Joint Camera Spectral Sensitivity Selection and Hyperspectral Image Recovery

Ying Fu[1[0000−0002−6677−694X]], Tao Zhang[1[0000−0002−7358−0603]], Yinqiang Zheng[2[0000−0001−7434−5069]], Debing Zhang[3[0000−0003−4048−0531]], and Hua Huang[1[0000−0003−2587−1702]]

[1] Beijing Laboratory of Intelligent Information Technology, School of Computer Science and Technology, Beijing Institute of Technology, Beijing, 100081,China
[2] National Institute of Informatics, Tokyo, 101-8430, Japan
[3] DeepGlint, Beijing, 100091, China

Abstract. Hyperspectral image (HSI) recovery from a single RGB image has attracted much attention, whose performance has recently been shown to be sensitive to the camera spectral sensitivity (CSS). In this paper, we present an efficient convolutional neural network (CNN) based method, which can jointly select the optimal CSS from a candidate dataset and learn a mapping to recover HSI from a single RGB image captured with this algorithmically selected camera. Given a specific CSS, we first present a HSI recovery network, which accounts for the underlying characteristics of the HSI, including spectral nonlinear mapping and spatial similarity. Later, we append a CSS selection layer onto the recovery network, and the optimal CSS can thus be automatically determined from the network weights under the nonnegative sparse constraint. Experimental results show that our HSI recovery network outperforms state-of-the-art methods in terms of both quantitative metrics and perceptive quality, and the selection layer always returns a CSS consistent to the best one determined by exhaustive search.

Keywords: Camera spectral sensitivity selection, hyperspectral image recovery, spectral nonlinear mapping, and spatial similarity

1 Introduction

Compared with ordinary panchromatic and RGB images, the hyperspectral image (HSI) of the natural scene can effectively describe the spectral distribution and provide intrinsic and discriminative spectral information of the scene. It has been proven beneficial to numerous applications, including segmentation [43], classification [49], anomaly detection [50], face recognition [39], document analysis [28], food inspection [48], surveillance [37], earth observation [6], to name a few.

Hyperspectral cameras are widely used for HSI acquisition, which needs to densely sample the spectral signature across consecutive wavelength bands for every scene point. Such devices often come with a high cost and tend to suffer from the degradation of spatial/temporal resolution.

Recently, some methods [38, 3, 42, 22] have been presented to directly recover the HSI from a single RGB image. Since the mapping from RGB to spectrum is three-to-many, some prior knowledge has been introduced. Examples include radial basis function network mapping [38], K-SVD based sparse coding [3], constrained sparse coding [42], and manifold-based mapping [22]. In particular, Jia *et al.* [22] disclose the nonlinear characteristics of natural spectra, and show that properly designing nonlinear mapping can significantly boost the recovery accuracy. Inspired by this observation, we propose a spectral convolutional neural network (CNN) to better approximate the underlying nonlinear mapping. In contrast to the pixel-wise operation in [38, 3, 22], we further propose to better utilize the spatial similarity in HSI via a properly designed spatial CNN. In addition, the input RGB image is employed to guide the HSI reconstruction and residual learning is used to further preserve the spatial structure in our network. Experimental results show that our recovery network outperforms state-of-the-art methods in terms of quantitative metrics and perceptive quality, and both the spectral CNN module and the spatial CNN module have contributed to this performance gain.

Existing methods [38, 3, 42, 22] mainly focus on the HSI recovery under a given camera spectral sensitivity (CSS) function, while [4] shows that the quality of spectral recovery is sensitive to the CSS used. For example, given a CSS dataset, the optimal CSS selection may improve the accuracy by 33%, as shown in [4]. Rather than using exhaustive search, an evolutionary optimization methodology is used in [4] to choose the optimal CSS, which still needs to train the recovery method for multiple times and results in high time complexity.

Through experiments, we have found that the performance of our CNN-based recovery method is dependent on the CSS as well. This motivates us to develop a CNN-based CSS selection method with a single training process, which can jointly work with our HSI recovery method and have low time complexity. In this work, we propose a novel CSS selection layer, which can automatically determine the optimal CSS from the network weights under nonnegative sparse constraint. As illustrated in Figure 1, this filter selection layer is appended to the recovery network, which jointly selects the proper CSS and learns the mapping for HSI recovery from a single RGB image captured with the algorithmically selected CSS. Experiment results show that the selection layer always gives a CSS that is consistent with the best one determined by exhaustive search. The spectral recovery accuracy can be further boosted by this optimal CSS, compared with using a casually selected CSS. To the best of our knowledge, this work is the first to integrate optimal CSS selection with HSI recovery via a unified CNN-based framework, which boosts HSI recovery fidelity and has much lower complexity.

Our main contributions are that we

- Design a CNN-based HSI recovery network to account for spectral nonlinear mapping and utilize spatial similarity in the image plane domain;
- Develop a CSS selection layer to retrieve the optimal CSS on the basis of the nonnegative sparse constraint onto the weight factors;

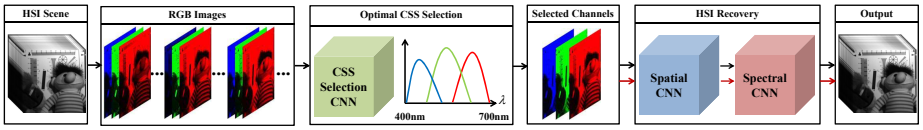


Fig. 1. Overview of the proposed method, which combines optimal CSS selection and HSI recovery into a unified CNN-based framework. The parameters for these two aspects are jointly learned first. Then, the RGB images are captured under the selected optimal CSS as the inputs and the underlying HSIs are reconstructed by using the HSI recovery network. The black arrow shows the training process and the red arrow denotes the testing process.

– Jointly determine the optimal CSS and learn an accurate HSI recovery mapping in a single training process.

2 Related Work

Hyperspectral imaging can effectively provide discriminative spectral information of the scene. To obtain HSIs, whiskbroom and pushbroom style scanning systems [41, 5] are widely used to capture the scene pointwisely or linewisely. RGB or monochromatic cameras with variant filters [51, 10, 9] or specific illuminations [40, 19] were also used to capture the HSI. But all these methods scanning along the spatial or spectral dimension result in low temporal resolution. To capture dynamic scenes, snapshot hyperspectral cameras [14, 16, 15, 8] were developed to capture full 3D HSIs, but they sacrificed spatial resolution.

