} \sum_{j \in C_i} \Phi(X)_i^j G(X)_j, \quad \forall i. \quad (2)$$

这里 $\delta_i(X) = j \in C_i \Phi(X)_j$ 和 C_i 表示 K_i 个选定补丁的索引集。因此，每行 $\Phi(X)_i$ 只有 K_i 个非零项。对于基于不同模型的非局部方法，嵌入 $G(X)$ 和非零元素各不相同。例如，在 WNNM[15]中， $j \in C_i \Phi(X)_j$ 对应于 X_i 在组特定子空间上的投影，作为所选补丁的函数。具体来说，用于计算 Z_i 的子空间由 $X_i^T C_i X_i$ 的特征向量 U_i 跨越。因此 $Z_i = X_i^T C_i U_i \text{diag}\{\sigma\} U_i^T X_i$ ，其中 $\text{diag}\{\sigma\}$ 是通过将与加权核范数[15]相关的收缩函数应用于 $X_i^T C_i X_i$ 的特征值而获得的。我们在补充材料中展示了关于更经典的非局部图像恢复方法的概括。

除了硬块匹配外，其他方法，例如非局部均值算法[2]，通过计算参考块与邻域中每个块之间的相关性来应用软块匹配。每个元素 $\Phi(X)_{ji}$ 仅由每个 $\{X_i, X_j\}$ 对确定，因此 $\Phi(X)_{ji} = \phi(X_i, X_j)$ ，其中 $\phi(\cdot)$ 由距离度量确定。在[2]中，采用高斯核的加权欧氏距离作为度量，使得 $\phi(X_i, X_j) = \exp\{-\|X_i - X_j\|_2^2/h^2\}$ ，其中 h 是高斯核的标准差。

此外，身份映射直接用作[2]中的嵌入，即 $G(X)_j = X_j$ 。在这种情况下，(1) 中的非局部框架简化为

$$Z_i = \frac{1}{\delta_i(X)} \sum_{j \in S_i} \exp\left\{-\frac{\|X_i - X_j\|_{2,a}^2}{h^2}\right\} X_j, \quad \forall i, \quad (3)$$

并且 S_i 是 X_i 邻域中的一组索引。请注意， a 和 h 都是常数，分别表示高斯核的标准差和滤波程度[2]。值得注意的是，软 BM 的 S_i 的基数比硬 BM 的 C_i 的基数大得多，这为使用相邻位置之间的特征相关性提供了更大的灵活性。

传统的非局部方法的缺点是参数要么是固定的[2]，要么是通过次优方法获得的[8, 27, 15]，例如，WNNM 的参数是基于低秩假设学习的，即次优，因为最终目标是最小化图像重建误差。

¹ 在我们的分析中，如果 A 是一个矩阵，则 A_i 、 A_j 和 A_{ij} 分别表示它的第 i 行、第 j 列以及第 i 行和第 j 列的元素。

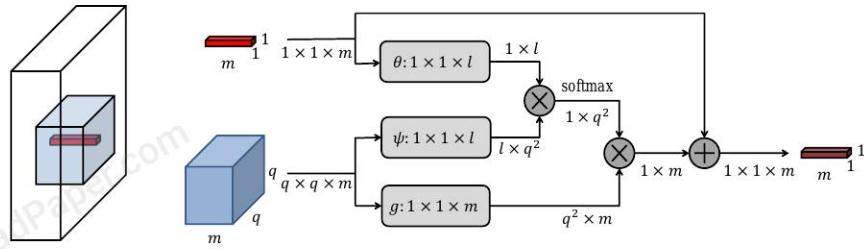


图 1：我们的非本地模块在单个位置上工作的图示。白色张量表示整个图像的深层特征表示。红色纤维是这个位置的特征，蓝色张量表示它附近的特征。 θ 、 ψ 和 g 通过 1×1 卷积后跟重塑操作来实现。

3.3 提议的非本地模块

基于 (1) 中的通用非局部框架，我们提出了另一种软块匹配方法，并应用欧几里得距离和线性嵌入高斯核[41]作为距离度量。线性嵌入定义如下：

$$\Phi(\mathbf{X})_i^j = \phi(\mathbf{X}_i, \mathbf{X}_j) = \exp\{\theta(\mathbf{X}_i)\psi(\mathbf{X}_j)^T\}, \quad \forall i, j, \quad (4)$$

$$\theta(\mathbf{X}_i) = \mathbf{X}_i \mathbf{W}_\theta, \quad \psi(\mathbf{X}_i) = \mathbf{X}_i \mathbf{W}_\psi, \quad \mathbf{G}(\mathbf{X})_i = \mathbf{X}_i \mathbf{W}_g, \quad \forall i. \quad (5)$$

