RGB to Spectral Reconstruction via Learned Basis Functions and Weights

Biebele Joslyn Fubara Mohamed Sedky Dave Dyke
Staffordshire University
Stoke-on-Trent, UK
biebele.fubara@research.staffs.ac.uk {m.h.sedky,david.dyke}@staffs.ac.uk

Abstract

Single RGB image hyperspectral reconstruction has seen a boost in performance and research attention with the emergence of CNNs and more availability of RGB/hyperspectral datasets. This work proposes a CNN-based strategy for learning RGB to hyperspectral cube mapping by learning a set of basis functions and weights in a combined manner and using them both to reconstruct the hyperspectral signatures of RGB data. Further to this, an unsupervised learning strategy is also proposed which extends the supervised model with an unsupervised loss function that enables it to learn in an end-to-end fully self supervised manner. The supervised model outperforms a baseline model of the same CNN model architecture and the unsupervised learning model shows promising results. Code will be made available online here.

1. Introduction

Humans are able to perceive the world in colour due to the cones in the human visual system that converts light in the visible spectrum. Images captured in RGB colour format are very well suited to humans as they are in a familiar format to the visual system of human beings. However, the visible spectral range of the electromagnetic spectrum contains information beyond the three RGB values generally expected from traditional colour images. This additional data encode hyperspectral colour information, as well as material information. Hyperspectral images are a rich source of scene information that has been explored in several applications including remote sensing [6, 7], astronomy [31], agriculture [24, 29], medical image analysis [8] and computer graphics [10].

Capturing hyperspectral images can be done using hyperspectral cameras, however, these devices are usually expensive, bulky, slow (non-realtime) and low in spatial resolution; making them unsuitable for several realtime and high resolution applications. There have been several attempts at improving the speed [15, 36] and spatial resolution[41, 40, 25] of hyperspectral capturing devices. There remains a trade-off between capture speed and spatial resolution; price and size also still remains a challenge.

Alternatively, spectral reconstruction from RGB images can be used as a hyperspectral image capturing means. This paradigm of hyperspectral image recording has seen an increase in interest recently [4, 14, 3, 37]. The idea is to learn a mapping between the RGB values and their corresponding spectral response. Several methods of learning this mapping has been reported in the literature including Radial Basis Functions (RBF) [26], sparse coding [21, 13, 3] and Convolutional Neural Networks (CNN) [33, 14, 39], with the best performing methods being CNN/deep learning-based.

Deep learning has become ubiquitous in computer vision, due to the surge in the performance of deep learning methods in object detection, depth estimation and semantic segmentation. This work follows this trend, and uses deep learning as a method to learn the transformation from RGB images to spectral response cubes. Most methods tackle this problem by designing a CNN to learn the conversion of RGB triplets to 31-channel spectral images directly, this is evident in the NTIRE 2018 challenge [4]. In this work, the problem is formulated differently. Instead of predicting a 31-channel spectral image, the model predicts weights for a set of basis functions which are learned at the same time as the weights. In classical spectral reconstruction literature, the spectrum was recovered by weighted combination of basis functions [1] or sparse coding [27]. The proposed method combines the simplicity of weighted basis functions and the performance and robustness of deep learning. The network predicts 10 weights for each pixel as well as learns a set of 10 basis functions which is then combined to form the final spectral image.

A second method is also proposed, which can be used when there is no ground truth spectral information to use during training. This method learns the spectral reconstruction in a completely unsupervised manner simply from the RGB input data. Similar to the first method, it leverages learned basis functions and weights for the reconstruction task.

2. Related work

Hyperspectral reconstruction from RGB image is a great alternative to the direct capture of hyperspectral cubes using hyperspectral imaging devices. This idea dates back to several early works [12, 16, 11, 3, 23, 1] that use PCA to extract basis functions from collected databases of spectral reflectances such as those from Munsell Colour chips. These works formulated the problem as a weighted combination of basis functions. Other studies [27, 21, 13] used sparse coding to recover spectral cubes from RGB images. Nguyen *et al*. [26] proposed the use of a Radial Basis Function (RBF) network to learn the mapping from RGB to spectral reflectance. They also proposed the use of white balancing before the RBF model to minimise the effects of illumination on the performance of the model. Arad and Ben-Shahar [3] tackled the problem by creating a sparse spectral dictionary along with their projections in RGB. The RGB reprojections are then used with new RGB images to reconstruct the spectral image with the help of the associated spectral signatures in the dictionary. In order to avoid learning a three-to-many mapping of RGB to spectrum, Jia *et al*. [18] proposed to learn an RGB to 3D embedding (three-to-three), which is then used with a manifold-based reconstruction method to generate the full spectrum; their approach produced more accurate results than the three-to-many approach. These methods mainly work on RGB vectors individually due to the nature of the models used.

An emerging method of spectral reconstruction from RGB involves the use of deep learning/CNNs. This approach is particularly interesting as the model uses the full image or patches of the image instead of working on individual RGB vectors to form the final spectral cube. This approach does let the model take spatial contextual information into account when training and during inference. HSCNN [38] is one of such CNN-based approaches. it works by upsampling the RGB image in the spectral dimension to the desired resolution before passing the upsampled data to a deep learning model that learns the mapping to the corresponding ground truth spectrum. Shi *et al*. [33] improved on this by removing the hand-engineered upsampling step and using a deep residual network and densely connected network to achieve more accurate results. Alvarez-Gila *et al*. [2] used a Generative Adversarial Network to learn hyperspectral signals from RGB images. Other deep learning based methods have been proposed that make use of different CNN model architectures such as UNet [39, 34] and Resnet [9, 22] to map RGB images to 31-channel spectral images.

