From Less to More: Spectral Splitting and Aggregation Network for Hyperspectral Face Super-Resolution

Junjun Jiang [†], Chenyang Wang [†], Xianming Liu [†], Kui Jiang [‡], Jiayi Ma [‡]
[†] Harbin Institute of Technology [‡] Wuhan University

Abstract

*High-resolution (HR) hyperspectral face image plays an important role in face related computer vision tasks under uncontrolled conditions, such as low-light environment and spoofing attacks. However, the dense spectral bands of hyperspectral face images come at the cost of limited amount of photons reached a narrow spectral window on average, which greatly reduces the spatial resolution of hyperspectral face images. In this paper, we investigate how to adapt the deep learning techniques to hyperspectral face image super-resolution (HFSR), especially when the training samples are very limited. Benefiting from the amount of spectral bands, in which each band can be seen as an image, we present a spectral splitting and aggregation network (SSANet) for HFSR with limited training samples. In the shallow layers, we split the hyperspectral image into different spectral groups. Then, we gradually aggregate the neighbor bands at deeper layers to exploit spectral correlations. By this spectral splitting and aggregation strategy (SSAS), we can divide the original hyperspectral image into multiple samples (*from less to more*) to support the efficient training of the network and effectively exploit the spectral correlations among spectrum. To cope with the challenge of small training sample size (S3) problem, we propose to expand the training samples by a self-representation model and symmetry-induced augmentation. Experiments show that SSANet can well model the joint correlations of spatial and spectral information. By expanding the training samples, SSANet can effectively alleviate the S3 problem.*

1. Introduction

Benefit from hyperspectral imagery which can capture the local spectral properties of human tissue, hyperspectral face analysis has attracted more and more attention from scholars in the field of face related computer vision tasks because of its satisfactory performance under uncontrolled conditions, such as low-light environment and spoofing attacks. However, the hyperspectral face imaging system is often compromised due to the limitations of the amount of the incident energy. There is always a tradeoff between the spatial and spectral resolution of the real imaging process. With the increase of the spectral bands, all other factors kept constant, to ensure a high signal-to-noise ratio (SNR) the spatial resolution will inevitably become a victim. Therefore, how to obtain a reliable hyperspectral face image with high spatial resolution remains a very challenging problem.

Super-Resolution (SR) reconstruction can infer a high-resolution (HR) image from a single low-resolution (LR) image or sequential observed LR images [1]. It is a post-processing technique that does not require hardware modifications, and thus could break through the limitations of the imaging systems. Since the pioneer work of Bake and Kanda [2], face SR, *a.k.i*, face hallucination, has received increased attention [3–6]. Especially in recent years, the emergence of deep learning technology has promoted a large number of face SR methods [7]. However, these methods are mainly for gray/RGB face images and cannot exploit the spectral correlation efficiently. This is mainly due to that the spectral dimensionality of hyperspectral face image is very high and the number of training samples of current hyperspectral face image dataset is extremely small, *e.g.*, around 100. There are not enough training samples to support the training of the complex deep network. In addition, these deep learning algorithms require large amounts of memory and are computationally expensive. Whether it is a traditional shallow learning-based method or a deep learning-based method, it is difficult to construct a powerful representation model from only dozens of images to reconstruct the target HR hyperspectral face images.

In summary, when we apply the existing learning-based face SR methods to hyperspectral images, the challenges are twofold: (i) the spectral correlations of hyperspectral face images cannot be fully utilized; (ii) hyperspectral face images are difficult to obtain, and the training samples of existing hyperspectral face image database is very small, thus cannot support these complex learning models.

To this end, in this paper we propose a hyperspectral face super-resolution (HFSR) method with limited training samples based on a spectral splitting and aggregation network

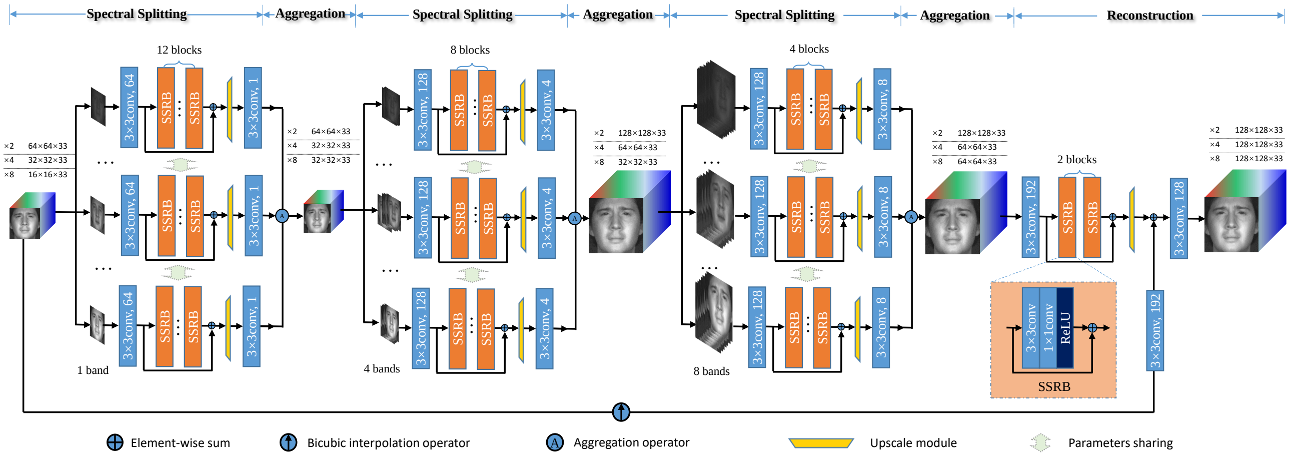

Figure 1. The flow chart of our proposed spectral splitting and aggregation network (SSANet) based hyperspectral face super-resolution (HFSR) method. SSANet includes three spectral splitting and aggregation modules followed by a reconstruction module.

