

# Hierarchical Regression Network for Spectral Reconstruction from RGB Images

Yuzhi Zhao<sup>1</sup>, Lai-Man Po<sup>1</sup>, Qiong Yan<sup>2</sup>, Wei Liu<sup>2,3</sup>, Tingyu Lin<sup>1</sup>

<sup>1</sup>City University of Hong Kong, Hong Kong SAR, China

<sup>2</sup>SenseTime Research

<sup>3</sup>Harbin Institute of Technology, China

## Abstract

*Capturing visual image with a hyperspectral camera has been successfully applied to many areas due to its narrow-band imaging technology. Hyperspectral reconstruction from RGB images denotes a reverse process of hyperspectral imaging by discovering an inverse response function. Current works mainly map RGB images directly to corresponding spectrum but do not consider context information explicitly. Moreover, the use of encoder-decoder pair in current algorithms leads to loss of information. To address these problems, we propose a 4-level Hierarchical Regression Network (HRNet) with PixelShuffle layer as inter-level interaction. Furthermore, we adopt a residual dense block to remove artifacts of real world RGB images and a residual global block to build attention mechanism for enlarging perceptive field. We evaluate proposed HRNet with other architectures and techniques by participating in NTIRE 2020 Challenge on Spectral Reconstruction from RGB Images. The HRNet is the winning method of track 2 - real world images and ranks 3rd on track 1 - clean images.*

## 1. Introduction

Hyperspectral (HS) imaging technology refers to the spectral signature is densely sampled to many narrow bands. It combines imaging technology with spectral technology to detect the two-dimensional geometric space and one-dimensional spectral information of the target to obtain continuous, narrow-band images with high spectral resolution. Normally, most of the civil cameras capture only three primary colors. However, HS spectrometers can obtain the spectrum of each pixel in the scene and collect the information into a set of images. To visualize HS images, a response function is adopted to transform HS images into RGB format. Conversely, we can acquire HS images from

the visible format by learning the inverse function. In this paper, we propose a general hierarchical regression network (HRNet) for spectral reconstruction from RGB images.

HS imaging technology has many advantages and particular characteristics. There have been many applications based on HS imaging technology, e.g, remote sensing technology [25], pedestrian detection [17, 23], food processing [29], medical imaging [2]. However, in recent years, the development of HS imaging has encountered a bottleneck since it mainly depends on spectrometers. The traditional spectrometers saves images with huge volume and need long operation time, which restricts HS imaging technology applied to portable platforms and high-speed moving scenes [28]. Although researchers have continuously optimized the traditional pipeline [7, 35], these hardware devices are still expensive and of high complexity. Thus, we present a low cost and automate approach only based on RGB cameras. To address the problem, we propose a HRNet that learns the process of RGB images to corresponding HS projections.

In general, spectral reconstruction is an ill-posed problem. Moreover, there is unknown noise in environment leading to degraded RGB images. However, there is dense correspondence between RGB images and HS images, making it possible to exploit the correlation from many RGB-HS pairs. Since the information of RGB image is much less than HS image, there may be many reasonable HS image combinations corresponding to a same RGB image. The algorithm needs to learn a reasonable mapping function that produces high-quality HS images. With the development of deep convolutional neural network (CNN), it is eligible to learn the blind mapping for spectral construction.

The previous methods [32, 21, 33, 6, 36] mainly utilize an auto-encoder structure with residual blocks [14]. The network often performs convolution at low spatial resolution since the features are more compact and the computation is more efficient. However, as the network goes deeper, it fails to remain the original pixel information due to per-

Corresponding author: yzzhao2-c@my.cityu.edu.hk



















