Fast-n-Squeeze: towards real-time spectral reconstruction from RGB images

Mirko Agarla Simone Bianco Marco Buzzelli Luigi Celona Raimondo Schettini
Department of Informatics Systems and Communication
University of Milano - Bicocca
Viale Sarca 336, Building U14, Milan, Italy
{first_name.second_name}@unimib.it

Abstract

We present an efficient method for the reconstruction of multispectral information from RGB images, as part of the NTIRE 2022 Spectral Reconstruction Challenge. Given an input image, our method determines a global RGB-to-spectral linear transformation matrix, based on a search through optimal matrices from training images that share low-level features with the input. The resulting spectral signatures are then adjusted by a global scaling factor, determined through a lightweight SqueezeNet-inspired neural network. By combining the efficiency of linear transformation matrices with the data-driven effectiveness of convolutional neural networks, we are able to achieve superior performance than winners of the previous editions of the challenge.

1. Introduction

Consumer digital imaging devices are typically characterized by three types of sensor, covering three bands in the spectrum of visible light that are mapped to what is referred to as “red”, “green”, and “blue” (RGB). Higher-end devices designed to capture finer bands within the light spectrum fall under the term of hyperspectral and multispectral imaging sensors. Acquiring this level of spectral detail produces additional information that allows for better discrimination of metamers, i.e. different spectral signatures that would appear as equivalent stimuli under given conditions. This discrimination capability, in turn, can facilitate and improve the performance of a variety of computer vision tasks, such as remote sensing, anomaly detection, and applications for medical imaging [11]. One downside of spectral imaging is the higher cost of the hardware involved, which is among the motivations that lead to the rise of spectral reconstruction: the task of estimating a multispectral or hyperspectral signature corresponding to a given input RGB triplet. By definition, this is an ill-posed problem, as it involves an increase in dimensionality. For this reason, methods for spectral reconstruction must exploit additional priors and contextual information, such as pixel neighborhoods and camera characteristics.

This paper describes our solution to the NTIRE 2022 Spectral Reconstruction Challenge [5], in which each participant is provided with a set of RGB-spectral image pairs for development, training, and tuning, and the final assessment is conducted on a blind test set. As we will discuss later on, existing methods for spectral reconstruction are traditionally based on defining sparse-representation dictionaries to encode the transformation from RGB to spectral domain. An intrinsic advantage of this type of solution is its relative efficiency, as the final reconstruction can be obtained through the application of one or more linear transformation matrices applied to pixel values. More recent solutions exploit deep convolutional neural networks (CNNs) in order to learn rich dependencies of RGB information, and to map them to the final spectral reconstruction. This approach allowed for a significant improvement in the reconstruction performance, in line with many other applications of deep learning in the latest years [6, 9], at the cost of higher computational complexity. Furthermore, image-to-image methods based on CNNs are prone to introducing artifacts, as we will show with a comparative analysis.

In this work we propose a hybrid solution to the problem of spectral reconstruction from RGB images, called Fast-n-Squeeze. Our approach is based on the definition of a global RGB-to-spectral linear transformation matrix, which is automatically defined using low-level image features, and efficiently applied. The output reconstruction is then corrected by a global scaling factor, determined through a lightweight convolutional neural network, whose computational complexity can be constrained thanks to the simplified nature of the task. Our overall solution is highly efficient (198.45 FPS in GPU), easily implemented, and it produces results that are comparable to the more computationally-demanding solutions that won previous editions of the challenge, while at the same time avoiding artifacts in the reconstruction process.

2. Related works

Aeschbacher et al. [1] introduced what is commonly referred to as A+, built upon the method by Arad et al. [3] and the method by Timofte et al. [13]. This technique for spectral reconstruction is based on sparse dictionary representation of the input RGB pixels into atoms, and on transferring said representation to the corresponding higher-dimensional spectral atoms for full spectrum reconstruction. During the training phase, a set of hyperspectral atoms is generated from the dataset and projected into the corresponding RGB using color matching functions. The neighbors of training signatures for each dictionary atom are used to compute RGB-to-spectral matrices, which are then exploited at reconstruction to process each RGB image pixel. In line with this work, we introduce a solution based on the application of RGB-to-spectral matrices to reconstruct the hyperspectral version of the input image in a highly efficient fashion. Furthermore, by using global as opposed to local matrices, we prevent the introduction of spatial reconstruction artifacts.

