

# 论文阅读笔记

我阅读的论文是 Dynamic Attentive Graph Learning for Image Restoration

## 图像恢复的相关工作与本文的贡献

在介绍中作者指出图像恢复的主要方式有3种，A)基于深度学习方法利用数据带来的外部信息；B)利用图像本身具有的先验信息；C)数据+利用图像先验知识的模型设计。本文就是从第三种方式出发。

问题的关键在于如何利用图像的先验信息设计对应的网络结构（这实际上是将网络结构进行人为规则化白盒化的设计）**本文重点研究如何充分利用图像的非局部自相关性（Non-local self-similarities）**

受到Non-Local Network的启发，可以**引入图结构化关系来表示局部与远距离节点的关系**，再细节一点，对于节点的定义有pixel level 和 patches level两种选择，论文中对认为图像退化直接作用于像素，因此如果直接选择pixel level会带来数据的偏差，所以选择patch-level。



我认为更具体的解释有两个方面：

- 1.从图像恢复任务来看，图像退化直接作用于像素，因此使用pixel-level很容易拟合具有相同退化特征的数据集，也就意味着这种方式的可迁移性会很差（当然这需要实验证明，在论文中提供了消融实验证明patch-level的高效性）
- 2.从运行效率看，patch-level的运行效率更高，计算量更少，合适的patch设计可以高效的处理多种尺寸的输入图像。（例如swin transformer的patch设计）

利用图像的Non-local similarities + patch-level的工作已经存在了，而本文重点聚焦到**节点聚合计算的优化设计**，之前的这些方法主要有两种方式计算节点聚合：A)完全图的计算方式，类似于plain self-attention，即每个节点都参与单个节点的聚合计算；B)固定的邻居个数，即单个节点只聚合计算其固定数目的邻居。

论文中指出对于完全图的方式，不仅效率低，而且没用充分利用不同patch间明显的相似性（与其期望网络自己去黑盒化学到这点，不如白盒化的设计对应的结构）；

另外，对于第二种方式，图像中不同内容的重复性是不同的，这意味应该区别对待它们，也就是说固定的邻居数可能过于僵化，不能充分高效的利用这个图像先验知识，这也是本文提出动态邻居的关键原因。

要实现动态邻居选择，需要确定每个候选节点的得分以及一个动态阈值。论文借鉴图注意力网络中节点权重计算的方式，并把它迁移到处理图像patch上，并设计了实现动态计算阈值的方法。



我的理解：节点聚合的关键在于邻接矩阵的设计，完全图的邻接矩阵过于密集，没有充分利用图像本身的相似性先验知识，固定邻居的方式可以提供一种稀疏的邻接矩阵，但是缺乏灵活性，对于不同类型的相似内容缺乏自适应性，因此本文设计了动态计算邻接矩阵的计算方式，并且在消融实验中分别对比了这几种方法，证明了提出方法的高效性，这个是本文的核心贡献。

接下来，将具体介绍模型的设计细节。

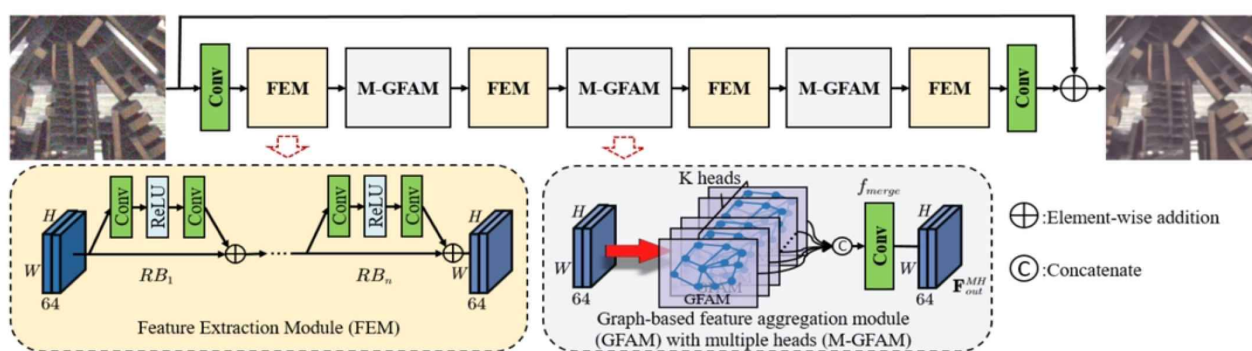
## 模型设计的具体细节

宏观的网络架构是两种模块(FEM&M-GFAM)的堆叠，**同时将输入的特征图连接到最后的输出** (论文中称为**global pathway**)



