Final Report

Introduction

The task of recovering the spectral image from a low dimensional spectral measurement has been of great interest for a variety of applications for a long time. Since spectral information is valuable to distinguish materials, hyperspectral imaging has been frequently used in scientific research and established in the professional sector. For example, industrial machine vision cameras could be boosted for increased food quality inspection using spectral features.

The technique of multi-spectral imaging is not widely used despite its advantages. Major reasons are simply the cost. The devices for acquiring HSIs are typically more complicated and expensive. In addition, most hyperspectral imagers rely on precise scanning to generate 3D data cube, which hinders them from being portable and cost-effective.

In order to obtain spectral information more effectively, a recent trend is to reconstruct the hyperspectral information from RGB-images. RDB-cameras are widely used and are most certainly capable of producing visually appealing images. They are available to anybody and are a baseline indicator multi-spectral imaging has to outperform.

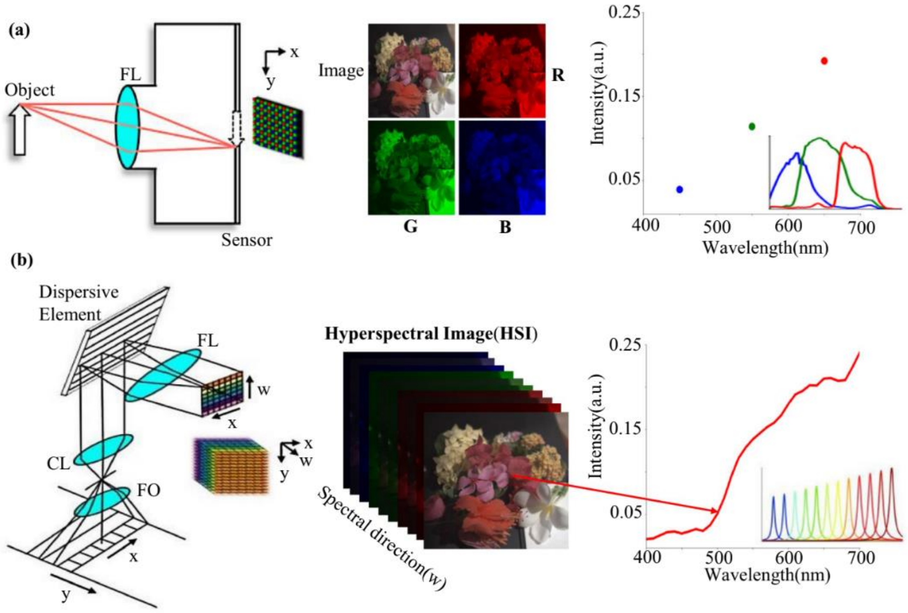

Figure 1. Schematic diagrams of an RGB camera (a) and a typical hyperspectral imager (b).

However, recovering hyperspectral information from RGB images is an ill-posed inverse problem. In order to encourage tackling this problem, two competitions have been launched. Various neural network architecture has been proposed to increase the reconstruction accuracy, such as convolution neural network (CNN) and Generative Adversarial Network (GAN).

Within this work, machine learning is applied to describe the mapping from RGB-images to spectral images. Our main motivation was to see if we could improve the accuracy of the classification problem by using hyperspectral information. Hyperspectral having rich spectral information which might be useful for computing better features and our task was to refute or confirm this hypothesis.

Dataset

The open-source datasets are not widely available as it is hardly captured and collected, especially with large number of samples. The most important attributes of the existing datasets are data amount, spatial resolution, spectral channels, and the diversity of the scenes. The Table 1 presents datasets that are publicly available and also have the same spectrum and channels number.

Dataset	Amount	Resolution	Spectral channels	Spectrum/(nm)	Featured scenes
CAVE[34]	32	512 × 512	31	400–700	Skin, hair, food and drink
ICVL[29]	203	1392 × 1300	31	400–700	Urban, rural, indoor and plant
BGU-HS[27]	286	1392 × 1300	31	400–700	Urban, rural, indoor and plant
ARAD-HS[28]	510	512 × 482	31	400–700	Statue, vehicle and paint

Table 1. Properties of four open-source hyperspectral datasets.

