

数字图像处理报告五：频域知识与 CNN 的结合

姓名：鲁国锐

学号：17020021031

专业：电子信息科学与技术

2020 年 4 月 8 日

目录

1	题目描述	2
2	问题背景	2
3	方法简述	2
3.1	将频域知识用于图像预处理	2
3.2	用 CNN 预测小波变换系数	3
4	总结	3

1 题目描述

试探讨如何将频率域知识用于卷积神经网络（可以从网上查阅相关资料）

2 问题背景

在计算机视觉领域，深度神经网络已经得到相当广泛的应用，并已经在诸多方向上取得了非凡的成果。然而，现存的神经网络主要实行的是空间域的操作，且所处理图像的尺寸大小必须是固定的 [2]；而在现实生活中获得的图像尺寸一般是多样的，同时其尺寸往往也大于神经网络规定的输入尺寸。另外，大尺寸图像还会使得计算成本和片间通信带宽大大增加 [2]。所以，基于可操作性的需要，图像通常都要被调整大小（下采样）以适应神经网络和降低计算成本及通信带宽 [2]。

但下采样又会带来新的问题。当图片尺寸被缩小后，分辨率降低，使得一些重要的信息被丢失 [2]，从而导致最终的结果不尽人意。例如，在人脸识别任务中，卷积神经网络（CNN）对低分辨率图像识别的准确度要明显低于高分辨率图像 [1]。

在这种情况下，将频域中的信息作为输入交给神经网络可以有效减少下采样带来的信息损失 [2]。除此之外，还可以用 CNN 对小波变换的系数进行预测，以此来构建超分辨率图像 [1]。

3 方法简述

3.1 将频域知识用于图像预处理

[2] 中提出了一种学习频域特征的通用方法，能够在减少计算量和通信带宽的同时尽可能多地保留信息，从而提高在分类、识别任务中的准确率。该方法大体分成两个部分：一个是提取频域特征的数据预处理方法，主要用来提升对图像信息的保留程度，增加分类、识别任务的准确度；另一个是对数据进行剪枝的方法，主要用于减少计算量和片间通信带宽。这里主要介绍前者。

该预处理方法大致分为如下几步：

1. 遵循图像预处理和数据增强的步骤，对其进行大小调整，裁剪和旋转；
2. 将图像转换到 $YCbCr$ 色彩空间，再将图像转换到频域——这里用的是二维余弦变换（DCT）；
3. 将相同频率的二维 DCT 系数组合成一个通道，构成一个三维的 DCT 图像块；
4. 选择对结果影响最大的频率通道子集；
5. 将于 $YCrCb$ 色彩空间中选出的通道连接在一起，形成一个张量；
6. 用从训练集中算出的均值和方差，来对每一个频率通道进行归一化。

操作流程也可见图1。

这种方法的好处在于，可以通过增加特征映射的通道数来缩短它的长和宽，同时还能保证总的点数不变，即将一张 $H \times W \times C$ 的图转换成一张 $\frac{H}{M} \times \frac{W}{N} \times (C \times M \times N)$ 的频域特征映射，从而维持着与输入图像相同的数据量。相应的，频域特征映射也可以跳过传统 CNN 的输入层；然后调整下一层的维度，使其能够直接接收频域特征映射。这样可以使对传统 CNN 结构的调整尽可能小。

[2] 中的后续章节，证明了特征映射中的大部分频率通道可以被剔除而不会影响神经网络的准确度，同时也提出了相应的实现方法：基于学习的频率通道选择和静态频率通道选择。时间紧迫，这里不对其进一步阐述。

