机器学习大作业—《Deep Bilateral Learning for Real-Time Image Enhancement》阅读报告

161910126 赵安

一、知识补充

1、双边格网(bilateral grid)

双边网格结合了图像二维的空间域信息以及一维的灰度信息，可认为其是一个三维的数组。

2、双边滤波（bilateral filtering)

双边滤波（Bilateral filter）是一种非线性的滤波方法，是结合图像的空间邻近度和像素值相似度的一种折中处理，同时考虑空域信息和灰度相似性，达到保边去噪的目的。具有简单、非迭代、局部的特点 。双边滤波器的好处是可以做边缘保存，一般过去用的维纳滤波或者高斯滤波去降噪，都会较明显地模糊边缘，对于高频细节的保护效果并不明显。

3、仿射变换（affine transformation)

是指二维坐标到二维坐标的线性变换。保持图像的“平直性”和“平行性”。但是角度可能会发生变换。任意一个仿射变换都可以表示为乘以一个矩阵（线性变换）再加上一个平移向量（平移）的形式

4、图像增强（image enhancement)

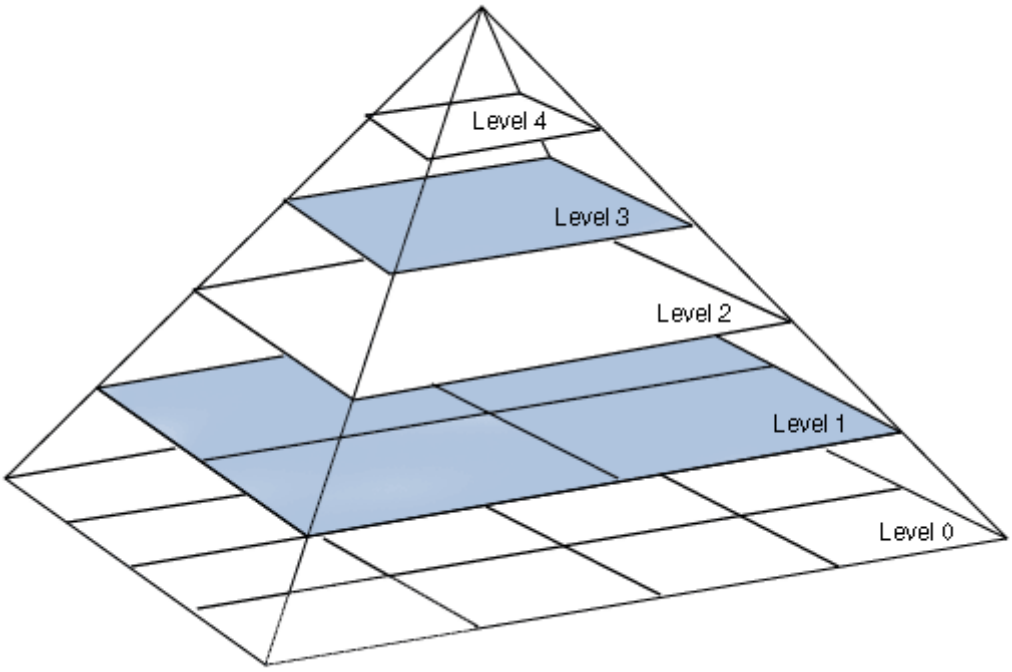
增强图像中的有用信息，它可以是一个失真的过程，其目的是要改善图像的视觉效果，针对给定图像的应用场合。

有目的地强调图像的整体或局部特性，将原来不清晰的图像变得清晰或强调某些感兴趣的特征，扩大图像中不同物体特征之间的差别，抑制不感兴趣的特征，使之改善图像质量、丰富信息量，加强图像判读和识别效果，满足某些特殊分析的需要。

5、高斯金字塔（Gaussian Pyramid)

图像金字塔是图像中多尺度表达的一种，最主要用于图像的分割，是一种以多分辨率来解释图像的有效但概念简单的结构。图像金字塔最初用于机器视觉和图像压缩，一幅图像的金字塔是一系列以金字塔形状排列的分辨率逐步降低，且来源于同一张原始图的图像集合。其通过梯次向下采样获得，直到达到某个终止条件才停止采样。金字塔的底部是待处理图像的高分辨率表示，而顶部是低分辨率的近似。我们将一层一层的图像比喻成金字塔，层级越高，则图像越小，分辨率越低。

高斯金字塔是通过高斯平滑和亚采样获得一些列下采样图像，也就是说第K层高斯金字塔通过平滑、亚采样就可以获得K+1层高斯图像，高斯金字塔包含了一系列低通滤波器，其截至频率从上一层到下一层是以因子2逐渐增加，所以高斯金字塔可以跨越很大的频率范围。金字塔的图像如下：