CAVE is an early and frequently used hyperspectral dataset. Unlike others, this dataset was captured using a tunable filter instead of a dispersive grating or prism. It contains 32 images with 512 × 512 pixels and 31 spectral bands between 400 nm and 700 nm. CAVE is a collection of various indoor objects with controlled illumination.

ICVL is collected and published by Arad and Ben Shahar in 2016. This dataset contains 203 images acquired by a line-scanning camera (Specim PS Kappa DX4 hyper-spectrometer). Various indoor and outdoor scenes are included to increase the diversity. The spatial resolution is 1392 × 1300, and 31 spectral channels from 400 nm to 700 nm in 10 nm interval are published for visible applications.

BGU-HS has been the largest natural HSI dataset so far. The dataset consists of ICVL dataset images. During the SR challenge NTIRE-2018, the dataset has been expanded to include 286 images, further divided into 256 training images, 10 verification images, and 20 test images. Each HSI has a spatial resolution of 1392 × 1300, and 31 spectral bands, ranging from 400 to 700 nm with an interval of 10 nm.

ARAD-HS is an HSI dataset for NTIRE-2020 with 510 images, further divided into 450, 30 and 30 images for training, validation, and testing respectively. The spatial resolution is 512 × 482, and the number of spectral bands is 31. This dataset was collected by a portable hyperspectral camera (Specim-IQ). A large variety of indoor and outdoor scenes are collected, such as statues, vehicles and paints.

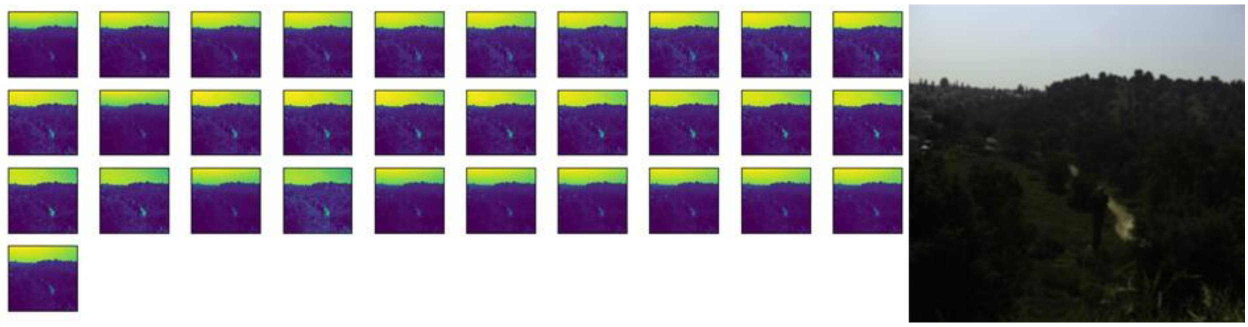

Figure 2. Hyperspectral reconstruction. 31 hyperspectral images for 1 RGB image.

Additionally, a RETINA dataset presented below was selected for classification.

RETINA is the Diabetic Retinopathy dataset for identifying signs of diabetic retinopathy in eye images. The dataset consists with a large set of high-resolution retina images taken under a variety of imaging conditions. A clinician has rated the presence of diabetic retinopathy in each image on a scale of 0 to 4, according to the following scale: 0 – No illness, 1 – Mild, 2 – Moderate, 3 – Severe, 4 – Proliferative illness.

