Stephan Göb*, Theresa Ida Götz, Thomas Wittenberg

Multispectral single chip reconstruction using DNNs with application to open neurosurgery

Abstract: Multispectral imaging devices incorporating up to 256 different spectral channels have recently become available for various healthcare applications, as e.g. laparoscopy, gastroscopy, dermatology or perfusion imaging for wound analysis. Currently, the use of such devices is limited due to very high investment costs and slow capture times. To compensate these shortcomings, single sensors with spectral masking on the pixel level have been proposed. Hence, adequate spectral reconstruction methods are needed. Within this work, two deep convolutional neural networks (DCNN) architectures for spectral image reconstruction from single sensors are compared with each other. Training of the networks is based on a huge collection of different MSI image-stacks, which have been subsampled, simulating 16-channel single sensors with spectral masking. We define a training, validation and test set (*'HITgoC')* resulting in 351 training (631.128 sub-images), 99 validation (163.272 sub-images) and 51 test images. For the application in the field of *neurosurgery* an additional testing set of 36 image stacks from the *Nimbus* data collection is used, depicting MSI brain data during *open surgery*. Two DCNN architectures were compared to bilinear interpolation (BI) and an intensity difference (ID) algorithm. The DCNNs (*ResNet-Shinoda)* were trained on *HITgoC* and consist of a preprocessing step using BI or ID and a refinement part using a ResNet structure. Similarity measures used were PSNR, SSIM and MSE between predicted and reference images. We calculated the similarity measures for *HitgoC* and *Nimbus* data and determined differences of the mean similarity measure values achieved with the *ResNet-ID* and baseline algorithms such as BI algorithm and *ResNet-Shinoda*. The proposed method achieved better results against BI in SSIM (.0644 vs. .0252), PSNR (15.3 dB vs. 9.1 dB) and 1-MSE*100 (.0855 vs. .0273) and compared to *ResNet-Shinoda* in SSIM (.0103 vs. .0074), PSNR (3.8 dB vs. 3.6 dB) and 1-MSE*100 (.0075 vs. .0047) for *HITgoC/Nimbus*. In this study, significantly better results for spectral reconstruction in MSI images of open neurosurgery was achieved using a combination of ID-interpolation and ResNet structure compared to standard methods.

Keywords: Spectral Reconstruction, Debayering, Demosaicing, DCNN, open neurosurgery

https://doi.org/10.1515/cdbme-2021-2010

1 Introduction

Multispectral imaging (MSI) or hyperspectral imaging (HSI) devices incorporating nine, sixteen or even 128 or 256 different spectral channels across wavelengths within and beyond the visual range have in the past years become available for various applications in the field of healthcare and biomedicine. Some typical clinical examples for such applications based on MSI and HSI are e.g., diagnostic laparoscopy [1] or gastroscopy [2], screening dermatology [3], or perfusion imaging for wound analysis [4], or open neurosurgery [6,7]. An overview of the possibilities for MSI and HSI possibilities for digital and computational pathology has recently been presented by *Ortega et al.* [5].

Nevertheless, currently routine use of such MSI or HSI devices is limited due to very high investment-costs as well as slow capture times, as e.g., a brush-broom sensor with 256 spectral lines has to be moved mechanically over the field of view while scanning and concatenating the different spectral data. Similar, steerable spectral filters need time to integrate a sufficient number of spectral images. In order to compensate these named shortcomings, alternative approaches have recently proposed. One suggested method is based on spectral masking on the pixel level using only one single sensor plane (similar to the well-known Bayer pattern for single-chip RGB cameras) and successive interpolation of the missing spectral values based on image processing or image analysis approaches. Typical such image processing methods for spectral reconstruction include BI and nonlinear filters. The advantages of such filters are fast and real-time implementation as well as the independence from any application. However, if a spectral reconstruction with a focus on specific applications is needed, higher-level reconstruction approaches are desirable. To this end, recently deep convolutional neural networks (DCNN) have been suggested.