(SSANet). Thanks to the amount of spectral bands of hyperspectral data, we split the hyperspectral image into different spectral groups and take each of them as an individual sample (in the sense that each group will be fed into the some network). By this splitting strategy, the training of the complex network can be guaranteed. To exploit the spectral correlations, which are destroyed by the splitting strategy but very important for the hyperspectral image reconstruction, we introduce an aggregation operator to gradually aggregate the neighbor bands at the deeper layers. As shown in Fig. 1, at the shallow layers, the complex network (with more blocks or channels) can be well-trained with 'adequate' data (each group can be seen as a sample because of the parameter sharing strategy). At deeper layer, we design some light structures (with less blocks or channels) because we have reduced the training samples in order to exploit the spectral correlations. In general, although we only have a small number of training samples, we can expect to alleviate the S3 problem through this carefully designed spectral splitting and aggregation network.

Additionally, in order to further cope with the challenge of S3 problem, we propose to expand the training samples by self-representation learning. In particular, based on the assumption that there is a path from the training sample itself to the mean face of all training samples, we develop a training sample expanding strategy by self-representation learning. Given one training sample, we leverage the remaining training samples to represent it. According to a predefined smooth parameter, we can obtain a large number of synthetic samples, thus forming a path from the training sample itself to the mean face of the training set. In addition, we also expand the training dataset by the symmetry of the human face, thus we can directly double the size of the training dataset.

To sum up, this paper makes the following contributions: (i) We propose the first SR method for hyperspectral face image. (ii) A novel spectral splitting and aggregation network is introduced to "generate" more training samples to alleviate the S3 problem *from less to more*, which is inevitable when we apply deep learning method to the HFSR task. (iii) We propose some schemes to expand the sample base based on self-representation and symmetry argumentation *from less to more*.

2. Related Work

Face SR is a very hot topic since the moment it was born. In recent years, a lot of methods have been proposed. In the following, we will revisit these approaches from global face based methods to local patch representation based methods, and then to the deep learning methods that have received much attention most recently. In addition, we also review some general hyperspectral image SR methods.

2.1. Global Model-based FSR

Face model based approaches leverage the global face statistical models, such as principal component analysis (PCA) [4], locality preserving projection (LPP) model [8], uniform space projection [9,10] and nonnegative matrix factorization (NMF) [11], to model the facial image and super-resolve the target HR face image globally. They can maintain the main structure of human face well. However, their reconstruction results lack detailed local facial features, and inevitably have ghosting artifacts.

2.2. Local Patch-based FSR

Considering that the human face is a highly structured subject, many face SR approaches try to exploit the prior knowledge by dividing the whole face image into small patches [12–14]. Among them, position-patch based methods have gained widespread attention in recent years. However, their solutions for the linear combination problem are unstable or not unique. To this end, our previous work [15] introduces the locality-constrained representation (LcR) technique to simultaneously incorporate sparsity and locality into the patch representation. With the introduced locality constraint, it obtains a stable and reasonable representation. In order to alleviate the inconsistency of the two spaces, LR and HR spaces, some work have been proposed to iteratively obtain the patch representation and perform neighbor embedding or learn the mapping in correlation spaces [16–21].

2.3. Deep Learning-based FSR

More recently, to reconstruct the latent HR image locally while thinking globally, DNNs, especially CNNs, have been applied to construct the mapping relationship between the LR images and their HR counterparts and shown strong learning capability and accurate prediction of HR images [22, 23]. For example, Dong *et al.* [22] developed a general image SR method based on SRCNN. This is the very first attempt to use deep learning tools for image SR reconstruction. The approach of Liu *et al.* [23] proposes to introduce the domain expertise to design a Sparse Coding based Network (SCN). Recently, R-DGN [24], CBN [25], LCGE [26], Attention-FH [27], FSRNet [28], and [29] are the most competitive approaches for face hallucination. They unitized very deep networks to model the relationship between the LR images and their HR counterparts, and verified that deeper networks can produce better results due to the large receptive field, which means considering more contextual information, *i.e.*, very large image regions.

2.4. Deep Learning-based HFSR

According to whether an additional guided image (such as panchromatic, RGB, or multispectral image) is utilized, hyperspectral image SR techniques can be divided into two categories: multi-source information fusion based hyperspectral image SR (sometimes called hyperspectral image pansharpening) and single hyperspectral image SR [30, 31]. The former is to leverage high-frequency spatial information from HR auxiliary image, and fuse them to the target HR hyperspectral image [32–35]. Though achieving good performance, they need a well co-registered auxiliary image which is arduous in real applications. Without co-registered auxiliary image, the latter single hyperspectral image SR methods have still attracted considerable attention. How to exploit the abundant spectral correlations among successive bands is the essential problem for them. Traditional methods try to incorporate the low-rank and group-sparse constraints to exploit the spectral-spatial prior [36, 37]. Due to its superior performance, deep learning techniques have also been introduced into the single hyperspectral image SR task. They often adopt a two-step strategy to get an intermediate result through a deep network firstly, and then use spectral decomposition constraint to ensure the accuracy of the reconstructed spectral information [38, 39]. Recently, there are also some approaches to directly learn an end-to-end deep networks to exploit the spatial-spectral prior. For example, Mei *et al.* [40] presented a 3D full convolutional neural network to represent the spectral-spatial information. Wang *et al*. [41] proposed to combine the advantages of 2D and 3D convolutions to exploit the spatial and spectral information. In [42], Li *et al.* developed a grouped deep recursive residual network (GDRRN) based single hyperspectral image SR method. The designed group-wise convolution and recursive structure can guarantee that it could yield very good performance. Inspired by the concept of group convolution, Jiang *et al*. [43] proposed the spatial-spectral prior network based super-resolution network (SSPSR). Most recently, inspired by the work of deep image prior [44], the approach of [45] presents an effective single hyperspectral image restoration algorithm. In general, these deep methods achieve better results than traditional methods. However, due to the limited hyperspectral training samples and the high dimensionality of spectral bands, it is difficult to fully exploit the spatial information and the correlation among the spectra of the hyperspectral data.