The NTIRE 2020 Spectral Reconstruction Challenge [4] collected a number of highly effective solutions, mostly based on the application of properly-designed Convolutional Neural Networks (CNNs). In the following, we describe the most effective solutions that participated to the 2020 edition of the challenge.

Li et al. [10] proposed an Adaptive Weighted Attention Network (AWAN). Their solution is a neural architecture implementing a number of novel weight attention mechanisms, combined with a prior on the known camera spectral sensitivity. More specifically, the backbone architecture is a sequence of dual residual attention blocks, characterized by interleaved long and short skip connections, and by a novel adaptive weighted channel attention module that exploits interdependencies among intermediate features. A final patch-level second-order non-local attention module is then used to capture long-range spatial contextual information at the end of the neural process. The overall loss is obtained as a weighted average between the direct effect on spectral reconstruction, and the indirect effect on RGB reconstruction, computed by applying the camera spectral sensitivity function.

Zhao et al. [15] proposed a Hierarchical Regression Network (HRNet). Their solution is a neural network composed of four parallel branches that process and combine the input RGB image at different levels of resolution. Specifically, each level is obtained by applying a non-learnable PixelUnShuffle layer, which redistributes the existing pixels by halving spatial resolution in each direction, and consequently increasing the channel dimension. Every neural branch performs a first inter-level integration step, where the output of the lower branch is upscaled via learnable PixelShuffle, concatenated, and reprocessed via convolution.

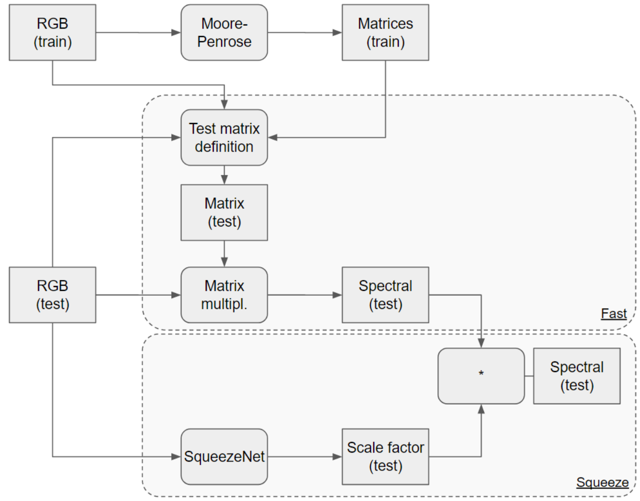

Figure 1. Schematic diagram of the method proposed: Fast-n-Squeeze.

Then, a sequence of residual dense blocks is applied for artifact reduction, followed by a residual global block that implements an attention mechanism through a multi-layer perceptron. The network is only trained via L1 loss.

3. The proposed method: Fast-n-Squeeze

A schematic representation of the proposed method is reported in Figure 1. From the diagram reported it is possible to see the two parts of which Fast-n-Squeeze is composed: the first part, Fast, estimates the best RGB-to-spectral linear transformation matrix to be used on the input image; the second part, Squeeze, refines the reconstruction provided by the transformation matrix by a global scaling factor. In the following we describe in details the two parts.

3.1. Fast

The first part of the proposed method allows to estimate the best RGB-to-spectral linear transformation matrix to be applied to the input RGB image. A Moore-Penrose pseudo-inverse matrix [12] is computed for each training image together with a feature vector representing the average, maximum and standard deviation of each color channel:

$$feat_i = [\mathrm{avg}(\mathrm{RGB_i})\,\max(\mathrm{RGB_i})\,\mathrm{std}(\mathrm{RGB_i})]^{\alpha} \quad (1)$$

which is then elevated to the power $\alpha = 0.9$. Given a test image, its feature vector is computed as in Equation 1, and then it is compared with all the vectors in the training set with the Chebychev distance. The images having a Chebychev distance lower than $1.05\times$ the minimum distance are identified, the corresponding pseudo-inverse matrices are selected, and their element-wise median is computed.

3.2. Squeeze

The second part of the proposed method uses a CNN model to predict a single global scaling factor to be used on the spectral image, recovered using the matrix computed by the first step. This part is named after the CNN model used, i.e. Squeezenet-v1.1 [8]. The model is modified by removing the Dropout layer and replacing the softmax classification layer with a fully-connected layer that estimates one single output. In order to allow for fast inference, the input image is resized to 256×256 pixels.

