The Academic College of Tel-Aviv

**THE SCHOOL OF COMPUTER SCIENCE**

# Spectrally aware Demosaicing

Thesis submitted in partial fulfillment of the requirements for the M.Sc. degree in the School of Computer Science of the Academic College of Tel Aviv University

By

Yahel Rony

The research work for the thesis has been carried out under the supervision of

Dr. Saabni Raid

# Contents

# 1 Introduction

Digital color cameras aim to record the light reflected/emitted from a scene in three color channels: Red, Green, and Blue (RGB). This is most commonly achieved using an optical system which includes a single CMOS sensor overlaid with a "Bayer filter mosaic" also known as an RGB Color Filter Array (CFA). This CFA allows a monochrome light sensor to record RGB color information at the expense of spatial sub-sampling. Figure 1.1 depicts a commonly used RGB Bayer filter pattern. While alternative camera configurations, such as 3-CCD cameras, can record RGB color information without spatial sub-sampling - their use is largely limited to scientific or other speciality applications.

Reproduction of color images at full sensor resolution from single-sensor CFA or "Mosaic" cameras, requires some form of interpolation to recover missing color information from spatially sub-sampled images. This process is commonly referred to as "demosaicing", Figure 1.2 depicts the spatially sub-sampled color channels of an image combined to form a full color image via demosaicing. Due to the ill-posed nature of the demosaicing problem, solutions will invariably suffer from inaccuracies and/or visual artifacts.

Over the years, demosaicing methods have evolved from simple interpolation through more complex and spatially-aware methods [18, 20], to machine-learning based solutions. Most recently, deep neural nets and particularly Convolutional Neural Nets (CNNs) have emerged as the leading methodologies for state-of-the-art demosaicing performance. As with most modern machine-learning systems, advanced demosaicing systems require significant amounts of data for training and testing.

This paper will review recent methodologies for demosaicing, the data-sets used to train them, and highlight possible shortcomings of current training methodologies. These shortcomings will be expiermentally demonstated and quantified. Finally, a data-set and training framework will be

proposed to reduce the dependence of demosaicing methods on camera-specific training data, while increasing their camera-specific performance.

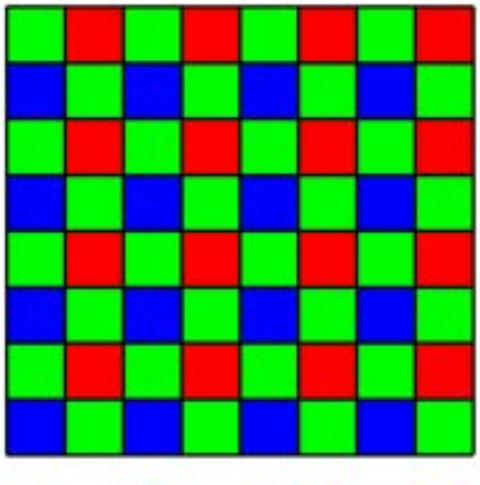


Figure 1.1: RGB Bayer Filter in a "GRBG" order of arrangement. Other common filter orders include "BGGR", "RGBG", and "RGGB".

(a) (b) (c) (d) (e) (f)

Figure 1.2: Depiction of a Bayer mosaic image (a) and its red (b), green (c), and blue (d) components, combined to form a full color image via demosaicing (e). The resulting image is somewhat degraded when compared to a fully spatially sampled image (f).

# 2 Previous Work

## 2.1 Demosaicing

Most Digital cameras capture only a single color for each pixel (commonly Red, Blue, Green) in a CFA pattern. Demosaicing is the process of interpolating the full color information for each pixel given these sub-samples.

Early demosaicing solutions suggested interpolations such as Bilinear interpolation [16]. Bilinear interpolation calculates the missing color based on averaging the neighboring pixels of the same color. Although simple and time efficient, this method works well mainly on flat areas, and tends to produce significant artifacting such as "zippper" patterns and moiré in more complex and textured areas.

In order to overcome these issues, more advanced interpolations methods were proposed, which leveraged inter-channel correlations and considered spatial features (such as edges and gradients) within the image [18]. Zhang *et al*. [27] also exploited the image nonlocal redundancy to improve the local color reproduction result. They first use local directional interpolation (LDI) within local window to create multiple estimates of the missing color sample. Than, nonlocal pixels that are similar to the estimated pixel are searched. Nonlocal adaptive threshholding is used to improve the estimate. This allows the reconstruction to use the structural qualities of the image and not only be applied on a pixel level. Many previous solutions have been based on the assumption of high spectral correlation, and were bound to fail around areas like edges, where this correlations are often weak. This method uses the fact that in natural images similar patterns can be found in further away areas of an image. These patterns can than be used to better interpolate a given similar patch.

In recent years, deep neural networks have achieved great results for solving the demosiacing problem, and many state of the art (SOTA) solutions

are based on this method. For example, Tan *et al*. [25] suggested a deep residual convolutional neural network that is trained end-to-end with two stage architecture: the first step aims to recover a guidance prior from the G channel by training on a large amount of training images. This is a common method for reconstruction that is chosen due to the fact that in a Bayer CFA the red and blue channels cover 25 percent of the images each, where the green channel covers 50 percent of the total pixels, allowing much better accuracy in its reconstruction. The seconds step uses the initial reconstructed G channel information to reconstruct the R and B channels. This can be done due to the high correlation between the R G B channels. The G channel is also refined in the second stage. The residual structure of the network can be used to further improve the results - a simple algorithm (such as bilinear interpolation) can first be applied to create an initial full color image, and the network will use it to predict the residual, meaning instead of creating the image from scratch it will produce the difference between the initial given image and the desired reconstructed result. The algorithm was trained and tested using the Waterloo Exploration Database (WED).

Cui *et al*. [9] further extends the multi-stage residual CNN processing by proposing a 3-stage method. Similar to Tan, the first stage is reconstructing the green chanel independently. Considering the the differences in the inter-correlations between G/B and G/R, in the second stage two separate networks are used to reconstruct the red and blue channels using the green channel as guidance. In the third stage, a high quality RGB images is reconstructed. They chose the training Data set to be WED as well.