Summary: As for the shallow learning based face SR methods, the local patch based models have much more strong representation ability than global face based models, rendering effectively the fine individual details to an input LR face. To obtain an accurate representation of the image patch, the current patch representation based methods all try to exploit the image priors (*e.g*., collaborative, local, sparse, and low-rank constraints) or add some geometric regularizations from the HR space. They can well deal with the noisy input, inconsistency between LR and HR manifold spaces. These hand-designed prior models may not be effective. Deep neural network based methods leverage large-scale training dataset to train the network, and thus they can obtain very good performance. However, they may be not suitable for our task, where only about 100 samples are available. Although many SR methods for general hyperspectral images have been proposed in recent years, they all rely on a large-scale training set. Therefore, how to adapt the deep learning techniques for the HFSR task, especially when the training samples are very limited, is an urgent and extremely challenging issue.

3. The Proposed Method

Different from gray/RGB face SR problem, in which large paired HR and LR face training samples can be collected, HFSR can only leverage a small size of training dataset. Therefore, the very limited data cannot support the representation and modeling of complex hyperspectral data. To this end, in this paper we carefully design a deep network based on gradually spectral splitting and aggregation to alleviate the S3 problem. In addition, we also propose two strategies to expand the training samples by a self-representation model and the symmetry-induced augmentation. In this section, we will first present the details of SSANet and then introduce our training sample expanding strategies.

3.1. Spectral Splitting and Aggregation Network

Spectral splitting module: As we discussed above, because the dimensionality of hyperspectral images is very high and the training samples are very limited, it is very difficult for existing deep neural network based approaches to effectively represent the hyperspectral data. Thanks to the amount of spectral bands of hyperspectral data, where each spectral band (neighbor spectral bands) can be seen as a training sample, we split the hyperspectral image into different groups (samples), thus largely expanding the size of the original training dataset. Note that we let each split group be fed into a branch network, and let different branches share the same parameters. Through this parameter sharing based splitting strategy, it will provide a strong data support for training a deep network.

As shown in Fig. 1, at the shallow layer (*e.g.*, the first spectral splitting and aggregation stage) we split the hyperspectral data into as many groups as possible (a band is a group). In this way, we can obtain multiple training samples, which makes it possible to design a complex network (*i.e.*, many spatial-spectral residual blocks (SSRBs)) to extract features of hyperspectral data. In the spectral splitting network, each group will be fed to a branch network to extract the features.

Aggregation module: However, these extracted features may overlook the correlations among the spectral bands, which are very import to model and reconstruct the hyperspectral data. Therefore, at the deeper layers, we gradually increase the size (band number) of the group (*i.e.*, the band number of each group, from 1 to 4, and then to 8, and finally to all 33 bands for different spectral splitting and aggregation stages). Meanwhile, we also let the neighbor groups overlap with each other. When we feed these overlapped groups to the splitting branch networks, we obtain the output by an aggregation operator, where the output is averaged according to their spectral indices. It should be noted that before feeding the outputs of different branch networks to the following aggregation module, we apply an additional *Conv* layer to reduce the channel dimension to the number of input bands. Thus, each branch network acts like a "reconstruction" network and the input and output have the same channels. As we know, with the increase of the group size, the samples available for training will decrease. Thus, at the deeper layers we design some light structures (with less SSRBs). For example, with the progress of spectral splitting and aggregation reconstruction, the number of SSRBs decreases from 12 to 8, 4, and 2 for different spectral splitting and aggregation stages, respectively.

The proposed SSANet includes three spectral splitting and aggregation modules followed by a reconstruction module. In order to effectively and efficiently represent and reconstruct the hyperspectral face images, we carefully set the numbers of input band as well as the number of SSRB at each spectral splitting network. In addition, to fully exploit the spectral information of hyperspectral data, we also change the feature channels of different spectral splitting networks. At the shallow layer, where the input has only one spectral band, we set the number of feature channel to 64, and then increase it to 128, 128, and then to 192. This is because the input has more spectral bands with the progress of spectral splitting and aggregation reconstruction (from 1 to 4, 8, and then to 33), it calls for more feature channels to represent the spectral information of hyperspectral data.

Upscaling strategy: For the proposed SSANet, another design worth mentioning is the upscaling strategy. The most commonly used upscaling strategy is the first upsampling (when the input is fed into the network) or the last upsampling (at the end of the network). In this paper, we are inspired by the pyramid SR network [46] and propose an upscaling search strategy. In particular, by inserting upscaling modules at different locations (before each aggregation module) of the network, we search for a set of optimized upscaling schemes for different upsampling factors. As shown in Fig. 1, above the input of each splitting network, we give the optimal upscaling scheme (and the size of the input image) under different upsampling factors, *e.g.*, $\times 2$, $\times 4$, and $\times 8$. More details can be found at Section 4.1.1.

Spatial-spectral residual block (SSRB): Different from the commonly used residual module, we introduce a spatial-spectral residual block [47]. It is a residual module like block, and includes one 3×3 *Conv* and one 1×1 *Conv* followed by an ReLU layer. We modify the residual module due to the following considerations. First, the added 1×1 *Conv* can be leveraged to reorganize and reweight the importance of spectral bands, thus efficiently exploiting the spectral correlations of hyperspectral face image. Second, the added 1×1 *Conv* can be seen as the purpose of information distillation and more useful information can be well preserved. Third, it deepens the depth of the network at a small cost in term of parameters.