The parameters of the SqueezeNet-v1.1 model are finetuned using the Mean Absolute Error (MAE) as loss function. The choice of this loss is driven by two considerations: i) we experimentally found that the competition target metric, i.e. Mean Relative Absolute Error (MRAE) was not able to drive the optimization, and ii) MAE shows a good correlation with angular error metrics [2]. We train the model for a total of 30 epochs by using Adam optimizer with starting learning rate of 1×10^{-4} which decays by a factor of 0.5 every 10 epochs, a batch-size equal to 16, and exponential decay rates β_1 and β_2 equal to 0.9 and 0.999. Each RGB image fed to the model is normalized by its global maximum value and then resized to 256×256 pixel resolution using bilinear interpolation. We randomly apply horizontal and vertical flip, and rotation by an angle between 0 and 360 degrees. At the end of each epoch, the MRAE is estimated on the validation set. The model that achieves the lowest MRAE is chosen as best model.

We implement the proposed method in Python 3.8 using the PyTorch package with CUDA-v11.6 as back-end. The proposed model is trained on a workstation equipped with an Intel i7-4770 CPU @3.40GHz, 16GB DDR4 RAM 2400MHz, NVIDIA GeForce GTX 1080 GPU with 2560 CUDA cores. The training process of the proposed solution lasts about 30 minutes.

4. Experiments

4.1. Datasets

We train and evaluate our Fast-n-Squeeze on the Arad_1K Hyperspectral Database provided by NTIRE 2022 challenge [5]. This dataset consists of 1000 RGB-HyperSpectral (HS) pairs divided into 900 training, 50 validation and 50 testing samples. Each spectral image is characterized by 31 bands in the 400 nm to 700 nm range, with a spatial resolution of 482×512 pixels. To generate the RGB image corresponding to the HS tensor, there is a fixed camera response function CRF applied to HS bands. The rendering process can be defined as:

$$RGB = HS \times CRF. \tag{2}$$

The Poisson noise is added to the RGB and then the normalized by $\frac{0.18}{\mathrm{mean(RGB)}}$. Two samples from the database are shown in Figure 2.

4.2. Evaluation metrics

The proposed Fast-n-Squeeze is objectively evaluated and compared with other methods by using the following metrics:

- **Root Mean Squared Error (RMSE).** It computes the root mean square error between the generated spectral images and the corresponding ground-truths. It is defined as:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\|G(x)_i - y_i\|^2}, \tag{3}$$

  where N denotes the total number of pixels in the spectral image, $G(x)$ and y are respectively the generated spectral image and the corresponding ground-truth.

- **Mean Relative Absolute Error (MRAE).** It estimates the pixel-wise disparity (mean absolute value) between all bands of generated spectral image and ground-truth. It is computed as follows:

$$\text{MRAE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|G(x)_i - y_i|}{y_i}. \tag{4}$$

- **Back Projection MRAE (BPMRAE).** It evaluates the colorimetric accuracy of the reconstructed RGB by applying a fixed camera response function to the generated and ground-truth spectral images. It is defined by the following equation:

$$\text{BPMRAE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|(CRF \times G(x)_i) - (CRF \times y_i)|}{(CRF \times y_i)}, \tag{5}$$

  where CRF is the camera response function.

- **Weighted accuracy.** As indicated in [4], it evaluates the reconstruction error taking into account which material the pixels belong to. It is estimated by firstly grouping similar spectra into 1000 clusters, then by individually calculating the mean MRAE of each group.

4.3. Ensemble strategy

The Squeeze part of our solution can be boosted by ensambling multiple training iterations. Specifically, the corresponding entry in the NTIRE 2022 challenge is computed as a weighted average:

$$\text{Squeeze} = \frac{1}{3}\text{Squeeze}_{(1)} + \frac{2}{3}\text{Squeeze}_{(2)} \tag{6}$$

where the weights were empirically set.

Figure 2. Samples from the NTIRE 2022 Arad_1K Hyperspectral Database. Each row displays the RGB image and 5 different bands of the corresponding HS image.

