

指导教师： 杨涛

提交时间： 2015.3.29

The task of  
**D**igital Image Processing

数字图像处理

School of Computer Science

No: 1

姓名： 张玉柱

学号： 2012302491

班号： 10011205



# 对于学习数据依赖的卷积滤波器森林

Sean Ryan Fanello<sup>1,2</sup> Cem Keskin<sup>1</sup> Pushmeet Kohli<sup>1</sup> Shahram Izadi<sup>1</sup>  
Jamie Shotton<sup>1</sup> Antonio Criminisi<sup>1</sup> Ugo Pattacini<sup>2</sup> Tim Paek<sup>1</sup>

微软研究院<sup>1</sup> iCub Facility - Istituto Italiano di Tecnologia<sup>2</sup>

## 摘要

我们提出“滤波森林”，对于给定一个信号和背景是一种有效的，新的具有识别能力的方法来预测连续变量。滤波森林可以被用于一般的信号恢复工作，这是通过滤波器卷积来解决的，它试图学习最佳过滤内核使其适用于每一个数据点。这个模型可以同时学习内核及其值的大小，以观察和空间及时间的背景条件下。根据效率和准确性，我们表明，滤波森林优于基于马尔可夫随机场和最近提出的以回归森林为基础的方法。特别地，我们将证明对于自然图像，滤波森林如何用来学习最佳去噪滤波器以及用于其他任务，如深度图像细化和信号量值估计。许多的试验和比较说明滤波森林在实现精确性上优于或者与最先进的艺术手法相媲美，并且快上几个数量级。

## 1. 介绍

概率模型比如两两马尔可夫随机场和条件随机场已经被用于解决在计算机视觉中遇到的逐像素标签问题，包括语义分割，光流，图像降噪和立体声。这些模型允许相互作用之间的关系变量（如那些对应于相邻像素）进行建模，并且通常导致的结果是使图像边缘变得平滑。

尽管结果令人印象深刻，基于场的模型有两个缺点：估计模型的结构和参数比较困难，在此模型下推理出最大后验概率的解决方案代价巨大。这导致研究人员研究更快的基于森林的替代品。

决策森林已经被用来做比如身体部位分类和器官检测的问题，当它们有效的预测时，也有一个重要的缺点。一般来说，基于森林有关的模型进行预测是建立在输出变量（比如像素标签）是独立的假设下的，

因此没有实施空间的平滑度。最近的一个例外是发生在作者研究迭代的使用，分类预测叠加技术（离散）的标签，当就空间背景而言时。

在这篇论文中，我们提出滤波森林，一个非迭代的基于森林的预报器为了连续变量的估计，这可以将学到的模型或者当前背景合并。我们在这里应用滤波森林是为了修复损坏的 $n$ 维信号的任务。

本文的主要贡献是使用一个高效的多尺度决策森林，训练有区别的预测滤波器应适用于任何给定的变量（如像素），滤波森林有如下的性能：(i)树递归地划分输入信号，这样一个简单的卷积核对于

每一个叶子节点来说都是合适的；(ii)我们训练森林用一个新的，正则最小二乘法使卷积核的每一个叶子节点误差最小；(iii)我们能够实现最高水平的精确度，比当前最先进的速度还要快上几个数量级；(iv)没有基于场的后处理是必需的。

我们对许多任务评估FFs包括自然图像去噪，深度图像的细化，一维信号幅度估计。流行的去噪图像数据集的实验表明我们的方法实现准确性优于最先进的水平，快四个数量级。改进的结果也证明了深度图像的细化和一维信号的幅度估计，暗示了我们的方法的通用性。我们相信滤波森林可以扩展到许多其他相关的任务，如学的边缘检测，图像锐化，甚至形态滤波离散型分类任务。

### 1.1. 相关工作

在图像处理中的卷积：对于图像去噪，托马西双边滤波和维纳外推插值是非常受欢迎的滤波器。前者取代了强度在具有加权像素 $i$ 附近的强度的组合，在像素 $i$ 权重根据强度的差异和空间的距离。后者定义了局部滤波器基于局部方差估计。两者都尽量维持边界，尽管它们很流行，但是它们不能总是获得好的去噪结果。

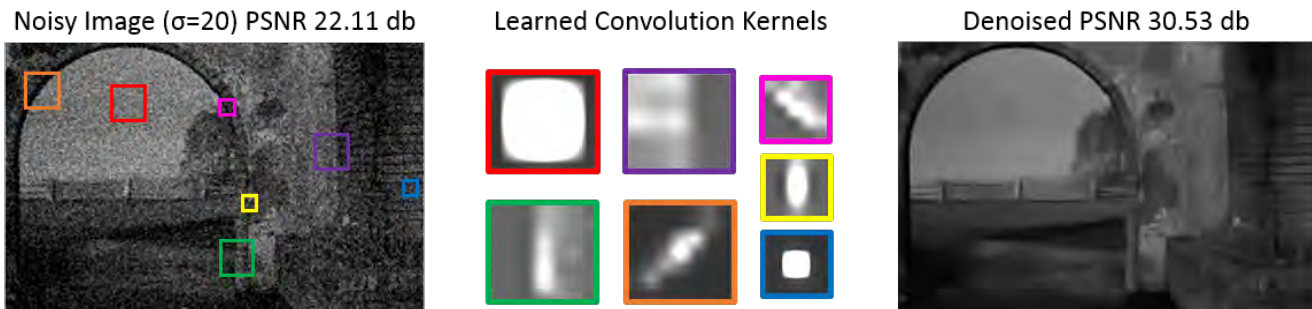


图1. 滤波森林学习空间变化的滤波器. 滤波森林学习最优卷积内核作为局部图像外观的函数. (左) 输入噪声图像, 叠加选择位置. (中) 相应的位置学习卷积滤波(颜色编码). (右) 去噪图像. 在无纹理的区域 (红色方块) 滤波森林学会使用大的, 各向同性平滑滤波器. 相反, 在对象边界(橙色或绿色方块)滤波森林学会使用定向, 边缘保持滤波器.