3.2. Training Sample Expanding

As we discussed above, the spectral dimensionality of hyperspectral face image is very high, but the training sample number of available hyperspectral face image dataset is extremely small. Therefore, to well represent the high-dimensional hyperspectral face image, one possible solution is to expand the training set by synthesizing new training samples.

In this paper, we develop a self-representation learning method to synthesize the face image. In particular, we assume that all face images are from the same source, *i.e.*, the mean face, and there is a path from the mean face to the individual face. Therefore, we can use some states on the path to obtain new samples. Mathematically, the self-representation based face synthesis method can be described as following,

$$\hat{x}_i = \sum_{j=1, j\neq i}^{N} w(x_i, x_j)x_j, \tag{1}$$

where x_i is the sample to be reconstructed, $\{x_j | j \neq i, 1 \leq j \leq N\}$ are the training samples except x_i itself, and $w(x_i, x_j)$ represents the reconstruction weight corresponding to x_j. From the above definitions, we can learn that the synthesized training sample is a linear combination of the training samples except itself. In particular, the combination weight is proportional to the similarity between current sample x_i and the remaining training samples x_j,

$$w(x_i, x_j) = \frac{\text{sim}(x_i, x_j)}{\sum_{j\neq i, j=1}^{N} \text{sim}(x_i, x_j)}. \tag{2}$$

Here, $\text{sim}(x_i, x_j)$ denotes the similarity between x_i and x_j and is defined as follow,

$$\text{sim}(x_i, x_j) = \exp\left(-\frac{||x_i - x_j||_2^2}{\sigma^2 G}\right), \tag{3}$$

where σ denotes the synthesis controlling parameter, and G is the mean of all elements of $\{||x_i - x_j||_2^2 | j \neq i, 1 \leq j \leq N\}$. Since the entire face image is high-dimensional, the globally reconstruction methods may be not accurate or cannot introduce additional information to the training dataset (please refer to the ***Remark 2***). Thus, we decompose the global face into small image patches, and apply the above-mentioned approach for each small image patch. Lastly, by integrating all the reconstructed image patches, we can synthesize a new training sample.

Remark 1: The human face, as a highly structured object, has two eyes, a nose, and a mouth. The global structure information of different objects is very similar. The differences between different people are mainly reflected in the detailed features. Therefore, we have reason to believe that through self-representation learning, we can synthesize some feature detail information that is not available in the training dataset. This information can actually be regarded as a kind of noise compared to the represented face image. Therefore, the proposed method based on self-representation learning can bring more information to the training dataset.

Table 1. Effectiveness analysis of the proposed spectral splitting and aggregation structure.

	Dataset	CC↑	SAM↓	RMSE↓	ERGAS↓	PSNR↑	SSIM↑
w/o SSAS	O	.9987	.8100	.0054	.4780	46.6650	.9911
	O+Self	.9990	.7088	.0048	.4179	47.9166	.9928
	O+Sym	.9990	.7271	.0049	.4279	47.6540	.9925
	O+Self+Sym	.9991	.6834	.0046	.4041	48.2217	.9932
SSAS	O	.9989	.7555	.0051	.4440	47.2672	.9923
	O+Self	.9992	.6552	.0044	.3825	48.6323	.9938
	O+Sym	.9992	.6549	.0043	.3746	48.7505	.9941
	O+Self+Sym	.9993	.6086	.0039	.3430	49.4840	.9948
Average Improvements		.0002↑	.0638↓	.0005↓	.0460↓	.9192↑	.0014↑

Remark 2: It seems that putting a reconstructed sample back to the training set will add no extra information to the dataset. If we reconstruct the face image globally, the above point is the truth. But when we reconstruct the face image locally (patch-wisely), this would be not true. In other words, for those patch-based methods, the reconstructed HR result does not have to be a linear combination of all the original training samples. We can imagine an extreme situation: if we set the patch size to very small level, *i.e.*, pixel-wise, we can reconstruct any image (any content) and dose not have to be a linear combination of training samples, *e.g.*, a cat or a dog image. Therefore, these patch-based reconstruction method can reconstruct some samples, which dose not have to be a linear combination of the original training samples. Therefore, when we put reconstructed HR samples back to the training set, we can actually introduce some additional information in the sense of that we can add an image (which cannot be linear combination with the original training) to the training dataset.

4. Experiments

We use Pytorch libraries[1] to implement and train the proposed SSANet method (the code will be available upon acceptance). We train different models to super-resolve the hyperspectral face images for upsampling factors 2, 4 and 8 with random initialization. We use the ADAM optimizer [48] with an initial learning rate of 1e-4 which decays by a factor of 10 at each 30 epochs. In our experiments, we find it will take 60 epochs to achieve the stable performance. The models are trained with a batch size of 4.