Level 4
Level 3
Level 2
Level 1
Level 0


input image
gaussian pyramid level0
gaussian pyramid level1
gaussian pyramid level2

6、图像信噪比（Image signal-to-noise ratio）

信噪比是信号均值与背景标准偏差的比值:

$$SNR = \frac{\mu_{sig}}{\sigma_{bg}}$$

对于图像，这里的" **信号值** "往往是灰度值。分母有时采用背景信号值的方差，代表的物理意义是噪声功率。对于高对比度黑背景图，上式直接计算的结果通常是无穷。所以我们改用信号均值与信号标准偏差来衡量：

$$SNR = \frac{\mu_{sig}}{\sigma_{sig}}$$

图像的信噪比应该等于信号与噪声的功率谱之比，但通常功率谱难以计算，有一种方法可以近似估计图像信噪比，即信号与噪声方差之比。首先计算图像所有象素的局部方差，将局部方差的最大值认为是信号方差，最小值是噪声方差，求出它们的比值，再转成dB数。

信噪比大，图像画面就干净，看不到什么噪波干扰（表现为"颗粒"和"雪花"），看起来很舒服；若信噪比小，则在画面上，可能满是雪花，严重影响图像画面。信噪比与图像质量之间具有如下对应关系：
（1）S/N为60dB(比率为1000：1)时，图像质量优良，不出现噪声；
（2）S/N为50dB(比率为316：1)时，图像有少量噪声，但图像质量算好；
（3）S/N为40dB(比率为100：1)时，图像有一定的精细颗粒或雪花，图像的精细结构受到一定的损失；
（4）S/N为30dB(比率为32：1)时，图像将是有大量噪声的劣质图像；
（5）S/N为20dB(比率为10：1)时，图像就不能使用。

二、问题

摄像器材的发展使得所拍摄的图像分辨率逐渐增高，给图像处理的相关算法带来巨大压力，算法增强所面对的图像的文件大小也在逐渐增大，这就会带来原本的程序逐渐开销过大，尤其是在移动端计算速度明显下降，图像增强效果同样较差。所以在移动端处理高分辨率的图像时的计算速度和处理效果是很重要的。

三、方法概述

图像增强不仅与图像局部特征有关，和图像的整体特征也有关，因此这里低分辨率的部分分成了局部特征和全局特征两个路径，最后融合成一个特征用来代表仿射变换。由于图像的变换在双边空间中是可以被近似成线性问题的，所以这篇论文使用了成对的输入输出图片，训练了一个卷积神经网络来预测bilateral space中的模型的参数,，运算过程中使用低分辨率图像以降低计算代价。具体步骤为：首先使用低分辨率图片生成仿射变换, 然后将这个仿射变换进行升采样使其能应用在正常分辨率图像上，最后利用升采样过的仿射变换来优化原本的图像。