Earlier methods which used PCA, sparse coding and RBFs worked well but did not take the spatial context of the image into account. Deep learning methods take image spatial contextual information into account and produce more accurate results. However, they all cast the problem as a 3-to-31-channel mapping problem. It has been shown in early PCA and sparse coding works that a weighted combination of basis functions/vectors can produce spectral signatures of RGB inputs. Given that knowledge, the proposed method is formulated as a weighted combination of basis functions; with the weights and basis functions being learned jointly using deep learning without hand-crafted basis functions. This method incorporates the advantages of PCA based methods and that of deep learning based methods to form a robust spectral reconstruction approach. A second method is also proposed which uses the same approach but introduces an unsupervised learning loss that is useful in the absence of ground truth hyperspectral training data.

3. Proposed method

3.1. Problem formulation

Images are formed by cameras using sensors to measure light reflection. The sensors can be in the form of a Charge-Coupled device (CCD) or Complementary Metal-Oxide-Semiconductor (CMOS) chips embedded in the imaging device. Much like human vision, the sensor detects light reflections like the cone cells in the human eyes. The mechanics of image formation by cameras involve the combination of several elements. These elements include the scene illuminant (source of light, e.g. sunlight), objects/surfaces in the scene and the imaging device [35, 32], amongst others. The illuminant is characterised by a combination of several wavelengths in the electromagnetic spectrum. This combination is known as the Spectral Power Distribution (SPD) of the illumination source and also what gives the illuminant its colour. When light is emitted or radiated by an illuminant, its characteristic wavelengths travel from the illuminant and are incident on an object's surface. The object, whose surface possesses a reflectance spectrum referred to as the Surface Spectral Reflectance (SSR), reflects the incident wavelength spectrum of the illuminant modulated by its SSR. The now modulated electromagnetic radiation reaches the imaging device and is captured by it. The captured image is a representation of the spectral response of the sensors in the imaging device. The imaging device is made up of sensors characterised by their spectral sensitivities which dictates how much of the reflected spectra is captured. Figure 1 depicts the image formation process.

Adopting the dichromatic reflection model and under the Lambertian assumption, the formation of an image can be described as:

$$I_c = \int_{\Omega} E(\lambda) S(\lambda) Q_c(\lambda) d\lambda \tag{1}$$

where $E(\lambda)$ is the spectral power distribution of the light source, $S(\lambda)$ is the surface spectral reflectance of the pixel,

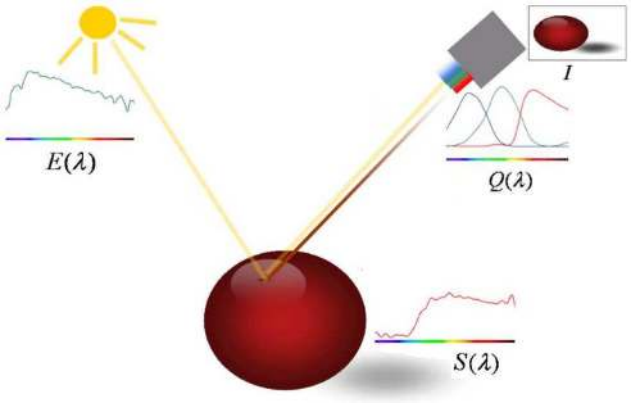


Figure 1: Image formation diagram

$Q_c(\lambda)$ is the camera sensor spectral response function, λ is the visible electromagnetic spectrum ranging from 400nm to 700nm and c represents the red, green and blue colour channels. $E(\lambda)$ and $S(\lambda)$ can be combined to form the scene spectral radiance, turning equation 1 into:

$$I_c = \int_{\Omega} R(\lambda) Q_c(\lambda) d\lambda \tag{2}$$

with $R(\lambda)$ being the spectral radiance.

The goal is to decompose an image into its constituent parts as shown in equation 2 so that the spectral radiance/signature can be recovered. However, the only known parts are the image pixel intensity values and the sensor spectral response, which makes this an under-determined problem. Due to the recent success by deep leaning models in image classification, object detection, object segmentation and depth estimation, this research adopts the use of deep learning to learn the mapping from RGB images to spectrum.

The problem is set as a supervised learning task, with a convolutional neural network learning to predict weights for a set of basis functions which are learned jointly with the weights from scratch. The final spectrum is then formed by taking a weighted combination of the basis functions using the learned weights (see figure 2 for the architecture of the proposed supervised learning method). To achieve this, the spectrum $R(\lambda)$ could be presented as:

$$R(\lambda) = \sum_{i=1}^{n} w_i \emptyset_i(\lambda) \tag{3}$$

where w_i is the weight, $\emptyset_i(\lambda)$ is the basis function, i is the basis function index and n is the number of basis functions. Previous research [12, 16, 11, 32] has shown that up to 10 basis functions are enough to accurately reproduce spectral signatures. Based on this, the number of basis functions automatically learned is set to 10.