Method	MRAE	RMSE
Fast	0.3381	0.0517
Fast (oracle glob. scaling factor)	0.2040	0.0289
Squeeze$_{(1)}$	0.3344	0.0567
Squeeze$_{(2)}$	0.3398	0.0612
Squeeze	0.3331	0.0579
Fast-n-Squeeze	0.3164	0.0516
Fast-n-Squeeze (oracle select.)	0.2753	0.0452

Table 1. Quantitative comparison results of the different parts of the proposed solution on the NTIRE 2022 validation set.

Our final solution, Fast-n-Squeeze, combines both the Fast and the Squeeze component, exploiting the uncorrelated nature of the underlying techniques to improve upon the spectral reconstruction of both:

$$\text{Fast-n-Squeeze} = \frac{1}{2}\text{Fast} + \frac{1}{2}\text{Squeeze} \tag{7}$$

The weights were also empirically set.

We also consider an hypothetical lower bound: for each test image, the best solution between Fast and Squeeze is selected assuming the availability of an oracle.

4.4. Testing result on NTIRE 2022 challenge

Official results on the validation and test set of the NTIRE 2022 challenge are reported, respectively, in Table 1 and Table 2.

The validation performance allows us to investigate in detail our solution. The Fast part, in particular, achieves 0.3381 MRAE. If, however, we assume the availability of an oracle that determines an optimal global scaling factor, this error decreases to 0.2040 MRAE. This behavior is illustrated visually in Figure 3.

Method	MRAE	RMSE
Fast	0.4629	0.0691
Squeeze	0.4160	0.0895
Fast-n-Squeeze	0.3647	0.0641
Fast-n-Squeeze (oracle select.)	0.2915	0.0541

Table 2. Quantitative comparison results of the different parts of the proposed solution on the NTIRE 2022 test set.

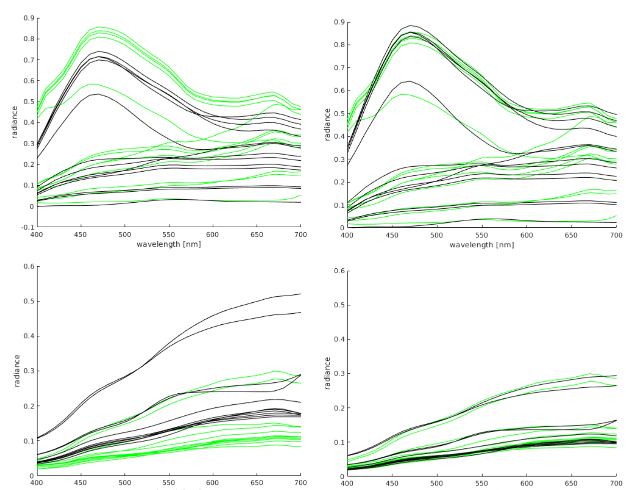

Figure 3. Effect of the application of the oracle scaling global factor to Fast. Groundtruth (green lines) and predicted spectra (black lines). First row corresponds to ARAD_1K_0901, and MRAE goes from 0.1993 (left) to 0.0939 (right); second row corresponds to ARAD_1K_0946, and MRAE goes from 0.7617 (left) to 0.0849 (right).

The definitive test set performance reported in Table 2 clearly highlights the individual contributions of each component. The Fast-n-Squeeze solution, with 0.3647 MRAE, allows for a relative improvement of 21% over Fast, and 12% over Squeeze. Finally, the oracle selection between Fast and Squeeze achieves 0.2915 MRAE, highlighting the potential of the proposed solution in case of a classifier trained to identify two different classes of images, and suggesting a direction for future developments.

4.5. Comparison with state-of-the-art methods

In this section we present the comparison of our Fast-n-Squeeze with the two best methods of the NTIRE 2020 Spectral Reconstruction Challenge, namely AWAN [10] and HRNet [15], and a sparse coding method presenting some similarities to the proposed one, i.e. A+ [1]. AWAN, Arad+, and HRNet have been trained on the NTIRE 2022 Arad_1k Hyperspectral Database using the original codes provided by the authors with the default settings. For training AWAN and HRNet we reduce the batch size (i.e. AWAN = 6 and HRNet = 4) due to the limited GPU memory. For the same reason, at inference time, we process the input RGB image by splitting it into 128×128 patches with an overlap of 64 pixels using the algorithm provided by the authors. The results for HRNet are obtained by generating the spectral images using a single model, while the original method gets the final spectral image by calculating the average of 8 spectral images obtained with an ensemble of models.