嵌入变换 \mathbf{W}_θ 、 \mathbf{W}_ψ 和 \mathbf{W}_g 都是可学习的，分别具有 $m \times 1$ 、 $m \times 1$ 、 $m \times m$ 的形状。因此，建议的非局部操作可以写成

$$\mathbf{Z}_i = \frac{1}{\delta_i(\mathbf{X})} \sum_{j \in S_i} \exp\{\mathbf{X}_i \mathbf{W}_\theta \mathbf{W}_\psi^T \mathbf{X}_j^T\} \mathbf{X}_i \mathbf{W}_g, \quad \forall i, \quad (6)$$

其中 $\delta_i(x) = \sum_{j \in S_i} \phi(x_i, x_j)$ 。与 [2] 类似，为了获得 \mathbf{Z}_i ，我们评估 x_i 和邻域 S_i 中的每个 x_j 之间的相关性。 $\phi(x_i, x_j)$ 的更多选择在第 5 节中讨论。

所提出的非局部操作可以通过常见的可微分操作来实现，因此在合并到神经网络中时可以共同学习。我们通过添加跳过连接将其包装为非本地模块，如图 1 所示，因为跳过连接使我们能够将非本地模块插入任何预训练模型，同时通过初始化 \mathbf{W}_g 保持其初始行为作为零。这样的模块仅引入有限数量的参数，因为 θ 、 ψ 和 g 是 1×1 卷积并且 $m = 128$, $l = 64$ 。实际上，该模块在每个位置上的输出仅取决于其 $q \times q$ 邻域，因此该操作适用于各种大小的输入。

与其他方法的关系：最近的工作结合了非局部 BM 和神经网络来进行图像恢复 [30, 23, 41]。Lefkimiatis [23] 建议首先将 BM 应用于噪声图像块。硬 BM 结果用于对补丁特征进行分组，CNN 对匹配的补丁进行可训练的协作过滤。乔等。[30] 将类似的非局部 BM 与 TNRD 网络 [7] 相结合用于图像去噪。然而，作为传统方法 [8, 27, 15]，这些作品 [23, 30] 直接在退化的输入块上进行硬 BM，这在严重退化的图像上可能不准确。相比之下，我们提出的非局部操作作为软 BM 应用于学习到的深度特征表示，这些特征表示对退化更稳健。此外，[23] 中的匹配结果与神经网络隔离，类似于传统方法，而所提出的非本地模块以端到端的方式与整个网络联合训练。

王等。[41] 使用类似的方法将非本地操作添加到用于高级视觉任务的神经网络中。然而，与我们的方法不同，Wang 等人。[41] 计算整个图像的特征相关性。这相当于在我们的方法中将邻域扩大到整个图像。我们凭经验表明，由于对退化输入图像的相关性估计不准确，增加邻域大小并不总能提高图像恢复性能。因此，必须选择适当大小的邻域以获得图像恢复的最佳性能。此外，[41] 中的非本地操作只能处理固定大小的输入图像，而我们在 (6) 中的模块可以灵活处理各种图像大小。最后，当我们的非局部模块被纳入 RNN 框架时，允许相邻状态之间的相关信息流动，以增强对不准确相关估计的鲁棒性。这是处理退化图像的新独特公式。接下来提供更多详细信息。

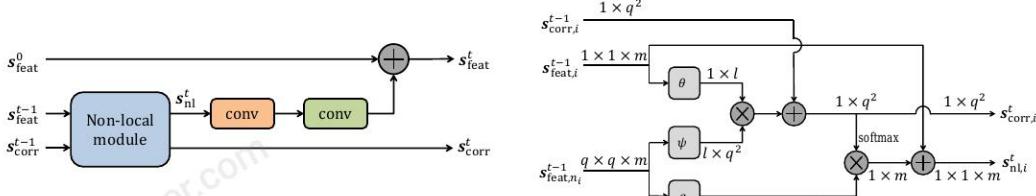


Figure 2: An illustration of the transition function
在拟议的 NLRN 中频繁出现。图 3: 针对单个位置 i 的操作

NLRN 中使用的非本地模块。

4 非局部循环网络

在本节中，我们描述了包含非本地模块以形成我们的 NLRN 的 RNN 架构。我们采用 RNN 的通用公式，它由一组状态组成，即输入状态、输出状态和循环状态，以及状态之间的转换函数。输入、输出和循环状态分别表示为 x 、 y 和 s 。在每个时间步 t ，RNN 接收输入 nxtts ， t 并且 RNN 的循环状态和输出状态递归更新如下：

$$s^t = f_{\text{input}}(x^t) + f_{\text{recurrent}}(s^{t-1}), \quad y^t = f_{\text{output}}(s^t), \quad (7)$$