这种设计的原因是最初的特征图保留更多的低频信息，而网络深层输出的特征图更多的是高频信息，在目标检测，分割等相关网络设计上也有类似的结构（也可以看作是多尺度信息融合）

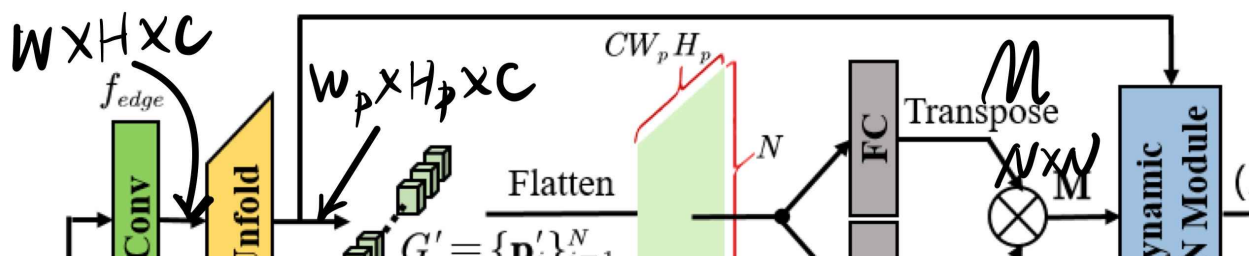
Feature Extraction Module(FEM)是常用的的residual block，本文的主要工作集中在Graph feature attention module（GFAM模块）的设计上。



将输入特征图分别通过两个1x1的卷积映射并通过Unfold操作得到两类向量。




一类用于直接表示图结构中的节点特征向量,另一类准备用于构建边关系特征向量。





后一类向量经过flatten后变成行向量（总共有N个patches）再经过线性映射后通过点积得到每对向量的相似性（pair-wise similarities）矩阵  $M_{N \times N}$ 。

在Dynamic KNN Module中首先取相似度矩阵中每个行向量的均值并经过映射作为阈值，相似度矩阵对应行减去对应阈值经过ReLU激活得到Adjacency matrix。

 相似度矩阵M的第  $i$  行可以看作是所有patches对第  $i$  个patch的相似度，作者希望可以自适应根据不同的patches确定不同的邻居节点，即基于这一行相似度向量自动确定一个阈值，大于阈值的对应节点作为邻居参与加权聚合，低于阈值则去掉（权重为0）。符合直觉的方式是取均值，经过映射后得到阈值，使用ReLU激活是为了结果可微。


对应的计算公式：

$$T_i = \frac{\phi_1(p'_i)}{N} \sum_{k=1}^N M_{i,k} + \phi_2(p'_i)$$

$$A_{i,:} = \text{ReLU}(M_{i,:} - T_i)$$

之后 Adjacency Matrix 经过Softmax作为节点间权重用于节点向量的聚合，之后通过折叠操作（Fold）得到与输入特征图相同大小的输出特征图（Fold中重叠区域取均值），因为输入输出尺寸相同，因此添加了一个全局连接：

