



**QUEEN'S
UNIVERSITY
BELFAST**

RGB-2-hyper-spectral image reconstruction for food science using encoder/decoder neural architectures

Williamson, R., Martinez Del Rincon, J., Koidis, A., & Reaño, C. (2023). RGB-2-hyper-spectral image reconstruction for food science using encoder/decoder neural architectures. In *IEEE Symposium on Computers and Communications (ISCC 2023): Proceedings* (pp. 872-875). (IEEE Symposium on Computers and Communications: Proceedings). Institute of Electrical and Electronics Engineers Inc..
<https://doi.org/10.1109/ISCC58397.2023.10218310>

Published in:

IEEE Symposium on Computers and Communications (ISCC 2023): Proceedings

Document Version:

Peer reviewed version

Queen's University Belfast - Research Portal:

[Link to publication record in Queen's University Belfast Research Portal](#)

Publisher rights

Copyright 2023 IEEE.

This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

General rights

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Open Access

This research has been made openly available by Queen's academics and its Open Research team. We would love to hear how access to this research benefits you. – Share your feedback with us: <http://go.qub.ac.uk/oa-feedback>

RGB-2-Hyper-Spectral Image Reconstruction for Food Science Using Encoder/Decoder Neural Architectures

Robert Williamson, Jesus Martinez del Rincon and Anastasios Koidis
Queen's University Belfast, UK
{rwilliamson15, j.martinez-del-rincon, t.koidis}@qub.ac.uk

Carlos Reaño
Universitat de València, Spain
carlos.reano@uv.es

Abstract—Hyper-spectral imaging captures spatial and spectral information of a subject. This is used for the identification of substances within a scene, and food analysis. Presented is an investigation into the capabilities of encoder/decoder deep learning architectures for hyper-spectral image reconstruction from RGB images. For this analysis state-of-the-art (SOTA) techniques for hyper-spectral image reconstruction and other architectures from different fields have been used. Our approach examines a food science case study, using a CPU-based server and different accelerators. An in-house multi-sensor setup was used to capture the dataset which contains hyper-spectral images of twenty slices of different Spanish Ham in the range of 400-1000 nm and their analogous RGB images. The results show no degradation in the output when moving outside of the visual range. This study shows that the SOTA methods for reconstructing from RGB do not produce the most accurate reconstruction of the spectral domain within the range of 400-1000 nm.

Index Terms—Hyper-spectral image reconstruction; RGB; deep learning; GPU accelerators;

I. INTRODUCTION

Hyper-spectral imaging (HSI) is an emerging field which aims to combine the advantages of optical spectroscopy and those of two-dimensional object visualization. The light which hits each pixel is broken down into multiple different spectral bands to provide more information about the image. The spectral information gathered is added to the spatial information of the two-dimensional image to create a three-dimensional data cube [21] [23]. This spectral information aims to facilitate finding objects, identifying materials, or detecting processes and is currently used in multiple applications such as food analysis [12], autonomous driving [11], and mapping [9], to name a few. While HSI has been proven valuable, the equipment used to acquire this HSI information is typically difficult to operate and is accompanied by an expensive price tag for the sensor.

To mitigate the previous disadvantages, there have been a number of proposed methods for reconstructing hyper-spectral images from a corresponding RGB image [29]. These methods allow for the exclusion of the spectral camera equipment needed to capture this data. This allows for an alternative to the traditional method which would enable the promotion of the imaging technique within fields such as food inspection and quality assurance, which are incapable of adapting this

technology due to the extremely high cost of entry. Hyper-spectral imagery has been reconstructed previously [3]. However, the majority of existing work in hyper-spectral imagery reconstruction is carried out within the visual range. This is due to the spectral representation within an RGB only covering this visual range of 400-700nm [1]. Most existing approaches to HSI reconstructions build on the advances of deep learning (DL), particularly on encoder/decoder convolutional neural network (CNN) [4].