其中 f 、输入 f 、输出和 $f_{\text{recurrent}}$ 在每个时间步重复使用。在我们的 NLRN 中，我们设置如下：

- s_0 是输入图像 I 的函数。
- $x^t = 0, \forall t \in \{1, \dots, T\}$, and $f_{\text{input}}(0) = 0$.
- 输出状态 y^t 仅在最终输出时计算。

我们添加了来自第一个状态的恒等路径，这有助于训练期间的梯度反向传播 [35]，以及每个位置与其前一状态的邻域之间的深度特征相关性的 $tdya$ 残差路径。因此， $s_t = \{s_{\text{feat}}, s_{\text{corr}}\}$ ， $s_{\text{corr}} = f(s_{\text{feat}}, s_{\text{corr}}, s_{\text{feat}})$ ， $s_{\text{feat}} = f(s_{\text{feat}}, s_{\text{corr}}, s_{\text{feat}})$ ， $\forall t \in \{1, \dots, T\}$ ，其中 s_{feat} 表示 tfreataitn 中的特征图深度特征相关性的集合。对于 f ，在从恒等路径添加特征 s_0 之前，首先采用递归的非局部 modcourlre ，然后是两个卷积层。非局部模块中的权重就像卷积层一样在循环状态之间共享，因此我们的 NLRN 仍然保持整体的高参数效率。

值得注意的是，在 non-local 模块内部，在 softmax normalization 之前，将来自先前状态的位置 i 的特征相关性 s_{cor}^i 添加到当前状态中的估计特征相关性，从而实现相关信息的传播在相邻状态之间进行更稳健的相关估计。具体可以看图 3。初始状态 s_0 设置为输入图像上一个卷积层之后的特征。 f_{output} 由另一个单个卷积层表示。除非本地模块外，所有层都有 128 个过滤器，内核大小为 3×3 。批量归一化和 ReLU 激活函数在每个卷积层之前执行 [11i8za]。我们采用残差学习，当 NLRN 展开 T 次时，NLRN 的输出是残差图像 $I^r = f_{\text{layer}}(os1Tlooutput)$ 。在训练期间，目标是

minimize the mean square error $\mathcal{L}(I, I^r) = \frac{1}{2} \|I - I^r\|^2$, where I denotes the ground truth image.

与其他 RNN 方法的关系：尽管之前已采用 RNN 进行图像恢复，但我们的 NLRN 是第一个将非局部操作合并到具有相关传播的 RNN 框架中的方法。DRCN 递归地将单个卷积层多次应用于输入特征图，而没有来自第一个状态的恒等路径。DRRN [a3151] 在每个状态中同时应用恒等路径和残差路径，但没有非局部操作，因此相邻状态之间没有相关信息流。MemNet 在几种类型的内存块之间建立密集连接，并且权重在同一类型的内存块中共享，但在不同类型之间是不同的。与 MemNet 相比，我们的 NLRN 具有高效但有效的 RNN 结构，有效深度较浅，参数较少，但获得了更好的恢复性能，这在第 5 节中有详细说明。

5 实验

数据集：对于图像去噪，我们采用两种不同的设置与最近基于深度学习的方法 wso[2d8i, f, f2e3re, e, n4t9s, s, e3tt6in]:g (1) As in]，we choose as

训练集来自训练集的 200 张图像和来自伯克利分割数据集 (BSD) [29] 的测试集的 200 张图像的组合，并在两个流行的基准上进行测试：Set12 和 Set68， $\sigma = 15, 25, 50$ 以下 [49]。 (2) 与 [28, 36] 一样，我们使用 BSD 中 train 集中的 200 张图像和 val 集中的 100 张图像的组合作为训练集，并在 Set14 和 200 张图像的 BSD 测试集上进行测试， $\sigma = [28, 36]$ 之后的 30, 50, 70。此外，我们在 Urban100 数据集 [19] 上评估了我们的 NLRN，该数据集包含丰富的结构模式和纹理，以进一步证明我们的 NLRN 使用图像自相似性的能力。训练集和测试集严格不相交，并且在每个实验设置中所有图像都转换为灰度。对于图像 SR，我们遵循 [20, 35, 36] 并使用 291 张图像的训练集，其中 91 张图像在 [46] 中提出，其他 200 张来自 BSD 训练集。我们采用四个基准集：Set5 [1]、Set14 [48]、BSD100 [29] 和 Urban100 [19]，使用三个放大因子： $\times 2, \times 3$ 和 $\times 4$ 进行测试。低分辨率图像通过双三次下采样合成。

训练设置：我们在训练期间从图像中随机抽取大小等于非局部操作邻域的补丁。我们使用翻转、旋转和缩放来增强训练数据。对于图像去噪，我们在训练过程中将零均值的独立同分布高斯噪声作为噪声输入添加到原始图像中。我们为每个噪声级别训练不同的模型。对于图像 SR，只有图像的亮度通道被超分辨，其他两个颜色通道通过双三次插值进行放大，遵循 [20, 21, 35]。此外，所有三个放大因子的训练图像： $\times 2, \times 3$ 和 $\times 4$ 通过双三次插值放大到所需的空间大小，并组合成一个训练集。我们使用这个集合为所有这三个放大因子训练一个单一模型，如 [20, 35, 36] 中所示。