To obtain high-resolution HSIs in real time, some coding-based hyperspectral imaging approaches have been presented, relying on the compressive sensing (CS) theory. CASSI [17, 44] employed a coding aperture with the disperser to uniformly encode the spectral signals into 2D space. DCCHI [45, 46] incorporated a co-located panchromatic camera to collect more information simultaneously with the CASSI measurement. SSCSI [33] jointly encoded the spatial and spectral dimensions in a single gray image. Besides, fusion-based approaches [25, 1, 32, 31, 12, 11, 35] were presented. These approaches were based on a hybrid camera system, in which a low spatial resolution hyperspectral camera and a high spatial resolution RGB camera were mounted in a coaxial system. The captured two images could be fused into a high resolution HSI, which has the same spatial resolution as the RGB image and same spectral resolution as the input HSI. All these coding-based and fusion-based hyperspectral imaging systems demand either high precision optical design or expensive hyperspectral camera.

To avoid using the above mentioned specialized devices, i.e. multiple illuminations, filters, coding aperture and hyperspectral cameras, HSI recovery from a single RGB image has attracted more attention. The spectral recovery from three values provided by the RGB camera arouses a three-to-many mapping, which is severely underdetermined in general.

To unambiguously determine the spectrum, some prior knowledge on the mapping is introduced. Nguyen *et al.* [38] learned the mapping between white balanced RGB values and illumination-free spectral signals based on a radial basis function network. Arad and Ben-Shahar [3] built a large hyperspectral dataset for natural scenes, and derived the mapping between hyperspectral signatures and their RGB values under a dictionary learned by K-SVD. Robles-Kelly [42] reconstructed the illumination-free HSI based on a constrained sparse coding approach by using a set of prototypes extracted from the training set. Jia *et al.* [22] proposed a two-step manifold-based mapping method, which highlighted the role of nonlinear mapping in spectral recovery. In this work, we present a spectral CNN module to better account for spectral nonlinear mapping, and a spatial CNN module to further incorporate the spatial similarity.

Arad and Ben-Shahar [4] first recognized that the quality of HSI recovery from a single RGB image was sensitive to the CSS selection. To avoid the heavy computational cost of exhaustive search, they proposed an evolutionary optimization based selection strategy. However, the training has still to be conducted multiple times under different CSS instances. In this work, we propose a CSS selection layer under the nonnegative sparse constraint, and jointly select the optimal CSS and learn the mapping for HSI recovery via a unified CNN-based framework. This can be achieved in only one training process, in contrast to repeated training operations in [4].

3 Joint Optimal CSS Selection and HSI Recovery

In this section, we present a CNN-based method for simultaneous optimal CSS selection and HSI recovery from a single RGB image. The overall framework of the proposed method is shown in Figure 1. In the training stage, given a large set of CSS functions and HSIs, we first synthesize multiple RGB images for each HSI under variant CSS functions, which are the input of the network. The designed optimal CSS selection network is utilized to select the best CSS and the corresponding RGB channels. In the HSI recovery network, we design a spectral CNN to approximate the complex nonlinear mapping between the RGB space and the spectra space, and a spatial CNN for the spatial similarity. The CSS selection network and the spectral recovery network are combined to recover the HSI, which should be close enough to its corresponding HSI in the training dataset. In the testing stage, the input RGB image is obtained under the selected CSS. A HSI will be obtained by feeding this input RGB image into the recovery network, which has been learned in the training stage.

In the following, we first describe the motivation of our network structure by digesting common approaches for HSI recovery from a single RGB image. Then, we introduce our CNN-based method for both HSI recovery and optimal CSS selection. Finally, the learning detail is provided.

3.1 Preliminaries and Motivation

Let $\mathbf{Y} \in \mathbb{R}^{3\times M}$ and $\mathbf{X} \in \mathbb{R}^{B\times M}$ denote the input RGB image and the recovered HSI, where M and B are the number of pixels and bands in the HSI. The relationship between $\mathbf{Y}$ and $\mathbf{X}$ can be described as

$$\mathbf{Y} = \mathbf{C}\mathbf{X}, \tag{1}$$

where $\mathbf{C} \in \mathbb{R}^{3\times B}$ denotes the RGB CSS function.

Most state-of-the-art methods assume that the CSS function is known and model HSI recovery from a single RGB image as

$$E(\mathbf{X}) = E_d(\mathbf{X}, \mathbf{Y}) + \lambda E_s(\mathbf{X}), \tag{2}$$

where the first term $E_d(\mathbf{X}, \mathbf{Y})$ is the data term, and it guarantees that the recovered $\mathbf{X}$ should be projected to $\mathbf{Y}$ under the CSS function $\mathbf{C}$. The second term $E_s(\mathbf{X})$ is the prior regularization for $\mathbf{X}$.

The models for the first term in the previous works [38, 3, 42, 22] can be generally described as

$$E_d(\mathbf{X}, \mathbf{Y}) = \|f_d(\mathbf{X}) - \mathbf{Y}\|_F^2. \tag{3}$$

where the function f_d is linear mapping for [3, 42] and spectral nonlinear mapping for [38, 22]. [22] shows that the nonlinear mapping can effectively assist HSI recovery, compared with the linear constraint in [3].

In addition, [3, 42] assume that the spectra can be sparsely described by several bases, which means

$$E_s(\mathbf{X}) = \|\mathbf{D}\boldsymbol{\alpha} - \mathbf{X}\|_F^2 + \|\boldsymbol{\alpha}\|_1, \tag{4}$$

where $\mathbf{D}$ is the learned spectral dictionary and $\boldsymbol{\alpha}$ is the corresponding spectral sparse coefficient. [38] and [22] implicitly assume that the spectral information lies in a low dimensional space in the model.

Furthermore, since the neighboring pixels in the recovered HSI $\mathbf{X}$ should be similar, [42] also has the spatial constraint as

$$E_s(\mathbf{X}) = \|f_s(\mathbf{X})\|_F^2, \tag{5}$$

where the function f_s denotes the local spatial operation.