## 2.2 Denoising

Images are inevitably contaminated by noise during acquisition process, which may visibly degrades the quality of an image. Denoising is the ill-posed task of trying to restore the true images by removing the noise. Similar to demosaicing, denoising algorithms also work well in flat areas of the images, but tend to smooth high textured more complex areas. This is due to the fact that noise, edges and texture are all high frequency and therefor are difficult to differentiate during the demosaicing process, leading to loss of detail in the more complex regions.

Early methods like total variation denoising [22] and wavelet coring [23] use hand-craft features. Later on, more advanced methods were suggested

using image priors such as self-similarity [7] and sparse representation [3]. Dabov *et al*. [10], proposed BM3D which is often regarded as a denoising benchmark.

In recent years, as is the case with many computer vision tasks, convolutional neural networks (CNNs) have become an increasingly common method for solving the denoising problem [1]. CNNs based methods are able to learn more complex spatial and color correlations than previously proposed methods and have shown state-of-the-art performance.

## 2.3 Joint ISP tasks

Another significant trend has been the integration of several steps in the image processing pipeline into a single machine learning system. In the conventional image processing pipeline, demosaicing precedes denoising. However, demosaicing can compound image noise, presenting a greater challenge to any denoising algorithm applied to the demosaiced image. For this reason, Gharbi *et al*. [11] proposed a joint demosaicing and denoising CNN, trained to overcome moire and other artifacts by training on images which are prone to them. This approach has become increasingly common, and has been presented as a distinct challenge track at a CVPR workshop where 24 methods were presented [1]. Different approaches suggested combining demosaicing and denoising, denoising and super resolution, super resolution and demosaicing, and all three - demosaicing, denoising and super resolution.

In 2019, Qian *et al*. [21] have proposed integrating demosaicing, denoising, and super-resolution, presenting both the large-scale date-set PixelShift200 (which will be further alaborated in the next section) and Trinity Enhancement Network (TENet) - a CNN architecture to achieve all three goals in an holistic perspective. After analyzing the charecteristics of each task and the interaction between them to try and determine the best order in which they should be combined, they proposed the following pipeline: denoising on the raw sensor image (a mosaiced image) $\rightarrow$ super resolution $\rightarrow$ demosaicing.

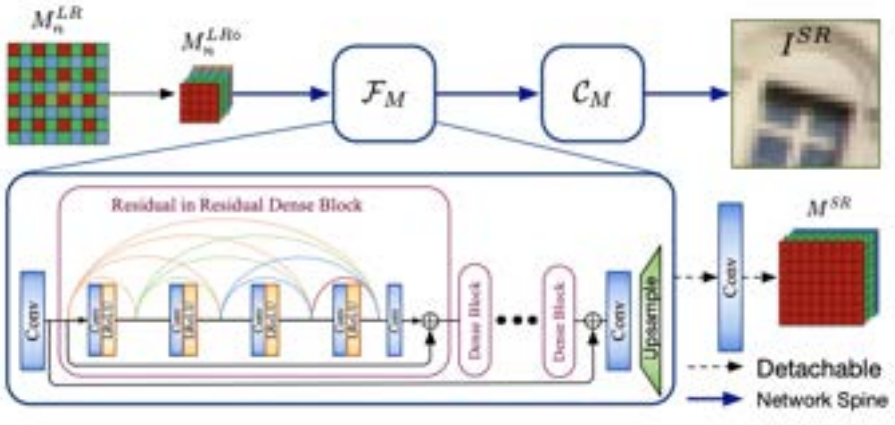



Figure 2.1: Trinity Enhancement Network (TENet) is comprised of a joint denoising and super-resolution module FM and a demosaicing module CM.

## 2.4 Data-sets

This section reviews the most commonly used data-sets, comparing them and discussing their advantages and shortcomings. A systemic review of literature on demosaicing dating back to 2004 has produced a list of demosaicing focused image data-sets available to researchers, these data-sets are detailed in Table 2.1. A further review of 25 learning-based demosaicing papers from the years 2019-2022 has highlighted the data-sets for most commonly used for training/testing in recent work. These data-sets and their frequency of use are depicted in Figure 2.2, note that published works may often use two or more data-sets for training and/or testing.

While data-sets for other computer vision tasks have rapidly grown in scale to hundreds of thousands or even millions of images, data-sets for RGB demosaicing remain well below the 10,000 image mark. To further compound the issue of scale, data-sets are often either camera-specific, or rely on inappropriate methods to generate ground truth data (such as unknown camera ISP demosaicing algorithms). Despite the increasing availability of larger data-sets with robust ground-truth data (e.g. Pixelshift[21]), the majority of recently published papers continue to rely on data-sets with spatially sub-samples ground truth data (e.g. WED, Flicker2K/500, MIT, ImageNet, MSR) or very small data-sets with less than 50 images (e.g. Kodak).

While the Kodak Data set has been vastly used and has been established

as one of the standard evaluation Data sets in the research community, multiple issues regarding it have been raised along the years, suggesting that it's unrepresentative qualities may have misled the research to some extent. Besides it's very small volume, it has been claimed that the Kodak Data set images have more spectral correlation and and less color saturation when compering with images in other Data sets. When compered with the target Data set for this problem, which is images captured by current digital cameras, the images are smoother and have less color saturation. The main issue arises where a basic solution can outperform a more complex solution that is better tailored for the target digital images, only due to the fact that it was tested on the Kodak Data set.

A basic requirement for any data-set used to train/test demosaicing systems is that its images should not be spatially sub-sampled. Early attempts, such as the Kodak data-set, utilized source images which could be considered fully spatially sampled (such as scanned film negatives[1]). Although modern solution for fully spatially sampled RGB images exist, such as 3CCD cameras or Sony "Pixel Shift" technology, the majority of commonly used demosaicing data-sets rely on images which were originally spatially sub-sampled (c.f. Table 2.1). While some of these researchers have applied mitigations, such as downsampling, to reduce the effect of spatial subsampling (e.g. MSR [13]), others rely solely on demosaicing performed by an often unknown ISP.

Commonly used data-sets can be categorized as follows:

### 2.4.1 Down-sampled RGB images from various cameras

This category contains some of the most commonly used data-sets for both training and testing demasicing problems. It includes Flickr500, DIV2K, Flickr2K and WED. The images in these data-sets were obtained by scraping the internet for various suitable images. After scraping the images they are filtered by licence, resolution, diversity etc. The advantages of such method are vast - the use of existing images from the internet saves the task of acquiring images from scratch and therefor can contain a larger, and more diverse data-set. Surprisingly, in practice, most these data-sets are relatively small, with the largest one being WED, containing 5K im-

[1] While camera film is indeed spatially sub-sampled due to film "grain" (i.e. small light sensitive particles which chemically record the itensity of light in different wavelengths), this subsampling is negligible when considering the scanning resolution used to digitize the film.

ages.