3.2. Network architecture

A modified UNet [30] network is used with skip connections to allow lower level features to flow to deeper layers. The 2x2 pooling layers are replaced with linear downsampling layers. Four contracting steps are done as using five did not show any improvements. The cropping step before concatenation in the expansive path is replaced with a direct concatenation as cropping might dispose of edge information which could be useful for robust prediction, especially around the edges of the image. The model was trained using patches from the RGB image and corresponding ground truth spectral cubes. The RGB image and spectral cubes were resized to 512x512, and 64x64 patches were extracted deterministically (next to each other) and not randomly, which were used for training. The use of batch normalisation [17] was experimented with, however it was found to be detrimental to the network's performance. SELU [20] activation function was also tested, but similar to batch normalisation, it was found to be harmful to the network's performance. This was expected as SELU is self normalising. To further boost robustness and regularisation, dropout was experimented with, however, it produced inferior results. The basis functions are learned as a 10x31 matrix variable during training without going through any neural network layer. During inference, the saved trained matrix is simply loaded into memory and used. At test time, the full RGB image is passed through the CNN. The spectral cube is then generated as a weighted combination of the basis functions, using the predicted weights. Mean Relative Absolute Error (MRAE) was used as the loss function to facilitate training. It was selected to mitigate overly weighting higher value spectral errors over lower valued ones and enable a more consistent weighting. The MRAE was computed as shown in equation 4.

$$MRAE = \frac{\sum_{i,c} \frac{|R_{ic} - R^{gt}_{ic}|}{R^{gt}_{ic}}}{R_n} \tag{4}$$

where R_{ic} and R^{gt}_{ic} denote the i-th pixel and c spectral channel values for the recovered spectrum and ground truth spectrum respectively; and the total number of datapoints in the spectral cube is represented by R_n.

3.3. Experimental setup

The traning batch size was 128, the learning rate used was 1e-4 and the Adam optimiser [19] was used during training. The model was trained for 200 epochs and random horizontal and vertical flips were used for data augmentation, with a weight decay of 1e-5. The use of weight decay proved helpful to avoid overfitting. Weight decay was only used for the basis functions weights predicting CNN and not for learning the basis functions; the intuition is to allow the optimiser to find the most appropriate basis functions

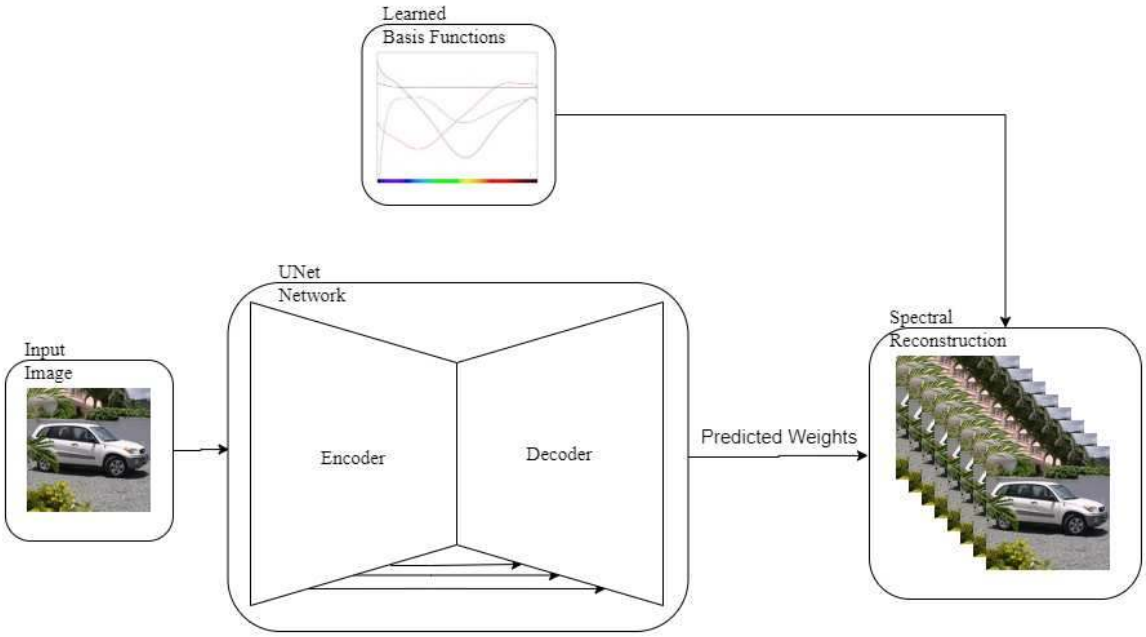


Figure 2: Architecture of proposed method using supervised learning

without the interaction of the additional weight decay on the loss function. The work was implemented in python using PyTorch [28] on a Linux machine. Training took roughly 2.7 hours on an Nvidia RTX 2080Ti and consumed about 5.8GB of VRAM. Inference uses 1473MB of VRAM, with single image inference time of 34.44ms on an Nvidia RTX 2080Ti. The runtime shows how efficient the proposed solution is as it can be deployed in real-time applications.

3.4. Unsupervised learning approach

An unsupervised learning method is also proposed, which can be used when there is no ground truth spectral signature data available. To tackle this case, two additional modules are added on to the supervised learning architecture which enables the model to be trained in an end-to-end fully unsupervised manner whilst still producing good spectral reconstruction from RGB images. The architecture for the unsupervised learning method is depicted in figure 3. It shows the extra modules that enable unsupervised training - "image formation" and "photometric reconstruction loss". The image formation module is responsible for reconstructing an image from the predicted spectral cube. To achieve this, a discretised version of equation 2 is used as shown in equation 5.

$$I_c = \sum_{\lambda} R(\lambda) Q_c(\lambda) \tag{5}$$

Substituting equation 3 into equation 5 produces the final formula (equation 6) for reconstructing the image.

$$I_c = \sum_{\lambda} (\sum_{i=1}^{n} w_i \emptyset_i(\lambda)) Q_c(\lambda) \tag{6}$$

with I_c being the reconstructed pixel value for the c-th colour channel, Q_c is the c-th camera sensor spectral response, w_i and $\emptyset_i$ are predicted weight for the i-th basis function and the i-th basis function respectively, n is the number of basis functions, λ the spectral wavelength and c the colour channels (R, G, B).