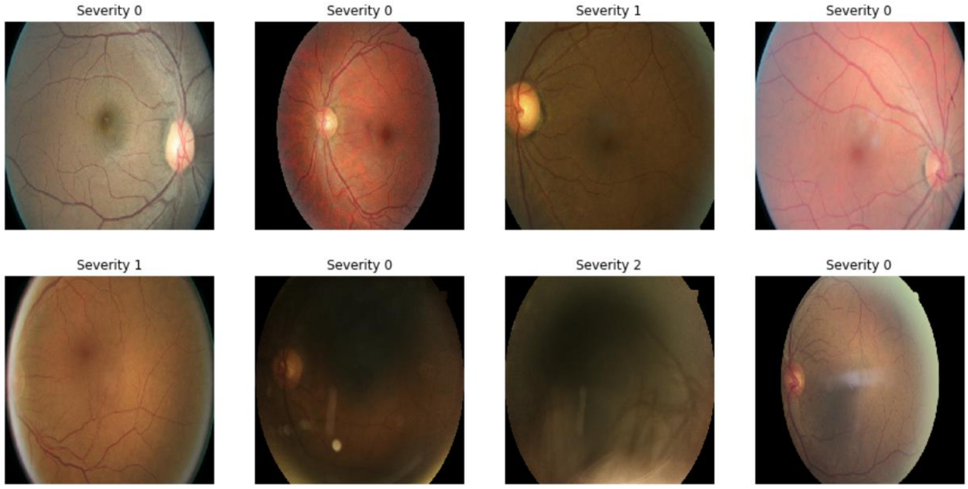

Figure 3. Images presented in Retina dataset.

Related work

In this section, we first outline research related to investigation the most optimal model for two different types of problems: retina classification and hyperspectral reconstruction. Then we review literature related to the work.

[U-Net: Convolutional Networks for Biomedical Image Segmentation.] Olaf Ronneberger, Philipp Fischer, and Thomas Brox.

For transformation task we decided to try U-net model, which considers small details of the image as well as more general patterns. This model is usually used for image segmentation. As our task is some kind similar, U-net fit well. We tried models with different depth and compared the results.

For retina classification we searched for advanced deep learning architectures that are used for solving complex problem especially related to image recognition. They should use data augmentation techniques which comprise of various reflections, patch extractions, and image translations. Based on the results of our research later we compare performance of several classification models on the Retina dataset.

Key Paper

[A survey on computational spectral reconstruction methods from RGB to hyperspectral imaging.] Nature Scientific Reports 2022.

The authors compare the spectral performance of different SR algorithms based on datasets presented above. The article also describes analysis of different methods applicable for reconstruction as well as accuracy comparison of the advanced deep learning architecture. Our task in this work is to confirm or disprove the research results, evaluating the accuracy of multiclass classification for reconstructed Retina dataset images.

Method

This article presents two possible algorithms of hyperspectral reconstruction from RGB images. In this paper, we studied only the data-driven methods. The structure of our approach is presented in Figure 2.

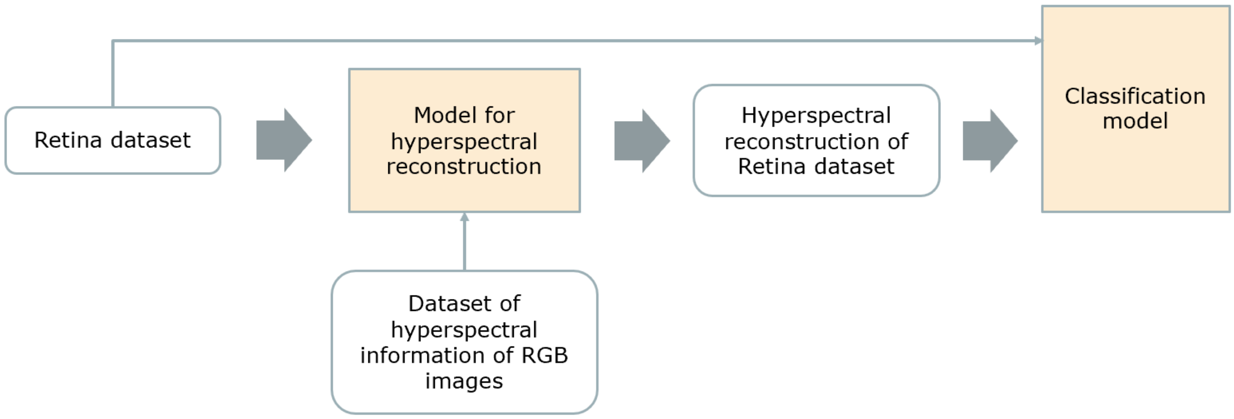

Figure 4. The structure of the presented project approach.