进一步自适应方法包括非局部的方法K-SVD, LSSC和CSR. 其中最精确的方法之一是三维块匹配(BM3D)算法, 使用含噪图像的参考块的局部性来获得一个单一预测.

空间相干性滤波, 相当数量的论文已经考虑过像素标记的任务使用滤波. 用卷积过滤器一元的可能性用于推断问题的解决方案, 如最佳流, 立体声, 图形-背景分割和Kontschieder等. 在一些顺序的分类方法综合这些想法, 从预测因子的单层过滤输出作为预测因子下一层的输入. 这些方法是有效的, 但他们的建模能力是有限的, 事实上只有一个固定形式的滤波函数是可以使用的. 相比之下, 我们的方法是使用一个森林来选择滤波器, 滤波器从一个潜在的无限制大的词典中被应用. 我们的方法可以被认为是一般化的卷积模型, 基于输入信号的局部外观, 其中的内核对于整个图像来

说不再是确定的, 但是空间上是变化的. 决策树结构化输出预测. 在随机领域结合决策森林的使用进行图形去噪的目的已经被研究, 还有, Nowozin等. 已经展示了如何任意的, 依赖于数据成对的潜在在CRF模型可以使用决策树编码, 并有效的学习. 决策树也已用于降低推理方法的计算复杂度. Shapovalov等人最近展示了如何使用推理机框架提出了结合决策森林有效分配类标签三维点. 我们的工作的不同之处在于, 它不要求这些方法明确的推论, 它是非迭代的并且包含一个简单的 (空间上变化的) 卷积运算.

## 2. 方法

在图像处理和计算机视觉上的许多问题可以被阐述

为一个回归问题, 我们想映射  $f_w: \mathcal{X} \rightarrow \mathcal{Y}$ , 从输入空间  $\mathcal{X}$  到产出空间  $\mathcal{Y}$  是受一些参数  $w$  参数化的, 学习这些预测一般被阐述为使用实验风险最小化的原则

$$w^* = \underset{w}{\operatorname{argmin}} \sum_{i \in T} L(y_i^{GT}, f_w(x_i)) \quad (1)$$

其中,  $T$  是训练数据组成的双输入-预期的 (地面真值) 输出对  $(x_i, y_i^{GT})$ , 当真正解决方案为  $y_i$  的时候函数  $L$  表示计算预测  $f_w(x_i)$  的损失

举个例子, 考虑信号处理的任务需要修复原始信号  $y$ , 或者是它的一些特征 (如边缘), 从可能的含噪声的观察序列  $x$ . 噪声的类型可以是附加的, 比如  $x = y + \xi$ , 或者是相乘的, 泊松分布等等.  $x$  与内核  $w$  的卷积被认为是得到一个近似原始信号  $y$  的有效方法. 如果我们假设预测函数  $f_w(x_i)$  的形式是卷积并且损失是绝对值, 然后我们可以发现最佳的核心是:

$$w^* = \underset{w}{\operatorname{argmin}} \sum_{i \in T} \|y_i^{GT} - x_i \overset{\wedge}{w}\|^2 \quad (2)$$

其中,  $i$  是数据点索引,  $x_i$  是  $k$  维数据点  $i$  的背景, 然而, 固定内核的应用均匀到整幅图像很少得到一个令人满意的结果, 因为在整个信号中信号和噪声特征是变化的, 我们的目的是从一个训练集的最优内核中训练一个模型使其能够高效的匹配一个数据点和背景.

我们提出了用随机树解决这个问题, 在叶节点储存最优滤波  $w$  并且有效的把它们分配给输入  $x_i$ . 值得注意的是, 这些滤波可以是局部的, 具体的和简单的, 但树仍然可以处理复杂类型的噪声 (像一个分段线性函数融合到一个连续函数). 例如FF可以通过近似处理乘性噪声的响应在每一叶节点加性噪声. 因此我们选择通过假设加性噪声所学习的存储线性过滤器, 其通过最小二乘优化得到有效的封闭形式解.

$$\mathbf{w}^* = \underset{\mathbf{w}}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{X}\mathbf{w}\|^2, \quad (3)$$

其中,  $\mathbf{y} \in \mathbb{R}^N$  是  $N$  标定的真实数据 (无噪音) 的向量, 并且  $\mathbf{X} \in \mathbb{R}^{N \times k}$  是矩阵  $N$  的观察值, 其每一行都是  $\mathbf{x}_i^T$ .

需要注意的是, 线性滤波  $\mathbf{X}\mathbf{w}$  可以替换为一个非线性函数  $f_{\mathbf{w}}(\mathbf{x})$  适用于其他类型的问题。例如一个中值滤波器对不同类型的噪声, 或者一个  $s$  形的函数, 其可以把一个离散的标记问题转化为连续的, 所以它可以通过回归得到解决。

现在, 我们将更加详细地描述这一模式, 从这点上, 我们将集中于一个特定的应用程序, 即图像降噪, 使说明更直观。

## 2.1. 滤波森林

我们使用回归森林去学习图像块  $\mathbf{x}$  的空间分割, 以便在每一个分区内简单的线性卷积可以产生 (可以看见的) 好的结果。我们将展示: i) 如何训练 FF 最大限度地已从标记的训练数据定义明确的去噪损失

ii) 我们的正则矩阵如何激励边缘保护滤波, 和 iii) 对于每个内核, 如何学习一个适当的规模。在第三专栏我们的实验中表明在测试时间内 FF 以很快的速度精确地去噪。下面, 在呈现更多的 FF 的细节之前, 先简短的回顾一下决策森林的标准。

### 2.1.1 决策森林卷积

森林是一个  $T$  的决策树的集成 [1, 2]。每棵树由非终止 (分割) 结点和终止 (叶子) 节点组成。在测试时间一个像素  $i$  被传递到根节点。在每个分割节点  $j$ , 评估分割函数  $f(i; \theta_j)$  这个计算二元判定基于环绕图像像素  $i$  的函数, 函数是基于学习参数  $\theta_j$ 。依据二元判定 像素通过向左或右子节点。当到达叶子节点时, 在类标签 (用于分类) 存储的直方图或连续变量 (回归) 的密度被输出。