The photometric reconstruction loss module takes the original RGB image and the reconstructed RGB image as inputs. It then determines the error or difference between both inputs and uses that residual as a training signal in the absence of ground truth hyperspectral spectrum data. Intuitively, if there is no error between the two RGB images, the network must have predicted the correct spectrum as it is a precursor to the image formation, however, if there is a large error, then the network must have predicted an incorrect spectrum for the image. L1 loss was used as the image reconstruction loss function as it was empirically found to be sufficient for unsupervised learning.

4. Experimental results

To assess the performance of the proposed supervised learning method, the model is trained using the training strategy described in section 3.3. A baseline CNN-based RGB to spectral reconstruction model of the same CNN model architecture is trained using the same training strat-

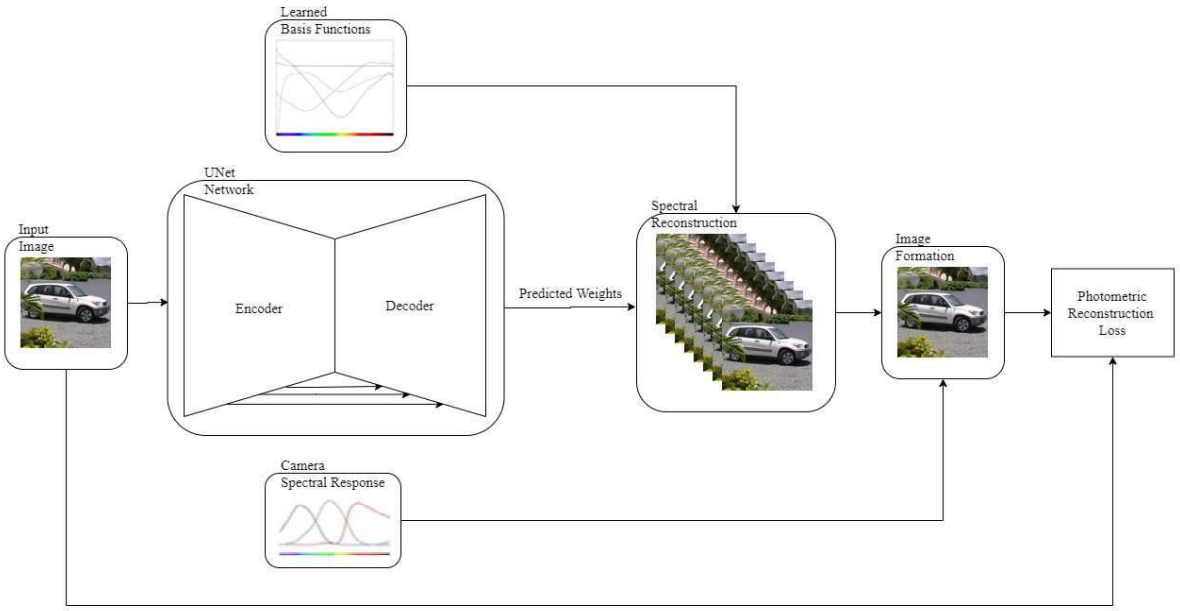


Figure 3: Architecture for unsupervised learning of spectral reconstruction

egy for fairness. The only difference between the two methods is that the baseline method directly outputs a 31-channel spectral image from the last layer whilst the proposed method uses the predicted weights and learned basis functions to create the spectral image before the loss function.

4.1. Dataset and evaluation premise

The dataset used for experimentation is the NTIRE 2020 Spectral Reconstruction Challenge [5] dataset. It is a hyperspectral image dataset with varying scenes and contents. It is made up of 450 hyperspectral cubes along with their corresponding RGB images for training. A further 10 RGB and hyperspectral image sets were also provided for validation purposes. As it is a challenge, another set of 20 RGB images were provided (the test set) which were split between two challenge tracks (10 each), however, their associated ground truth hyperspectral cubes were not provided. The challenge comprises of two tracks - Clean track and Real World track. The clean track involves hyperspectral data reconstruction from uncompressed 8-bit RGB images in *png* format. The images have been formed using the CIE 1964 colour matching function. On the other hand, the real world track involves recovering hyperspectral data from RGB images which have gone through mosaicing, noise injection, demosaicing and *jpeg* compression. This makes it a tougher problem to solve. The evaluation in this work follows the same approach and assesses the performance of the proposed method against the baseline using the validation data from the Clean and Real World tracks of the NTIRE 2020 Spectral Reconstruction Challenge. All models are trained on the challenge training dataset.

4.2. Evaluation metrics

Quantitative evaluation of the performance of the proposed method is done using the MRAE and root mean square error (RMSE) metrics. The MRAE metric has already been introduced in section 3.2 (equation 4). RMSE is an error metric that squares the residuals, takes and averages and finally a root of the result. it is written as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(R_i - R_i^{gt})^2} \tag{7}$$

where n is the total number of datapoints, R_i^{gt} and R_i are the i-th pixel of the ground truth and reconstructed hyperspectral cubes respectively.