The contributions of this paper are:

- A selection and adaptation of state-of-the-art (SOTA) encoder/decoder neural architectures for HSI reconstruction.
- Exhaustive comparison and evaluation of 10 neural architectures for a challenging HSI reconstruction outside of the visual range.
- Development of an RGB-2-HSI reconstruction model for Spanish Ham analysis.
- Evaluation on a CPU-based server and a selection of different accelerators (i.e. edge devices and GPU-based servers) to identify the capabilities of each model when deployed in a variety of environments.

II. RELATED WORK

Existing methods for spectral reconstruction can be split into prior-based and data-driven methods [29]. Prior-based methods are those that narrow down the solution space through the use of prior knowledge as constraints. However, these methods suffer from poor spatial structure similarity and correlation between spectra. On the contrary, data-driven methods rely on recent advances in various neural network architectures, which take advantage of abstract automatically-derived features, and the availability of increasingly large-scale data sets. Zhang et al. [29] found that on a number of open-source data sets, data-driven methods constantly outperformed prior-based methods. The data-driven methods can be broken down into the following categories, according to the neural network backbone architecture: CNN [8], U-Net [19], GAN [20], Dense Network [15], Attention Network [10], and Multi-branch Network [28].

In [29], the authors performed a systematic review of more than twenty-five SOTA spectral reconstruction methods. The methods were compared using 5 different data sets, comprising

both indoor and outdoor scenarios with natural and controlled lighting. All of these data sets cover the range of 400-700 nm. It was found that the models that were categorised as Attention Networks performed the best due to their increase in network complexity over the other model types. However, the data sets used are not a great representation of how hyper-spectral data is normally presented, as there is a lack of spectral bands compared to the number seen in a hyper-spectral scene [18], and the target reconstruction does not stretch further than the visual range. In this regard, Zhang et al. [29] state that it may be challenging for Attention Networks models to perform greater amounts of spectral up-sampling and maintain this high accuracy reported.

Further extensions and complex models such as transformer-based networks like MST++ [4] and attention-based networks like HRNet [30] can potentially be utilised to perform similar or better reconstruction outside the visual range, although this has not been performed in the literature. Similarly, other related U-net [19], attention [10] and transformer networks [14] existing in related disciplines (deblurring, super-resolution, etc) have the potential to perform highly in our out-of-visual-range HSI reconstructions. While these methods have not been explored in this task, we aim to validate their performance and compare them against HSI-specific as part of this study.

III. RESEARCH METHODOLOGY

A. Ham Dataset

In order to validate our approach in a real food analysis test case, a data set of Spanish ham images was recorded. This data set consists of 20 images containing ham, which were all captured under the same illumination scenario and on the same background. The background was specially chosen to give a consistent illumination and reflectance level for the entire image. Two cameras, one RGB and one hyper-spectral camera were used. For the hyper-spectral sensor, the spatial resolution of these images is 780 x 424 pixels and the spectral resolution is 5 nm from 400-1000 nm giving 121 spectral bands. An integration time of 15.0 ms was used with an acquisition distance of 550 mm. Corresponding RGB images have been generated from the hyper-spectral images with a resolution of 780 x 424 pixels.

B. Encoder/Decoder Deep Learning Techniques

The following SOTA models will be analyzed: HSCNN+ [22], HRNet [30], EDSR [17], HDNet [13], HINet [7], MIRNet [26], Restormer [25], MPRNet [27], MST [3], and MST++ [4]. These models encompass a variety of different structures and techniques, including U-Net-based models, transformer-based solutions and a collection of models which incorporate a residual approach.

1) *Hyper-spectral Reconstruction*: HSCNN+ [22] HSCNN-D replaces the residual block which was implemented within HSCNN-R with dense ones, which utilise a path-widening fusion scheme that helps in the handling of the gradient issues seen in HSCNN-R. As such, HSCNN-D is capable of learning a better map relation than the other implementations

of HSCNN, making this version our choice in the comparison due to it taking feature fusion as a concatenation block.