All the experiments are conducted on the UWA Hyperspectral Face Database (UWA-HSFD)[2] [54]. It contains 145

[1] `https://pytorch.org`

[2] It should be noted that we only use the database of UWA-HSFD due to the following reason. Currently, there are only three widely used hyperspectral face databases, the Hong Kong Polytechnic University Hyperspectral Face Database (PolyU-HSFD) [52], the CMU Hyperspectral Face Database (CMU-HSFD) [53], and UWA-HSFD used in our paper. The face images in PolyU-HSFD are very noisy and not suitable for the supervised learning. As for CMU-HSFD, the related project is unavailable at present,

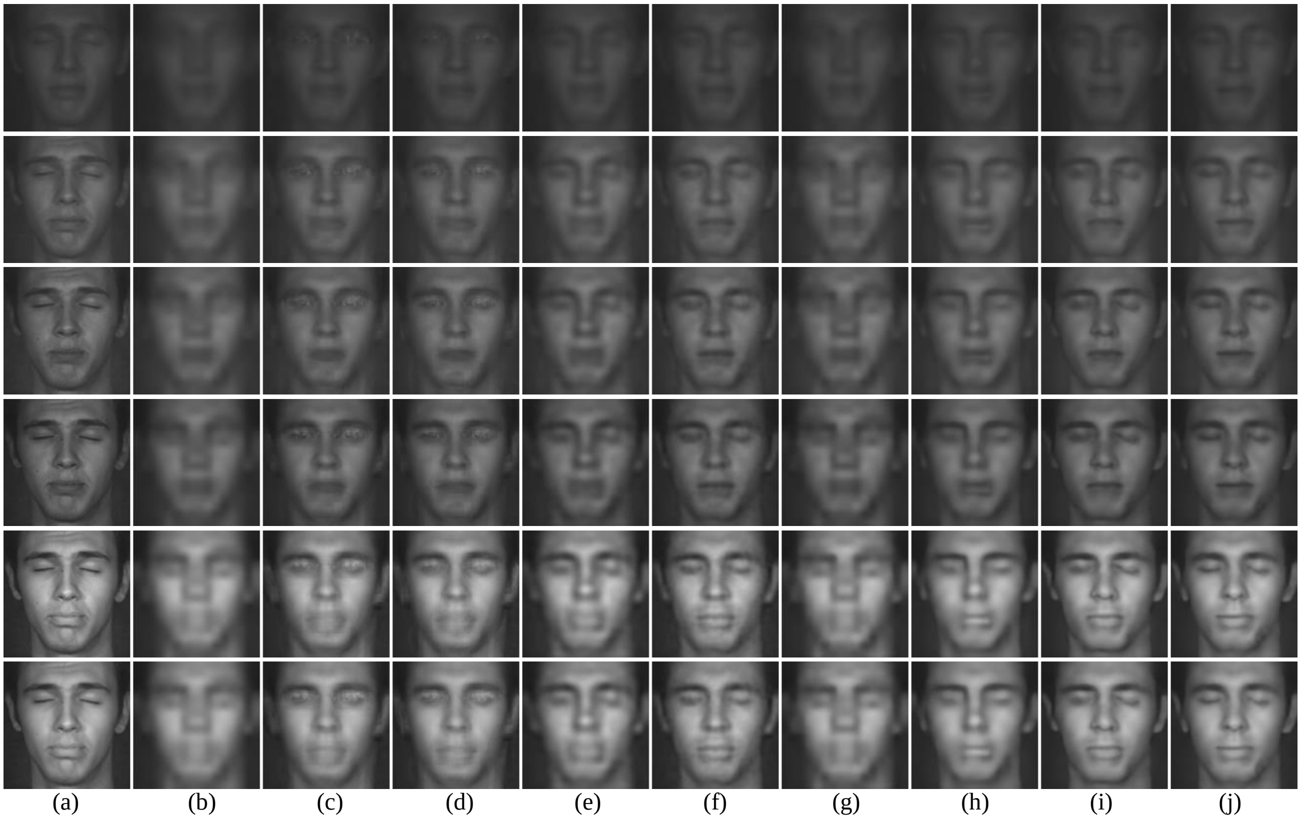

Figure 2. 8 × SR results of one object of different methods at bands 5, 10, 15, 20, 25, and 30. (a) HR faces, (b) Bicubic interpolation, (c) results of LSR [49], (d) results of LcR [15], (e) results of EDSR [50], (f) results of SAN [51], (g) results of 3DCNN [40], (h) results of GDRRN [42], (i) results of SSPSR [43], (j) Our results.

hyperspectral face images of 80 subjects collected in four sessions over time. The face images are acquired by an indoor imaging system using a CRI's VariSpec LCTF filter integrated with a Photon focus camera. In UWA-HSFD [54], the spectral range, in which the images have been acquired, is from 400 nm to 720 nm with a step size of 10 nm, resulting 33 spectral bands. Alignment errors are present between individual bands due to subjects movements and eye blinking during image acquisition. To simulate the situation that the testing face image is not in the training phases, we randomly selecting one cube from one session for each of the 70 subjects (includes 10% evaluation samples). The remaining 10 subjects are used for testing. In our experiments, all the input LR images are obtained by 2×, 4×, or 8× Bicubic downsampling.

Evaluation measures. Six widely used quantitative picture quality indices (PQIs) are employed to evaluate the performance of our method, including cross correlation (CC) [30], spectral angle mapper (SAM) [55], root mean squared error (RMSE), erreur relative globale adimensionnelle de synthese (ERGAS) [56], peak signal-to-noise ratio (PSNR), and structure similarity (SSIM) [57]. For PSNR and SSIM of the reconstructed hyperspectral images, we report their mean values of all spectral bands. CC, SAM, and ERGAS are three widely adopted quality indices in HS fusion task, while the remaining three indices are commonly used quantitative image restoration quality indices. The best values for these indices are 1, 0, 0, 0, $+\propto$, and 1, respectively.

4.1. Investigation to the Proposed SSANet

4.1.1 Progressively upsampling

The simplest upsampler is to do upsampling at the beginning by Bicubic interpolation or perform pixel shuffle at the end. They either increase the parameters and computational complexity of the network or increase the difficulty of neural network training. Inspired by the Laplacian pyramid SR network [46], we perform upsampling at the intermediate layers to progressively upsample the input image. We insert the upsampling module to each spectral splitting and aggregation module, as shown in Fig. 1. Similar to the Neural Architecture Search (NAS) [58], in our experiments we manually search the optimal structure. More details can be found *in the supplementary*.

and we cannot obtain the database.