4.3. Results and discussion

The results of the quantitative evaluation of the baseline and proposed method for both the clean and real world tracks are shown in table 1. The figures show that the proposed supervised learning method outperforms the baseline methods on both MRAE and RMSE metrics on both tracks. It produced a 0.0014 and 0.0008 reduction in MRAE and RMSE respectively in the clean track; and 0.0018 and 0.0009 reduction in MRAE and RMSE respectively in the real world track. Despite the modest improvements, it should be noted that the proposed method achieves it by predicting fewer parameters (10 vs 31). The performance on

Table 1: MRAE and RMSE results on the validation dataset

	Baseline - Clean		Proposed - Clean		Baseline - Real World		Proposed - Real World	
	MRAE	RMSE	MRAE	RMSE	MRAE	RMSE	MRAE	RMSE
ARAD_HS_0453	0.0469	0.0164	0.0518	0.0175	0.0677	0.0180	0.0659	0.0214
ARAD_HS_0465	0.0435	0.0050	0.0473	0.0059	0.0807	0.0081	0.0815	0.0081
ARAD_HS_0463	0.0716	0.0378	0.0646	0.0332	0.0858	0.0388	0.0810	0.0357
ARAD_HS_0455	0.0703	0.0337	0.0755	0.0366	0.1022	0.0364	0.0980	0.0320
ARAD_HS_0462	0.0542	0.0105	0.0439	0.0078	0.0889	0.0119	0.0850	0.0106
ARAD_HS_0457	0.0450	0.0083	0.0400	0.0065	0.0804	0.0101	0.0825	0.0101
ARAD_HS_0451	0.0386	0.0145	0.0325	0.0119	0.0764	0.0288	0.0689	0.0252
ARAD_HS_0456	0.0264	0.0104	0.0246	0.0104	0.0419	0.0132	0.0421	0.0130
ARAD_HS_0459	0.0209	0.0067	0.0225	0.0067	0.0440	0.0111	0.0460	0.0112
ARAD_HS_0464	0.0302	0.0158	0.0309	0.0147	0.0523	0.0209	0.0508	0.0203
Average	0.0448	0.0159	0.0434	0.0151	0.0720	0.0197	0.0702	0.0188

the clean dataset is much better than the performance on the real world dataset. This is expected as the real world model has to learn noise removal and *jpg* compression levels in additional to its ultimate task of spectral reconstruction. Looking at the metrics on individual images, it can be seen that on both tracks the baseline methods performs better than the proposed method on some images, however, the proposed method outperforms the baseline on more images and therefore on overall performance. The reconstructed spectra for some pixels are shown in figure 4. It can be seen that the proposed method, especially on the clean track matches the ground truth spectra better than the baseline method. This goes to show the fidelity of the reconstruction capablity of the proposed method. The models appear to perform best in the blue spectral range as well as the range between green and red. They exhibited less reconstruction accuracy in the high red spectral range, perhaps suggesting the availability of fewer ground truth spectra in that range or a deficiency in the CNN model architecture employed. Despite this, the proposed model appears to cope better with the issue as can be seen in figure 4 (d, e, f). This display of robustness can be attributed to the problem formulation, which does not learn individual spectrum band details but rather a global set of basis functions with weights to form the spectrum. Additionally, the proposed method is advantageous because it is able to reconstruct the spectral cube using fewer parameters than would normally be required (i.e. predicting 10 weights per pixel instead of 31, a 67.74% reduction in predicted output). This becomes even more significant when predicting 301 spectral bands (96.68% reduction in this case).

The trained models for both tracks were submitted to the NTIRE 2020 Spectral Reconstruction Challenge [5]. They achieved RMAEs of 0.0440 and 0.0714 on the clean and real world tracks respectively.

4.4. Unsupervised learning method results

The same CNN model architecture used for the supervised learning method is employed for the unsupervised learning strategy. The reconstructed spectral cube is passed to the image formation module which creates an image from it and then passes the recreated image to the image reconstruction loss module where the photometric error between the reconstructed image and the original image is used for backpropagation and thus, learning.

Table 2 provides figures of the evaluation metrics used to assess the model's performance. It is clear when comparing table 2 against table 1 that the supervised learning method outperforms the unsupervised learning method. This is somewhat expected as the unsupervised model did not interact with any ground truth spectrum during training. What the table shows is that the model did in fact learn to reconstruct spectral images from RGB images as MRAE and RMSE are 0.1056 and 0.0275 respectively. These are relatively low numbers for a model that has not learned the task, especially when none of these error metrics was used as the training loss function. Qualitatively, figure 5 indicates that the model does a good job at reconstructing spectra. Figure 5 (a and b) show good reconstruction quality, with the reconstruction spectra appearing very similar to the ground truth and showing competitive results against the proposed supervised learning method. However, figure 5 (c) shows an example of a poor reconstruction. While the

Table 2: Clean track validation dataset result for the unsupervised learning model