HRNet [30] is a network which consists of four hierarchical stages, with each stage being convolutional layers, residual dense blocks and residual global blocks.

HDNET [13] is a high-resolution dual-domain learning network. This network utilises a spatial-spectral attention module and frequency-domain learning by implementing a supervisor module.

MPRNet [27], Multi-Stage progressive image restoration network is a multi-stage architecture which involves information exchange between the different stages from early to late stages along with lateral connections between the feature processing blocks to avoid loss of information. A supervised attention module is used between every two stages of the network and learns to refine the features of one stage before they are passed to the next stage.

MST [3] or Mask-guided spectral-wise transformer which is a novel framework that utilises a CASSI system and spectral-wise multi-head self-attention (S-MSA) modules. A mask-guided mechanism (MM) is used to direct the attention of S-MSA to the spatial regions of the scene which have a high-fidelity spectral representation. A U-Shaped structure is followed, which incorporates an encoder, bottleneck and decoder.

MST++ [4] is a multi-stage spectral-wise transformer, which is constructed of single-stage spectral-wise transformers (SSTs). The U-Shaped SST consists of an encoder, a bottleneck and a decoder. These sections of the SST are constructed from Spectral-wise attention blocks (SABs), which consist of a Feed Forward Network and a Spectral-wise multi-head self-attention (S-MSA) module.

2) *Super Resolution and Denoising*: HINet [7] or Half Instance Normalization Network consists of two sub-networks, with each being based on U-Net. A convolutional layer is used within each stage to extract the initial features, which are then passed down the encoder/decoder structure of the U-Net which performs down-sampling and up-sampling. The sub-networks are connected through the use of a cross-stage feature fusion (CSFF) module and a supervised attention module (SAM).

EDSR [17] builds upon the results seen by Ledig et al.'s [16] SRResNet by employing a better ResNet structure. This was done by removing the batch normalization layers from the network.

MIRNet [26] is a network built from the core approach of a multi-scale residual block (MRB). MRB consists of parallel multi-resolution convolution streams, which extract multi-scale features and exchange information across these streams. The full-convolutional streams which are incorporated within MRB consolidate the high-resolution features with the help of the low-resolution features, and vice versa. The selective kernel feature fusion (SKFF) which is used, performs adjustments through two operations, fuse and select.

Restormer [25] or Restoration Transformer, is an efficient Transformer model built upon multi-head attention and feed-forward network blocks. The transformer blocks which make

up this model consist of multi-Dconv head transposed attention (MDTA) and gated-Dconv feed-forward network (GDFN).

IV. EXPERIMENTAL EVALUATION

This section describes the performed experiments and discusses the results obtained. First, we detail the experimental setup, next we describe the evaluation methodology and then we present and discuss the results.

A. Experimental Setup

For training the models, we used a GPU server with two 24-core Intel Xeon Platinum 8168 CPUs, 512GB of RAM, and an NVIDIA Tesla V100 GPU. For inference, apart from the GPU server used for training, we also used:

- Another GPU server with an AMD FX-8350 CPU, 16GB of RAM, and an NVIDIA GTX 1050ti GPU with 4GB of VRAM.
- A CPU server with an AMD EPYC 7702 dual 64-core (i.e. 128 core in total) processor, 786GB of RAM, and without GPU.
- An edge device Jetson Nano with a 4-core ARM A57 CPU, a 128-core Maxwell GPU, and 4GB RAM shared by CPU and the GPU.