我们使用 Adam 优化器来最小化损失函数。我们将初始学习率设置为 $1e-3$ 并在训练期间将其减少一半五次。我们对权重使用 Xavier 初始化。我们将梯度限制在 0.5 的范数以防止梯度爆炸，这在经验上显示可以加速训练收敛，并且我们在训练期间采用 16 作为小批量大小。使用 Titan Xp GPU 训练模型大约需要 3 天。对于非本地模块，我们对输入补丁外的邻域使用圆形填充。对于卷积，我们用零填充特征图的边界以保持特征图的空间大小。

5.1 模型分析

在本节中，我们从以下几个方面分析我们的模型。首先，我们进行了在非局部模块中使用不同距离度量的消融研究。表 1 比较了在 12 个展开步骤的 NLRN 中使用的实例化，包括欧氏距离、点积、嵌入点积、高斯、对称嵌入高斯和嵌入高斯。Embedded Gaussian 实现了最好的性能，并在以下实验中采用。

我们在表 2 的 PSNR 方面比较了 NLRN 及其变体。我们有一些观察结果。首先，具有无限制权重的相同模型比其权重共享对应模型表现更差。我们推测具有未绑定权重的模型容易出现模型过度拟合并且训练收敛速度慢得多，这两者都会破坏其性能。为了研究非局部模块的功能，我们实现了一个与 NLRN 参数数量相同的基线 RNN，发现它比 NLRN 差了大约 0.2 dB，显示了使用非局部图像属性进行图像恢复的优势。此外，我们实现了每隔一个状态或每三个状态使用非本地模块的 NLRN，并观察到如果 NLRN 中使用非本地模块的频率降低，性能会相应降低。我们通过在恢复精度方面与对应部分进行比较，展示了在相邻状态之间传播相关信息的好处。为了进一步分析非局部模块，我们在图 4 中可视化了非局部操作的特征相关图。可以看出，随着循环状态数量的增加，位置

表 1：我们提出的模型与 Set12 上各种距离度量的图像去噪比较，噪声级别为 25。

Distance metric	$\phi(\mathbf{X}_i, \mathbf{X}_j)$	PSNR
Euclidean distance	$\exp\{-\ \mathbf{X}_i - \mathbf{X}_j\ _2^2/h^2\}$	30.74
Dot product	$\mathbf{X}_i \mathbf{X}_j^T$	30.68
Embedded dot product	$\theta(\mathbf{X}_i)\psi(\mathbf{X}_j)^T$	30.75
Gaussian	$\exp\{\mathbf{X}_i \mathbf{X}_j^T\}$	30.69
Symmetric embedded Gaussian	$\exp\{\theta(\mathbf{X}_i)\theta(\mathbf{X}_j)^T\}$	30.76
Embedded Gaussian	$\exp\{\theta(\mathbf{X}_i)\psi(\mathbf{X}_j)^T\}$	30.80

表 2：我们的 NLRN 及其在 Set12 上的变体的图像去噪比较，噪声级别为 25。

Model	PSNR
NLRN w/o parameter sharing	30.65
RNN with same parameter no.	30.61
Non-local module in every other state	30.76
Non-local module in every 3 states	30.72
NLRN w/o propagating correlations	30.78
NLRN	30.80

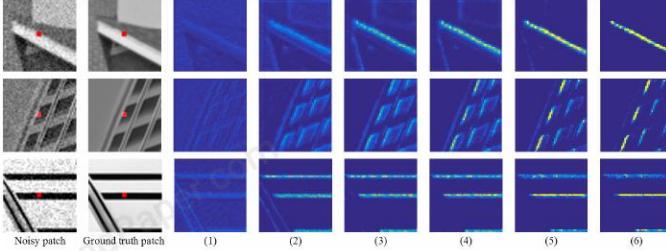


图 4: 用于图像去噪的非局部操作的相关图示例。Noisy patch/ground truth patch: 非局部操作中使用的红色中心像素的邻域。(1)–(6): 来自 NLRN 的循环状态 1–6 的相关图, 展开长度为 6。

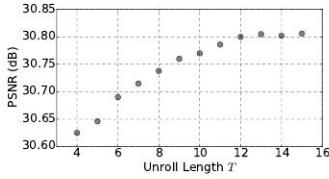


图 6: 我们提出的模型在 Set12 上的展开长度与图像去噪性能对比, 噪声级别为 25。

具有相似特征的特征在地图中逐渐显示出更高的相关性, 这证明了非局部模块在利用图像自相似性方面的有效性。

图 5 研究了非局部模块中邻域大小对图像去噪结果的影响。性能在 $q = 45$ 时达到峰值。这表明限制邻域有助于将相关性计算集中在空间附近的相关特征上并增强相关性估计。因此, 有必要选择合适的邻域大小 (而不是整个图像) 进行图像恢复。除非另有说明, 否则我们在本文的其余部分选择 $q = 45$ 。