MRAE	RMSE
0.1056	0.0275

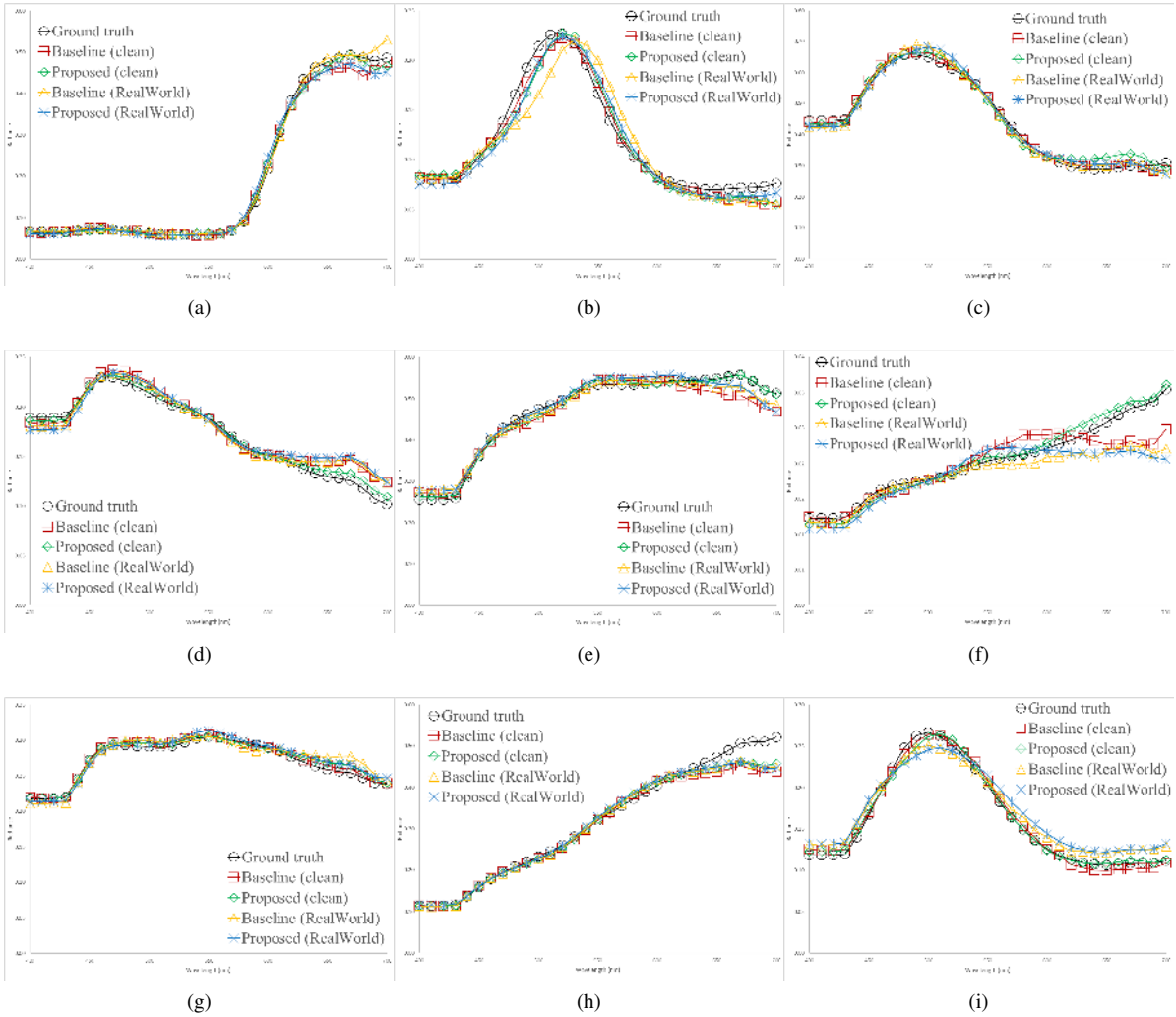


Figure 4: Comparison between ground truth and reconstructed spectra for some randomly selected pixels

shapes are somewhat similar, the magnitudes differ by up to 0.2 and the unsupervised reconstructed spectrum has negative points which are non existent in the ground truth. The worse performance of the unsupervised method can be attributed to issues like this. Despite that, it is believed that training with more data and better CNN model architectures can make unsupervised learning a viable methods for spectral reconstruction.

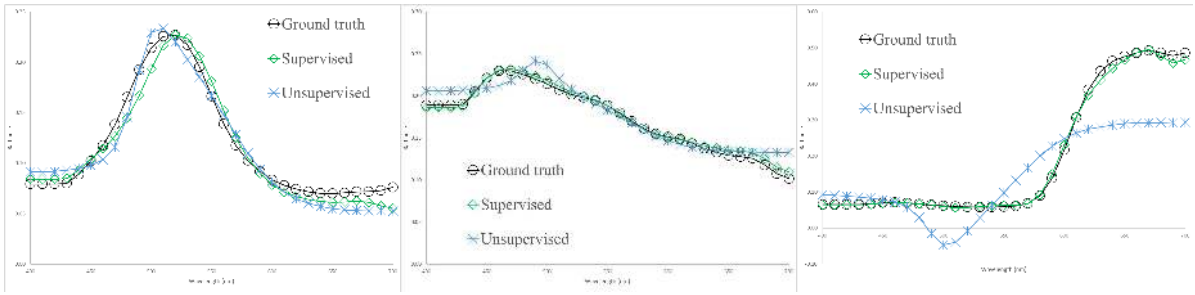
4.5. Significance of basis functions for learning

To further assess the performance of the proposed supervised learning method and show the importance of the use of learned basis functions, three additional models were trained for both the baseline and proposed methods using reduced contraction layers in the UNet backbone architecture. Contraction layers of 3, 2 and 1 were used. Table 3 shows the results of the additional training and indicates that the proposed method which uses learned basis functions performs better even as the CNN architecture becomes shallower, thereby indicating the importance of the learned basis functions in the proposed method.

Table 3: Results for different numbers of contraction layers

Contraction layers	MRAE	
	Baseline	Proposed
4	0.0448	0.0434
3	0.0465	0.0462
2	0.0485	0.0470
1	0.0586	0.0539

To check the significance of the basis functions for learning in the unsupervised learning context, another model us-



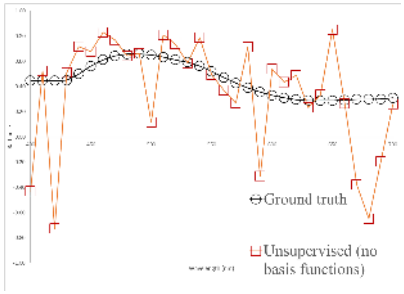
(a) (b) (c)

Figure 5: Output spectra of selected pixels using the proposed unsupervised learning method

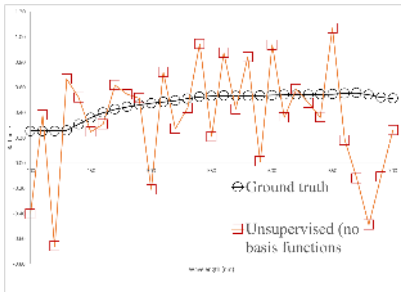
ing the same unsupervised learning strategy was trained but without the use of basis functions (essentially the baseline model with the unsupervised learning image formation and loss attached to it). The idea was to check if the basis functions contributed to the stability, accuracy and learning capability of the unsupervised learning model. Table 4 and figure 6 indicate that the answer to that question is a yes. The table shows a rise of 1.1041 and 0.2263 in MRAE and RMSE respectively when compared with the unsupervised learning method that uses learned basis functions in table 2. Figure 6 shows that the model does not learn how to reconstruct spectra. The reconstructed spectra are very dissimilar to the ground truth spectra and probably does not form the liking of any known spectral signature. This once again highlights the importance of the basis functions for unsupervised/self supervised learning of spectra as it provides a global strategy to learning and removes the 3-to-many mapping problem, essentially providing global constraints on the model and leading the optimiser down the right path.