森林中的每一棵树训练是独立的。提供一组示例像素指数  $i$  和标定好的真实数据。从根节点开始, 随机地提出一组候选的二元分割函数的参数  $\theta$ 。对于每一个候选的, 该组训练的像素被分割为左子结点和右子节点的集合。考虑到每个候选分区, 评估目标函数 (典型的信息增益对于分类问题) 并且选择最好的  $\theta$ 。训练然后继续往下分割树, 递归分区原始集成分成先后更小的子集。当一个节点到达最大深度  $D$  或者包含太少的示例停止训练。

<sup>1</sup> 函数  $f$  可以是任意函数的图像区域周围的像素  $i$  并不仅限于只使用图像块  $\mathbf{x}_i$ 。

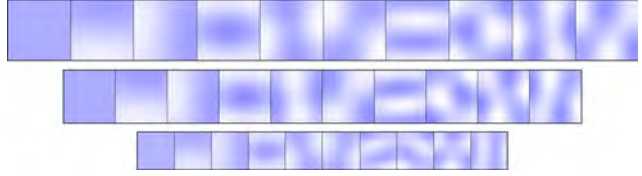


图2. 基于主成分分析的图像特征. 前10个主成分 (伪彩色显示) 从噪声训练块中计算得到, 块的分辨率为  $11 \times 11$ ,  $7 \times 7$ , 和  $3 \times 3$ . 他们代表平滑区域, 边角, 以及更复杂的图案。注意这些主成分分析组件只用来计算决策森林用于分区的外观空间的特征; 最终的滤波输出图像从噪声的输入直接计算。

### 2.1.2 滤波森林的细节

训练数据. 对于滤波森林, 我们使用成对的含噪图像块  $\mathbf{x}_i$  与其相应的无噪声的标定好的真实数据值为  $y_i$  (仅对于中央像素)。<sup>2</sup> 由于是标准的去噪技术, 我们的实验, 我们通过向无噪声的标定好真实数据的图像添加特定特性的合成噪声生成训练数据 (如高斯  $\sigma = 20$  灰度级)。

特征类型. 我们在滤波森林中使用两种类型的图像特征。符号  $\theta$  用于封装使用哪种类型的特征, 特征的参数, 应用于标量值的阈值导致  $a$  必须的二元决策。  $\theta$  的各个方面通过随机优化在每一个分割点学习 [8]。

第一种类型的特征是多尺度滤波器组 (正如图2所阐述的) 在每个分辨率为  $3 \times 3$ ,  $5 \times 5$ ,  $7 \times 7$ ,  $9 \times 9$ , 和  $11 \times 11$  的块, 主成分分析是在所有的噪声训练块上执行。为训练和测试, 该滤波器组在每个分辨率上应用图像卷积。我们允许这个特性的两个子类型: 一元和配对。一元特征需要简单的响应在其中之一分辨率上像素  $i$  的最高之一的10个主成分分析, 而配对的特征在其中之一分辨率上, 在两个不同的主成分分析组件会有不同的响应 (在像素  $i$ )。分割函数  $f$  然后简单地应用学到的阈值以此来得到标量值。一元/配对, 主成分分析特征组件, 分辨率, 最佳阈值的选择是通过最优节点优化火的。我们采用主成分分析本征载体, 而不是预先定义的滤波器组 [14] 因为主成分分析的不同类型信号的通用性。此外, 主成分分析确保主方向表示合适的特征, 对于内部节点与坐标轴对称的分割。

第二类型的特征是基于在每一个象素, 局部的估计的均匀性。在每个尺度为  $3 \times 3$ ,  $5 \times 5$ ,  $7 \times 7$ ,  $9 \times 9$ , 和  $11 \times 11$ , 我们计算每个图像块的方差。

<sup>2</sup> 当我们使用  $\mathbf{x}_i$  去方便的表示一个特别的噪音图像块, 对邻近的像素这些块会共享值, 个体  $\mathbf{x}_i$  是不需要显示计算的

我们只使用这种类型的特征的一元变量. 函数  $f$  适用于像素  $i$  的响应方差在学习阈值时的尺度之一。

训练目标 我们工作的主要贡献之一是训练目标的形式, 让  $S_j$  表示训练数据的子集, 其可以到达树的任意节点  $j$ , 分割函数在节点  $j$  被参数化为参数, 分割  $S_j$  为左右子集分别为  $S_j^L(\theta)$  和  $S_j^R(\theta)$ 。训练算法通过最小化能量  $E(\theta, S_j)$  作为选择分割函数的参数

$$\theta_j = \operatorname{argmin} E(\theta, S_j), \quad (4)$$

$E$  被定义为两个孩子节点的能量加权和。

$$E(\theta, S_j) = \sum_{c \in \{L, R\}} |S_j^c(\theta)| E^c(S_j^c(\theta)). \quad (5)$$

每个孩子节点的能量计算公式

$$E^c(S_j^c) = \|\mathbf{y}_j^c - \mathbf{x}_j^c \mathbf{w}^*\|^2, \quad (6)$$

其中

$$\mathbf{w}^* = \operatorname{argmin}_{\mathbf{w}} (\|\mathbf{y}_j^c - \mathbf{x}_j^c \mathbf{w}\|^2 + \|\Gamma(\mathbf{x}_j^c, \mathbf{y}_j^c) \mathbf{w}\|^2), \quad (7)$$

在这里,  $\mathbf{x}_j^c$  和  $\mathbf{y}_j^c$  是从  $S_j^L$  和  $S_j^R$

计算得到的, 它代表噪声数据块和节点  $j$  分别地到达左子节点 ( $c = L$ ) 或者右子节点 ( $c = R$ ) 无噪声标定真实数据值的集合。此外,  $\Gamma(\mathbf{x}, \mathbf{y})$  代表依靠数据的正则加权矩阵。