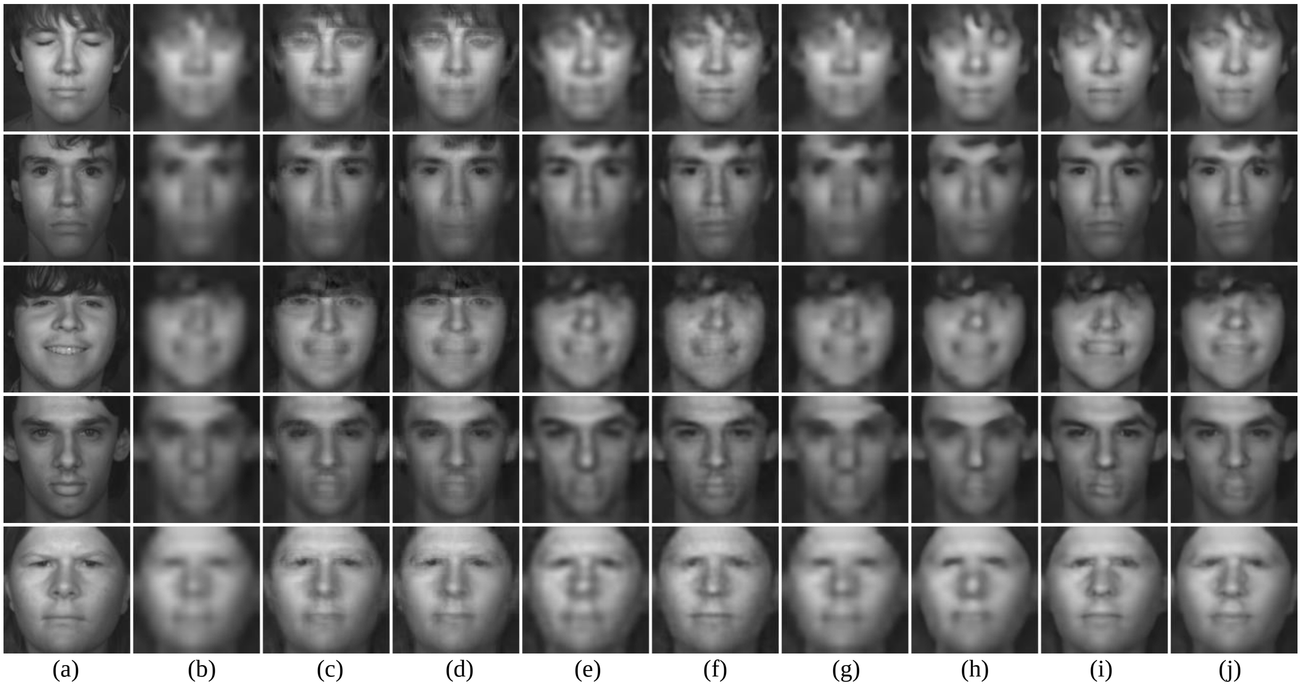

Figure 3. 8 × SR results of five objects of different methods at band 30. The meanings of (a)-(j) are the same as Fig. 2.

Table 2. Effectiveness analysis of the proposed training sample expanding strategy.

Dataset	r	CC↑	SAM↓	RMSE↓	ERGAS↓	PSNR↑	SSIM↑
O	2	0.9989	0.7555	0.0051	0.4440	47.2672	0.9923
O+Self	2	0.9992	0.6552	0.0044	0.3825	48.6323	0.9938
O+Sym	2	0.9992	0.6549	0.0043	0.3746	48.7505	0.9941
O+Self+Sym	2	0.9993	0.6086	0.0039	0.3430	49.4840	0.9948
O	4	0.9957	1.2985	0.0110	0.9269	41.4541	0.9673
O+Self	4	0.9960	1.2496	0.0107	0.9002	41.8074	0.9687
O+Sym	4	0.9958	1.2734	0.0109	0.9195	41.6118	0.9678
O+Self+Sym	4	0.9960	1.2409	0.0107	0.9003	41.8483	0.9689
O	8	0.9825	2.0828	0.0230	1.8961	35.2369	0.8901
O+Self	8	0.9838	1.9939	0.0221	1.8281	35.6496	0.8956
O+Sym	8	0.9828	2.0371	0.0228	1.8804	35.3614	0.8909
O+Self+Sym	8	0.9842	1.9934	0.0219	1.8095	35.6852	0.8973

4.1.2 Spectral splitting and aggregation

To alleviate the S3 problem, we carefully design the spectral splitting and aggregation strategy (SSAS). It splits the hyperspectral face image into as many groups as possible at shallow layers to guarantee the training of a complex network, and gradually aggregates the neighbor bands at deeper layers to fully exploit spectral correlations. To demonstrate the effectiveness of SSAS, we report the comparison results of the proposed deep neural network with and without SSAS in Table 1 when the upsampling factor is 2. To simulate the network without SSAS, we maintain the primary structure of SSANet and set the band size of each branch network to 33 (the number of spectral bands for UWA-HSFD database). In this situation, there is no spectral splitting and aggregation process for SSANet. From the tabulated results, we can clearly learn that SSANet with SSAS can stably improve the performance regardless of the training datasets (the original training dataset and expanded training dataset). In all cases, the average improvement (in term of PSNR) of SSANet with SSAS over SSANet without SSAS is nearly 1.0 dB, and the spectral similarity has also increased by nearly 10% (please refer to the SAM index).