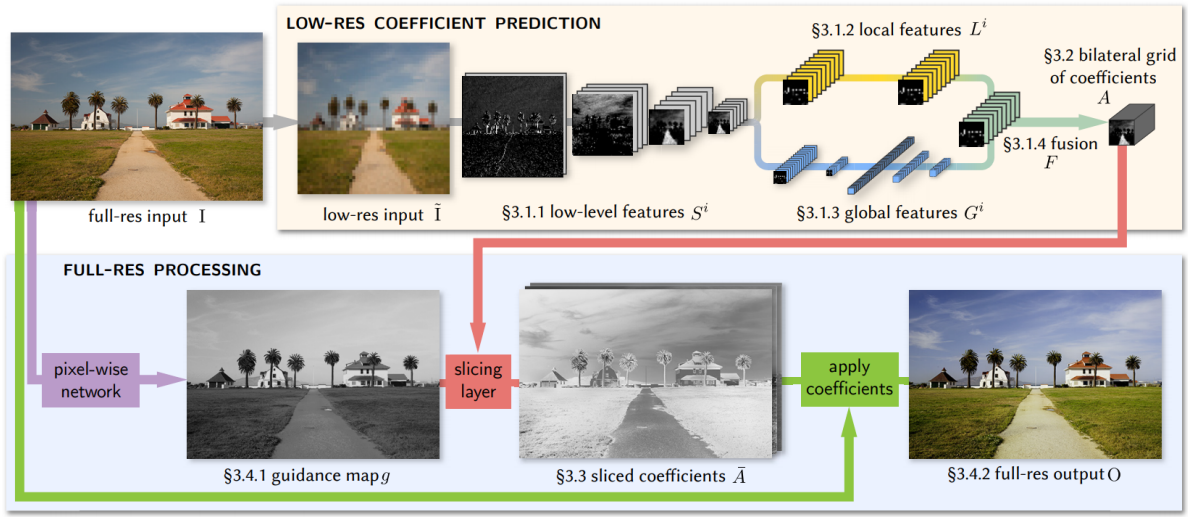
四、算法特点

1、该算法的学习过程主要是学习一幅图像到另一幅图像之间的变换方式，而不是学习一幅图像，变换方式比单纯的输出一幅图像更容易学习。

2、主要在低分辨率的图像的双边网格上运行预测算法

3、虽然大部分学习过程是在低分辨率下运行的,但是损失函数是在全分辨率下评估的，使得低分辨率转换直接优化其对高分辨率图像的影响。

五、神经网络的架构介绍



1 、低分辨率特征

对图像进行下采样，公式如下：

$$S_c^i[x,y] = \sigma(b_c^i + \sum_{x',y',c'} w_{cc'}^i[x',y']S_{c'}^{i-1}[sx+x', sy+y'])$$

上式中，i=1,.....n_S 为每个卷积层的索引，c，c' 为卷积层的通道的索引，w' 为卷积核权重矩阵，b' 为 bias，激活函数σ采用 ReLU ，卷积时采用zero-padding。

2、局部特征

低层特征 $\mathbf{S^i}$ 输入一个 $\mathbf{n_L}$ = 2层卷积层得到局部特征 $\mathbf{L^i}$ ，即对上部分特征做进一步处理，通过n_L=2的卷积层进一步提取特征，这里设定步长stride=1,使得这部分分辨率和通道数不再改变。加上前面的下采样那里用到的卷积的话，总共是 $\mathbf{n_L}$+$\mathbf{n_S}$ 层卷积，如果要获得一个更高空间的分辨率,可以通过减小 $\mathbf{n_S}$ ，增大 $\mathbf{n_L}$ 实现。如果没有局部特征，预测出来的系数会失去空间的位置信息。

3、全局特征

全局特征层有2个卷积层，stride=2，之后接3个全连接层组成，层数为 $\mathbf{n_G}$ = 5。全局特征效果:

(a) **ours**, with global features　　(b) without global features

4、融合与线性预测

使用一个逐点的仿射变换和一个ReLU激活函数来融合全局和局部特征，公式如下：

$$F_e[x,y] = \sigma(b_c + \sum_{c'} w'_{cc'} G^{n_G}_{c'} + \sum_{c'} w_{cc'} L^{n_L}_{c'}[x,y])$$

这样得到了一个16×16×64的特征矩阵,将其输入1×1的卷积层得到大小为16×16，输出通道是96的特征：

$$A_c[x,y] = b_c + \sum_{e'} F_{e'}[x,y] w_{cc'}$$

5、将图像特征作为双边网格

将上述得到的最终的特征图A作为双边网格的第三维

$$A_{dc+z}[x,y] \longleftrightarrow A_c[x,y,z]$$

其中d = 8是网格的深度。A可以看作是一个16 × 16 × 8双边网格，每个网格单元包含12个数字，每个数字对应一个3 × 4仿射颜色变换矩阵的系数。这种变换让我们将公式

$$S^i_c[x,y] = \sigma(b^i_c + \sum_{x',y',c'} w^i_{cc'}[x',y'] S^{i-1}_{c'}[sx+x', sy+y'])$$

中的卷积步长解释为双边域，其中它们对应于(x,y)维度上的卷积，并表示z和c维度上的完全连通性。在网格中简单的应用3维卷积会导致局部连接，因此，这种操作更加有表现力。

6、使用可训练的slicing layer进行上采样

以上描述了如何从低分辨率图像中学习预测双边网格的系数A

现在需要将这些信息传输回原始输入的高分辨率空间，以产生最终的输出图像。为此，我们引入了基于双边网格的slicing layer，该层以单通道引导图 g 和特征图 A 作为输入，对于A和 g 该层是次可微的，这使得我们在训练时可以进行反向传播。

利用引导图 g 对 A 进行上采样，是利用A的系数进行三次线性插值，位置由g决定：

$$\bar{A}_c[x,y] = \sum_{i,j,k} \tau(s_x x - i)\tau(s_y y - i)\tau(d \cdot g[x,y] - k) A_c[i,j,k]$$