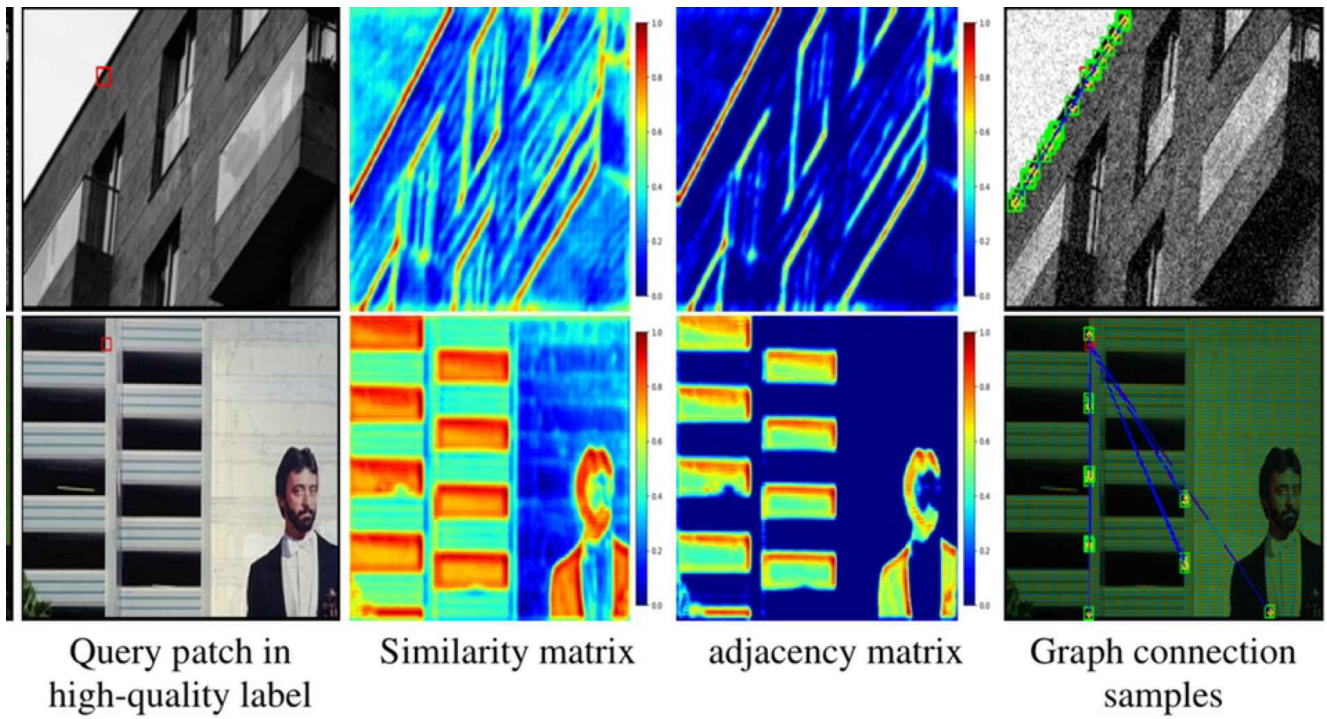
$$F_{out} = Fin + \text{Fold}(\hat{p}_{i=1}^N)$$

 通过添加残差连接，可以使深度叠加的Module可以更容易收敛，从直觉上也增强了不同类型（高/低频）信息的交互补充。之后的多头结构（M-GFAM）也是为了增强模型的学习能力，直觉上也是一种分而治之的思想。