We report the comparison on the validation image pairs of NTIRE 2022 as the HS images for the testing set are not publicly available. The results are summarized in Table 3. We also visualize the output of each method in Figure 4 by pseudo-color map. For all the considered metrics, the proposed Fast-n-Squeeze outperforms the competitors by a large gap, namely -0.08 MRAE on the second best method, i.e. AWAN. There are three reasons the proposed method outperforms the other three methods. The first is that the Fast part of our Fast-n-Squeeze provides a good approximation of the shape of the ground-truth spectra for each pixel. The second is that the Squeeze part of the proposed method effectively scales the spectra based on the global statistics of the generated spectral image. The third is that compared to AWAN and HRNet, Fast-n-Squeeze does not introduce spatial artifacts, as can be seen in rows 2, 3 and 4 of Figure 4. We will further discuss the latter in the next subsection.

We compare the performance of the proposed Fast-n-Squeeze with previous methods also on the validation set of NTIRE 2020 Spectral Reconstruction Challenge. Table 4 reports the quantitative comparison. For all the considered metrics the best model is AWAN, followed by HRNet and Arad+. These methods are pixel based algorithms trained on a not processed database. This approach allows these methods to better reconstruct the spectra associated to the RGB pixels without being diverted by the image normalization factor and noise (See Section 4.1 for details). Our method, despite being the least performing among the analyzed solutions on this database, represents a good balance between efficiency and effectiveness (see Section 4.5.2).

4.5.1 Spatial reconstruction artifacts comparison

In this section we focus on the problem of spatial reconstruction artifacts. For this analysis we consider two perceptually motivated distance metrics, such as the Peak Signal-to-Noise Ratio (PSNR) and the Structural SIMilarity (SSIM) index [14]. The PSNR is defined as follows:

$$
\begin{aligned}
MSE &= \frac{1}{N}\sum_{i=1}^{N} \|G(x)_i - y_i\|^2, \\
PSNR &= 20 \times \log_{10}\left(\frac{p_{max}}{MSE}\right),
\end{aligned}
\tag{8}
$$

where $p_{max} = 1$ is the maximum possible value of pixels in HS images. On the other hand, SSIM is computed in the following way:

$$
\begin{aligned}
SSIM_{i,n} &= \frac{(2\mu\mu^* + C_1)(2\hat{\sigma} + C_2)}{(\mu^{*2} + \mu^2 + C_1)(\sigma^{*2} + \sigma^2 + C_2)}, \\
SSIM &= \frac{1}{M \times W \times H}\sum_{i=1}^{M}\sum_{n=1}^{W \times H} SSIM_{i,n},
\end{aligned}
\tag{9}
$$

where μ^* and σ^{*2} are the mean and variance for the n^{th} $N \times N$ window in the i^{th} band on the generated spectral image. Similarly, μ and σ^2 account for the mean and variance of the window in the ground-truth spectral image. Also, $C_1 = k_1 L$, and $C_2 = k_2 L$ are introduced to avoid division by zero when the mean or covariance values are close to zero. M is the number of bands, W and H are width and height of the image. We set $N = 47$ following [7], where it is shown that a larger local window is better for distortions with larger signal-error cross-correlation.

The PSNR and SSIM results on the NTIRE 2022 validation set are reported in the second part of Table 3. The proposed Fast-n-Squeeze achieves the best performance for both the metrics with respect to the state-of-the-art methods. The worst results are achieved by HRNet, i.e. 0.1266 lower for SSIM and 5.88 lower for PSNR than Fast-n-Squeeze. A qualitative comparison of how spatial artifacts manifest for different bands of a sample image is provided in Figure 4. As it is possible to see, AWAN introduces very noticeable blocking artifacts likely due to patch-wise processing. HRNet generates a spectral image in which there are many discontinuities in the pixel intensities that are not present in the ground-truth image. Finally, for Arad+ and our Fast-n-Squeeze the generated spectral image presents no spatial artifacts and it is very similar to the ground-truth.