Table 4: Clean track validation dataset result for the unsupervised learning model without basis functions

MRAE	RMSE
1.2097	0.2538



(a)



(b)

Figure 6: Output spectra from unsupervised learning without the use of basis functions

5. Conclusion

This paper presents novel methods for recovering hyperspectral image cubes from RGB images. The novelty comes from the use of basis function in a neural network environment and the unique method of training a CNN to predict weights and learn a set of basis functions jointly. The strategy is implemented using a modified UNet CNN model architecture and trained in a supervised manner with the NTIRE 2020 Spectral Reconstruction Challenge dataset. The model is compared with a baseline model of the same architecture without the use of basis functions and found to outperform the baseline model whilst requiring fewer predicted outputs. A further innovation is introduced in the form of a method for learning the mapping from RGB data to spectra in a completely unsupervised manner. The model can be trained end-to-end and leverages the physics of image formation and photometric reconstruction error as enablers. The importance of the basis functions in this context is explored and shown to provide significant boost in training stability and learning ability for the unsupervised learning model and also enhances the robustness of the supervised learning model. This goes to show the significance of formulating the RGB to spectrum mapping in the way introduced in this paper.

References

[1] Farnaz Agahian, Seyed Ali Amirshahi, and Seyed Hossein Amirshahi. Reconstruction of reflectance spectra using weighted principal component analysis. *Color Research & Application*, 33(5):360–371, oct 2008. 1, 2

[2] Aitor Alvarez-Gila, Joost Van De Weijer, and Estibaliz Garrote. Adversarial Networks for Spatial Context-Aware Spectral Image Reconstruction from RGB. In *Proceedings - 2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017*, volume 2018-Janua, pages 480–490, 2017. 2

[3] Boaz Arad and Ohad Ben-Shahar. Sparse recovery of hyperspectral signal from natural RGB images. In *European Conference on Computer Vision*, volume 9911 LNCS. Springer Verlag, 2016. 1, 2

[4] Boaz Arad, Ohad Ben-Shahar, Radu Timofte, Luc Van Gool, Lei Zhang, and Ming Hsuan Yang. NTIRE 2018 challenge on spectral reconstruction from RGB images. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2018-June:1042–1051, 2018. 1

[5] Boaz Arad, Radu Timofte, Ohad Ben-Shahar, Yi-Tun Lin, Graham Finlayson, and Others. NTIRE 2020 Challenge on Spectral Reconstruction from an RGB Image. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, jun 2020. 5, 6

[6] José M. Bioucas-Dias, Antonio Plaza, Gustavo Camps-Valls, Paul Scheunders, Nasser M. Nasrabadi, and Jocelyn Chanussot. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geoscience and Remote Sensing Magazine*, 1(2):6–36, 2013. 1

[7] M Borengasser, WS Hungate, and R Watkins. *Hyperspectral remote sensing: principles and applications*. CRC press, Boca Raton, 2007. 1

[8] Mihaela Antonina Calin, Sorin Viorel Parasca, Dan Savastru, and Dragos Manea. Hyperspectral imaging in the medical field: Present and future. *Applied Spectroscopy Reviews*, 49(6):435–447, aug 2014. 1

[9] Yigit Baran Can and Radu Timofte. An efficient CNN for spectral reconstruction from RGB images. *arXiv preprint arXiv:1804.04647*, 2018. 2

[10] Jean Claude Iehl and Bernard Péroche. An Adaptive Spectral Rendering with a Perceptual Control. *Computer Graphics Forum*, 19(3):291–300, sep 2000. 1

[11] D Connah, S Westland, and M G A Thomson. Recovering spectral information using digital camera systems, 2001. 2, 3

[12] Jae Kwon Eem, Hyun Duk Shin, and Seung Ok Park. Reconstruction of surface spectral reflectances using characteristic vectors of Munsell colors. In *Color and Imaging Conference*, volume 5, pages 127–131, 1994. 2, 3

[13] Ying Fu, Yongrong Zheng, Lin Zhang, and Hua Huang. Spectral Reflectance Recovery From a Single RGB Image. *IEEE Transactions on Computational Imaging*, 4(3):382–394, jul 2018. 1, 2

[14] Silvano Galliani, Charis Lanaras, Dimitrios Marmanis, Emmanuel Baltsavias, and Konrad Schindler. Learned Spectral Super-Resolution. *CoRR*, abs/1703.0, mar 2017. 1

[15] Liang Gao, Robert T. Kester, Nathan Hagen, and Tomasz S. Tkaczyk. Snapshot Image Mapping Spectrometer (IMS) with high sampling density for hyperspectral microscopy. *Optics Express*, 18(14):14330, jul 2010. 1

[16] Ron Gershon, Allan D Jepson, John K Tsotsos, and Toronto Toronto. From [R,G,B] to surface reflectance: computing color constant descriptors in images. In *IJCAI'87: Proceedings of the 10th international joint conference on Artificial intelligence*, pages 755–758, 1987. 2, 3