Overall model is like basic U-net from paper. The only difference here is that our last conv layer outputs shape with 31 dimensions.

The optimizer is basic – Adam that shows better results in most cases. We start with loss 0.01 and reduce it, if the loss doesn't improve 2 epochs in a row. As a loss we used the MSE (mean squared error) and PSNR (peak signal-to-noise ratio) as a metric.

Comparing 3 and 4 encoder-decoder model, the one with bigger depth showed bigger metric values on validation dataset.

To determine the optimal classifier, we compared the accuracy of neural network models: InceptionV3, ResNet50, VGG19 and CNN. After analysis we have decided to use ResNet50 architecture and final model structure presented below.

```
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 resnet50 (Functional)       (None, 2048)              23587712

 flatten (Flatten)           (None, 2048)              0

 dense (Dense)               (None, 512)               1049088

 dropout (Dropout)           (None, 512)               0

 dense_1 (Dense)             (None, 5)                 2565

=================================================================
Total params: 24,639,365
Trainable params: 1,051,653
Non-trainable params: 23,587,712
_________________________________________________________________
```

Figure 5. Model structure used for classification.

For optimizer we also decided to use Adam. Loss is multi-class cross-entropy loss as it is the default loss function to use for multi-class classification problems. Metric was simple accuracy.

The next step was to create a classifier using selected algorithm. Model should classify based on the reconstruction of RGB dataset images.

Results and experiments

The hyperspectral reconstruction model got PSNR of 21.72 on validation dataset and about 25 while training, which isn't good enough. I'd like to admit that these results we got from training just on BGU dataset. When we merged datasets, the loss jumped to 200 hundred and wasn't decreasing. The reason for that in my opinion is that the datasets were created on different equipment and maybe in different ways, so the labels where different and our model couldn't find dependency. The other problem was that this model needs a lot of computational resources to run. Because of limited GPU memory we couldn't increase the size of the batch to more than 8. Otherwise, we got OOM (Out of Memory) error. The other problem with dataset was that it couldn't fit into RAM, so we needed to get data using generators, which slowed down our model training.

As for retina classification problem. First, it's worth noting that there were a lot of problems with the retina dataset. The data is highly imbalanced, which later affected the final accuracy.

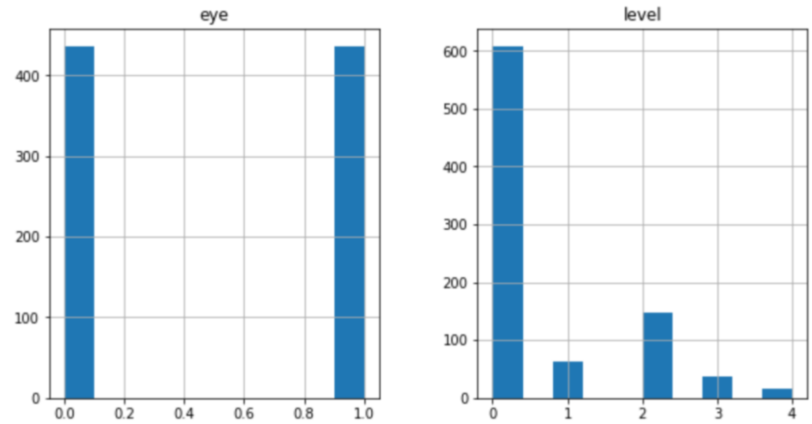

Figure 2. Picture 1 - balanced number of retina images for left and right respectfully. Picture 2 - Imbalanced class amount presented in the dataset.

The final accuracy is 71.27%. The accuracy of the other models ranges approximately from 65 to 70. Using class weights or oversampling/undersampling methods, prediction accuracy drops and starts to vary from 8 to 20 percent, depending on the method used.

Discussion and conclusion

The task turned out to be more complicated than we expected. At some points, due to the huge size of the datasets, we did not have enough computing power to analyze the results and continue to work. In addition, we can note that there are no analogs of the datasets presented in the open sources because of the complexity of capturing and processing this type of data.

A lot of time was lost for the processing the data and creating the image generators for the selected models. Otherwise, we would have been able to experiment with different hyperparameters to improve the results.