Spatial artifacts are more evident when the generated spectral image is backprojected to RGB using Eq. 2. Figure

Method	MRAE	RMSE	BPMRAE	Weighted accuracy	PSNR	SSIM	Infer. speed FPS
A+ [1]	0.3906	0.0785	0.3570	0.4004	29.1255	0.8420	0.31 (CPU)
AWAN [10]	0.3551	0.0704	0.3312	0.3660	29.7714	0.8619	0.14 (GPU)
HRNet [15]	0.5810	0.0848	0.5436	0.5565	27.5177	0.8037	40.26 (GPU)
Fast-n-Squeeze (this paper)	**0.2753**	**0.0452**	**0.2107**	**0.2986**	**33.4023**	**0.9303**	**198.45** (GPU)

Table 3. Quantitative comparison on the NTIRE 2022 validation set. Best result for each metric is reported in **bold**.

Method	MRAE	RMSE	BPMRAE
A+ [1]	0.0467	0.0155	**0.0001**
AWAN [10]	**0.0312**	**0.0111**	–
HRNet [15]	0.0423	0.0135	0.0061
Fast-n-Squeeze (this paper)	0.0882	0.0247	0.0177

Table 4. Quantitative comparison on the NTIRE 2020 validation set. Best result for each metric is reported in **bold**.

5 shows a validation image of Arad_1k generated with the different methods considered. It is highlighted in the detail of the sky how AWAN and HRNet introduce some artifacts of which Arad+ and the proposed method are instead lacking. The presence of such artifacts is reflected in the lower values of PSNR and SSIM for AWAN and HRNet with respect to those achieved by Arad+ and our Fast-n-Squeeze.

4.5.2 Inference speed comparison

For spectral reconstruction methods, efficiency is also crucial. In this section, we complement the part of performance estimation with that of computational efficiency. To this end, we measure the inference speed in terms of Frame-Per-Seconds (FPS) by running all the methods on the same workstation in GPU (apart from Arad+ which only runs in CPU). The inference speed estimated on the Arad_1k images having spatial resolution of 482×512 pixels is reported in the last column of Table 3. As is possible to see, the proposed method with 198.45 FPS is considerably faster than the other methods. This number is higher than the 104.71 FPS reported in [5] because we optimize the code efficiency by asynchronously running the forward pass for the two SqueezeNet models. Compared to deep learning-based methods that run in GPUs, it is an order of magnitude faster than HRNet and three orders of magnitude faster than AWAN. The proposed Fast-n-Squeeze is also faster than Arad+ in CPU, in fact the former has an inference speed of 50 FPS, while the latter of 0.31 FPS.

For our Fast-n-Squeeze we deepen the inference speed analysis by running the method in both GPU and CPU on input images with various standard image resolutions, namely that of the Arad_1k (482×512), Standard Definition (SD, 720×576), High Definition (HD, 1280×720), Full-HD (FHD, 1920×1080), 2K 2048×1080, Ultra-HD (UHD, 3840×2160), and Digital Cinema Initiatives 4K (DCI 4K, 4096×2160). On the basis of the results shown in Figure 6, we can claim that the proposed method is suitable for real-time spectral reconstruction of images even at very high resolution. Indeed, it has an upper limit of around 200 FPS for SD images and a lower limit of around 20 FPS for DCI 4K images in GPU. The CPU inference speed range is instead between 50 FPS for SD images and just under 4 FPS for DCI 4K.

5. Conclusion

We have proposed an efficient method for the reconstruction of multispectral information from RGB images, as a contribution to the NTIRE 2022 Spectral Reconstruction Challenge. Our approach defines a global RGB-to-spectral linear transformation matrix, estimated using low-level image features, and subsequently applies a global scaling factor, determined through a lightweight convolutional neural network. The matrix-based reconstruction was found to be effective at correctly estimating the spectral signature disregarding the global magnitude, while an oracle-based scaling factor showed a significant potential improvement on the MRAE-based evaluation. Based on these results, we consider for future developments the integration of additional shooting information (such as aperture and exposition) for a more accurate estimation of the appropriate scaling factor and, consequently, for a more accurate spectral reconstruction.

Furthermore, We have also presented a comparative analysis of our proposed solution against state-of-the-art methods, focusing on spatial reconstruction artifacts as well as inference speed, showing how the proposed Fast-n-Squeeze methods obtains positive results according to both criteria. In the future, we will consider a spatially-varying extension of our proposal, aiming for a good balance between efficiency and effectiveness.