依赖数据正则化训练。上面提到的优化的任务将是一个标准的正则化最小二乘问题如果不使用  $\Gamma(\mathbf{x}, \mathbf{y})$ 。为以后将变得更清楚,  $\Gamma$  激励边缘保持正则化。最小化问题的封闭形式的解决方案 (7) 可以表述为:

$$\mathbf{w}^* = (\mathbf{X}^T \mathbf{X} + \Gamma^T \Gamma)^{-1} \mathbf{X}^T \mathbf{y}. \quad (8)$$

如果  $\Gamma$  被设置为标识, 然后规范化项激励内核有较小的规范。在这项工作中, 我们代替研究依赖于数据的规范化使用, 仅当它们不同于在块的中心的像素; 时, 鼓励  $w$  的小条目。

为了实现这项工作, 我们使用一个大小为  $p^2 \times p^2$  的对角矩阵  $\Gamma$  计算作为数据矩阵  $\mathbf{X}$  的函数, 并且标定好的数据值如下:

$$\gamma_d(\mathbf{x}, \mathbf{y}) = \frac{1}{N} \sum_{i=1}^N (x_{i,d} - y_i)^2, \quad (9)$$

其中  $d \in \{1, \dots, p^2\}$  块中的索引像素,  $N$  是在  $\mathbf{X}$  中样本的总数(比如行数),

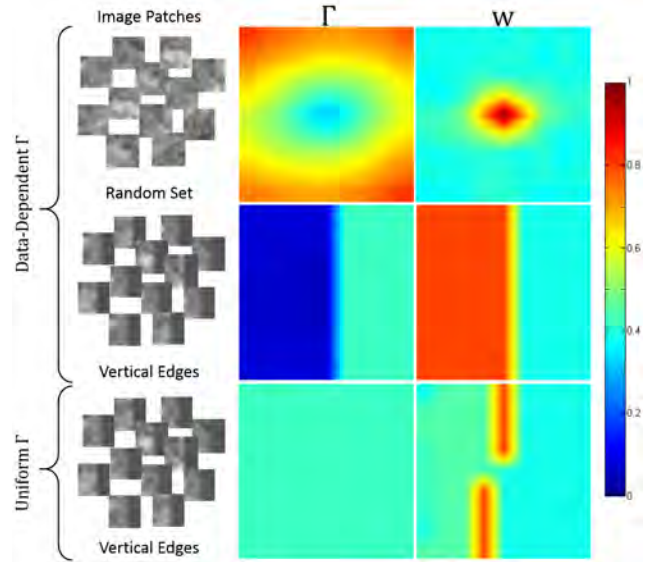


图 3.  $\Gamma$  加权矩阵的重要性. 第1列: 图像块的不同子集组成  $\mathbf{X}$ 。第2列: 产生加权矩阵  $\Gamma(\mathbf{x}, \mathbf{y})$ 。第3列: 滤波学习  $\mathbf{w}^*$ 。第1行和第2行根据公式 (9) 使用  $\Gamma$  计算, 而第3行使用  $\Gamma = \mathbf{I}$ 。行1使用一组随机选择的块, 而行2和3使用具有垂直边缘刚到中心像素右边的一组块。我们的加权正规化鼓励边缘滤波器。详情见文章。

$x_{i,d}$  和  $y_i$  分别是  $\mathbf{x}$  和  $\mathbf{y}$  的条目。直观的说, 矩阵  $\Gamma$  表明哪一个噪声图像块区域最有可能描述在块的中心标定好的真实数据  $y_i$ 。有更大值  $\gamma_d$  的像素  $d$  阻止滤波器  $w$  使用像素  $d$ , 而具有较小值的  $\gamma_d$  会鼓励滤波器使用像素  $d$ 。

为了阐述正则化项的重要性, 我们在图3展示了两组块: 随机生成 (第一行) 和 垂直边缘 (第二行和第三行)。从图中我们计算  $\Gamma$  (两列) 并且使用 (8) 滤波器  $w$  (三列)。正如我们所预料的, 使用随机块我们得到滤波器具有各相同性。在第二行, 我们选择包含垂直边缘的块, 其中中央像素在左边缘。正如所预料的, 正则矩阵鼓励滤波器考虑那些像素的左边缘, 他们有可能成为中央像素的良好预测因子这是一个正则化鼓励边缘保持滤波器的例子。

图 3 的第三行展示了通过设置  $\Gamma = \mathbf{I}$  对包含垂直边缘的块获得解决方案。在最后的情况, 滤波器所有的  $d$  维同样的正则化 (没有数据依赖性), 其结果是一个欠佳的滤波器, 不大可能归纳的很好。

多尺度滤波器学习. (7)的最优化是通过测试不同大小的滤波器, 从 $3 \times 3$  到  $11 \times 11$ , 并选择对应最小方差的一个. 这个简单的方案允许我们根据像素周围背景来计算每个叶子节点合适的内核大小. 数据依赖的内核的例子如图1所示. 平滑区域很有可能被大的内核 (红色) 恢复, 而细节区域则要小的或者定向滤波器来恢复 (橙色)。

当数达到一定的深度时或者样例的数量太少时, 训练阶段结束。

每个叶子节点存储一个滤波核函数 $w_{j,i}$ , 学习使用最小平方误差计算到达叶子节点 $j(i)$ 标记训练样例的子集, 使其最小。

测试信号滤波时间. 在测试时间, 对于每个树, 该测试像素 $i$ 被放到根节点. 然后每个分割节点将它发送到孩子节点之一直到其达到叶子节点  $j(i)$ . 得益于节点层次结构, 其下降是很高效的。核函数  $w_{j(i)}$  存储在被搜索的叶子节点  $j(i)$ , 并且在 $i$ 上作为一种标量积应用在噪声图像块, 来产生去噪输出