According to these analyzes, we present a CNN-based HSI recovery method from a single RGB image, which can effectively learn the nonlinear spectral mapping and the spatial structure information to improve the recovered HSI.

Besides, from Equation (1), we can see that the quality of recovered HSI $\mathbf{X}$ is influenced by both the input RGB image $\mathbf{Y}$ and the CSS function $\mathbf{C}$. Meanwhile, [4] shows that the selection of CSS significantly affects the quality of HSI recovery. To boost the accuracy of recovered HSI, it is essential to select the optimal CSS as well. Therefore, our method models the optimal CSS selection and the HSI recovery via a unified CNN-based framework.

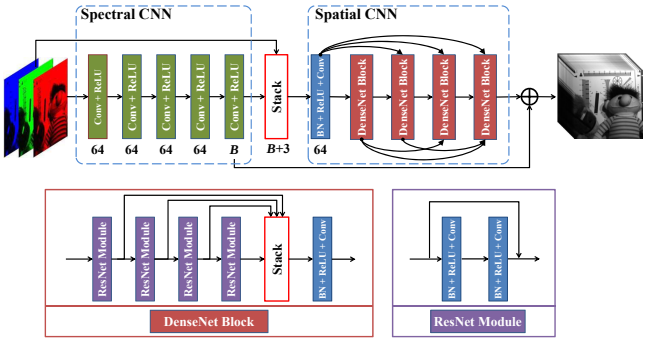


Fig. 2. The architecture of CNN-based HSI recovery from a single RGB image.

3.2 HSI Recovery

Previous works have shown that effectively exploiting the underlying characteristics of the HSI — spectral nonlinear mapping [38, 22] and spatial similarity [42] — can reconstruct high quality HSI from a single RGB image. Compared with these approaches, our method utilizes multiple CNN layers in spectral CNN to deeply learn the nonlinear mapping between spectra and RGB space, and employs DenseNet blocks [21] and ResNet modules [20] in the spatial CNN to enlarge the receptive field and obtain more spatial similarity in the space domain. Besides, our method uses the input RGB image to further guide the HSI recovery and residual learning to preserve the spatial structure. Figure 2 shows the HSI recovery network.

Spectral CNN Previous works for the spectral recovery from a single RGB image [38, 3, 42, 22] mainly consider the spectral mapping between the input RGB image and the recovered HSI. It is well-known that CNN can effectively learn the nonlinear mapping. Thus, we design a spectral CNN to learn the spectral nonlinear mapping between the RGB values and the corresponding spectrum, which consists of L layers. The output of the l-th layer is expressed as

$$\mathbf{F}_l = \mathrm{ReLU}(\mathbf{W}_l * \mathbf{F}_{l-1} + \boldsymbol{b}_l), \quad \text{where} \quad \mathbf{F}_0 = \mathbf{Y}, \tag{6}$$

where $\mathrm{ReLU}(x) = \max\{x, 0\}$, denoting a rectified linear unit [36]. $\mathbf{W}_l$ and $\boldsymbol{b}_l$ represent the filters and biases for the l-th layer, respectively. For the first layer, we compute a_0 feature maps using an $s_1 \times s_1$ receptive field, where $a_0 = 64$ in our network. Filters are of size $3 \times s_1 \times s_1 \times a_0$, when the input is an RGB image. In the 2-nd to $(L-1)$-th layers, we also compute a_0 feature maps using an $s_1 \times s_1$ receptive field and a rectified linear unit. The filters are of size $a_0 \times s_1 \times s_1 \times a_0$. Finally, the last layer uses the same receptive field and the filters are of size $a_0 \times s_1 \times s_1 \times B$. In the experiments, we set $L = 5$.

To perform the spectral nonlinear mapping, $s_1 = 1$ and the receptive field is 1×1 in the spatial domain. It implies that only the spectral nonlinear mapping is learned without any spatial structure.

RGB Guidance Many researches for pan-sharping [2, 52] employ the panchromatic image to preserve the structural information, as the two input images should have similar spatial structure for the same scene. Inspired by this, we use the input RGB image to guide the spatial information reconstruction, which is modeled by stacking the input RGB image and the initialized HSI $\mathbf{F}_L$ from the spectral CNN mentioned above. Thus, the output of the $(L+1)$-th layer can be expressed as

$$\mathbf{F}_{L+1} = \mathcal{C}(\mathbf{W}_{L+1} * \text{stack}(\mathbf{Y}, \mathbf{F}_L) + \boldsymbol{b}_{L+1}), \tag{7}$$

where $\mathcal{C}$ denotes the activation function. It is batch normalization (BN) [34] followed by a ReLU [36].

Spatial CNN Due to abundant self-repeating patterns in natural images [7, 13], the spatial information is usually similar in the neighboring area. To effectively exploit the spatial similarity, we need to obtain abundant spatial correlation in much larger area, which can be carried out by using several DenseNet blocks [21]. We employ N DenseNet blocks into the designed network, and the output of the n-th DenseNet block can be expressed as

$$\mathbf{S}_n = \mathcal{D}_n(\text{stack}(\mathbf{S}_0, \cdots, \mathbf{S}_{n-1})), \text{ where } \mathbf{S}_0 = \mathbf{F}_{L+1}, \tag{8}$$

where $\mathcal{D}_n$ denotes the n-th DenseNet block function.

In each DenseNet block, there are K ResNet modules [20]. For the n-th DenseNet block, the input is $\mathbf{S}_{n-1}$ and the k-th ResNet module can be expressed as

$$\mathbf{H}_n^k = \mathcal{R}_n(\mathbf{H}_n^{k-1}) + \mathbf{H}_n^{k-1}, \text{ where } \mathbf{H}_n^0 = \mathbf{S}_{n-1}. \tag{9}$$

In our spatial CNN, we set $N = 4$ and $K = 4$.

In addition, since the spatial structural information mainly exists in the high-pass components, we employ the residual learning to efficiently reconstruct detail information like [27]. Thus, the final output can be described as

$$\hat{\mathbf{X}} = \mathbf{S}_N + \mathbf{F}_L. \tag{10}$$

3.3 Optimal CSS Selection

Previous work [4] shows that the quality of HSI recovery is sensitive to the CSS used to generate the input RGB image. As will be shown in Section 4.3, we perform our HSI recovery method by using the synthetic RGB images under different CSS functions in a brute force way. From Figure 5, we can see that the accuracy of HSI recovery has about 10%∼25% difference. It means that properly choosing a CSS will contribute to the HSI recovery accuracy as well.