其中

$$\tau(\cdot) = max(1 - |\cdot|, 0)$$

表示线性插值。

S_x 和 **S_y** 分别表示网格和全分辨原图的高度和宽度比例，特别的，每个像素都被分配了一个仿射变换的系数系数，其在网格里对应的深度由图像灰度值g(x,y)决定，也就是**A_c[i, j, g[x,y]]**。这里的slicing使用 **OpenGL** 库完成，通过这个操作使得输出图的边缘遵循输入图的边缘，达到保边的效果，这个效应相对于反卷积来说更加明显，如下图所示。通过这个操作，可以将全分辨下复杂的操作转换成许多简单的局部操作（也就是在每个网格对应图像的操作）。

7、实现全分辨率的最终输出

后面的操作都是在全分辨率下进行，对于输入图像 **I**,提取其特征 **ϕ** 实现两个作用：1、它们可以用来获得引导图 **g**；2、可以用来给上述得到的全分辨局部仿射模型做回归。

（1）获得引导图的辅助网络

对原始图像三个通道操作后相加得到引导图：

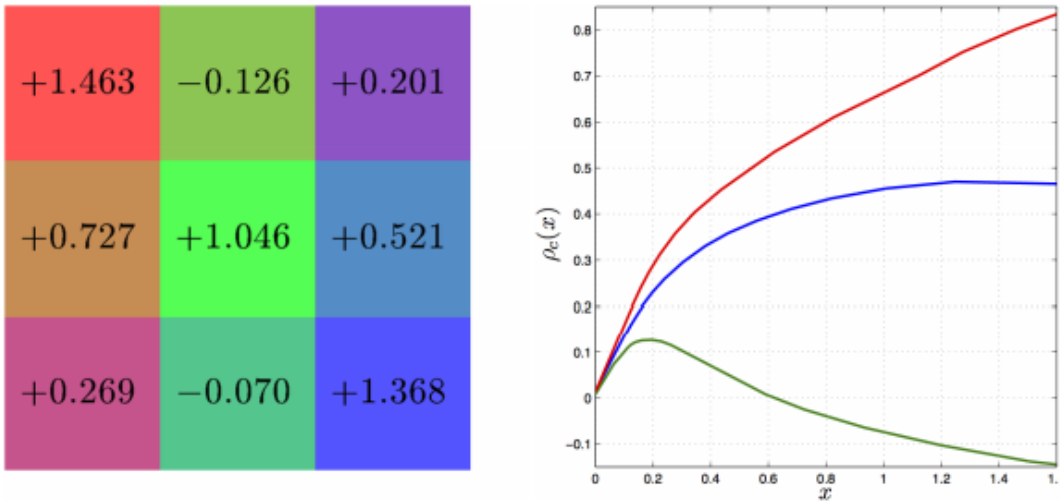
$$g[x,y] = b + \sum_{c=0}^{2} \rho_c(M_c^T \cdot \phi_c[x,y] + b_c')$$

$$\rho_c(x) = \sum_{i=0}^{15} a_c, imax(x - t_c, i, 0)$$

M，a，t，b，b'是需要学习的参数

（2）将最后的输出结果进行组装

$$O_c[x,y] = \bar{A}_{n\phi+(n\phi+1)c} + \sum_{c'=0}^{n_\phi-1} \bar{A}_{c'+(n_\phi+1)_c}[x,y]\phi_{c'}[x,y]$$



上图左边为颜色转换矩阵，右边为使用学习得到的引导图的效果

8、训练过程

对这个神经网络的训练是在全分辨率下进行的，通过最小化L_2损失来优化权重和偏置项，损失函数如下：

$$\mathcal{L} = \frac{1}{|D|}\sum_i ||I_i - O_i||^2$$

六、个人思考

可以通过对输入图做特征进一步的提取特征来增强其表达效果，即增加神经网络的层数，增加仿射变换的系数，网格特征的数量等，但是速度会变慢。另外，据查阅相关资料，将该网络用在图像增强外的其他任务上，如色彩化、去雾、深度估计等效果较差，这是因为其有较强的假设即输出是由输入的局部仿射变换得到的。

七、实验验证

该文在多个模型上进行评估，使用了HDR+ and the face brightening dataset，the MIT-Adobe “FiveK” dataset等多种数据集进行模型的评估和训练。实验阶段保留了500张图像用于验证和测试，并对其余的4500进行训练，同时使用随机裁剪，翻转和旋转来增加数据。

为了得到较好的效果，该文通过在各种型号及不同性能的移动端处理上进行实验，对网格深度，通道数，空间尺度，扩张卷积层的数量等多个参数进行反复修改调整。

该文作者还使用该架构来学习除图像增强之外的任务，如消光，着色，去雾和单眼深度预测，但这些实验的成功有限。