展开长度 T 决定了 NLRN 的最大有效深度 (即最大卷积层数)。展开长度对图像去噪结果的影响如图 6 所示。性能随着展开长度的增加而增加, 但在 $T = 12$ 后达到饱和。考虑到恢复精度和推理时间之间的权衡, 我们对 NLRN 采用 $T = 12$ 在所有的实验中。

5.2 与最先进方法的比较

我们分别将我们提出的模型与许多最近的图像去噪和图像 SR 竞争者进行了比较。PSNR 和 SSIM [42] 用于测量定量恢复性能。

图像去噪: 为了与基于深度网络的其他方法进行公平比较, 我们在两种设置下训练我们的模型: (1) 我们使用 TNRD [7]、DnCNN [49] 和 NLNet [23] 中的训练数据, 以及结果如表 4 所示。我们在原始论文 [23] 中引用了 NLNet 的结果, 因为没有可用的公共代码或模型。(2) 我们使用 RED [28] 和 MemNet [36] 中的训练数据, 结果如表 5 所示。我们注意到 RED 使用多视图测试 [43] 来提高恢复精度, 即 RED 处理每个测试图像及其旋转和翻转版本, 然后对所有输出进行平均以形成最终的去噪图像。因此, 我们对 NLRN 执行相同的程序, 并发现其性能 (称为 NLRN-MV) 得到持续改进。此外, 我们在比较中包括了最近基于非深度学习的方法: BM3D [8] 和 WNNM [15]。我们没有列出平均性能比 DnCNN 或 MemNet 差的其他方法 [52, 3, 45, 6, 50]。我们的 NLRN 在 Urban100 上明显优于所有竞争对手, 并在几乎所有噪声水平和数据集上产生最佳结果。

为了进一步展示 NLRN 网络设计的优势, 我们将不同版本的 NLRN 与几种最先进的网络模型 (即表 3 中的 DnCNN、RED 和 MemNet) 进行了比较。NLRN 使用的参数最少, 但性能优于所有竞争对手。具体来说, NLRN 的好处

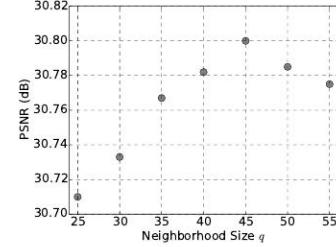


图 5: 我们在 Set12 上提出的模型的邻域大小与图像去噪性能的关系, 噪声级别为 25。

	DnCNN	RED	MemNet	NLRN
Max effective depth	17	30	80	38
Parameter sharing	No	No	Yes	Yes
Parameter no.	554k	4,131k	667k	330k
Multi-view testing	No	Yes	No	No Yes
Training images	400	300	300	400 300 300
PSNR	27.18	27.33	27.38	27.64 27.60 27.66

表 3: 我们提出的模型与噪声水平为 50 的 Set12 上最先进的网络模型的图像去噪比较。还比较了模型的复杂性。

表 4: 基准图像去噪结果。遵循[49]中的培训和测试协议。Set12、BSD68 和 Urban100 上各种噪声水平的平均 PSNR/SSIM。最佳表现以粗体显示。

噪声数据集 BM3D WNNM TNRD NLNet DnCNN NLRN								
	15	32.37/0.8952	32.70/0.8982	32.50/0.8958	-/-	32.86/0.9031	33.16/0.9070	
Set12	25	29.97/0.8504	30.28/0.8557	30.06/0.8512	-/-	30.44/0.8622	30.80/0.8689	
	50	26.72/0.7676	27.05/0.7775	26.81/0.7680	-/-	27.18/0.7829	27.64/0.7980	
	15	31.07/0.8717	31.37/0.8766	31.42/0.8769	31.52/-	31.73/0.8907	31.88/0.8932	
BSD68	25	28.57/0.8013	28.83/0.8087	28.92/0.8093	29.03/-	29.23/0.8278	29.41/0.8331	
	50	25.62/0.6864	25.87/0.6982	25.97/0.6994	26.07/-	26.23/0.7189	26.47/0.7298	
	15	32.35/0.9220	32.97/0.9271	31.86/0.9031	-/-	32.68/0.9255	33.45/0.9354	
Urban100	25	29.70/0.8777	30.39/0.8885	29.25/0.8473	-/-	29.97/0.8797	30.94/0.9018	
	50	25.95/0.7791	26.83/0.8047	25.88/0.7563	-/-	26.28/0.7874	27.49/0.8279	