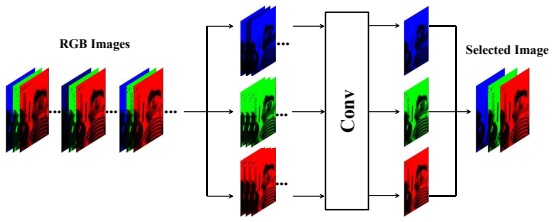


Fig. 3. The illustration of the optimal CSS selection.

The exhaustive search method and the evolutionary optimization method in [4] are not appropriate for our CNN-based HSI recovery method, since they need to train the HSI recovery network multiple times, which is extremely slow. Thus, we propose to design a selection convolution layer to retrieve the optimal CSS, which only needs to train once and has much lower time complexity.

To select the optimal CSS function, RGB images for each HSI are first synthesized with all CSS functions in a candidate dataset. Let $\mathbf{C}_j$ $(j = 1, \cdots, J)$ denote the j-th CSS function. Each synthetic RGB image via the j-th CSS function and the t-th HSI in the training dataset can be described as

$$\mathbf{Y}_{j,t} = \mathbf{C}_j \mathbf{X}_t. \tag{11}$$

Thus, for each scene, the input RGB images are obtained by stacking all RGB images under different CSS functions, i.e.,

$$\mathcal{Y}_t = \text{stack}(\mathbf{Y}_{1,t}, \cdots, \mathbf{Y}_{j,t}, \cdots, \mathbf{Y}_{J,t}). \tag{12}$$

Since our work mainly focuses on the HSI recovery from a single RGB image, we will select a CSS from an existing camera, without considering the combination of channels from different cameras. Note that the simulated RGB image cannot be negative, so the weight to select CSS should be nonnegative. To correctly localize the promising CSS, we further add a sparsity constraint. Thus, we design a convolution layer for the optimal CSS selection, which acts as the largest weight learned in the network under the nonnegative sparse constraint. To enforce the nonnegative constraint, the weights in this convolution layer for CSS selection are set to be positive. As for the sparse constraint, the weights are learned under the sparsity promoting l_1-norm.

The optimal CSS selection is equivalent to an RGB image selection in $\mathcal{Y}_t$, which is synthesized with the selected optimal CSS. As shown in Figure 3, we first separate RGB bands into three branches, which share the same convolution filter $\boldsymbol{V}$. The size of this filter is $J \times 1 \times 1 \times a_1$, where $a_1 = 1$ is the number of the output band for each branch. Thus, the output of this optimal CSS selection network can be expressed as

$$\hat{\mathbf{Y}}_t = \text{stack}(\boldsymbol{V} * \mathcal{Y}_t(R), \boldsymbol{V} * \mathcal{Y}_t(G), \boldsymbol{V} * \mathcal{Y}_t(B)), \tag{13}$$

where $\mathcal{Y}_t(R)$, $\mathcal{Y}_t(G)$, and $\mathcal{Y}_t(B)$ denote all the red, green, and blue channels in $\mathcal{Y}_t$, respectively.

The values in $\boldsymbol{V}$ can be determined by minimizing the mean squared error (MSE) under the nonnegative sparse constraint between the selected RGB image $\hat{\mathbf{Y}}$ and the corresponding ground truth image

$$\mathcal{L}_c(\boldsymbol{V}) = \frac{1}{T}\sum_{t=1}^{T} \|\hat{\mathbf{Y}}_t(\boldsymbol{V}) - \mathbf{Y}_t\|^2 + \|\mathbf{V}\|_1, \;\; s.t. \;\; \mathbf{V} \geq 0, \tag{14}$$

where $\hat{\mathbf{Y}}_t$ is the t-th output, $\mathbf{Y}_t$ is the t-th corresponding selected optimal CSS, and T is the number of training samples. A larger value in $\boldsymbol{V}$ represents that its corresponding CSS is better for HSI recovery. Consequently, the CSS corresponding to the largest value in $\boldsymbol{V}$ is selected as the optimal CSS.

3.4 Learning Details

The parameters for the HSI recovery network are denoted as $\boldsymbol{\Theta}$, and can be achieved by minimizing the MSE between the reconstructed HSI $\hat{\mathbf{X}}$ and the corresponding ground truth image,

$$\mathcal{L}_s(\boldsymbol{\Theta}) = \frac{1}{T}\sum_{t=1}^{T} \|\hat{\mathbf{X}}_t(\hat{\mathbf{Y}}_t, \boldsymbol{\Theta}) - \mathbf{X}_t\|^2 + \|\boldsymbol{\Theta}\|_2^2, \tag{15}$$

where $\hat{\mathbf{X}}_t$ is the t-th output, $\mathbf{X}_t$ is the corresponding ground truth, and $\hat{\mathbf{Y}}_t$ is the corresponding selected RGB image by the CSS selection network.

In our method, since the output of CSS selection network in Equation (14) is the input of HSI recovery network, it depends on the HSI recovery network training. Thus, we first append the optimal CSS selection network onto the HSI recovery network to select the optimal CSS and learn the mapping for spectral recovery together. In this joint training phase, we train the entire network by minimizing the loss

$$\mathcal{L} = \mathcal{L}_c(\boldsymbol{V}) + \tau\mathcal{L}_s(\boldsymbol{\Theta}), \tag{16}$$

where τ is a predefined parameter. Please note that $\mathbf{Y}_t$ needs not to be explicitly labeled in this joint training process, and thus $\|\hat{\mathbf{Y}}_t(\boldsymbol{V}) - \mathbf{Y}_t\|^2$ can be ignored in Equation (14). Then, the CSS corresponding to the largest value in $\boldsymbol{V}$ is selected as the optimal CSS, which is used to synthesize RGB images. These synthetic RGB images act as the input of the recovery network to recover HSIs.