[17] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *32nd International Conference on Machine Learning, ICML 2015*, volume 1, pages 448–456, 2015. 3

[18] Yan Jia, Yinqiang Zheng, Lin Gu, Art Subpa-Asa, Antony Lam, Yoichi Sato, and Imari Sato. From RGB to Spectrum for Natural Scenes via Manifold-Based Mapping. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2017-Octob, pages 4715–4723, 2017. 2

[19] Diederik P Kingma and Jimmy Lei Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015. 3

[20] Günter Klambauer, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter. Self-Normalizing Neural Networks. In *Advances in neural information processing systems*, pages 971–980, 2017. 3

[21] Steven Lansel, Manu Parmar, and Brian A Wandell. Dictionaries for sparse representation and recovery of reflectances. In *Computational Imaging VII*, volume 7246, page 72460D, 2009. 1, 2

[22] Yi-Tun Lin and Graham D Finlayson. Physically Plausible Spectral Reconstruction from RGB Images. *ArXiv*, abs/2001.0, 2020. 2

[23] Yonghong Long, Haixia Luo, and Aichun Yi. Color Recovering Based on Dichromatic Reflection Model and Finite Dimensional Linear Model. In *2009 International Conference on Measuring Technology and Mechatronics Automation*, pages 441–444. IEEE, 2009. 2

[24] S. Mahesh, D. S. Jayas, J. Paliwal, and N. D.G. White. Hyperspectral imaging to classify and monitor quality of agricultural materials, mar 2015. 1

[25] Shaohui Mei, Xin Yuan, Jingyu Ji, Yifan Zhang, Shuai Wan, and Qian Du. Hyperspectral Image Spatial Super-Resolution via 3D Full Convolutional Neural Network. *Remote Sensing*, 9(11):1139, nov 2017. 1

[26] Rang M H Nguyen, Dilip K Prasad, and Michael S Brown. Training-Based Spectral Reconstruction from a Single RGB Image. In *European Conference on Computer Vision*, 2014. 1, 2

[27] Manu Parmar, Steven Lansel, and Brian A. Wandell. Spatio-spectral reconstruction of the multispectral datacube using sparse recovery. In *Proceedings - International Conference on Image Processing, ICIP*, pages 473–476, 2008. 1, 2

[28] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison,

Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems*, 2019. 4

[29] Lankapalli Ravikanth, Digvir S. Jayas, Noel D.G. White, Paul G. Fields, and Da Wen Sun. Extraction of Spectral Information from Hyperspectral Data and Application of Hyperspectral Imaging for Food and Agricultural Products. *Food and Bioprocess Technology*, 10(1):1–33, jan 2017. 1

[30] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer Verlag, 2015. 3

[31] James F. Scholl, E. Keith Hege, Michael Hart, Daniel O'Connell, and Eustace L. Dereniak. Flash hyperspectral imaging of non-stellar astronomical objects. In Mark S. Schmalz, Gerhard X. Ritter, Junior Barrera, and Jaakko T. Astola, editors, *Mathematics of Data/Image Pattern Recognition, Compression, and Encryption with Applications XI*, volume 7075, page 70750H. SPIE, aug 2008. 1

[32] Mohamed Sedky, Mansour Moniri, and Claude C. Chibelushi. Spectral-360: A physics-based technique for change detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 405–408. IEEE, jun 2014. 2, 3

[33] Zhan Shi, Chang Chen, Zhiwei Xiong, Dong Liu, and Feng Wu. HSCNN+: Advanced CNN-based hyperspectral recovery from RGB images. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, volume 2018-June, pages 1052–1060, 2018. 1, 2

[34] Tarek Stiebei, Simon Koppers, Philipp Seltsam, and Dorit Merhof. Reconstructing spectral images from RGB-images using a convolutional neural network. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, volume 2018-June, pages 1061–1066, 2018. 2

[35] Shoji Tominaga. Surface reflectance estimation by the dichromatic model. *Color Research & Application*, 21(2):104–114, 1996. 2

[36] Ashwin A. Wagadarikar, Nikos P. Pitsianis, Xiaobai Sun, and David J. Brady. Video rate spectral imaging using a coded aperture snapshot spectral imager. *Optics Express*, 17(8):6368, apr 2009. 1

[37] Jiqing Wu, Jonas Aeschbacher, and Radu Timofte. In Defense of Shallow Learned Spectral Reconstruction from RGB Images. In *Proceedings - 2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017*, volume 2018-Janua, pages 471–479. Institute of Electrical and Electronics Engineers Inc., jul 2017. 1

[38] Zhiwei Xiong, Zhan Shi, Huiqun Li, Lizhi Wang, Dong Liu, and Feng Wu. HSCNN: CNN-Based Hyperspectral Image Recovery from Spectrally Undersampled Projections. In *Proceedings - 2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017*, volume 2018-Janua, pages 518–525, 2017. 2

[39] Yiqi Yan, Lei Zhang, Jun Li, Wei Wei, and Yanning Zhang. Accurate spectral super-resolution from single RGB image using multi-scale CNN. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 11257 LNCS, pages 206–217, 2018. 1, 2

[40] Naoto Yokoya, Claas Grohnfeldt, and Jocelyn Chanussot. Hyperspectral and multispectral data fusion: A comparative review of the recent literature, jun 2017. 1

[41] Lei Zhang, Wei Wei, Chengcheng Bai, Yifan Gao, and Yanning Zhang. Exploiting Clustering Manifold Structure for Hyperspectral Imagery Super-Resolution. *IEEE Transactions on Image Processing*, 27(12):5979–5982, dec 2018. 1