While there are some advantages, the use of such data-sets holds many problems - the main one being that the images have already been processed by a full ISP (Image Signal Processors) pipeline. They have already been demosaiced, denoised, color corrected etc. Additionally, the processes and camera information are completely untraceable to us since they were taken on different, unknown cameras. The images are then down-sampled in ordered to obtain ground truth for the RGB channels, resulting in further degradation of the image resolution.

The use of multiple unknown cameras as a source of ground truth is likely to create multiple biases in the ground truth data. Different cameras may have different CFAs (c.f. Section 3.3, Figures 3.1, 3.2), different ISP settings (e.g. color tuning, tone mapping, sharpening, etc.), and other differences which affect the resulting images. Images collected from online sources may have even undergone post-capture editing and adjustments which could create additional biases. Arguably, the inclusion of images from multiple sources may somewhat mitigate individual biases, but detailed information on the progeny of each image it is quite difficult to assess the extent of such biases. Furthermore, the test or validation image subset of such a data-set may contain a different distribution of image sources than the training subset - this could severely skew either training, evaluation, or both.

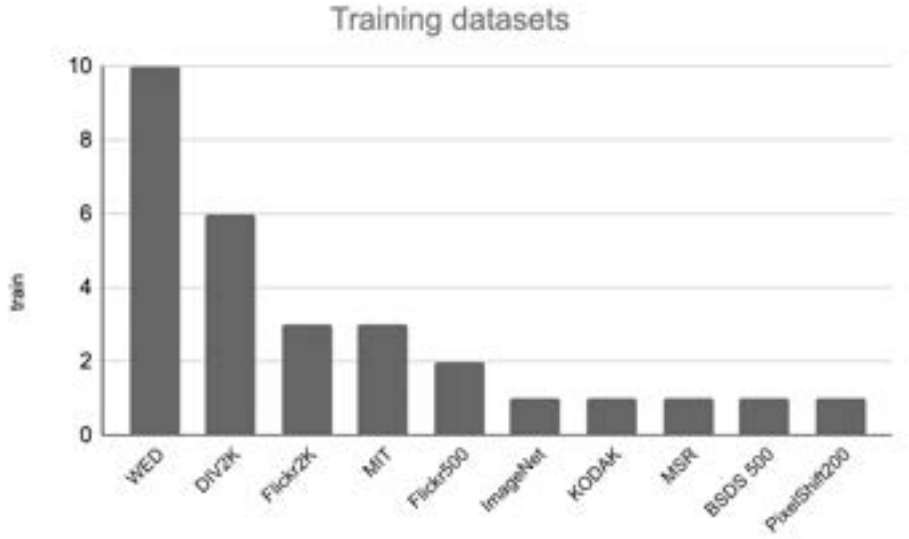



Figure 2.2: data-set used to train 25 sampled learning-based algorithms. Note that some papers utilized more than one training data-set.

| data-set | Year | # Images | Resolution | Camera | GT Generation Method |
|---|---|---|---|---|---|
| PixelShift [21] | 2019 | 200 | 4K | Sony ILCE-7RM3 | Pixel shift |
| Flickr500 [24] | 2018 | 500 | $640 \times 480$ | Various (Web) | Camera Demosaic[1] |
| DIV2K [2] | 2017 | 1,000 | 2K | Various (Web) | Camera Demosaic |
| Flickr2K [15] | 2017 | 2,650 | 2K | Various (Web) | Camera Demosaic |
| Kodak[14] | 1993 | 25 | $768 \times 512$ | Film | Scanned Film |
| WED[17] | 2016 | 4,744 | $< 0.4$MP | Various | Camera Demosaic |
| MSR [13] | 2014 | 557 | $210 \times 318$ | Various (below) | Downsampling |
| Canon | 2014 | 57 | $210 \times 318$ | Canon EOS 550D | Downsampling |
| Panasonic | 2014 | 500 | $210 \times 318$ | Panasonic Lumuix DMC-LX3 | Downsampling |
| McMaster[27] | 2011 | 8 | $2310 \times 1814$ | Film | Scanned Film |
| MIT Moire [11] | 2016 | 2M patches | 128x128 | Various (Web) | Camera Demosaic |
| BSDS | 2001 | 500 | $481 \times 321$ | N/A | N/A |

Table 2.1: Overview of data-sets for training and testing demosaicing systems. Note that many of these data-sets rely on inaccurate methods for producing fully sampled RGB images such as downsampling and the application existing demosaicing algorithms.
[1] Images collected from the internet, most likely demosaiced by source camera ISP.

### 2.4.2 Down-sampled RGB images from a specific camera

This category includes data-sets such as Kodak, MSR, and McMaster.

Similar to the previous category, the images on this data-set have already been through the ISP process, and are down-sampled to obtain GT information. While holding the advantage of having the full camera information, this data-sets are small, ranging from 25 to 500 images, and, as in all the other data-sets, are camera specific, resulting in different distribution between the training data-set and the final input.

### 2.4.3 Full data from a specific camera

The PixelShift200 data-set was introduced trying to solve some of the problems mentioned above. The pixelshift200 data-set contains 200 high-quality 4K resolution full color sampled real-world pairs of raw images and color ones. This data-set was obtained using the pixel shift technique, in which the same frame is taken several times, each time moving the camera sensor in precisely one pixel, allowing each pixel in the image to have the full information of all three channels - R, G, B. The images were collected using the Sony ILCE-7RM3 digital camera. This data-set, in many ways, offers a great improvement over the previous - these images have the full color information, have not yet been processed and were not down-sampled, allowing for high-resolution images and fewer artifacts. Another advantage

it holds over most of the commonly used data-sets is the camera used for obtaining the images is known. Despite these very prominent improvements, this data-set is still posing a lot of downsides - it contains a very small amount of images, only 200, and it is camera specific.