The loss is minimized with the adaptive moment estimation method [29]. For all designed network modules, we set the mini-batch size to 16, momentum parameter to 0.9, and weight decay to 10^{-4}. To fit the nonnegative constraint for the optimal CSS selection, its convolution layer's weights are initialized as random positive numbers and all negative weights are set to zero during the forward and backward propagation. In the HSI recovery network, all convolution layer's weights are initialized by the method in [18]. The network has been trained with the deep learning tool Caffe [23] on a NVIDIA Titan X GPU.

4 Experimental Results

In the following, we will first introduce the datasets used for training and testing of all methods, and the metrics for quantitative evaluation. Then, we compare our method with several state-of-the-art HSI recovery methods under a typical CSS. In addition, the effectiveness of our optimal CSS selection method is evaluated on two CSS datasets.

4.1 Datasets and Metrics

We evaluate our joint CSS selection and CNN-based HSI recovery from a single RGB image on three public hyperspectral datasets, including the ICVL dataset [3], the NUS dataset [38], and the Harvard dataset [9]. The ICVL dataset consists of 201 images, which is by far the most comprehensive natural hyperspectral dataset. We randomly select 101 images in this dataset for training and use the rest for testing. The NUS dataset contains 41 HSIs in the training set and 25 HSIs in the testing set. The Harvard dataset consists of 50 outdoor images captured under daylight illumination. We remove those 6 images with strong highlights, and randomly use 35 images for training and 9 images for testing. All HSIs in these datasets have 31 bands. Two CSS datasets are used to evaluate the optimal CSS selection. The first dataset [24] contains 28 CSS curves and the second dataset [26] contains 12 CSS curves. Both datasets cover different camera types and brands.

We uniformly extract the patch pairs from each HSI and its corresponding RGB images under variant CSS functions with the size of 64×64 and the stride of 61. We randomly select 90% pairs for training and 10% pairs for validation. The RGB patches are regarded as the network's input and the HSI patches are regarded as the ground truth.

Three image quality metrics are utilized to evaluate the performance of all methods, including root-mean-square error (RMSE), structural similarity (SSIM) [47], and spectral angle mapping (SAM) [30]. RMSE and SSIM are calculated on each 2D spatial image, which measure the spatial fidelity between the recovered HSI and the ground truth. SAM is calculated on the 1D spectral vector, which shows the spectral fidelity. Smaller values of RMSE and SAM suggest better performance, while a larger value of SSIM implies better performance.

4.2 Evaluation on HSI Recovery

Here, we first compare our CNN-based HSI recovery method with three state-of-the-art HSI recovery methods from a single RGB image under the known CSS, including radial basis function network based method (RBF) [38], the sparse representation based method (SR) [3], and manifold-based mapping (MM) [22]. The original HSIs in datasets serve as ground truth. To fairly compare with these methods, we use the CSS function of Canon 5D Mark II to synthesize RGB values, which is the same as [22].

Table 1. RMSE, SSIM, and SAM results for different HSI recovery methods on there HSI datasets.

Dataset	Metrics	RBF	SR	MM	Ours
ICVL	RMSE	7.7152	3.0223	2.1245	**1.3533**
	SSIM	0.9546	0.9582	0.9946	**0.9975**
	SAM	0.1419	0.0645	0.0470	**0.0277**
NUS	RMSE	14.8785	8.9766	6.1825	**5.2426**
	SSIM	0.8648	0.8701	0.9555	**0.9649**
	SAM	0.3239	0.2358	0.2114	**0.1712**
Harvard	RMSE	11.8101	4.9534	4.9616	**2.1923**
	SSIM	0.9539	0.9126	0.9541	**0.9924**
	SAM	0.1723	0.1848	0.2273	**0.0956**

Table 2. Time complexity for different HSI recovery methods. (Unit: second)

Size	RBF	SR	MM	Ours(CPU)	Ours(GPU)
$256 \times 256 \times 31$	0.20	2.08	4.13	18.94	**0.09**
$512 \times 512 \times 31$	0.69	8.38	16.27	75.87	**0.36**
$1024 \times 1024 \times 31$	2.55	33.36	65.67	299.04	**1.58**

Table 1 provides the average results over all HSIs in the testing sets from three HSI datasets, for quantitative comparison of RBF, SR, MM and our method. The best results are highlighted in bold. We observe that MM [22] outperforms the other two methods. The reason is that MM effectively approximates the spectral nonlinear mapping between the RGB values and spectral signatures, compared with RBF and SR. This demonstrates that the spectral nonlinear mapping is much relevant for HSI recovery from a single RGB image. Our method provides substantial improvements over all these methods, in terms of RMSE, SSIM and SAM. This reveals the advantages of deeply exploiting the intrinsic properties of HSIs and verifies the effectiveness of our HSI recovery network.

To visualize the experimental results for all methods, several representative recovered HSIs and the corresponding recovered spectral errors on three datasets are shown in Figure 4. The ground truth, our results, error images for RBF/SR/MM/our methods, and RMSE results along spectra for all methods are shown from top to bottom. The ground truth and our results are the 16-th band for all scenes. The error images are the average absolution errors between the ground truth and the recovered results across spectra. We can observe that the recovered images from our method are consistently more accurate for all scenes, which verifies that our method can provide higher spatial accuracy. The RMSE results along spectra for all methods show that the results of our method are much closer to the ground truth than other compared methods along spectra, which demonstrates that our approach obtains higher spectral fidelity.