B. Evaluation Methodology

We compare the methodologies following standard evaluation metrics such as Peak Signal to Noise Ratio (PSNR) [24] and Root Mean Square Error (RMSE) [6]. RMSE, MSE and PSNR are defined in Equation 1, Equation 2 and Equation 3, respectively.

$$RMSE = \sqrt{\left(\frac{1}{n}\right) \sum_{i=1}^n (y_i - x_i)^2} \quad (1)$$

$$MSE = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n (x_{ij} - y_{ij})^2 \quad (2)$$

$$PSNR(x, y) = 10 \log_{10} \left(\frac{R^2}{MSE} \right) \quad (3)$$

In Equation 2, m is the number of rows in the prediction image, n is the number of columns in the prediction image, x_{ij} is the pixel value from the prediction image and y_{ij} is the pixel value from ground truth image. In Equation 3, R is the maximum pixel value of the images and MSE is Mean Squared Error as calculated in Equation 2.

The Spanish ham data set (see Section III) was split into 16 testing and 4 validation images. Each method was trained for 100 epochs with an initial learning rate of 0.00004 and the Cosine Annealing scheme [5]. The loss function of Mean Relative Absolute Error is used between the predicted hyper-spectral image and the ground truth. The results of the methods prediction for the four validation images can be seen in Table I. In the next sections, we discussed these results.

In addition to the Spanish ham data set, we also validate all methods in a standard data set ARAD1K [2] to validate

the correct functioning and performance for RGB to hyper-spectral reconstruction within the visual range of 400-700 nm. The setup for this experiment was the same as the previous one carried out for the Spanish Ham data set. The results of this experiment can be also seen in Table I, this gives context to the results of the Spanish Ham data set experiment by creating a baseline.

TABLE I
RMSE AND PSNR OF ALL MODELS, TESTED IN THE ARAD1K AND SPANISH HAM DATASETS

Models	Results			
	ARAD1k		Spanish Ham	
	RMSE	PSNR	RMSE	PSNR
MST++	0.024760	18.96000	0.031600	30.309150
MST	0.025500	19.03200	0.032200	30.098575
HSCNN+	0.058800	26.36200	0.030250	28.850250
MPRNET	0.027000	33.49600	0.028000	31.348475
HDNet	0.031700	32.12950	0.028475	31.091650
Restormer	0.027400	33.40270	0.033800	29.666400
MIRNET	0.027360	33.29350	0.031100	30.499300
HINet	0.030346	32.50730	0.036325	28.934750
EDSR	0.043725	28.29130	0.028575	31.084375
HRNet	0.054950	26.89298	0.033425	29.827650

C. Spectral Analysis

As we step further from the visual range towards the near-infrared (NIR) range, we expect the spectral reconstruction to deviate from the ground truth further due to the lack of representation within the initial RGB image. However, once we look at the PSNR of the spectral bands outside the visual range and their RMSE as well, we can see there is no great difference between these spectral bands and those within the visual range, indicating that all methods successfully predict outside the visual range. There is no great change in the RMSE when moving outside the visual range, this can be seen in Fig. 1.

D. Comparison of Techniques

From the results shown in Table I, we can see that MPRNET [27] on average performs the best when recreating a hyper-spectral image of 121 bands from 400 - 1000 nm from an RGB image. It is also the best in the visual range, as per ARAD1k dataset, as indicated by PSNR and similar RMSE to MST++. The results seen on the ARAD1k match that presented [2] Based on the measurement of RMSE, on average the worst-performing method is HINet [7]. MPRNET [27] is not the most complex of the models put forward in this review. The notable difference between MPRNET [27] and other models is the nested inclusion of U-Net models within the overall network. This is in comparison to other models such as MST and MST++ which follow a U-Net-based structure. With respect to PSNR, on average HSCNN+ [22] was the worst performer, despite being the only model presented which has been tested outside of the visual range. This can be explained due to the simplicity of the model in comparison to the other models in the review. HINet [7] on average has the highest RMSE on the validation set. This is unusual as HINet [7]