表 5: 基准图像去噪结果。遵循[36]中的培训和测试协议。14 张图像、BSD200 和 Urban100 上各种噪声水平的平均 PSNR/SSIM。红色是最好的，蓝色是第二好的表现。

Dataset	Noise	BM3D	WNNM	RED	MemNet	NLRN	NLRN-MV
14 图片	30	28.49/0.8204	28.74/0.8273	29.17/0.8423	29.22/0.8444	29.37/0.8460	29.41/0.8472
	50	26.08/0.7427	26.32/0.7517	26.81/0.7733	26.91/0.7775	27.00/0.7777	27.05/0.7791
	70	24.65/0.6882	24.80/0.6975	25.31/0.7206	25.43/ 0.7260	25.49/0.7255	25.54/0.7273
BSD200	30	27.31/0.7755	27.48/0.7807	27.95/0.8056	28.04/0.8053	28.15/0.8423	28.20/0.8436
	50	25.06/0.6831	25.26/0.6928	25.75/0.7167	25.86/0.7202	25.93/0.7214	25.97/0.8429
	70	23.82/0.6240	23.95/0.6346	24.37/0.6551	24.53/0.6608	24.58/0.6614	24.62/0.6634
Urban100	30	28.75/0.8567	29.47/0.8697	29.12/0.8674	29.10/0.8631	29.94/0.8830	29.99/0.8842
	50	25.95/0.7791	26.83/0.8047	26.44/0.7977	26.65/0.8030	27.38/0.8241	27.43/0.8256
	70	24.27/0.7163	25.11/0.7501	24.75/0.7415	25.01/0.7496	25.66/0.7707	25.71/0.7724

表 6: 基准 SISR 结果。数据集 Set5、Set14、BSD100 和 Urban100 上比例因子 $\times 2$ 、 $\times 3$ 和 $\times 4$ 的平均 PSNR/SSIM。最佳表现以粗体显示。

Dataset	Scale	SRCNN	VDSR	DRCN	LapSRN	DRRN	MemNet	NLRN
Set5	$\times 2$	36.66/0.9542	37.53/0.9587	37.63/0.9588	37.52/0.959	37.74/0.9591	37.78/0.9597	38.00/0.9603
	$\times 3$	32.75/0.9090	33.66/0.9213	33.82/0.9226	33.82/0.923	34.03/0.9244	34.09/0.9248	34.27/0.9266
	$\times 4$	30.48/0.8628	31.35/0.8838	31.53/0.8854	31.54/0.885	31.68/0.8888	31.74/0.8893	31.92/0.8916
Set14	$\times 2$	32.45/0.9067	33.03/0.9124	33.04/0.9118	33.08/0.913	33.23/0.9136	33.28/0.9142	33.46/0.9159
	$\times 3$	29.30/0.8215	29.77/0.8314	29.76/0.8311	29.79/0.832	29.96/0.8349	30.00/0.8350	30.16/0.8374
	$\times 4$	27.50/0.7513	28.01/0.7674	28.02/0.7670	28.19/0.772	28.21/0.7721	28.26/0.7723	28.36/0.7745
BSD100	$\times 2$	31.36/0.8879	31.90/0.8960	31.85/0.8942	31.80/0.895	32.05/0.8973	32.08/0.8972	32.19/0.8992
	$\times 3$	28.41/0.7863	28.82/0.7976	28.80/0.7963	28.82/0.797	28.95/0.8004	28.96/0.8001	29.06/0.8026
	$\times 4$	26.90/0.7101	27.29/0.7251	27.23/0.7233	27.32/0.728	27.38/0.7284	27.40/0.7281	27.48/0.7306
Urban100	$\times 2$	29.50/0.8946	30.76/0.9140	30.75/0.9133	30.41/0.910	31.23/0.9188	31.31/0.9195	31.81/0.9249
	$\times 3$	26.24/0.7989	27.14/0.8279	27.15/0.8276	27.07/0.827	27.53/0.8378	27.56/0.8376	27.93/0.8453
	$\times 4$	24.52/0.7221	25.18/0.7524	25.14/0.7510	25.21/0.756	25.44/0.7638	25.50/0.7630	25.79/0.7729

来自固有的参数共享，仅使用不到 RED 参数的 $1/10$ 。与 RNN 的竞争对手 MemNet 相比，NLRN 仅使用一半的参数和更浅的深度来获得更好的性能，这显示了我们的非局部循环架构的优越性。