While data-sets without spatial sub-sampling have obvious advantages, a camera-specific data-set cannot be easily used to train demosaicing systems for different cameras. Given the rate of innovation in the image sensor market, it is quite unreasonable to collect such a data-set for each new camera system. The next section intends to demonstrate that it is critical to train the demosaicing network on a the same camera from which the mosaic images are received, and suggest a method to do so without repeated, laborious, data collection efforts.

# 3 Framework for Spectrally Aware Demosaicing

While the issue of spatial sub-sampling has often been addressed in previous works, it would seem that no significant attempt has been made to tackle the issue of camera spectral responses, and the effect it might have on both training and inference. It has been long established that cameras may differ significantly in their spectral response, and that this difference can affect their output [12]. One could then surmise that any machine-learning-based demosaicing system could suffer degraded performance in the response function of the camera(s) used to generate training data is significantly different from the camera response function of the camera used for inference.

In this section, a framework and data-set for spectrally aware demosaicing is presented. This framework facilitates the automatic creation of a "camera-native" (i.e. which matches the spectral response of the target camera) data-set for training CNNs or other machine learning systems for demosaicing. The framework consists of: a hyperspectral **data-set** which provides the ground spectrally adaptable ground truth information, **software** which provides the means to adapt the data-set to the response function of any camera, and a **methodology** for training and evaluating demosaicing algorithms.

## 3.1 Hyperspectral Source data-set

In contrast to conventional RGB cameras, Hyperspectral Imaging Systems (HISs) can record the amount of light reflected/emitted from a scene over a large number of narrow wavelength bands, often over 10s or 100s of bands. Traditional HISs employ either spatial or spectral scanning to avoid

spatial or spectral sub-sampling at the expense of long acquisitions time, often 10s of seconds or minutes. Hyperspectral images are ideal candidates for training demosaicing systems, as they are fully spatially sampled, and spectrally *over-sampled* when compared to RGB images. This section will describe how the spectral over-sampling of hyperspectral images can be utilized to easily create camera-specific training data for demosaicing algorithms.

While HISs have been common for over a half a century, they have most commonly been used in remote sensing, agriculture, geology, astronomy, earth sciences (cf. [6]), and others [4]. Interest in hyperspectral images of *natural* scenes has only recently began to grow, with early data-sets including less than a hundred images [8]. Table 3.1 details notable data-sets of natural hyperspectral images and their scope. With the increasing availability and scope of natural hyperspectral image data-sets, they have now become a viable option for training demosaicing algorithms.

| Dataset | Year | # Images | Spectral Resolution | Spatial Resolution |
|---|---|---|---|---|
| CAVE [26] | 2010 | 32 | 31 Bands | 512x512 |
| Chakrabarti [8] | 2011 | 50 | 31 Bands | 2048x2048 |
| TokyoTech [19] | 2015 | 30 | 31 Bands | 1392x1040 |
| ICVL [4] | 2016 | 201 | 519 Bands | 1392x1300 |
| TT59 | 2018 | 201 | 59 Bands | 2048x2048 |
| Hytexila | 2018 | 112 | 186 Bands | 1024x1024 |
| BGU HS | 2018 | 256 | 31 Bands | 1392x1300 |
| ARAD | 2020 | 510 | 31 Bands | 480x512 |
| ARAD 1K [5, 6] | 2022 | 1000 | 31 Bands | 480x512 |

Table 3.1: Notable data-sets of natural hyperspectral images.

The "ARAD 1K" hyperspectral data-set [5, 6] contains 1,000 natural hyperspectral images. The data-set was collected using Specim IQ mobile hyperspectral camera, allowing for both indoors and outdoors scenes and a wide and diverse data-set. The images collected are RAW $480 \times 512$px with 31 spectral bands in the 400-700nm range. The images provided are fully spatially sampled and can be used to produce RGB mosaic images for any known camera response function. The data-set also easily lends itself to various augmentations, for example: scene location relative to the virtual camera mosaic can be perturbed, generating x4 images per scene. The following section will present a training framework which adapts the ARAD 1K data-set to become a flexible, camera-specific training/testing data-set for demosaicing algorithms.

## 3.2 Training Framework

The data-set can be found at: `https://github.com/boazarad/ARAD_1K` The model used to training and evaluating the data-sets is: TENet [21]. For the sake of keeping the experiment as noise free as possible, the model was run as-is with no hyper parameter tuning or other interference.

## 3.3 Experimental Evaluation

To demonstrate the benefits of spectrally aware demosiaicing, two training/testing data-sets for RGB demosaicing were created from the ARAD 1K spectral data-set:

1. Data-set simulating the spectral response of Camera A: a Canon 1D Mark III.
2. Data-set simulating the spectral response of Camera B: a Point Grey Grasshopper 50S5C.

Figure 3.1 depicts the spectral response function for each camera. Using the tools described in Section 3.2, a set of train/test/validation mosaic images were generated from the ARAD_1K spectral image data set. The resulting camera specific "RAW" mosaic image were used to train a TENet CNN-based demosaicing system. The results of training on camera-native and non-native data-sets are reported in the following sections.

### 3.3.1 Camera-Native Demosaicing

#### Point Grey Grasshopper 50S5C

Training the TENet CNN on simulated Point Grey Grasshopper 50S5C "RAW" mosaic images achieved a PSNR of 43.96 over the corresponding Point Grey test set. Demosacied imaged show overall high image quality (c.f. Figure 3.3.B) and no noticeable artifacts (c.f. Figure 3.5.B).

#### Canon 1D Mark III

Training the TENet CNN on simulated Canon 1D Mark III "RAW" mosaic images achieved a PSNR of 30.17 over the corresponding Canon test

set. While both the architechture and training procudure were identical to those used in the Point Grey experiment, the trained TENet model produced images with periodic mosaic artifacts (c.f. Figure 3.6.B) which has a noticable effect on overall image quality (c.f. Figure 3.4.B). Future work is required to explore the significant performance gap of TENet over different camera types.

### 3.3.2 Non-Native Demosaicing

#### Point Grey Grasshopper 50S5C

Training the TENet CNN on simulated Canon 1D Mark III "RAW" mosaic images achieved a PSNR of 27.11 over the Point Grey Grasshopper 50S5C test set. While degraded in comparison to the camera-native demosaicing setting, overall image quality remains high (c.f. Figure 3.3.C) demosaicing artifacts are not clearly visible (c.f. Figure 3.5.C).