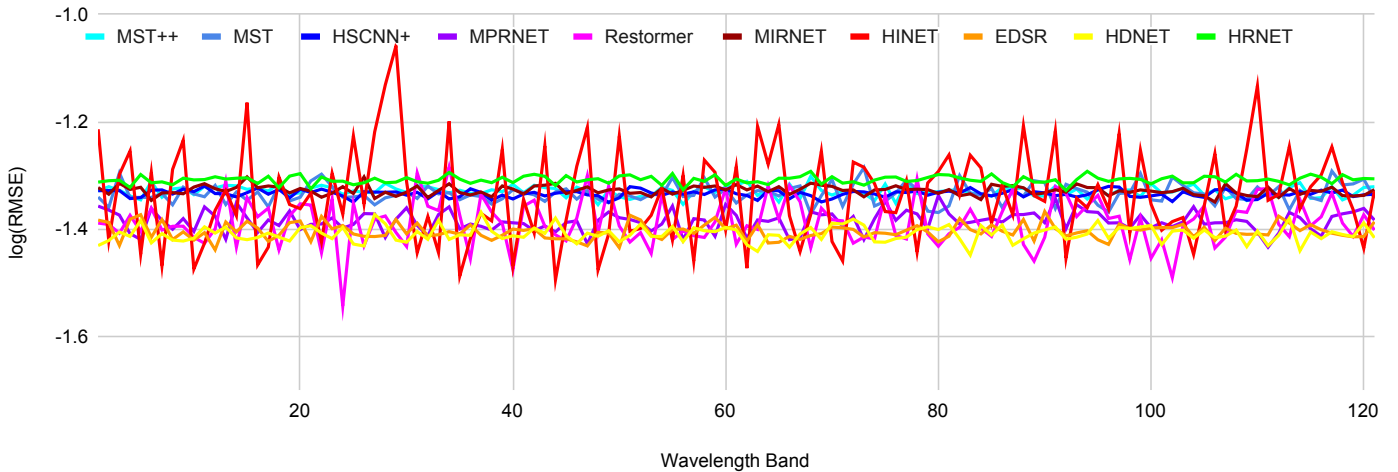


Fig. 1. RMSE of each model over wavelength band on the Spanish ham Dataset.

has a parameter count of 79 million, the model was originally created to boost performances on low-level image restoration tasks such as de-noising. De-noising data sets are easy to create through the addition of artificial noise. This means large data sets are easy to create which leads to large inefficient models being created. Whereas, this application has a limited data set due to the nature of the application.

E. Comparison of Accelerators

As commented in Section IV-A, inference was executed on a selection of different accelerators to identify the capabilities of each model when deployed in a variety of environments and to be deployed in an embedded device. The execution time of the inference on these hardware types can be seen in Table II.

Depending upon the use case of the model there may be limitations or specific qualities needed from the model, e.g. faster inference time and lower memory utilization. As expected, an edge device like the Jetson Nano needs more time than the rest. The two GPU servers perform similarly, while the CPU server is the one obtaining the best results in our experiments. Note that contrary to what usually happens on training, CPU performs better than GPU on inference, because the overhead of moving data to/from the GPU is not compensated by the reduction in computing time.

TABLE II
EXECUTION TIME ON THE VALIDATION IMAGES

Models	Execution Time in Seconds				No. Params (M)
	Nano	1050ti	V100	CPU	
MST++	13.93	2.44	1.84	1.45	24
MST	14.83	2.69	1.99	2.16	35
HSCNN+	11.34	1.84	1.72	0.57	4.7
MPRNet	12.61	2.61	2.10	2.40	55
HDNet	11.19	1.82	1.68	0.69	2.7
Restormer	12.10	2.13	2.18	1.05	151
MIRNET	16.67	4.70	3.96	2.78	57
HINet	17.25	3.44	4.81	3.44	79
EDSR	9.15	1.75	1.60	0.47	2.5
HRNet	11.35	2.22	2.90	1.44	79

From this table, some analysis can be derived. EDSR, HSCNN+ and HDNet are the fastest predicting methods. These computational performances are justified by the number of parameters that each network has. When comparing with the reconstruction error, we can discard EDSR and HSCNN+ given their poor reconstruction. We can conclude then that HDNet shows the best balance between reconstruction (Table I) and computational performance (Table II). Therefore, it would be a good choice for an embedded or CPU implementation. However, if computational resources are not constrained, MPRNet running on a GPU server is still a suitable choice.