$y_i = \mathbf{x}_i^\top \mathbf{w}_{j(i)}$ . 因此, 不同的核应用在每个输入块以块的外观为条件. 在森林中 每棵树被轮流应用并且结果取平均. 为了增加效率, 不同的树都能被独立和并行处理。

算法复杂度. 算法的执行时间的复杂性线性依赖信号的大小如  $O(|\mathbf{x}|)$ . 特别地, 对于图形降噪, 每个像素的运算数目是:

$$10O(p^2) + T(O(p^2)) + O(d) \quad (10)$$

其中 $p^2$ 表示最大的块的大小 (我们的情况是  $11 \times 11$ ). 第一项 $10O(p^2)$  由于主成分分析投影运算,  $T O(p^2)$ 是预测滤波器 $w$ 和块的标量积. 最后一项  $T O(d)$  是深度为 $d$ 的T树 降序森林的代价. 这些代价可以被忽略, 我们可以近似的认为每个像素运行时间为 $(T+10)O(p^2)$ . 这些运算可以并行, 因为最终结果不依赖于其他图像像素的结果. 这进一步降低了整体运行代价. 值得注意的是, 每个像素的代价远低于艺术方法最快状态之一 [9] (运行时间比较见表 1)。

实施细则. 在训练过程中, 对每棵树节点, 我们使用容器取样[28] 去选择样例像素 $i$ 这加快了训练过程同时确保一定程度的随机性. 只有在多个尺度的方差特征被允许在树的第一层中选择. 这有助于在早期填充中区分平滑区域。

### 3. 结果与比较

本节介绍结果FFs应用于以下任务:(i)图像去噪, (ii)深度图像细化(iii) 一维动力系统滤波. 对于第一个问题 我们使用流行BSDS500标准[2]进行详尽的评估, 通过比较我们的方法与当前先进的算法. 深度去噪的任务已经从[25]利用7场景数据集进行评估. 对于第三个实验中, 我们设计了一个真实的嘈杂的动力系统, 并且用从卡尔曼标准和最小均方 (LMS) 滤波器与滤波森林预测进行比较. 这些实验证明在不同的应用领域, 该框架的灵活性和有效性。

#### 3.1. 图像去噪

对图像进行去噪实验中, 我们使用相同的协议就像[14]; 图像以0.5的因子重新缩放, 验证集是用于参数调优, 在测试集进行最终模型的评估A 为了形成训练图像, 高斯噪声是零均值并且标准偏差  $\sigma \in \{20, 30, 40, 50\}$  被添加添加到图像中的每个独立像素A 我们从训练集和验证集中取300幅图像训练我们的方法, 试验集是由200图像组成的. 结果如表1所示. 我们提出的基于FF的方法获得了有竞争力的峰值信噪比得分, 每幅图像平均运行时间代价仅为0.025s. 为了进一步提高精度, 我们应用协同维纳滤波器[9]正如在后处理步骤 (维纳滤波如表所示) 中提出的. 这以一个小的代价而大大提高了精确度, 使它比除了RTF<sub>ALL</sub>[14]的其他所有的方法都要好, RTF使用具有四种最先进方法的降噪图像来计算其输入. 对于输入噪声  $\sigma=20$  我们匹配该算法的准确性, 算法的运行时间0:125秒相比RTF<sub>ALL</sub>所需的1275秒. 因此我们得出结论, 对于更多的现实的噪声级, FF表现的与当前最先进的算法一样好, 而且还要快上4个数量级. 在我们的实验中, 应用相同的后处理方法在精确性没有产生有效的增长. 相关比较如图 4.

#### 3.2. 深度图像细化

我们进行试验时, 在7个场景数据集中[25], 每个场景的第一个序列总共有6500帧. 合成的“标定好的真实数据”深度计算使用 KinectFusion [13] 场景重建. 通过添加与深度有关的训练图像生成每个像素的泊松噪声. 我们训练FF (四棵深度为12具有多尺度块的树) 从象棋场景中抽取1000帧, 并在其他场景中测试模型. 定量研究结果如表 2, 表明本文提出的方法优于对比方法。

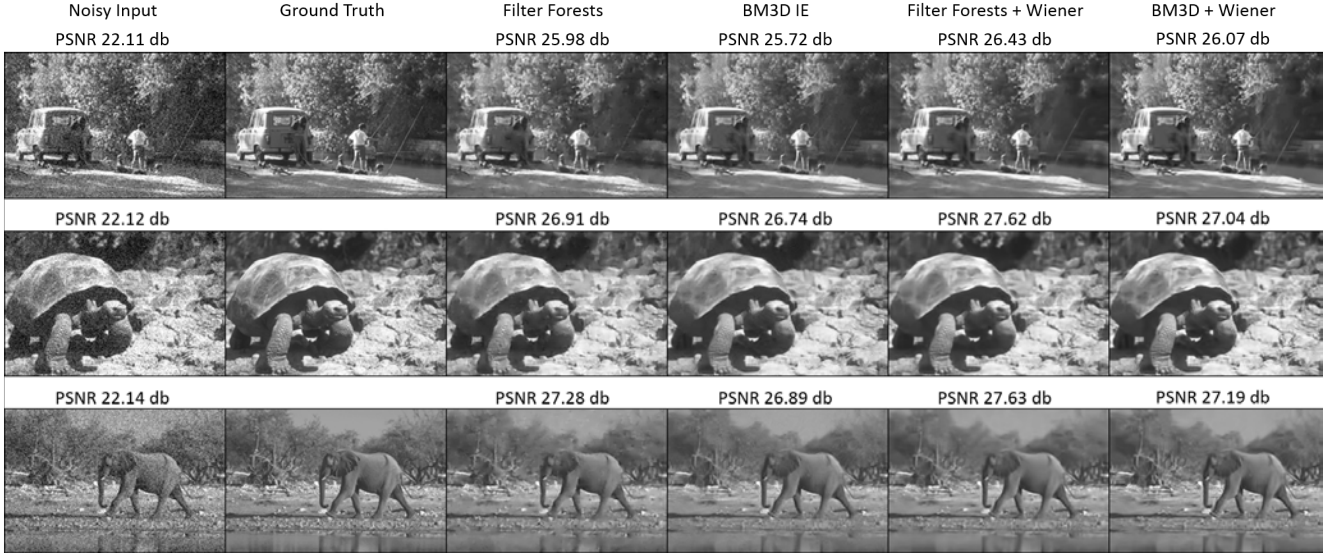


图4. 自然图像去噪定性结果. 从左到右: 输入噪声, 标定真实数据, 滤波森林输出, 三维块匹配第一阶段, 维纳滤波的滤波森林, 维纳滤波的三维块匹配. Best viewed digitally at high zoom.