#### Canon 1D Mark III

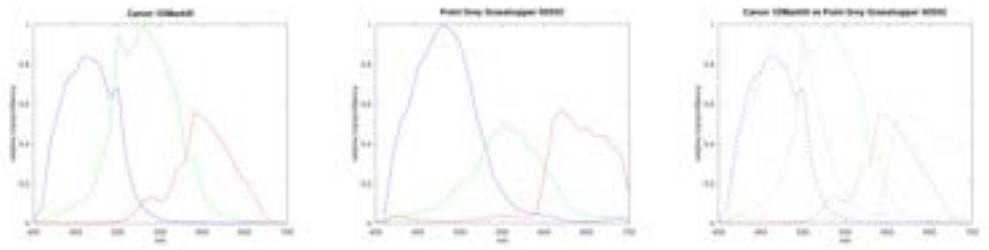
Training the TENet CNN on simulated Point Grey Grasshopper 50S5C "RAW" mosaic images achieved a PSNR of 40.3 over the over the Canon 1D Mark III test set. Similarly to the camera-native experiment periodic mosaic artifacts are visible in the results (c.f. Figure 3.6.C) and overall image quality is degraded even further (c.f. Figure 3.4.C)

### 3.3.3 Result Comparison

As evident by the results described in the previous section, the camera response function used to generate training data has a crucial impact on both overall performance, as well as performance over a camera-native or non-native test set. Table 3.2 details the results over each test set as a function of the training data used.

| Training data-set \Testing data-set | 1D Mark III | Grasshopper 50S5C |
|---|---|---|
| 1D Mark III | 30.17 | 27.11 |
| Grasshopper 50S5C | 40.3 | 43.96 |

Table 3.2: Performance of deep demosaicing algorithm in PSNR as a function of camera response function of train and test images.



(a) (b) (c)

Figure 3.1: Response function of the Canon 1D Mark III (a), the Point Grey Grasshopper 50S5C (b), and both response functions overlaid for comparison (c) where the Canon 1D Mark III is represented by dashed line, and the Point Grey Grasshopper 50S5C represented by a dotted line. Note the marked differences between camera response functions.

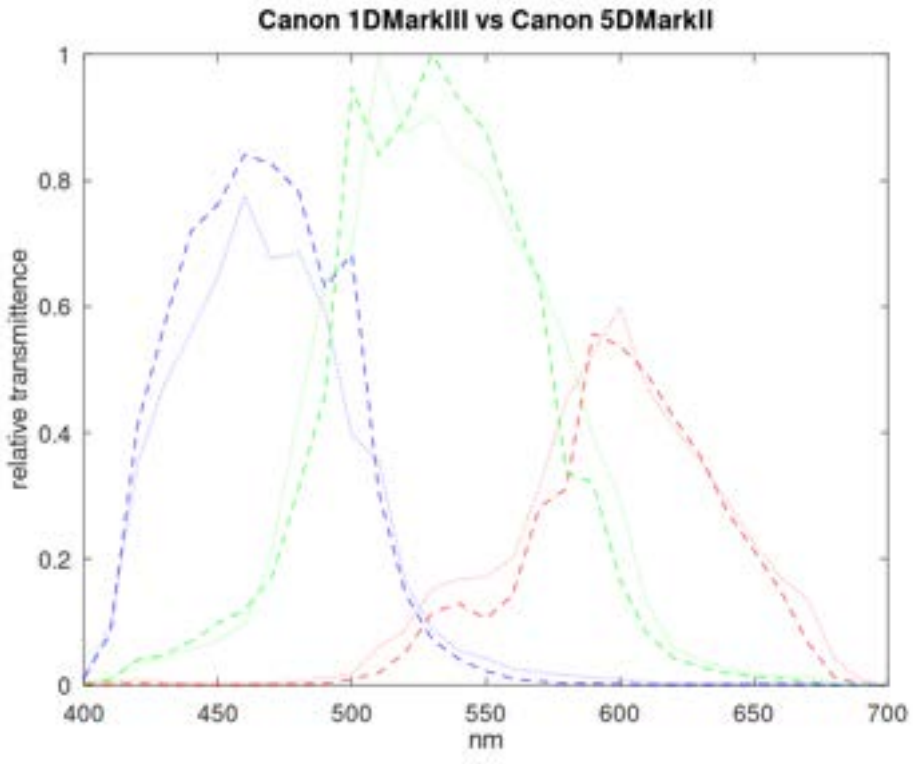



Figure 3.2: Response function of the Canon 1D Mark III (dashed) compared to the the Canon 1D Mark II (dotted). Despite being successive cameras in the same product line, the camera sensors' response functions show marked differences.