***Corresponding author: Stephan Göb:** Fraunhofer Institute for Integrated Circuits IIS, Am Wolfsmantel 33, Erlangen, Germany, goebsn@iis.fraunhofer.de, **T Götz, T Wittenberg,** Fraunhofer IIS

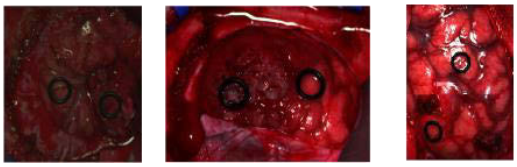

Figure 1: Images from the hyperspectral *Nimbus* data collection [6,7] as synthetic red-green-blue (RGB) representation.

Within this work, two DCNN architectures with the goal of spectral reconstruction from single sensor MSI are compared with each other, and furthermore some extensions for improvements are introduced and evaluated. The training of the networks is based on a large collection of different MSI image-stacks, which have been subsampled, simulating a 16-channel spectral pixel. For testing and evaluation of the proposed approach a set of 36 image stacks (and 572 sub-images) from *open neurosurgery* is used, related to the hyperspectral *Nimbus* data collection [6,7], see Figure 1.

2 Related Work

Procedures for spectral reconstruction can coarsely be divided into two groups, namely analytical algorithms and the application of DCNNs. Most cameras make use of standard analytical spectral reconstruction methods, such as BI. *Mihoubi et al.* [8] have further developed a new algorithm called *Intensity Difference* (ID). It is based on an analytical algorithm to detect the intensity of a multispectral image combined with the weighted bilinear algorithm. The algorithm was evaluated on the *CAVE* [9] database and achieved an average improvement of 1 dB compared to the common bilinear algorithm. *Shinoda et al.* [10] examined the spectral reconstruction process using analytical algorithms for pre-processing and a DCNN-ResNet structure for refinement. The network was trained and tested on the *CAVE* data set with using 8-fold cross validation. The reconstruction of the 16-band multi spectral filter array (MSFA) with the proposed method improved by 8.47 dB compared to the conventional method using BI.

3 Material and Methods

In order to train and validate the networks proposed in this work, a large HSI data set was built from different data collections, including *CAVE* [9], *TokyoTech* [11], *ICVL* [12], *HyTexiLa* [13] and *google* [14] data sets. This data collection was broken down into 70 percent training, 20 percent validation and 10 percent test data, resulting in 351 training (631.128 sub images), 99 validation (163.272 sub images) and 51 test images. We refer to this data collection as '*HITgoC*'. Additionally, the *Nimbus* [6,7] data set was used to test the proposed DCNN, including 36 images from open neurosurgery (see Fig. 1). All datasets have different local resolutions and a spectral resolution in the range of 400-1000 nm, with approximately every 1 nm occupied.

To work with HSI data of the *HITgoC* and *Nimbus* data sets, a multi spectral filter array (MSFA) had to be simulated. Thus, we selected 16 wavelengths in the range of 480-630 nm with a delta of 10 nm from the HSI data to yield MSI images. The used wavelengths and the pattern of the MSFA are shown in Figure 2. During pre-processing the training data consisted of MSFA data and related known original spectral channels.

For the analytical algorithms, we implemented BI and ID suggested by *Mihoubi et al.* [8]. Furthermore, we implemented a DCNN architecture suggested by the *Shinoda et al.* [10] and trained it on the joined *HITgoC* data collection. This approach (referred to as '*ResNet-Shinoda*') consists of two parts: A pre-processing step using BI and a refinement step using a ResNet based DCNN structure (Fig. 2), where the refinement is made up of 5 ResNet blocks (Fig. 3).

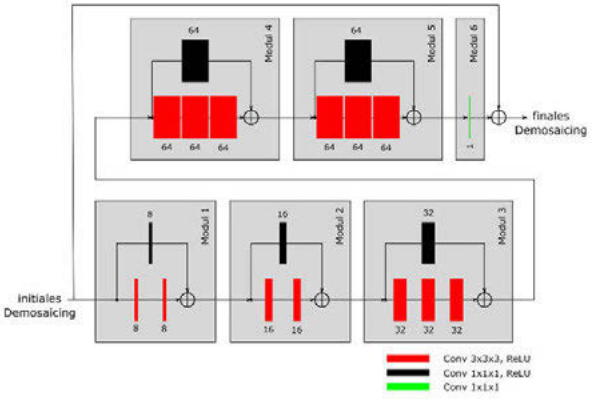

Figure 2: ResNet-based refinement of *ResNet-Shinoda*.