## 分析与讨论

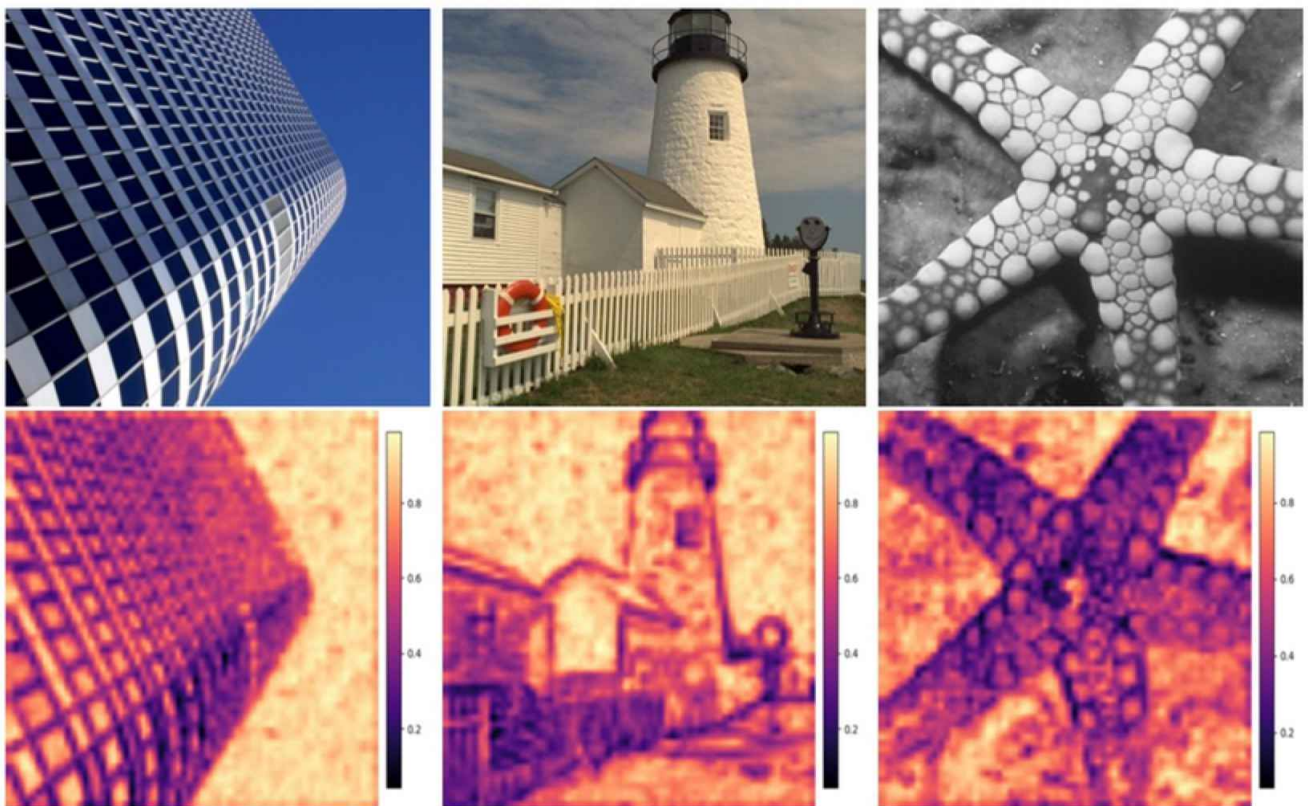
论文中讨论了模型是否有效捕获到长距离相关性（self-similarities是图像重要的先验之一）






图中红色框内的patch对应的绿色patches即为重要性较大的前几位邻居节点，可以看到模型有效的捕获到远距离的自相关性。

同时对比相似度矩阵和邻接矩阵的热力度差异，可以看出所学习到的阈值可以有效的过滤那些不相关的内容。



上图展示了不同patch处的邻居分配数量的归一化热力图，可以看出在图像的边缘高频部分，分配的邻居数较少，低频部分分配的邻居较多。说明模型可以根据不同的图像内容设定不同的邻居节点分

布。

 此小节的两个实验，证明了在intro中所指出本文的两个核心要素，第一是有效学习到self-similarities，第二是可以自适应确定不同的相邻节点（以前的方法都是完全图连接，或者固定邻居数量的聚合）

## 实验

文中首先在四个典型任务上与现有的方法进行对比，都取得了最先进的效果同时参数量也是可以接受的。

消融实验：A)分别研究了pixel-level 与 patch-level下非局部相关性的对比实验，将DAGL中的patch换成Non-local network(pixel level),并也设计了单头与多头（NL&MHNL）。

B)为了证明图注意力加权机制的高效性，将注意力加权直接换成聚合取平均，进行比较（GAT）。

C)为了证明动态邻居选择的高效性，将原来的dynamic KNN Module 换成了固定的邻居个数（THD），进行比较。

Table 5. Evaluation results about variants of DAGL on Urban100 test set ( $\sigma = 25$ ). NL and MHNL represent replacing our M-GFAM with non-local neural networks and multi-heads non-local neural networks, respectively. (w/o THD) and (w/o GAT) refer to removing dynamic KNN module and removing graph attention mechanism, respectively.

Mode	NL	MHNL	w/o THD	w/o GAT	DAGL
PSNR	30.73	30.92	31.28	30.77	31.39

实验结果证明了所提出策略的高效性。最后文章对不同结构超参数设定下的模型性能进行比较，在性能与参数量的权衡下选择最终的模型。

## 可能的新思路

我们发现模型是多个模块的重复堆叠，是否可以使得它们相互借鉴？例如深层的模块利用浅层模块的先验信息，这里不妨从互信息的角度去动态选择邻居节点，空间上每个位置的patch表示位置变量，选择不同位置变量间互信息最小的节点最为邻居节点，深层模块可以利用浅层模块已知的节点互信息作为先验计算后验互信息，从而更好的选择邻居或者有望减少模型参数量。