The average testing time among 10 independent trials for HSIs with different sizes of all compared methods is included in Table 2. All results performed on an Intel Core i7-6800K CPU are provided. We can see that our method have

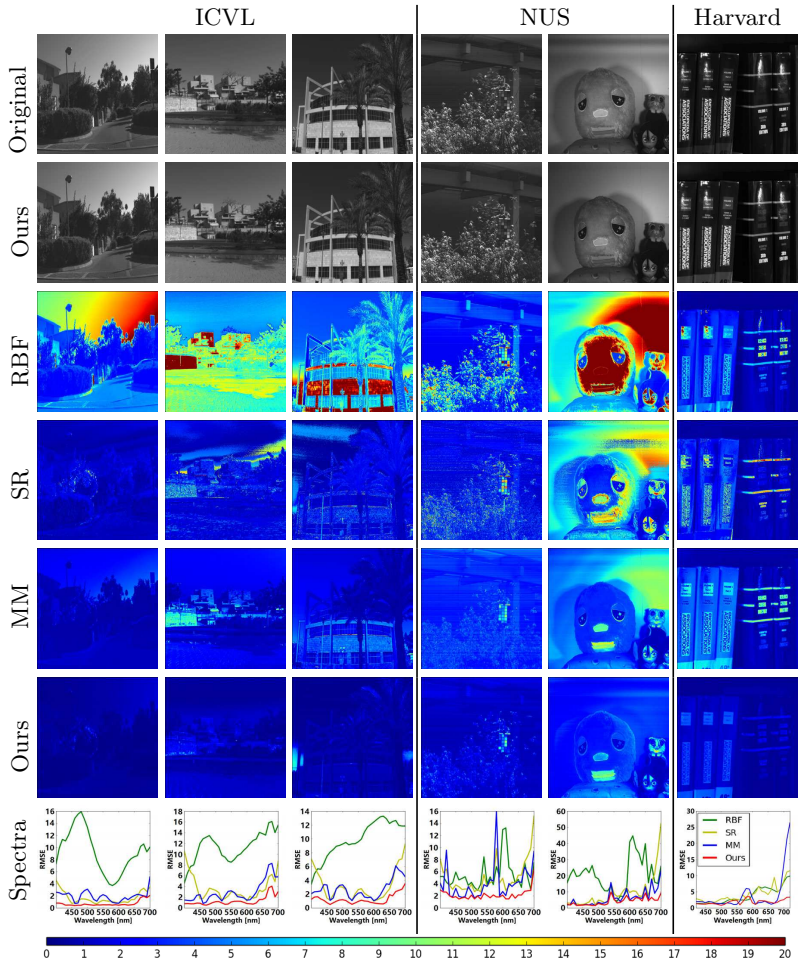
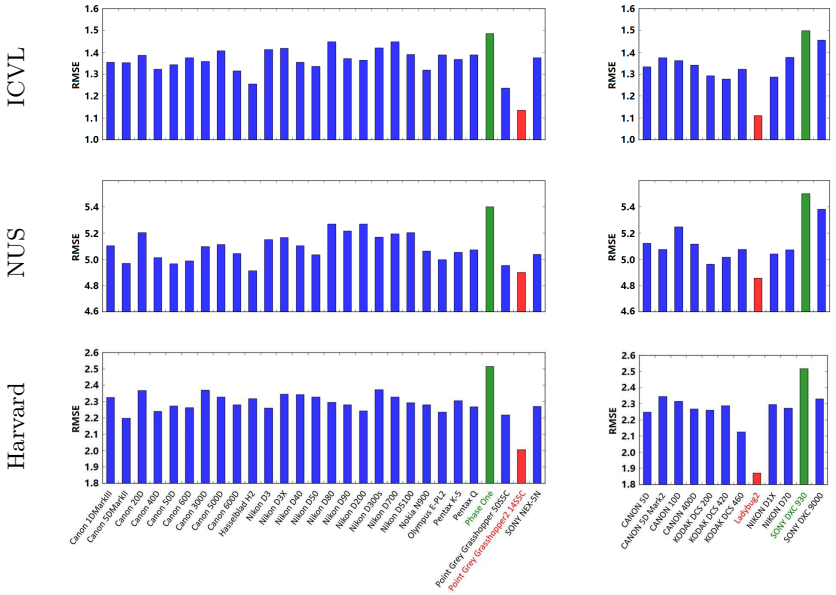


Fig. 4. Visual quality comparison on six typical scenes in HSI datasets. The ground truth, recovered HSI by our method, the error map for RBF/SR/MM/our results, and RMSE results along spectra for all methods are shown from top to bottom.

higher time complexity on CPU, yet the running time on GPU is much shorter. Compared with the other methods, our method on GPU can reconstruct HSIs more than 10 frames per second for size of $256 \times 256 \times 31$.

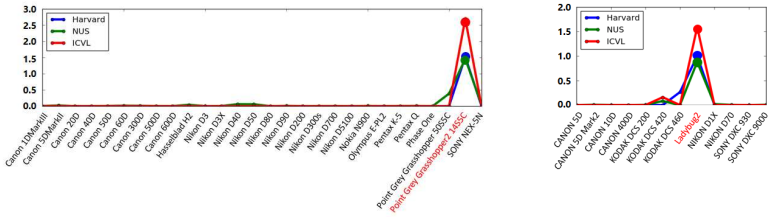
4.3 Evaluation on CSS Selection

To evaluate the effect of CSS functions in our spectral recovery network, we have conducted experiments to evaluate the performance of HSI recovery on both CSS datasets [24, 26]. First, we perform all methods on the synthetic RGB images by different CSS functions in a brute force way. As shown in Figure 5, we



(a) The CSS in [24] (b) The CSS in [26]

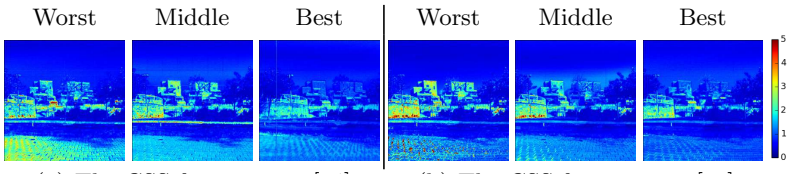
Fig. 5. The RMSE results of our HSI recovery network on three HSI and two CSS datasets. The red and green bars indicate the best and worst CSS functions for the HSI recovery in a brute force way, respectively.



(a) The CSS in [24] (b) The CSS in [26]

Fig. 6. The selected optimal CSS by our method on three HSI datasets.

can indeed observe that our method is also dependent on the CSS selection. In spite of that, our method is still superior even with an improper CSS. To obtain an optimal CSS for improved HSI recovery by training once, our method uses a convolutional layer to select the optimal CSS. In Figure 6, we can see that our method can effectively select the optimal CSS, which is consistent with the best one determined by exhaustive search. In addition, the selected CSS keeps the same for all three HSI dataset, which seems to indicate that this CSS properly encodes the intrinsic spectral information of the physical world.