Method	$\sigma = 20$	$\sigma = 30$	$\sigma = 40$	$\sigma = 50$	sec.
Wiener	25.73	24.79	23.95	23.19	<b>0.001</b>
FF	28.75	26.55	25.32	24.51	<b>0.025</b>
FFWiener	<b>29.65</b>	<b>27.48</b>	<b>26.15</b>	<b>25.25</b>	<b>0.125</b>
RTFPLAIN [14]	28.95	26.97	25.71	24.76	0.7
BM3D [9]	29.25	27.32	25.98	25.09	0.9
EPLL [31]	29.38	27.44	26.17	25.22	38
CSR [10]	29.17	27.24	25.91	24.99	124
LSSC [18]	29.40	27.39	26.08	25.09	172
RTFALL [14]	<b>29.67</b>	<b>27.72</b>	<b>26.43</b>	<b>25.51</b>	1275

表 1. 在BSD500基准, 去噪结果 (峰值信噪比) 和比较. 滤波森林与先进方法的去噪算法比较, 最后一列展示一个图像(241 × 161)的平均运行时间(单位秒). 所有试验均在4核因特尔至强机 (2.4GHz).

这也证明了滤波森林可以处理除了加性噪声的噪声类型. 在这个特定的应用程序中, 协同维纳滤波(FFW)没有显著提高准确性. 定性分析的例子如图. 5.

FF	FFw	Wiener	Bilateral	LMS	BM3D [9]
35.61	35.63	32.29	30.95	24.37	35.46

表 2. 测试为深度细化实验设置峰值信噪比 (db). 我们用我们的方法与其他流行的滤波器相比较: 维纳, 双边, 最小均方和三维匹配块 [9]

### 3.3. 一维动态信号滤波

这里, 我们展示应用滤波森林到非图像的形态. 考虑一个离散动力系统的形式:

$$\begin{aligned} \mathbf{x}_k &= \mathbf{A}_k \mathbf{x}_{k-1} + \mathbf{B}_k \mathbf{u}_k + \boldsymbol{\omega}_{k-1} \\ \mathbf{z}_k &= \mathbf{C}_k \mathbf{x}_k + \boldsymbol{\nu}_k \end{aligned} \quad (11)$$

其中:  $\mathbf{x}_k \in \mathbb{R}^n$  是在时刻  $k$  的内部形态, 其是从它以前的状态即时刻  $k-1$  进化得到的;  $\mathbf{A}_k \in \mathbb{R}^{n \times n}$  是转移矩阵

;  $\mathbf{B}_k \in \mathbb{R}^{n \times n}$  是一个说明控制向量  $\mathbf{u}_k$  的贡献的输入矩阵;  $\mathbf{z}_k \in \mathbb{R}^m$  是当前状态  $\mathbf{x}_k$  的观测,  $\mathbf{C}_k \in \mathbb{R}^{m \times n}$  是一个映射到观察状态的转移矩阵.

$\boldsymbol{\omega}_{k-1} \in \mathbb{R}^n$  和  $\boldsymbol{\nu}_k \in \mathbb{R}^m$  是零均值过程噪声, 它们的协方差矩阵分别是  $\mathbf{Q}_k \in \mathbb{R}^{n \times n}$  和  $\mathbf{R}_k \in \mathbb{R}^{m \times m}$ . 在这些背景下, 我们认为  $t$  时间序列的观测值为  $\mathbf{z}_k = \mathbf{z}_{k-t}, \dots, \mathbf{z}_k$ ,

其目的是学习滤波  $\mathbf{w}$  比如  $\mathbf{x}_k = \mathbf{z}_k^T \mathbf{w}$ . 因此, 用于图像去噪方法的模型可以应用于这种不加任何修改的情况下.

作为方程 11 实际的例子, 我们提到的直流电机的模型, 其由下面的状态空间表示法来描述

$$\begin{aligned} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}_k &= \begin{pmatrix} 0 & 1 \\ 0 & -a \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}_{k-1} + \begin{pmatrix} 0 \\ b \end{pmatrix} V_k + \mathbf{w}_{k-1} \\ z_k &= (1 \ 0) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}_k + \nu_k \end{aligned} \quad (12)$$

其中参数  $a = \frac{1}{\tau}$  和  $b = \frac{K}{\tau}$  可以从电机的数据表获得

(详情见 [15]). 这里  $x_1$  和  $x_2$  分别是电机的位置 [弧度] 和速率 [弧度/秒],  $V_k$  是电机的电压 [伏特]. 从负载编码器上给定一系列观测值  $z_1, \dots, z_k$ , 我们想要预测系统  $x_k$  的当前状态, 即电机编码器的去噪位置 [弧度]. 第一个实验, 信号被高斯噪声损坏 (零均值, 方差  $\sigma = 0.01$ ), 并且所提出的模型与卡尔曼滤波器相比. 在此设置中如果噪声是白噪声且协方差是已知的, 那么卡尔曼滤波器是最优估计量 [29] (i.e. 它使估计输出的均方误差最小)