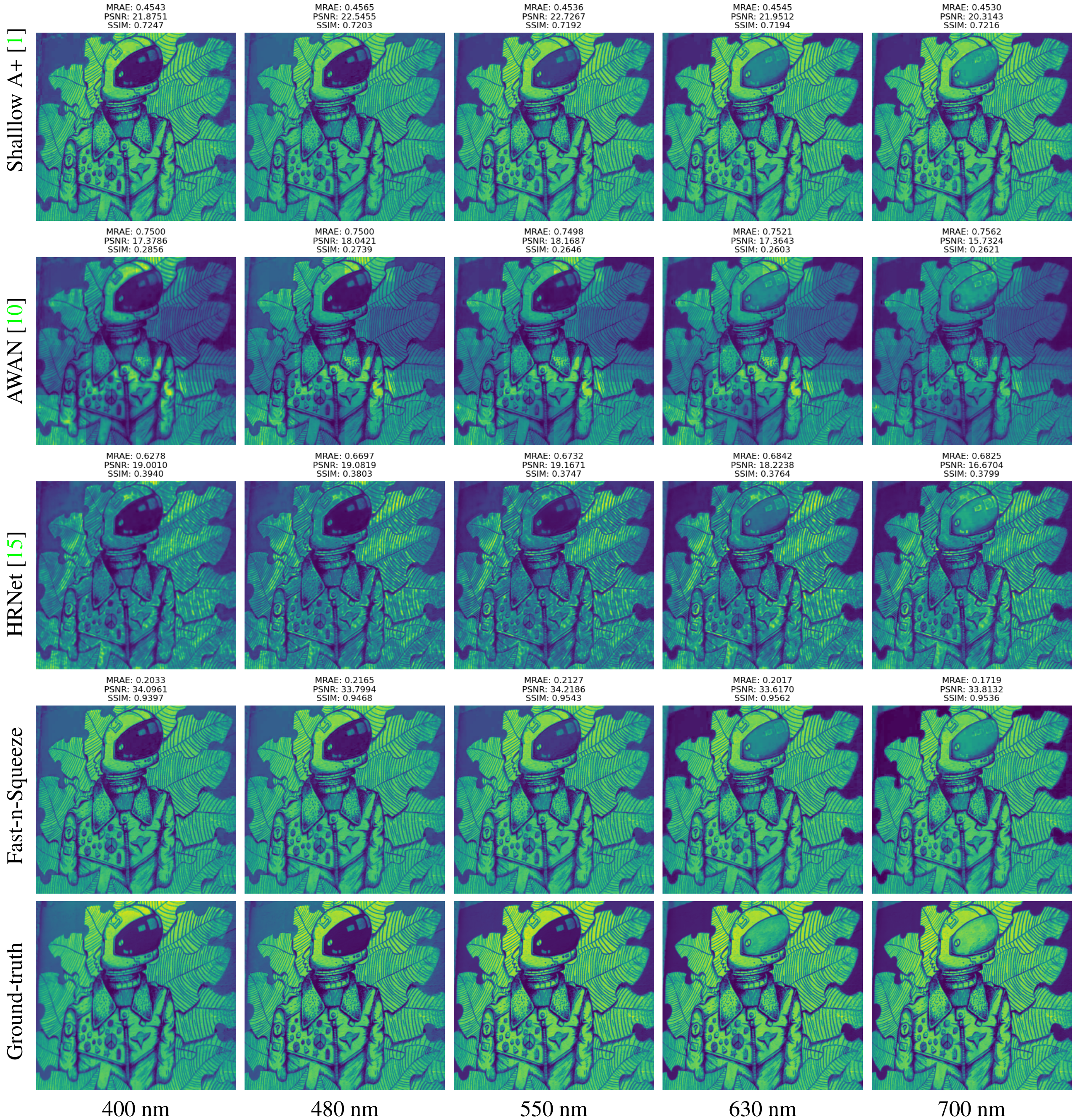

Figure 4. Comparison of generated results from A+ [1], AWAN [10], HRNet [15], and the proposed Fast-n-Squeeze on NTIRE 2022 HS challenge. On top of each HS image we report MRAE, PSNR and SSIM metrics. Best viewed in color on a screen.