4.2. Effectiveness of Training Sample Expanding

In this paper, we introduce two strategies to expand the size of training dataset. To demonstrate the effectiveness of these strategies, we show the comparison results of our method with different training datasets. As shown in Table 2, "O" denotes the original training dataset, "Self" demotes the expanded training dataset with self-representation based data argumentation, and "Sym" is the expanded training dataset using the symmetry nature of humane to flip the face image to double the size of training dataset. From the comparison results, we can clearly learn that the proposed self-representation and symmetry based data argumentation methods can improve the performance of our proposed deep neural network separately, no matter what the upsampling factor r is. The combination of these two strategies can achieve the best performance.

4.3. Comparison Results

Since the proposed SSANet is the first hyperspectral image SR method for human faces, there is no direct hyperspectral image SR method for comparison. Therefore, to demonstrate the effectiveness of our method, we adjust two traditional shallow learning based face SR methods, LSR [49] and LcR [15], two deep learning based general image SR methods, EDSR [50] and SAN [51], and one most recently proposed deep learning based single hyperspectral image SR method, SSPSR [43] for the hyperspectral image SR task. Specifically, as for LSR [49] and LcR [15], we super-resolve each band of the hyperspectral face image separately. For EDSR [50] and SAN [51], we retrain their networks by changing the number of input channels with the UWA-HSFD database. SSPSR [43] is also retrained by the UWA-HSFD database for fair comparison.

Fig. 2 shows the super-resolved results of one object of different SR approaches at five specific bands, and in Fig. 3 we also present the results of five objects at band 30. From these comparison results, we learn that Bicubic interpolation lost most of the face detailed features. LSR [49] and LcR [15] require the accurate alignment of face images. However, face images cannot be completely aligned due to non-rigid transformations such as expressions and poses. Therefore, ghosting effects often appear at the regions where the eyes and mouth cannot be completely aligned. EDSR [50] and SAN [51] are not specifically designed for hyperspectral faces, and they cannot fully exploit the rich spectral information of hyperspectral faces. Their results are much blur than SSPSR [43], which is a general hyperspectral image SR method. Results reconstructed by our proposed method are more credible (more similar to the ground truth), and at the same time our method greatly reduces the artificial effects of the reconstructed eyes, mouth, and nose.

It should be noted that almost all these comparison approaches cannot accurately reconstruct the slightly opened eyes. Actually, this shows a weakness of these learning-based techniques which depend highly upon the consistency (similarity) between the supporting testing image and training dataset. This is mainly because the training data cannot cover the changes caused by all factors (such as posture, expression, lighting, *etc*.), especially when the size of the training dataset is limited.

Table 3 tabulates the average objective results of seven comparison methods on the 10 testing images from UWA-HSFD dataset when the upsampling factor is set to 2, 4, and 8. The best and second best results are highlighted in boldface and underlined, respectively. Note that SSANet is the model trained by the original training samples, while SSANet+ denotes the model trained by the expanded training samples with "Self" and "Sym". These objective results also demonstrate the better performance of our method.

Table 3. Quantitative comparisons of different approaches.

	r	CC↑	SAM↓	RMSE↓	ERGAS↓	PSNR↑	SSIM↑
LSR [49]	2	0.9977	1.2394	0.0076	0.6526	43.7839	0.9784
LcR [15]	2	0.9980	1.0751	0.0073	0.6214	44.4102	0.9808
EDSR [50]	2	0.9988	0.7777	0.0053	0.4616	46.9760	0.9917
SAN [51]	2	0.9985	0.8484	0.0055	0.4921	46.3874	0.9906
3DCNN [40]	2	0.9986	0.8676	0.0061	0.5218	46.2634	0.9895
GDRRN [42]	2	0.9976	0.7848	0.0054	0.4730	46.8245	0.9911
SSPSR [43]	2	0.9987	0.7947	0.0052	0.4614	46.9406	0.9916
SSANet	2	**0.9989**	**0.7555**	**0.0051**	**0.4440**	**47.2672**	**0.9923**
LSR [49]	4	0.9937	1.6073	0.0134	1.1214	39.6160	0.9472
LcR [15]	4	0.9939	1.5774	0.0134	1.1144	39.7621	0.9480
EDSR [50]	4	0.9949	1.4242	0.0118	0.9995	40.6932	0.9628
SAN [51]	4	0.9948	1.4723	0.0118	1.0048	40.6052	0.9629
3DCNN [40]	4	0.9942	1.4482	0.0132	1.0984	40.0966	0.9556
GDRRN [42]	4	0.9950	1.3897	0.0118	0.9939	40.7684	0.9626
SSPSR [43]	4	0.9953	1.3750	0.0114	0.9656	41.0340	0.9651
SSANet	4	**0.9957**	**1.2985**	**0.0110**	**0.9269**	**41.4541**	**0.9673**
LSR [49]	8	0.9795	2.3873	0.0246	2.0349	34.4513	0.8782
LcR [15]	8	0.9810	2.3680	0.0240	1.9775	34.8670	0.8835
EDSR [50]	8	0.9820	2.2191	0.0232	1.9213	35.1152	0.8849
SAN [51]	8	0.9815	2.3889	0.0234	1.9420	34.9323	0.8855
3DCNN [40]	8	0.9741	2.2846	0.0283	2.3150	33.5437	0.8470
GDRRN [42]	8	0.9785	2.1964	0.0254	2.0910	34.3277	0.8704
SSPSR [43]	8	0.9816	2.1979	0.0234	1.9356	34.9966	0.8856
SSANet	8	**0.9825**	**2.0828**	**0.0230**	**1.8961**	**35.2369**	**0.8901**

5. Conclusions

In this paper, we present a hyperspectral face super-resolution (HFSR) approach based on spectral splitting and aggregation network (SSANet). This is the first face SR work focusing on hyperspectral images. Different from traditional face and general image SR tasks, in which there are enough training samples to support the training of a complex network, HFSR has to face the small training sample size (S3) problem. To this end, on the one hand we carefully design a spectral splitting and aggregation network to "generate" more training samples to alleviate the S3 problem *from less to more*, and simultaneously to fully make use of multiple spectral information. On the other hand we introduce two strategies to expand the training samples by a self-representation and the symmetry augmentation. Experimental results on public hyperspectral face database demonstrate that our proposed SSANet method and the self-representation and symmetry based training sample expanding strategy are effective for the HFSR task. In addition, we also report the comparison results on hyperspectral images of natural scenes, which demonstrate the generalization of our method.