图像超分辨率: 我们将我们的模型与表 6 中的几种最新 SISR 方法进行了比较，包括 SRCNN [10]、VDSR [20]、DRCN [21]、LapSRN [22]、DRRN [35] 和 MemNet [36]。我们在计算 PSNR 和 SSIM 之前裁剪图像边界附近的像素，如 [10, 33, 20, 21] 中所述。我们没有列出其他方法 [19, 33, 25, 34, 16]，因为它们的性能比 DRRN 或 MemNet 差。此外，我们在比较中不包括 SRDenseNet [39] 和 EDSR [24]，因为这两个网络模型中的参数数量比我们的 NLRN 大两个数量级以上，并且它们的训练数据集明显大于我们的。可以看出，NLRN 在所有放大因子和数据集中产生了最好的结果。补充材料中提供了视觉结果。

六，结论

我们提出了一种新的有效的循环网络，它结合了用于图像恢复的非本地操作。所提出的非本地模块可以与循环网络进行端到端的训练。我们已经研究了在封闭邻域内针对整个图像计算可靠特征相关性的重要性，并展示了在相邻循环阶段之间传递特征相关性消息的好处。对图像去噪和超分辨率基准的综合评估证明了 NLRN 优于现有方法。

References

- [1] M. Bevilacqua、A. Roumy、C. Guillemot 和 M. L. Alberi-Morel。低复杂度的单图像超基于非负邻居嵌入的分辨率。2012. [2] A. Buades、B. Coll 和 J.-M.羊肚菌。一种非局部图像去噪算法。在 CVPR, 2005 年。[3] H. C. Burger、C. J. Schuler 和 S. Harmeling。图像去噪：普通神经网络能否与 bm3d 竞争？在 CVPR, 2012 年。[4] S. Chandra、N. Usunier 和 I. Kokkinos。使用深度嵌入的密集和低阶高斯 crfs。在 ICCV, 2017 年。[5] H. Chang, D.-Y. Yeung 和 Y. Xiong。通过邻居嵌入的超分辨率。在 CVPR, 2004 年。[6] F. Chen、L. Zhang 和 H. Yu。用于图像去噪的外部补丁先验引导内部聚类。In ICCV, 2015. [7] Y. Chen 和 T. Pock。可训练的非线性反应扩散：用于快速有效图像恢复的灵活框架。IEEE TPAMI, 2017. [8] K. Dabov、A. Foi、V. Katkovnik 和 K. Egiazarian。通过稀疏 3-d 变换域进行图像去噪
协同过滤。IEEE TIP, 2007. [9] A. Danielyan、V. Katkovnik 和 K. Egiazarian。Bm3d 帧和变分图像去模糊。TIP, 21(4):1715 - 1728, 2012. [10] C. Dong, C. C. Loy, K. He, and X. Tang。学习用于图像超分辨率的深度卷积网络。在 ECCV, 2014 年。[11] C. Dong、C. C. Loy 和 X. Tang。加速超分辨率卷积神经网络。在 ECCV, 2016 年。[12] G. Freedman 和 R. Fattal。来自本地自身示例的图像和视频升级。ACM 图形交易 (TOG), 2011. [13] J. Gehring、M. Auli、D. Grangier、D. Yarats 和 Y. N. Dauphin。卷积序列到序列
学习。在 ICML, 2017 年。[14] D. Glasner、S. Bagon 和 M. Irani。来自单个图像的超分辨率。In ICCV, 2009. [15] S. Gu, L. Zhang, W. Zuo, and X. Feng。应用于图像去噪的加权核范数最小化。In CVPR, pages 2862 - 2869, 2014. [16] W. Han, S. Chang, D. Liu, M. Yu, M. Witbrock, and T. S. Huang。通过双态递归网络实现图像超分辨率。在 CVPR, 2018 年 6 月。[17] A. W. Harley、K. G. Derpanis 和 I. Kokkinos。使用局部的分段感知卷积网络
注意面具。In ICCV, 2017. [18] K. He, X. Zhang, S. Ren, and J. Sun。深度残差网络中的身份映射。在 ECCV, 2016. [19] J.-B. Huang、A. Singh 和 N. Ahuja。来自转换后的自样本的单图像超分辨率。在 CVPR, 2015 年。[20] J. Kim、J. Kwon Lee 和 K. Mu Lee。使用非常深的卷积网络进行精确的图像超分辨率。在 CVPR, 2016 年。[21] J. Kim、J. Kwon Lee 和 K. Mu Lee。用于图像超分辨率的深度递归卷积网络。在 CVPR, 2016 年。[22] W.-S. 赖, J.-B. Huang、N. Ahuja 和 M.-H. 阳。快速准确的深度拉普拉斯金字塔网络
超分辨率。在 CVPR, 2017 年。[23] S. Lefkimiatis。使用卷积神经网络的非局部彩色图像去噪。在 CVPR, 2017 年。[24] B. Lim、S. Son, H. Kim、S. Nah 和 K. M. Lee。用于单幅图像超分辨率的增强深度残差网络。在 CVPR 研讨会上, 2017 年。[25] D. Liu、Z. Wang、B. Wen、J. Yang、W. Han 和 T. S. Huang。鲁棒的单图像超分辨率通过
具有稀疏先验的深层网络。TIP, 25(7):3194 - 3207, 2016. [26] D. Liu, B. Wen, X. Liu, Z. Wang, and T. S. Huang。当图像去噪遇到高级视觉任务时：一种深度学习方法。在 IJCAI, 2018 年。[27] J. Mairal、F. Bach、J. Ponce、G. Sapiro 和 A. Zisserman。用于图像恢复的非局部稀疏模型。In ICCV, 2009. [28] X. Mao, C. Shen, and Y.-B. 阳。使用具有对称跳过连接的非常深的卷积编码器-解码器网络进行图像恢复。在 NIPS, 2016 年。[29] D. Martin、C. Fowlkes、D. Tal 和 J. Malik。人体分割自然图像数据库及其
应用于评估分割算法和测量生态统计。In ICCV, 2001. [30] P. Qiao, Y. Dou, W. Feng, R. Li, and Y. Chen。学习用于图像去噪的非局部图像扩散。在 ACM 多媒体会议上, 2017 年。[31] L. I. Rudin 和 S. Osher。具有自由局部约束的基于全变差的图像恢复。在 ICIP, 1994 年。[32] A. Santoro、D. Raposo、D. G. Barrett、M. Malinowski、R. Pascanu、P. Battaglia 和 T. Lillicrap。
一个简单的
用于关系推理的神经网络模块。在 NIPS, 2017 年。[33] S. Schulter、C. Leistner 和 H. Bischof。使用超分辨率森林快速准确地放大图像。在 CVPR, 2015 年。