The starting number of filters is 8 and is doubled in each ResNet block except for the last one with 64 filters. The conv-layers use a 3^3 kernel and a ReLU activation function. To improve this architecture, we replaced the BI in the *ResNet-Shinoda* by the above-mentioned ID-Interpolation. The new processing pipeline is also shown in Figure 2.

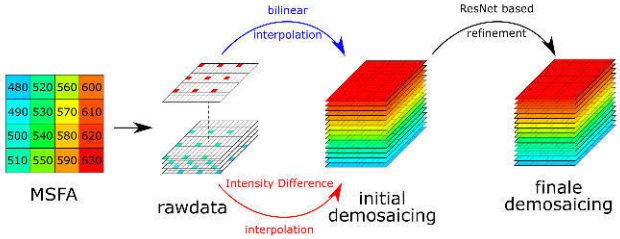

Figure 3: Processing pipeline for *ResNet-Shinoda* and *ResNet-ID*. *ResNet-Shinoda* use the bilinear interpolation (blue) and *ResNet-ID* use the intensity difference interpolation (red).

To make comparisons between the results of the networks and the analytic algorithms, 3 similarity measures were selected. The *image reconstruction quality* was determined by the *Peak-Signal-to-Noise Ratio* (PSNR), the *Structural Similarity* (SSIM) and the *Mean-Squared Error* (MSE) of the restored MSFA and MSI data. To analyze differences between the similarity measures of the proposed methods, the Welch-t-Test was used which (according to *Rasch et al.* [15]) is preferred to the Student-t-Test. If the P value is smaller than $P < .05$ the mean of two samples is seen to be significantly different.

4 Results

We calculated the similarity measures for the *HitgoC* and *Nimbus* data and determined the differences of the mean similarity measure values achieved with the *ResNet-ID* and baseline algorithms such as BI and *ResNet-Shinoda*. The resulting similarity measures are of all four methods are presented in box plots in Figures 4-6. T-tests are included in the graphs as lines between methods. The Figures are separated in two columns (left *HITgoC* and right *Nimbus*). Results for analytical algorithms are separated in the left and for the neuronal networks in the right of the column.

The proposed method achieved better results against BI in SSIM (.0644 vs. .0252), PSNR (15.34 dB vs. 9.12 dB) and in the 1-MSE*100 (.0855 vs. .0273) as well as compared to the *ResNet-Shinoda* in SSIM (.0103 vs. .0074), PSNR (3.87 dB vs. 3.66 dB) and 1-MSE*100 (.0075 vs. .0047) for *HITgoC/ Nimbus*. The proposed network *ResNet-ID* is significantly better than the compared *ResNet-Shinoda*, in the SSIM (*HITgoC*: t = -3.387, P <.001; *Nimbus*: t = -5.44, P < .001), in the PSNR (*HITgoC*: t = -4.34, P <.001; *Nimbus*: t = -3.74, P < .001) and in the MSE (*HITgoC*: t = -2.30, P = .0232; *Nimbus*: t = -3.21, P = .003). Both neural networks delivered significantly better results than all of the analytical algorithms (see Fig. 4-6).

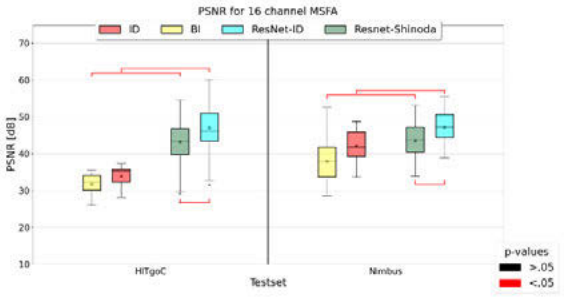

Figure 4: PSNR for 16 channel MSFA and the datasets *HITgoC* and *Nimbus* for the algorithms BI and ID and the neural networks *ResNet-Shinoda* and *ResNet-ID*.