References

[1] Jonas Aeschbacher, Jiqing Wu, and Radu Timofte. In defense of shallow learned spectral reconstruction from rgb images. In *International Conference on Computer Vision Workshops*, pages 471–479, 2017. 2, 5, 6, 7, 8

[2] Mirko Agarla, Simone Bianco, Luigi Celona, Raimondo Schettini, and Mikhail Tchobanou. An analysis of spectral similarity measures. In *Color and Imaging Conference*, volume 2021, pages 300–305. Society for Imaging Science and Technology, 2021. 3

[3] Boaz Arad and Ohad Ben-Shahar. Sparse recovery of hyperspectral signal from natural rgb images. In *European Conference on Computer Vision*, pages 19–34. Springer, 2016. 2

[4] Boaz Arad, Radu Timofte, Ohad Ben-Shahar, Yi-Tun Lin,

Figure 5. Comparison of some details of the reconstructed images with the considered methods. The images are obtained by applying Eq. 2 to the generated spectral images.

and Graham D Finlayson. Ntire 2020 challenge on spectral reconstruction from an rgb image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 446–447, 2020. 2, 3

[5] Boaz Arad, Radu Timofte, Rony Yahel, Nimrod Morag, Amir Bernat, et al. NTIRE 2022 spectral recovery challenge and dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 1, 3, 6

[6] Marco Buzzelli, Federico Belotti, and Raimondo Schettini. Recognition of edible vegetables and fruits for smart home appliances. In *2018 IEEE 8th International Conference on Consumer Electronics-Berlin (ICCE-Berlin)*, pages 1–4. IEEE, 2018. 1

[7] Hossein Bakhshi Golestani and Mohammad Ghanbari. Window size influence on ssim fidelity. In *International Symposium on Telecommunications (IST'2014)*, pages 355–360. IEEE, 2014. 5

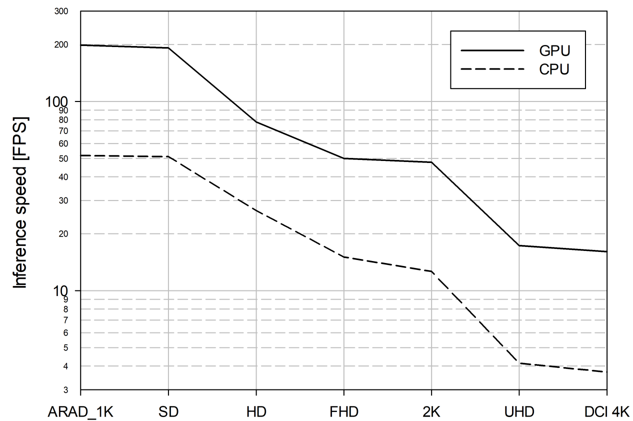

Figure 6. Inference time of the proposed Fast-n-Squeeze method when executed in GPU (solid line) and CPU (dashed line) for different standard image resolutions.

[8] Forrest N Iandola, Song Han, Matthew W Moskewicz, Khalid Ashraf, William J Dally, and Kurt Keutzer. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <0.5 mb model size. *arXiv preprint arXiv:1602.07360*, 2016. 3

[9] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015. 1

[10] Jiaojiao Li, Chaoxiong Wu, Rui Song, Yunsong Li, and Fei Liu. Adaptive weighted attention network with camera spectral sensitivity prior for spectral reconstruction from rgb images. In *Conference on Computer Vision and Pattern Recognition Workshops*, pages 462–463. IEEE/CVF, 2020. 2, 5, 6, 7, 8

[11] Yi-Tun Lin and Graham D Finlayson. Physically plausible spectral reconstruction from rgb images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 532–533, 2020. 1

[12] Roger Penrose. A generalized inverse for matrices. In *Mathematical proceedings of the Cambridge philosophical society*, volume 51, pages 406–413. Cambridge University Press, 1955. 2

[13] Radu Timofte, Vincent De Smet, and Luc Van Gool. A+: Adjusted anchored neighborhood regression for fast super-resolution. In *Asian conference on computer vision*, pages 111–126. Springer, 2014. 2

[14] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 5

[15] Yuzhi Zhao, Lai-Man Po, Qiong Yan, Wei Liu, and Tingyu Lin. Hierarchical regression network for spectral reconstruction from rgb images. In *Conference on Computer Vision and Pattern Recognition Workshops*, pages 422–423. IEEE/CVF, 2020. 2, 5, 6, 7, 8