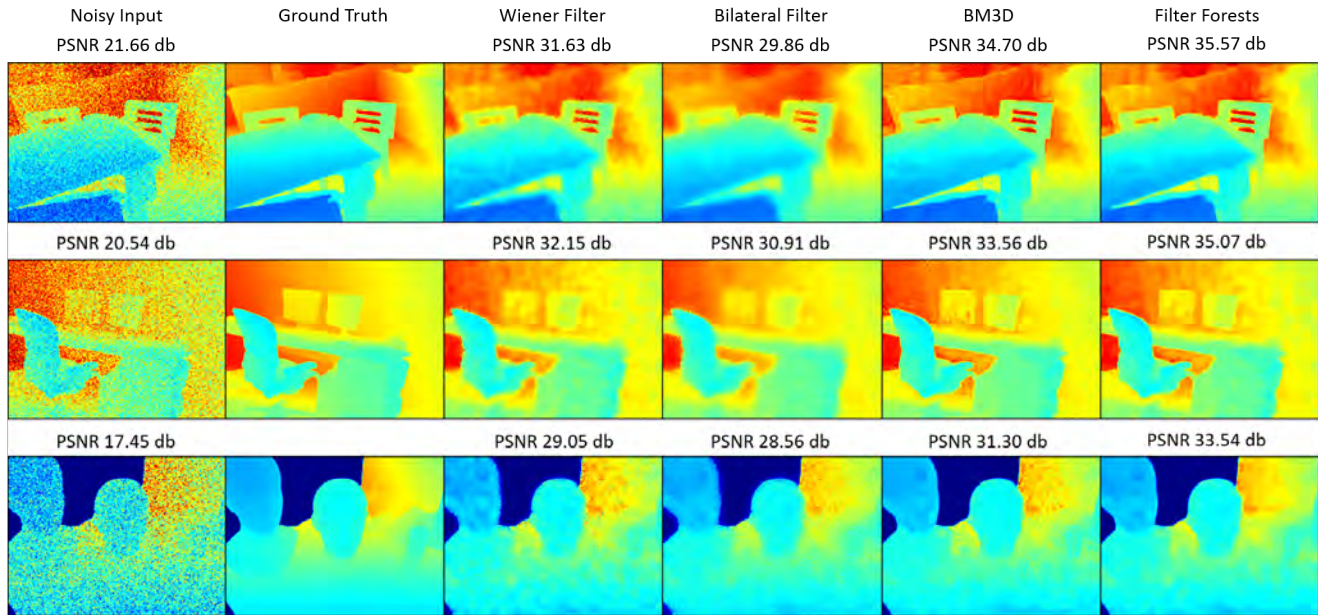


图 5. 定性的深度细化的结果. 从左到右: 噪声输入, 三维块匹配, 维纳滤波, 双边滤波, 最小均方滤波, 滤波森林. 滤波森林在视觉上产生更好的结果。

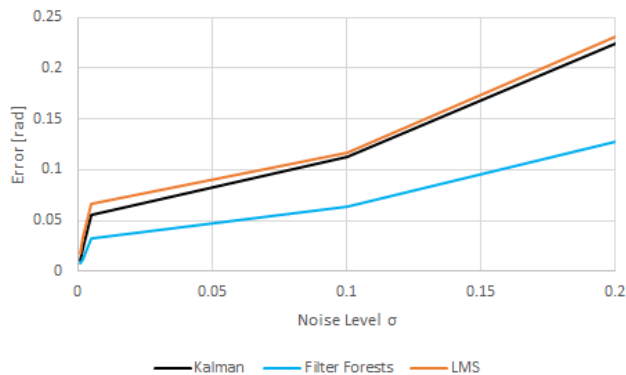


图6. 动力系统存在的结构性噪声滤波。我们误差显示以噪声水平 $\sigma$ 为变量.森林滤波系统优于最小方差滤波和卡尔曼滤波。

我们在训练阶段生成2000个样例的100个序列，并且在100个新的噪声序列上测试。一个具有0.0097弧度均方误差的噪声信号,卡尔曼滤波器具有了0.0018弧度的误差，在没有任何系统或噪声模型假设的情况下，滤波森林能收敛到最优滤波器具有0.00184弧度的错误。在图7, 左图, 我们展示了一个去噪后的序列. 值得注意的是, 最小均方滤波器w达到0.0033弧度的误差。

在第二个实验中, 我们生成形式  $w_k = \nu x_k$  的输入相关噪声 其中  $\nu$  具有零均值且方差  $\sigma \in \{0.01, 0.02, 0.05, 0.1, 0.2\}$  的随机变量。当噪声是结构化的, 卡尔曼滤波器不能保证是最优的。

在这里,森林滤波系统在每一个噪声级别要优于卡尔曼和最小均方滤波器。定量的结果如图6所示, 定性分析的结果如图7(右)。

#### 4. 结论

我们提出的滤波森林对于信号滤波和重建是一种高效的, 新颖的方法。滤波森林是一个非线性的, 基于卷积滤波数据自适应扩展的, 但是可以进一步扩展到其他类型的过滤器或非线性函数。一个新的依照数据处理的, 正则化的训练目标允许我们适应决策森林复原连续变量的任务, 同时利用其时空背景。

无数次的实验和定量比较结果表明滤波森林实现的精确度比当前最优的算法更优或者相等, 并且快上几个数量级。我们相信滤波森林很容易的扩展到其他应用,如边缘检测、锐化、修复、超高分辨率等等。

#### 参考

- [1] Y. Amit and D. Geman. Shape quantization and recognition with randomized trees. *Neural Computation*, 9(7), 1997. 1, 3
- [2] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. *IEEE Trans. PAMI*, 2011. 5
- [3] A. Blake, P. Kohli, and C. Rother. *Markov Random Fields for Vision and Image Processing*. The MIT Press, 2011. 1