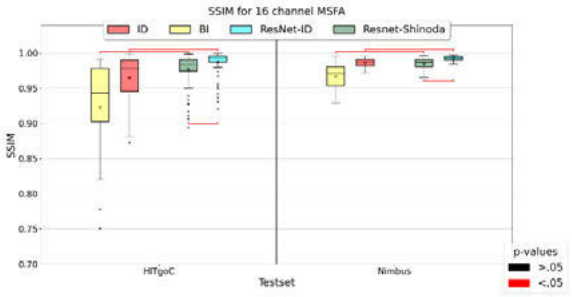

Figure 5: SSIM for 16 channel MSFA and the datasets *HITgoC* and *Nimbus* for the algorithms BI and ID and the neural networks *ResNet-Shinoda* and *ResNet-ID*.

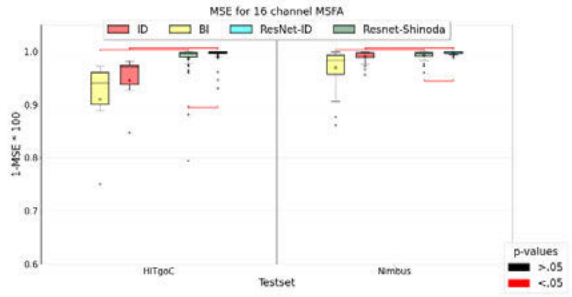

Figure 6: MSE for 16 channel MSFA and the datasets *HITgoC* and *Nimbus* for the algorithms BI and ID and the neural networks *ResNet-Shinoda* and *ResNet-ID*.

Table 1: PSNR values of beads image from *CAVE* of BI and the *ResNet-Shinoda* according to *Shinoda et al.*. Additionally, the PSNR improvement to BI is evaluated.

Algorithm	PSNR in dB	Improvement to BI in dB
BI	26.81	---
ResNet-Shinoda	32.69	5.88

For qualitative evaluation, single image from *CAVE* and from *Nimbus* (see Fig. 7) were extracted, and the spectral reconstruction for the algorithms was carried out on the images. Reconstruction and an error-image were calculated with *L*1-Norm of ID-interpolation, *ResNet-Shinoda* and *ResNet-ID*. PSNRs, the improvement to BI calculated by *Shinoda et al.* (Table 1) and by us for the beads image are summarized in Table 2.

Table 2: PSNR values of beads image from *CAVE* of BI, the *ResNet-Shinoda* and *ResNet-ID* according to our implementation. Additionally, the PSNR improvement to BI is evaluated.

Algorithm	PSNR in dB	Improvement to BI in dB
BI	23.33	---
ResNet-Shinoda	28.05	4.72
ResNet-ID	31.71	**8.38**

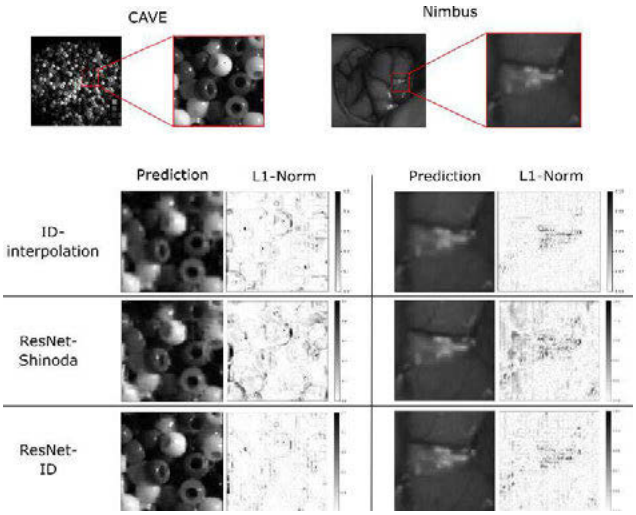

Figure 7: Prediction and error image of the reconstructions from the selected picture of *CAVE* database (error image scaled to .5) and *Nimbus* database (error image scaled to .1) for ID-interpolation, *ResNet-Shinoda* and *ResNet-ID*.