V. CONCLUSION

In this paper, a systematic evaluation and comparison of different methods that could be used for reconstruction from RGB images to hyper-spectral images has been performed, including HSCNN+, HRNet, EDSR, HDNet, HINet, MIRNet, Restormer, MPRNet, MST, and MST++. An ambitious RGB-2-HSI reconstruction outside of the visual range has been tested for the first time, achieving competitive results. Some of the proposed methods have never been applied to HSI reconstruction and still showed a competitive performance.

These methods have been applied to a novel data set, in addition to experiments on a standard dataset, in order to compare their performance. Each model has been evaluated based on RMSE and PSNR. The capabilities of these models to reconstruct a hyper-spectral image from a single RGB image, outside of the visual range and towards the near-infrared range have been shown.

The use of MPRNET, to reconstruct hyper-spectral images of ham shows that it could be a promising approach for other fields. There is a potential for MPRNET to be used to reconstruct hyper-spectral images in both a higher spatial and a higher spectral resolution. The use of the model for different domain reconstructions would require a new library of hyper-spectral images to train. Although the reconstruction model is built on a ham data set, it can probably be transferred to another similar domain with a different lighting situation.

Our approach was also evaluated using a CPU-based server and a selection of different accelerators. As expected, the edge devices perform worst but, depending on the use case of the model, it could be useful. MPRNET and HDNet were highlighted in this comparison as a good balance between reconstruction and computational performance.

In the future, we plan to extend the experimentation to other data sets and other methods which present a more robust technique to handle a greater variety of scenes. Experimentation to find the limitation of the spectral range of reconstruction before the spectral integrity is lost will also be carried out. This approach to RGB-2-HSI reconstruction can help lead to a more sophisticated approach, which can be supplemented with the use of a lower-cost instrument such as a near-infrared probe to further the reconstruction range (1000-2500 nm).