另外，该论文也证明了卷积神经网络对低频通道更加敏感，这一点与人类视觉系统（HVS）相似。这里同样不多做阐述。

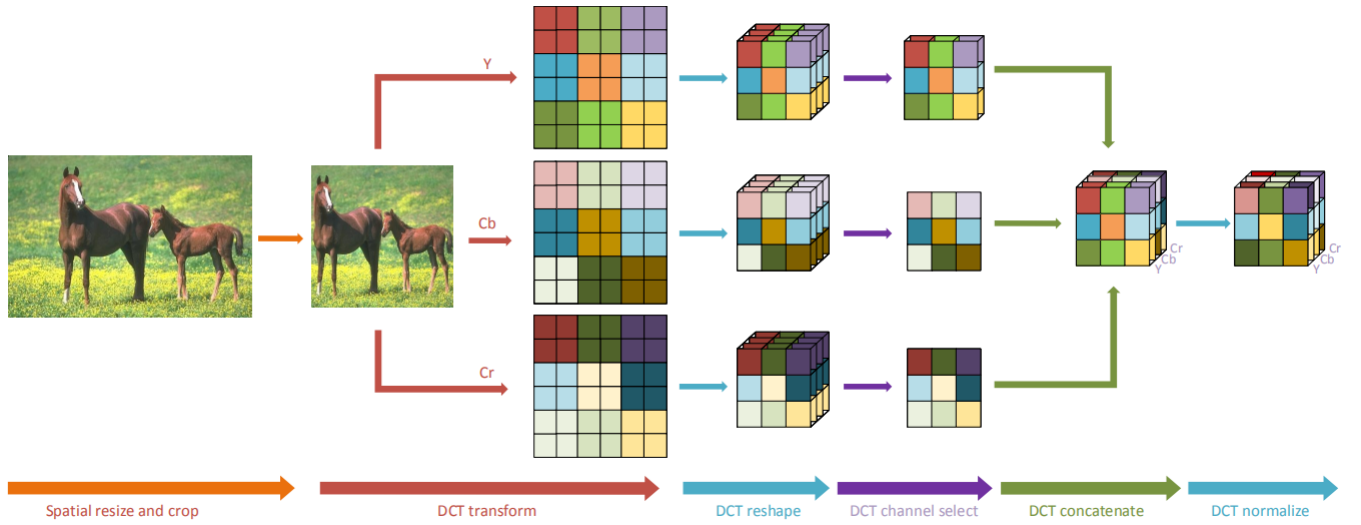


图 1: 预处理流程示意图

3.2 用 CNN 预测小波变换系数

[1] 中提出了另一个思路：用卷积神经网络来预测图像的小波变换系数。这篇文章的作者们假设：只要一幅图像的小波变换系数能被准确地预估出来，就能根据一张低分辨率图像生成一张带有丰富纹理细节和全局拓扑结构信息的高分辨率图像。更进一步地说，强调高频小波系数有助于复现纹理细节；而加强对低频小波系数重建的约束则可以加强全局拓扑信息的一致性。因此，如何生成高分辨率图像这一问题就归结为了如何准确预测小波系数的问题。

为了最大限度地发挥小波变换的优势，作者专门为其针对人脸识别这一任务设计了一个卷积神经网络结构。该网络结构又由三个子网络构成：

1. 嵌入网络 (*embedding net*)：输入一张低分辨率图像，输出一组特征映射；
2. 小波预测网络 (*wavelet prediction net*)：它由一系列平行独立的子网组成，其中每一个子网的目的是用嵌入网络输出的特征映射来学习一个特定的小波系数；
3. 重建网络 (*reconstruction net*)：由给定的小波系数恢复出一张高分辨率的图片。

由于时间紧张，没有时间再去深入学习小波变换，故对该论文的基本思想就只能介绍到这里。

4 总结

在本次报告中介绍了 [1] 和 [2] 两篇论文。由于水平有限，再加上时间紧迫，关于这两篇文章还有很多问题没有解决，如：为何要先将原图转换到 $YCbCr$ 色彩空间；如何把相同频率的小波系数组合成一个通道，同时还能保证整个特征映射在各频率分量上的通道数一致等。

参考文献

- [1] Huaibo Huang, Ran He, Zhenan Sun, and Tieniu Tan. Wavelet-srnet: A wavelet-based cnn for multi-scale face super resolution. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), pages 1689–1697, 2017. 2, 3
- [2] Kai Xu, Minghai Qin, Fei Sun, Yuhao Wang, Yen-Kuang Chen, and Fengbo Ren. Learning in the frequency domain, 2020. 2, 3