5 Discussion

In this contribution, two analytical algorithms and two neural networks for spectral reconstruction have been compared. In the work by *Shinoda et al.* a network was validated and tested on *CAVE* with 8-fold cross-validation and they yield an average PSNR of 43.05 dB for all 32 images, being an improvement of 8.47 dB to BI. By reimplementation of the *ResNet-Shinoda* our results of 28.05 dB in PSNR for the beads image differ from the results of 32.69 dB in PSNR from the original work (see Tables 1 & 2). In the evaluation of the BI algorithm, we yielded 23.33 dB in PSNR compared to 26.81 dB from Shinoda. It should be noted that Shinoda's paper does not describe exactly how their PSNR was calculated. Presumably, the PSNR in Shinoda's paper is not based on the 16 spectral data cube, but on the reconstruction of the RGB image derived by CIE D65. Therefore, our values are smaller because we are comparing the entire data cube. While Shinoda achieves an improvement of 5.88 dB, we get an improvement with our proposed method of 8.38 dB over the respective implemented variant of the bilinear interpolation. The smaller improvement of 4.77dB in PSNR between BI and our *ResNet-Shinoda* is due to the small number of training epochs, which was chosen due to a lack of time and resources. The prediction and the error pattern clearly show that the *ResNet-ID* yields better results, i.e., a lower deviation from the original image, especially in the area of the edges.

6 Conclusion

In this study, a new spectral reconstruction method is presented which combines two existing methods from literature, the ID algorithm and a ResNet architecture. Two analytical standard algorithms as well as the neural networks are evaluated in PSNR, MSE and SSIM on the datasets *HITgoC* and *Nimbus*. The proposed method – *ResNet-ID* – performs best in all tests.

Acknowledgment

The deep learning cluster were funded by the Federal Ministry of Education and Research under the project reference numbers 16FMD01K, 16FMD02 and 16FMD03.

References

[1] Zhang et al.Tissue classification for laparoscopic image understanding based on multispectral texture analysis. J Med Imaging. 2017
[2] Yoon et al. A clinically translatable hyperspectral endoscopy system for imaging the gastrointestinal tract. Nat Commun 10, 2019
[3] Vasefi et al. Hyperspectral and multispectral imaging in dermatology. In: Hamblin MR Pinar A, Gupta GK (eds): Imaging in dermatology. 2016, pp. 187-201
[4] Holmer et al. Hyperspectral imaging in perfusion & wound diagnostics - methods and algorithms for the determination of tissue parameters. Biomed Tech 25;63(5):547-556, 2018
[5] Ortega et al. Hyperspectral and multispectral imaging in digital and computational pathology: a systematic review. Biomed Opt Express. 2020, 11(6):3195-3233.
[6] Fabelo et al. Spatio-spectral classification of hyperspectral images for brain cancer detection during surgical operations. PLoS ONE 13(3), 2018
[7] Fabelo et al., In-vivo hyperspectral human brain image database for brain cancer detection. IEEE Access 7, pp. 39098-39116, 2019
[8] Mihoubi et al. Multispektral demosaicing using intensity-based spectral correlation. Proc's Int. Conf. on Image Processing Theory, Tools & Applications (IPTA), pp. 461-6, 2015
[9] Yasuma et al., Generalized Assorted Pixel Camera: Post-Capture Control of Resolution, Dynamic Range and Spectrum, Tech. Rep., Dept. Comp. Science, Columbia U. 2008.
[10] Shinoda et al. Deep demosaicking for multispectral filter arrays, in arXiv, 2018
[11] Monno et al., A practical one-shot multispectral imaging system using a single image sensor. IEEE Trans. Image Processing, pp 3048-3059, 2015
[12] Arad et al. Sparse Recovery of Hyperspectral Signal from Natual RGB Images, in Europ. Conf. on Comp. Vision, 2016
[13] Khan et al., HyTexiLa: High Resolution Visible and Near Infrared Hyperspectral Texture Images. Sensors 2018
[14] Prasad et al. Metrics and statistics of frequency of occurrence of metamerism in consumer cameras for natural scenes, in J. Optical Soc. America, pp. 1390-1402, 2015
[15] Rasch et al., The two-sample t test: pre-testing its assumptions does not pay off, Statistical Papers 52, p. 219-31, 2011