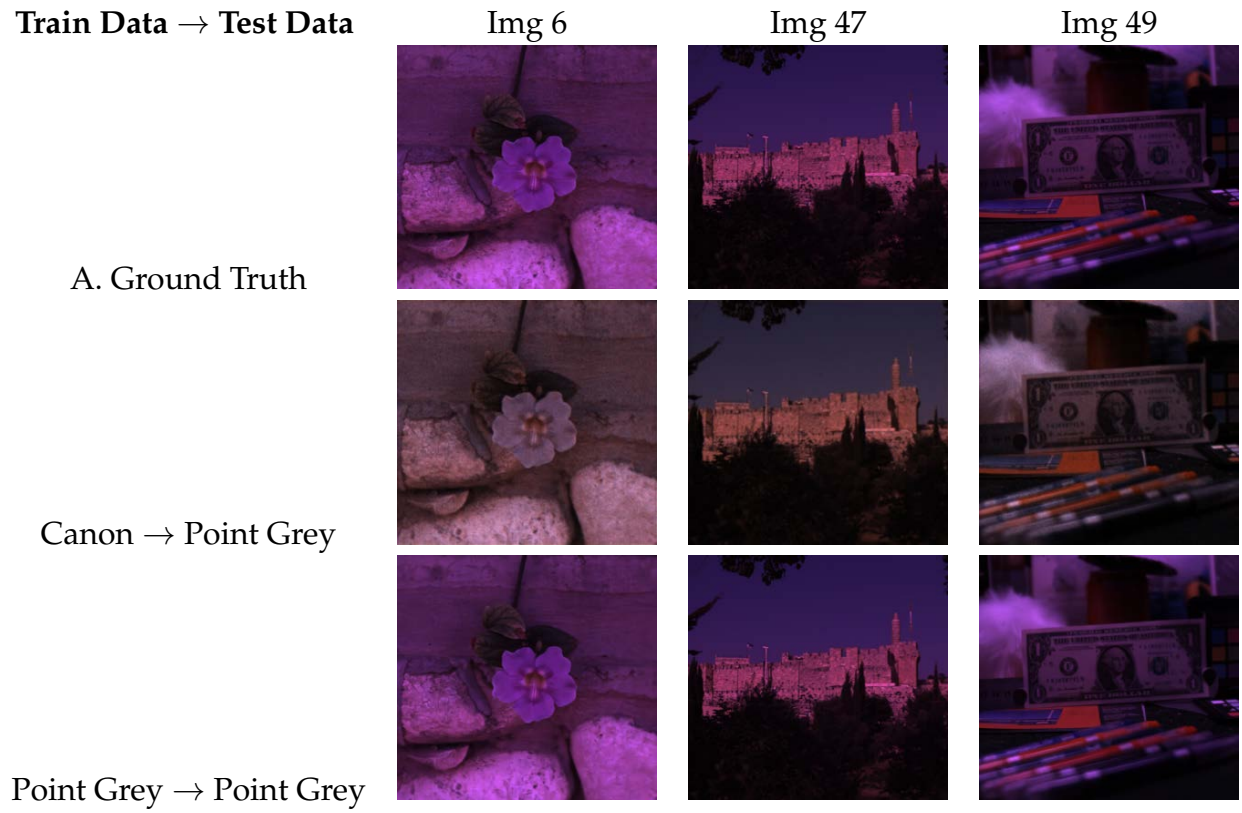



Figure 3.3: Demosaicing results compared to ground truth for the Point Grey Grasshopper 50S5C camera.

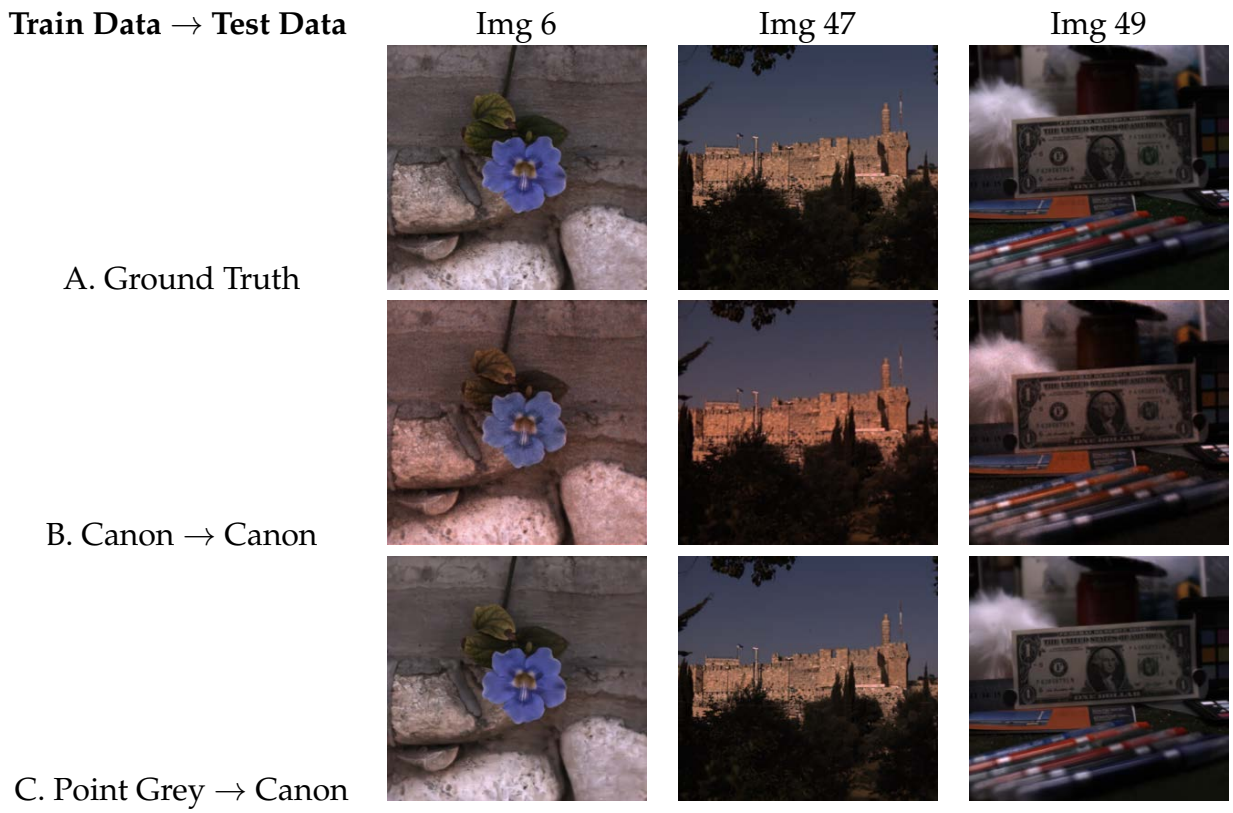



Figure 3.4: Demosaicing results compared to ground truth for the Canon 1D Mark III camera.