[34] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert 和 Z. Wang. 即时的

使用高效的亚像素卷积神经网络的单个图像和视频超分辨率。在 CVPR, 2016 年。 [35] Y. Tai, J. Yang 和 X. Liu. 通过深度递归残差网络实现图像超分辨率。在 CVPR, 2017 年。 [36] Y. Tai, J. Yang, X. Liu 和 C. Xu. Memnet: 用于图像恢复的持久内存网络。在 ICCV, 2017 年。 [37] R. Timofte, V. De 和 L. Van Gool. 锚定邻域回归, 用于快速基于示例的超级

解决。在 ICCV, 2013 年。 [38] C. Tomasi 和 R. Manduchi. 灰度和彩色图像的双边滤波。 In ICCV, 1998. [39] T. Tong, G. Li, X. Liu, and Q. Gao. 使用密集跳跃连接的图像超分辨率。在 ICCV, 2017 年。 [40] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser 和 I. Polosukhin.

注意力就是你所需要的。在 NIPS, 2017 年。 [41] X. Wang, R. Girshick, A. Gupta 和 K. He. 非局部神经网络。 arXiv 预印本 arXiv:1711.07971, 2017. [42] Z. Wang, A. C. Bovik, H. R. Sheikh 和 E. P. Simoncelli. 图像质量评估: 来自错误可见性

到结构相似性。 IEEE TIP, 2004. [43] Z. Wang, D. Liu, J. Yang, W. Han 和 T. Huang. 具有稀疏图像超分辨率的深度网络

事先的。在 ICCV, 2015 年。 [44] B. Wen, S. Ravishankar 和 Y. Bresler. 结构化过完备稀疏变换学习

收敛保证和应用。 IJCV, 2015. [45] J. Xu, L. Zhang, W. Zuo, D. Zhang, and X. Feng. 基于补丁组的非局部自相似先验学习图像去噪。 In ICCV, 2015. [46] J. Yang, J. Wright, T. S. Huang, and Y. Ma. 通过稀疏表示的图像超分辨率。 IEEE TIP, 2010. [47] R. Yin, T. Gao, Y. M. Lu 和 I. Daubechies. 两个基础的故事: 图像的局部非局部正则化

带有卷积小框架的补丁。 SIAM 影像科学杂志, 10(2):711 - 750, 2017. [48] R. Zeyde, M. Elad 和 M. Protter. 使用稀疏表示对单个图像进行放大。在国际

曲线与曲面会议, 2010. [49] K. Zhang, W. Zuo, Y. Chen, D. Meng 和 L. Zhang. 超越高斯降噪器: 用于图像降噪的深度 cnn 的残差学习。 IEEE TIP, 2017. [50] K. Zhang, W. Zuo, S. Gu 和 L. Zhang. 在图像恢复之前学习深度 cnn 降噪器。在 CVPR, 2017 年。 [51] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang 和 P. H. Torr. 有条件的

随机场作为递归神经网络。在 ICCV, 2015 年。 [52] D. Zoran 和 Y. Weiss. 从自然图像块的学习模型到整个图像恢复。在 ICCV, 2011 年。