REFERENCES

- [1] Naveed Akhtar and Ajmal Mian. Hyperspectral recovery from rgb images using gaussian processes. *IEEE transactions on pattern analysis and machine intelligence*, 42(1):100–113, 2018.
- [2] Boaz Arad, Radu Timofte, Rony Yahel, Nimrod Morag, Amir Bernat, Yuanhao Cai, Jing Lin, Zudi Lin, Haoqian Wang, Yulun Zhang, et al. Ntire 2022 spectral recovery challenge and data set. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 863–881, 2022.
- [3] Yuanhao Cai, Jing Lin, Xiaowan Hu, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. Mask-guided spectral-wise transformer for efficient hyperspectral image reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17502–17511, June 2022.
- [4] Yuanhao Cai, Jing Lin, Zudi Lin, Haoqian Wang, Yulun Zhang, Hanspeter Pfister, Radu Timofte, and Luc Van Gool. Mst++: Multi-stage spectral-wise transformer for efficient spectral reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 745–755, 2022.
- [5] Tristan Cazenave, Julien Sentuc, and Mathurin Videau. Cosine annealing, mixnet and swish activation for computer go. In *Advances in Computer Games: 17th International Conference, ACG 2021, Virtual Event, November 23–25, 2021, Revised Selected Papers*, pages 53–60. Springer, 2022.
- [6] Tianfeng Chai and Roland R Draxler. Root mean square error (rmse) or mean absolute error (mae). *Geoscientific model development discussions*, 7(1):1525–1534, 2014.
- [7] Liangyu Chen, Xin Lu, Jie Zhang, Xiaojie Chu, and Chengpeng Chen. Hinet: Half instance normalization network for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 182–192, 2021.
- [8] Leon O Chua. *CNN: A paradigm for complexity*, volume 31. World Scientific, 1998.
- [9] Abbey Douglas, Gabor Kereszturi, Lauren N Schaefer, and Ben Kennedy. Rock alteration mapping in and around fossil shallow intrusions at mt. ruapehu new zealand with laboratory and aerial hyperspectral imaging. *Journal of Volcanology and Geothermal Research*, 432:107700, 2022.
- [10] Ian C Fiebelkorn and Sabine Kastner. Functional specialization in the attention network. *Annual review of psychology*, 71:221–249, 2020.
- [11] Jon Gutiérrez-Zaballa, Koldo Basterretxea, Javier Echanobe, M Victoria Martínez, and Inés del Campo. Exploring fully convolutional networks for the segmentation of hyperspectral imaging applied to advanced driver assistance systems. In *Design and Architecture for Signal and Image Processing: 15th International Workshop, DASIP 2022, Budapest, Hungary, June 20–22, 2022, Proceedings*, pages 136–148. Springer, 2022.
- [12] Fatemeh Sadat Hashemi-Nasab and Hadi Parastar. Vis-nir hyperspectral imaging coupled with independent component analysis for saffron authentication. *Food Chemistry*, 393:133450, 2022.
- [13] Xiaowan Hu, Yuanhao Cai, Jing Lin, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. Hdnet: High-resolution dual-domain learning for spectral compressive imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17542–17551, 2022.
- [14] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. *Advances in neural information processing systems*, 28, 2015.
- [15] Mahmoud Kamel, Walaa Hamouda, and Amr Youssef. Ultra-dense networks: A survey. *IEEE Communications surveys & tutorials*, 18(4):2522–2545, 2016.
- [16] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017.
- [17] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017.
- [18] Marena Manley. Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials. *Chemical Society Reviews*, 43(24):8200–8214, 2014.
- [19] Gabriel Meseguer-Brocal and Geoffroy Peeters. Conditioned-u-net: Introducing a control mechanism in the u-net for multiple source separations. *arXiv preprint arXiv:1907.01277*, 2019.
- [20] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.
- [21] Armin Schneider and Hubertus Feussner. Chapter 5 - diagnostic procedures. In Armin Schneider and Hubertus Feussner, editors, *Biomedical Engineering in Gastrointestinal Surgery*, pages 87–220. Academic Press, 2017.
- [22] Zhan Shi, Chang Chen, Zhiwei Xiong, Dong Liu, and Feng Wu. Hscnn+: Advanced cnn-based hyperspectral recovery from rgb images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 939–947, 2018.
- [23] F. Vasefi, N. MacKinnon, and D.L. Farkas. Chapter 16 - hyperspectral and multispectral imaging in dermatology. In Michael R. Hamblin, Pinar Avci, and Gaurav K. Gupta, editors, *Imaging in Dermatology*, pages 187–201. Academic Press, Boston, 2016.
- [24] Wang Yuanji, Li Jianhua, Lu Yi, Fu Yao, and Jiang Qinzong. Image quality evaluation based on image weighted separating block peak signal to noise ratio. In *International Conference on Neural Networks and Signal Processing, 2003. Proceedings of the 2003*, volume 2, pages 994–997. IEEE, 2003.
- [25] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5728–5739, 2022.
- [26] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for real image restoration and enhancement. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16*, pages 492–511. Springer, 2020.
- [27] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14821–14831, 2021.
- [28] Hongyang Zhang, Junru Shao, and Ruslan Salakhutdinov. Deep neural networks with multi-branch architectures are intrinsically less non-convex. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 1099–1109. PMLR, 2019.
- [29] Jingang Zhang, Runmu Su, Qiang Fu, Wenqi Ren, Felix Heide, and Yunfeng Nie. A survey on computational spectral reconstruction methods from rgb to hyperspectral imaging. *Scientific reports*, 12(1):11905, 2022.
- [30] Yuzhi Zhao, Lai-Man Po, Qiong Yan, Wei Liu, and Tingyu Lin. Hierarchical regression network for spectral reconstruction from rgb images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 422–423, 2020.