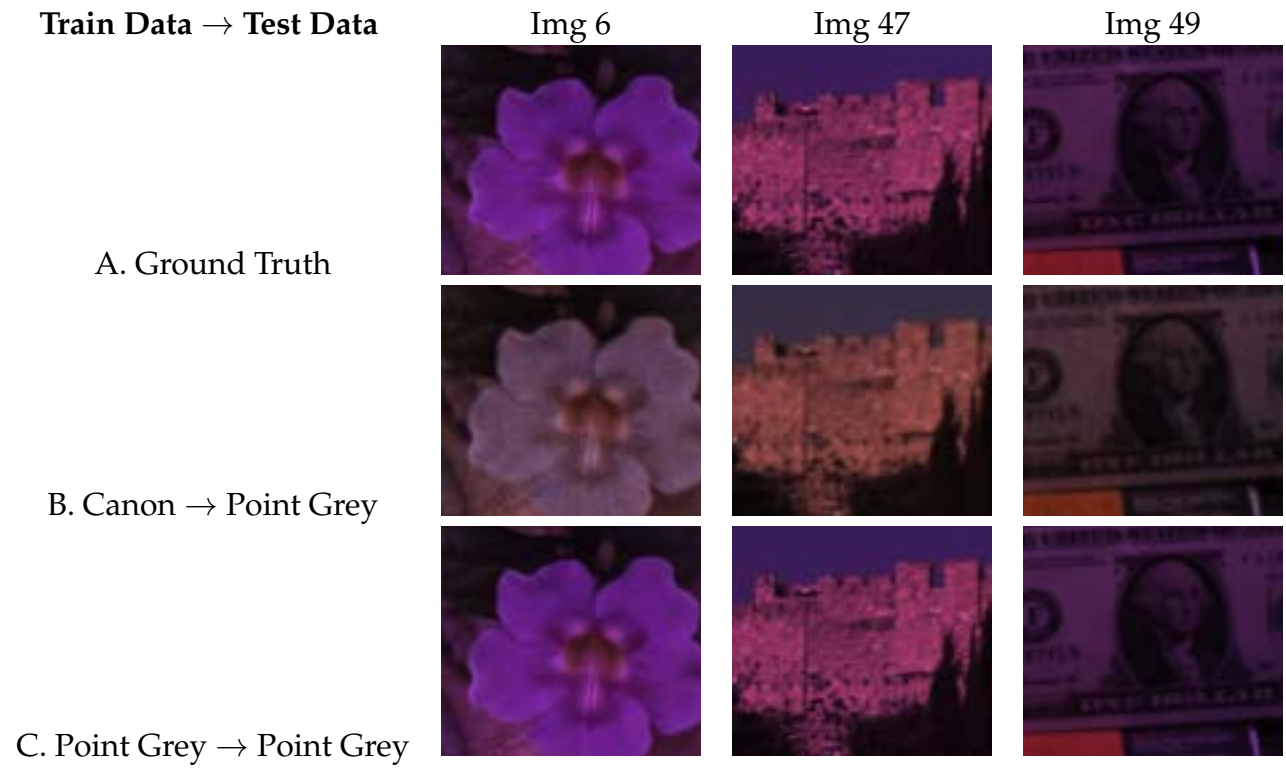



Figure 3.5: Detail of demosaicing results compared to ground truth for the Point Grey Grasshopper 50S5C camera.

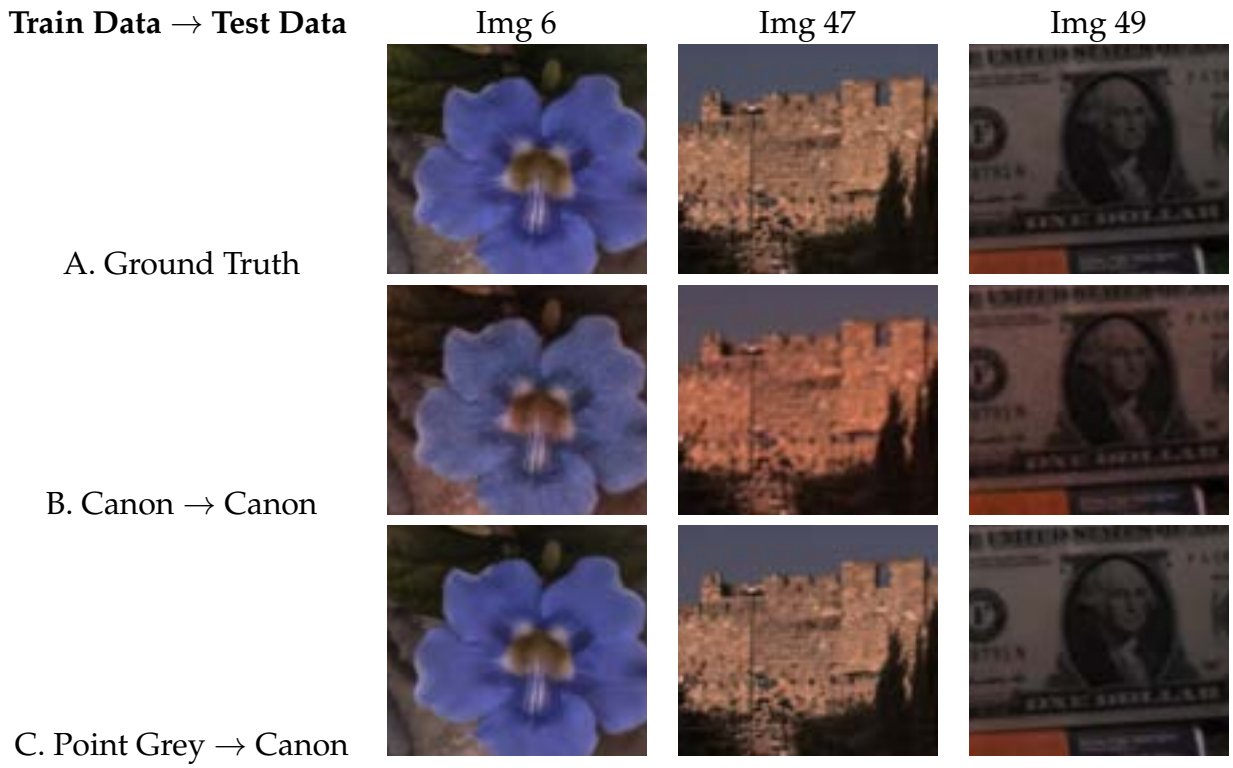



Figure 3.6: Detail of demosaicing results compared to ground truth for the Canon 1D Mark III camera.

# 4 Discussion

The use of spectrally aware demosaicing has clear benefits as demostrated in section 3.3. To date, applying these principals in practice would require a dedicated collection effort for each target camera. Given the rate of innovation in the image sensor and optics market, this would no doubt be impractical if not infeasible. However, the increasing availability of natural hyperspectral images, combined with the methodology presented in section 3.2, provide the opportunity to easily and efficiently create a large-scale, camera-native, training data-set for any current or future camera.

The demosaicing data-set presented here is larger than any other widely-used demosaicing data-set which does not rely on demosaicing/downsampling for ground truth generation, and novel in it's adaptability to an unlimited amount of target cameras. As hyperspectral imaging hardware and software improve, the methodology presented here could be applied to even larger data-sets of natural hyperspectral images, facilitating the training of increasingly complex machine learning systems.