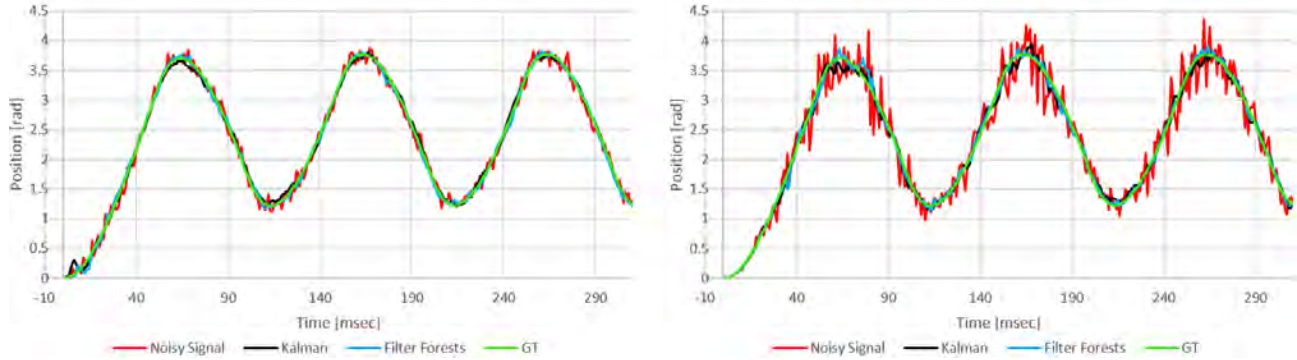


Figure 7. 动力系统滤波例子。卡尔曼滤波(黑线表示)和森林滤波(蓝绿色线表示)被用于从噪声观察结果(红色线)估计一个直流电机(绿色线)的编码器位置。左图: 添加高斯白噪声。在此背景下, 卡尔曼滤波器是最优估计量, 森林滤波学习与其很接近。右图: 输入相关的乘性噪声。在这种情况下, 我们更灵活的方法优于卡尔曼滤波。

- [4] L. Breiman. Random forests. *Machine Learning*, 45(1), 2001. 1, 3
- [5] A. Buades, B. Coll, and J. M. Morel. A non-local algorithm for image denoising. In *Proc. CVPR*, 2005. 2
- [6] A. Criminisi, D. Robertson, E. Konukoglu, J. Shotton, S. Pathak, S. White, and K. Siddiqui. Regression forests for efficient anatomy detection and localization in computed tomography scans. *Medical image analysis*, 2013. 1
- [7] A. Criminisi, T. Sharp, C. Rother, and P. Perez. Geodesic image and video editing. *Proc. ACM SIGGRAPH*, 2011. 2
- [8] A. Criminisi and J. Shotton. *Decision Forests for Computer Vision and Medical Image Analysis*. Springer, 2013. 1, 3
- [9] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Trans. Image Processing*, 2007. 2, 5, 6
- [10] W. Dong, X. Li, L. Zhang, and G. Shi. Sparsity-based image denoising via dictionary learning and structural clustering. In *Proc. CVPR*, 2011. 2, 6
- [11] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Processing*, 2006. 2
- [12] S. Haykin and B. Widrow. *Least-Mean-Square Adaptive Filters*. Wiley, 2003. 2
- [13] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, and A. Fitzgibbon. KinectFusion: Real-time 3D reconstruction and interaction using a moving depth camera. In *ACM UIST*, 2011. 5
- [14] J. Jancsary, S. Nowozin, and C. Rother. Loss-specific training of non-parametric image restoration models: a new state of the art. In *Proc. ECCV*, 2012. 2, 3, 5, 6
- [15] T. Kara and I. Eker. Nonlinear modeling and identification of a {DC} motor for bidirectional operation with real time experiments. *Energy Conversion and Management*, 2004. 6
- [16] P. Kotschieder, P. Kohli, J. Shotton, and A. Criminisi. GeoF: Geodesic forests for learning coupled predictors. In *Proc. CVPR*, 2013. 1, 2
- [17] J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proc. ICML*, 2001. 1
- [18] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Non-local sparse models for image restoration. In *Proc. ICCV*, 2009. 2, 6
- [19] S. Nowozin, C. Rother, S. Bagon, T. Sharp, B. Yao, and P. Kohli. Decision tree fields. In *ICCV*, pages 1668–1675, 2011. 2
- [20] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz. Fast cost-volume filtering for visual correspondence and beyond. In *CVPR*, pages 3017–3024, 2011. 2
- [21] S. Ross, D. Munoz, M. Hebert, and J. A. Bagnell. Learning message-passing inference machines for structured prediction. In *CVPR*, pages 2737–2744, 2011. 2
- [22] R. Shapovalov, D. Vetrov, and P. Kohli. Spatial inference machines. In *CVPR*, pages 2985–2992, 2013. 2
- [23] T. Sharp. Implementing decision trees and forests on a GPU. In *Proc. ECCV*. Springer, 2008. 1
- [24] J. Shotton, R. Girshick, A. Fitzgibbon, T. Sharp, M. Cook, M. Finocchio, R. Moore, P. Kohli, A. Criminisi, A. Kipman, and A. Blake. Efficient human pose estimation from single depth images. *IEEE Trans. PAMI*, 2013. 1
- [25] J. Shotton, B. Glocker, C. Zach, S. Izadi, A. Criminisi, and A. Fitzgibbon. Scene coordinate regression forests for camera relocalization in rgb-d images. In *Proc. CVPR*, 2013. 5
- [26] M. W. Tao, J. Bai, P. Kohli, and S. Paris. Simpleflow: A non-iterative, sublinear optical flow algorithm. *Comput. Graph. Forum*, 31(2):345–353, 2012. 2
- [27] C. Tomasi. Bilateral filtering for gray and color images. In *Proc. ICCV*, 1998. 1
- [28] J. Vitter. Random sampling with a reservoir. *ACM Trans. Math. Softw.*, 1985. 5
- [29] G. Welch and G. Bishop. An introduction to the kalman filter. Technical report, University of North Carolina at Chapel Hill, 1995. 6
- [30] N. Wiener. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. The MIT Press, 1964. 1
- [31] D. Zoran and Y. Weiss. From learning models of natural image patches to whole image restoration. In *Proc. ICCV*, 2011. 6