(a) The CSS functions in [24] (b) The CSS functions in [26]

Fig. 7. Visual quality comparison under different CSS functions on a typical scene. The recovered HSIs under the worst/middle/best CSS functions in [24] and [26] are shown in (a) and (b), respectively.

Figure 7 shows the HSI recovery results of a typical scene under different CSS functions. The recovered HSIs under the worst/middle/best CSS functions in [24] and [26] in terms of Figure 5 are shown in (a) and (b), respectively. We can see that the results under the selected optimal CSS are much close to the ground truth, which further demonstrates the effectiveness of joint optimal CSS selection and the accuracy of the learned CNN nonlinear mapping in HSI recovery.

5 Conclusion

In this paper, we have presented an effective CNN-based method to jointly select the optimal camera spectral sensitivity function and learn an accurate mapping to reconstruct hyperspectral image from an RGB image. We first propose a spectral recovery network with properly designed modules to account for the underlying characteristics of the HSI, including spectral nonlinear mapping and spatial similarity. Meanwhile, a camera spectral sensitivity selection layer is developed to append onto the recovery network, which can automatically retrieve the optimal sensitivity functions by using the nonnegative sparse constraint. Experimental results show that our method can provide substantial improvements over the current state-of-the-art methods in terms of both objective metric and subjective visual quality.

Our current network selects the optimal sensitivity from off-the-shelf cameras. Therefore, there is no need to produce a new filter array, which is known to be extremely expensive. However, filter array makers are indeed able to produce novel filter arrays, which might be better for this spectral recovery task than all existing RGB cameras. It is thus worth investigating the limitations of current filter manufacturing techniques and exploring how to incorporate these limitations in a more relaxed filter response designing (rather than selection) process.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grants No. 61425013 and No. 61672096.

References

1. Akhtar, N., Shafait, F., Mian, A.: Sparse spatio-spectral representation for hyperspectral image super-resolution. In: Proc. of European Conference on Computer Vision (ECCV). pp. 63–78 (Sep 2014)
2. Aly, H.A., Sharma, G.: A regularized model-based optimization framework for pan-sharpening. IEEE Trans. Image Processing **23**(6), 2596–2608 (2014)
3. Arad, B., Ben-Shahar, O.: Sparse recovery of hyperspectral signal from natural rgb images. In: Proc. of European Conference on Computer Vision (ECCV). pp. 19–34 (Oct 2016)
4. Arad, B., Ben-Shahar, O.: Filter selection for hyperspectral estimation. In: Proc. of International Conference on Computer Vision (ICCV). pp. 3172–3180 (Oct 2017)
5. Basedow, R.W., Carmer, D.C., Anderson, M.E.: Hydice system: Implementation and performance. In: SPIE's Symposium on OE/Aerospace Sensing and Dual Use Photonics. pp. 258–267 (1995)
6. Bioucas-Dias, J.M., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N.M., Chanussot, J.: Hyperspectral remote sensing data analysis and future challenges. IEEE Geoscience and remote sensing magazine **1**(2), 6–36 (2013)
7. Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising. In: Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). vol. 2, pp. 60–65 (Jun 2005)
8. Cao, X., Du, H., Tong, X., Dai, Q., Lin, S.: A prism-based system for multispectral bideo acquisition. IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI) **33**(12), 2423–2435 (2011)
9. Chakrabarti, A., Zickler, T.: Statistics of real-world hyperspectral images. In: Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 193–200 (Jun 2011)
10. Chi, C., Yoo, H., Ben-Ezra, M.: Multi-spectral imaging by optimized wide band illumination. International Journal of Computer Vision (IJCV) **86**(2-3), 140–151 (2010)
11. Dian, R., Fang, L., Li, S.: Hyperspectral image super-resolution via non-local sparse tensor factorization. In: Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5344–5353 (Jun 2017)
12. Dong, W., Fu, F., Shi, G., Cao, X., Wu, J., Li, G., Li, X.: Hyperspectral image super-resolution via non-negative structured sparse representation. IEEE Trans. Image Processing **25**(5), 2337–2352 (May 2016)
13. Dong, W., Zhang, L., Shi, G.: Centralized sparse representation for image restoration. In: Proc. of International Conference on Computer Vision (ICCV). pp. 1259–1266 (Nov 2011)
14. Ford, B.K., Descour, M.R., Lynch, R.M.: Large-image-format computed tomography imaging spectrometer for fluorescence microscopy. Optics Express **9**(9), 444–453 (2001)
15. Gao, L., Kester, R.T., Hagen, N., Tkaczyk, T.S.: Snapshot image mapping spectrometer (ims) with high sampling density for hyperspectral microscopy. Optics Express **18**(14), 14330–14344 (2010)
16. Gat, N., Scriven, G., Garman, J., Li, M.D., Zhang, J.: Development of four-dimensional imaging spectrometers (4d-is). In: Proc. of SPIE Optics + Photonics. vol. 6302, pp. 63020M–63020M–11 (2006)
17. Gehm, M.E., John, R., Brady, D.J., Willett, R.M., Schulz, T.J.: Single-shot compressive spectral imaging with a dual-disperser architecture. Optics Express **15**(21), 14013–27 (2007)

18. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: Proc. of International Conference on Artificial Intelligence and Statistics. pp. 249–256 (May 2010)
19. Han, S., Sato, I., Okabe, T., Sato, Y.: Fast spectral reflectance recovery using DLP projector. International Journal of Computer Vision (IJCV) **110**(2), 172–184 (2014)
20. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In: Proc. of European Conference on Computer Vision (ECCV). pp. 630–645 (Oct 2016)
21. Huang, G., Liu, Z., Maaten, L.v.d., Weinberger, K.Q.: Densely connected convolutional networks. In: Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2261–2269 (Jul 2017)
22. Jia, Y., Zheng, Y., Gu, L., Subpa-Asa, A., Lam, A., Sato, Y., Sato, I.: From rgb to spectrum for natural scenes via manifold-based mapping. In: Proc. of International Conference on Computer Vision (ICCV). pp. 4715–4723 (Oct 2017)
23. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: Convolutional architecture for fast feature embedding. In: Proc. of ACM Multimedia Conference (MM). pp. 675–678 (Nov 2014)
24. Jiang, J., Liu, D., Gu, J., Ssstrunk, S.: What is the space of spectral sensitivity functions for digital color cameras? In: IEEE Workshop on Applications of Computer Vision (WACV). pp. 168–179 (2013)
25. Kawakami, R., Wright, J., Tai, Y.W., Matsushita, Y., Ben-Ezra, M., Ikeuchi, K.: High-resolution hyperspectral imaging via matrix factorization. In: Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2329–2336 (Jun 2011)
26. Kawakami, R., Zhao, H., Tan, R.T., Ikeuchi, K.: Camera spectral sensitivity and white balance estimation from sky images. International Journal of Computer Vision (IJCV) **105**(3), 187–204 (2013)
27. Kim, J., Lee, J.K., Lee, K.M.: Accurate image super-resolution using very deep convolutional networks. In: Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1646–1654 (Jun 2016)
28. Kim, S.J., Deng, F., Brown, M.S.: Visual enhancement of old documents with hyperspectral imaging. Pattern Recognition **44**(7), 1461–1469 (2011)
29. Kingma, D.P., Ba, J.L.: Adam: a method for stochastic optimization. In: Proc. of International Conference on Learning Representations (ICLR) (May 2015)
30. Kruse, F.A., Lefkoff, A.B., Boardman, J.W., Heidebrecht, K.B., Shapiro, A.T., Barloon, P.J., Goetz, A.F.H.: The spectral image processing system (sips)–interactive visualization and analysis of imaging spectrometer data. Remote Sensing of Environment **44**(2-3), 145–163 (May 1993)
31. Kwon, H., Tai, Y.W.: Rgb-guided hyperspectral image upsampling. In: Proc. of International Conference on Computer Vision (ICCV). pp. 307–315 (Dec 2015)
32. Lanaras, C., Baltsavias, E., Schindler, K.: Hyperspectral super-resolution by coupled spectral unmixing. In: Proc. of International Conference on Computer Vision (ICCV). pp. 3586–3594 (Dec 2015)
33. Lin, X., Liu, Y., Wu, J., Dai, Q.: Spatial-spectral encoded compressive hyperspectral imaging. ACM Trans. on Graph. (Proc. of SIGGRAPH Asia) **33**(6), 233:1–233:11 (Nov 2014)
34. Loffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. In: Proc. of International Conference on Machine Learning (ICML). pp. 448–456 (Jun 2015)

35. Ma, C., Cao, X., Tong, X., Dai, Q., Lin, S.: Acquisition of high spatial and spectral resolution video with a hybrid camera system. International Journal of Computer Vision (IJCV) **110**(2), 141–155 (Nov 2014)
36. Nair, V., Hinton, G.E.: Rectified linear units improve restricted boltzmann machines. In: Proc. of International Conference on Machine Learning (ICML). pp. 807–814 (Jun 2010)
37. Nguyen, H.V., Banerjee, A., Chellappa, R.: Tracking via object reflectance using a hyperspectral video camera. In: IEEE Conference on Computer Vision and Pattern Recognition - Workshops. pp. 44–51 (Jun 2010)
38. Nguyen, R.M.H., Prasad, D.K., Brown, M.S.: Training-based spectral reconstruction from a single rgb image. In: Proc. of European Conference on Computer Vision (ECCV). pp. 186–201 (Sep 2014)
39. Pan, Z., Healey, G., Prasad, M., Tromberg, B.: Face recognition in hyperspectral images. IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI) **25**(12), 1552–1560 (2003)
40. Park, J.I., Lee, M.H., Grossberg, M.D., Nayar, S.K.: Multispectral imaging using multiplexed illumination. In: Proc. of International Conference on Computer Vision (ICCV). pp. 1–8 (Oct 2007)
41. Porter, W.M., Enmark, H.T.: A system overview of the airborne visible/infrared imaging spectrometer (aviris). In: Annual Technical Symposium. pp. 22–31 (1987)
42. Robles-Kelly, A.: Single image spectral reconstruction for multimedia applications. In: Proc. of ACM Multimedia Conference (MM). pp. 251–260 (Oct 2015)
43. Tarabalka, Y., Chanussot, J., Benediktsson, J.A.: Segmentation and classification of hyperspectral images using watershed transformation. Pattern Recognition **43**(7), 2367–2379 (2010)
44. Wagadarikar, A., John, R., Willett, R., Brady, D.: Single disperser design for coded aperture snapshot spectral imaging. Applied Optics **47**(10), 44–51 (Apr 2008)
45. Wang, L., Xiong, Z., Gao, D., Shi, G., Wu, F.: Dual-camera design for coded aperture snapshot spectral imaging. Applied Optics **54**(4), 848–858 (Feb 2015)
46. Wang, L., Xiong, Z., Shi, G., Wu, F., Zeng, W.: Adaptive nonlocal sparse representation for dual-camera compressive hyperspectral imaging. IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI) **39**(10), 2104–2111 (Oct 2017)
47. Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Processing **13**(4), 600–612 (Apr 2004)
48. Wu, D., Sun, D.W.: Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A reviewpart i: Fundamentals. Innovative Food Science & Emerging Technologies **19**, 1–14 (2013)
49. Xu, X., Li, J., Huang, X., Dalla Mura, M., Plaza, A.: Multiple morphological component analysis based decomposition for remote sensing image classification. IEEE Trans. Geoscience and Remote Sensing **54**(5), 3083–3102 (2016)
50. Xu, X., Wu, Z., Li, J., Plaza, A., Wei, Z.: Anomaly detection in hyperspectral images based on low-rank and sparse representation. IEEE Trans. Geoscience and Remote Sensing **54**(4), 1990–2000 (2016)
51. Yamaguchi, M., Haneishi, H., Fukuda, H., Kishimoto, J., Kanazawa, H., Tsuchida, M., Iwama, R., Ohyama, N.: High-fidelity video and still-image communication based on spectral information: Natural vision system and its applications. In: Electronic Imaging. pp. 60620G–60620G–12 (2006)
52. Yang, J., Fu, X., Hu, Y., Huang, Y., Ding, X., John, P.: Pannet: A deep network architecture for pan-sharpening. In: Proc. of International Conference on Computer Vision (ICCV). pp. 1753–1761 (Oct 2017)