# Bibliography


[1] Abdelhamed, A., Afifi, M., Timofte, R., and Brown, M. S. (2020). NTIRE 2020 challenge on real image denoising: Dataset, methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 496–497.

[2] Agustsson, E. and Timofte, R. (2017). Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 126–135.

[3] Aharon, M., Elad, M., and Bruckstein, A. (2006). K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing*, 54(11):4311–4322.

[4] Arad, B. and Ben-Shahar, O. (2016). Sparse recovery of hyperspectral signal from natural rgb images. In *European Conference on Computer Vision*, pages 19–34. Springer.

[5] Arad, B., Timofte, R., Yahel, R., Morag, N., Bernat, A., et al. (2022a). NTIRE 2022 spectral demosaicing challenge and dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

[6] Arad, B., Timofte, R., Yahel, R., Morag, N., Bernat, A., et al. (2022b). NTIRE 2022 spectral recovery challenge and dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

[7] Buades, A., Coll, B., and Morel, J.-M. (2005). A non-local algorithm for image denoising. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 2, pages 60–65. Ieee.

[8] Chakrabarti, A. and Zickler, T. (2011). Statistics of real-world hyperspectral images. In *CVPR 2011*, pages 193–200. IEEE.

[9] Cui, K., Jin, Z., and Steinbach, E. (2018). Color image demosaicking using a 3-stage convolutional neural network structure. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 2177–2181. IEEE.

[10] Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K. (2007). Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8):2080–2095.

[11] Gharbi, M., Chaurasia, G., Paris, S., and Durand, F. (2016). Deep joint demosaicking and denoising. *ACM Transactions on Graphics (ToG)*, 35(6):1–12.

[12] Jiang, J., Liu, D., Gu, J., and Süsstrunk, S. (2013). What is the space of spectral sensitivity functions for digital color cameras? In *2013 IEEE Workshop on Applications of Computer Vision (WACV)*, pages 168–179. IEEE.

[13] Khashabi, D., Nowozin, S., Jancsary, J., and Fitzgibbon, A. W. (2014). Joint demosaicing and denoising via learned nonparametric random fields. *IEEE Transactions on Image Processing*, 23(12):4968–4981.

[14] Kodak, E. (1993). Kodak lossless true color image suite (photocd pcd0992). *URL http://r0k. us/graphics/kodak*, 6.

[15] Lim, B., Son, S., Kim, H., Nah, S., and Lee, K. M. (2017). Enhanced deep residual networks for single image super-resolution. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

[16] Longere, P., Zhang, X., Delahunt, P. B., and Brainard, D. H. (2002). Perceptual assessment of demosaicing algorithm performance. *Proceedings of the IEEE*, 90(1):123–132.

[17] Ma, K., Duanmu, Z., Wu, Q., Wang, Z., Yong, H., Li, H., and Zhang, L. (2016). Waterloo exploration database: New challenges for image quality assessment models. *IEEE Transactions on Image Processing*, 26(2):1004–1016.

[18] Malvar, H. S., He, L.-w., and Cutler, R. (2004). High-quality linear interpolation for demosaicing of bayer-patterned color images. In *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 3, pages iii–485. IEEE.

[19] Monno, Y., Kikuchi, S., Tanaka, M., and Okutomi, M. (2015). A practical one-shot multispectral imaging system using a single image sensor. *IEEE Transactions on Image Processing*, 24(10):3048–3059.

[20] Pei, S.-C. and Tam, I.-K. (2003). Effective color interpolation in ccd color filter arrays using signal correlation. *IEEE Transactions on Circuits and Systems for video technology*, 13(6):503–513.

[21] Qian, G., Wang, Y., Dong, C., Ren, J. S., Heidrich, W., Ghanem, B., and Gu, J. (2019). Rethinking the pipeline of demosaicing, denoising and super-resolution. *arXiv preprint arXiv:1905.02538*.

[22] Rudin, L. I., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268.

[23] Simoncelli, E. P. and Adelson, E. H. (1996). Noise removal via bayesian wavelet coring. In *Proceedings of 3rd IEEE International Conference on Image Processing*, volume 1, pages 379–382. IEEE.

[24] Syu, N.-S., Chen, Y.-S., and Chuang, Y.-Y. (2018). Learning deep convolutional networks for demosaicing. *arXiv preprint arXiv:1802.03769*.

[25] Tan, R., Zhang, K., Zuo, W., and Zhang, L. (2017). Color image demosaicking via deep residual learning. In *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, pages 793–798.

[26] Yasuma, F., Mitsunaga, T., Iso, D., and Nayar, S. K. (2010). Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum. *IEEE transactions on image processing*, 19(9):2241–2253.

[27] Zhang, L., Wu, X., Buades, A., and Li, X. (2011). Color demosaicking by local directional interpolation and nonlocal adaptive thresholding. *Journal of Electronic imaging*, 20(2):023016.

# תקציר

ברוב המצלמות הדיגיטליות הקיימות כיום, כאשר המצלמה מצלמת תמונה, עבור כל פיקסל בתמונה המידע נתפס על ידי החיישן באחד הצבעים הבאים: אדום, כחול, ירוק. בפורמט הסופי של התמונה, כל פיקסל מיוצג בעזרת שלושת הצבעים יחד. כלומר, לפני הצגת התמונה יש להשלים שניים מהצבעים עבור כל פיקסל. למשימה זו קוראים דימוזאיקה.

כיום, רוב האלגוריתמים הקיימים פותרים את בעית הדימוזאיקה בעזרתת רשת לומדת. לצורך כך נדרש דאטה סט מדויק וגדול לאימון. אחת הבעיות הקיימות בתחום היא הקושי להשיג דאטה סט שכזה לבעיה זו. במחקר זה נסקור את הדאטה סטים הקיימים, נציג את הבעיות הפוטנציאליות בפתרונות הקיימים, ונציע דאטה סט הייפרספקטרלי תוך הדגמת ופירוט יתרונותיו.

-

המכללה האקדמית תל-אביב

בית הספר למדעי המחשב

# דימוזאיקה בעזרת דאטה ספקטרלי

חיבור זה הוגש כחלק מהדרישות לקבלת התואר "מוסמך " – M.Sc.

במכללה האקדמית תל-אביב

על ידי

**רוני יהל**

העבודה הוכנה בהדרכתו של

**דר' ראיד סעאבנה**