Acknowledgments: The research was supported by the National Natural Science Foundation of China (61971165), in part by the Fundamental Research Funds for the Central Universities (FRFCU5710050119), the Natural Science Foundation of Heilongjiang Province (YQ2020F004).

References

[1] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang. Super-resolution image reconstruction: a technical overview. *IEEE Signal Process. Magazine*, 20(3):21–36, may 2003.

[2] S. Baker and T. Kanade. Hallucinating faces. In *FG*, pages 83–88, 2000.

[3] Ce Liu, Heung-Yeung Shum, and Chang-Shui Zhang. A two-step approach to hallucinating faces: global parametric model and local nonparametric model. In *CVPR*, volume 1, pages 192–198, 2001.

[4] X. Wang and X. Tang. Hallucinating face by eigentransformation. *IEEE Trans. Syst. Man Cybern. Part C-Appl. Rev.*, 35(3):425–434, 2005.

[5] Yueting Zhuang, Jian Zhang, and Fei Wu. Hallucinating faces: Lph super-resolution and neighbor reconstruction for residue compensation. *Pattern Recogn.*, 40(11):3178–3194, 2007.

[6] Jeong-Seon Park and Seong-Whan Lee. An example-based face hallucination method for single-frame, low-resolution facial images. *IEEE Trans. Image Process.*, 17(10):1806–1816, Oct 2008.

[7] Junjun Jiang, Chenyang Wang, Xianming Liu, and Jiayi Ma. Deep learning-based face super-resolution: A survey. *ACM Computing Surveys*, 55(1):1–36, 2023.

[8] Sung Won Park and M. Savvides. Breaking the limitation of manifold analysis for super-resolution of facial images. In *ICASSP*, volume 1, pages I–573–I–576, Apr 2007.

[9] Hua Huang, Huiting He, Xin Fan, and Junping Zhang. Super-resolution of human face image using canonical correlation analysis. *Pattern Recogn.*, 43(7):2532–2543, 2010.

[10] Le An and Bir Bhanu. Face image super-resolution using 2D CCA. *Signal Proc*, 103:184–194, 2014.

[11] Jianchao Yang, John Wright, Thomas Huang, , and Yi Ma. Image super-resolution via sparse representation. *IEEE Trans. Image Process.*, 19(11):2861–2873, 2010.

[12] Chih-Yuan Yang, Sifei Liu, and Ming-Hsuan Yang. Structured face hallucination. In *CVPR*, pages 1099–1106, 2013.

[13] Yongchao Li, Cheng Cai, Guoping Qiu, and Kin-Man Lam. Face hallucination based on sparse local-pixel structure. *Pattern Recogn.*, 47(3):1261–1270, 2014.

[14] Zhuo Hui, Wenbo Liu, and Kin-Man Lam. A novel correspondence-based face-hallucination method. *Image and Vision Computing*, 2017.

[15] J. Jiang, R. Hu, Z. Wang, and Z. Han. Noise robust face hallucination via locality-constrained representation. *IEEE Trans. Multimedia*, 16(5):1268–1281, Aug 2014.

[16] Qiang Zhang, Fei Zhou, Fan Yang, and Qingmin Liao. Face super-resolution via semi-kernel partial least squares and dictionaries coding. In *DSP 2015*, pages 590–594, 2015.

[17] R. A. Farrugia and C. Guillemot. Face hallucination using linear models of coupled sparse support. *IEEE Trans. Image Process.*, 26(9):4562–4577, Sept 2017.

[18] Guangwei Gao, Yi Yu, Jin Xie, Jian Yang, Meng Yang, and Jian Zhang. Constructing multilayer locality-constrained matrix regression framework for noise robust face super-resolution. *Pattern Recognition*, page 107539, 2020.

[19] L. Liu, C. L. P. Chen, S. Li, Y. Y. Tang, and L. Chen. Robust face hallucination via locality-constrained bi-layer representation. *IEEE Trans. Cybern.*, 48(4):1189–1201, 2018.

[20] Jingang Shi, Xin Liu, Yuan Zong, Chun Qi, and Guoying Zhao. Hallucinating face image by regularization models in high-resolution feature space. *IEEE Trans. Image Process.*, 27(6):2980–2995, 2018.

[21] Liang Chen, Jinshan Pan, Junjun Jiang, Jiawei Zhang, and Yi Wu. Robust face super-resolution via position relation model based on global face context. *IEEE Trans. Image Process.*, 29:9002–9016, 2020.

[22] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.*, 38(2):295–307, 2016.

[23] Ding Liu, Zhaowen Wang, Bihan Wen, Jianchao Yang, Wei Han, and Thomas S Huang. Robust single image super-resolution via deep networks with sparse prior. *IEEE Trans. Image Process.*, 25(7):3194–3207, 2016.

[24] Xin Yu and Fatih Porikli. Ultra-resolving face images by discriminative generative networks. In *ECCV*, pages 318–333. Springer, 2016.

[25] Shizhan Zhu, Sifei Liu, Chen Change Loy, and Xiaoou Tang. Deep cascaded bi-network for face hallucination. In *ECCV*, pages 614–630. Springer, 2016.

[26] Yibing Song, Jiawei Zhang, Shengfeng He, Linchao Bao, and Qingxiong Yang. Learning to hallucinate face images via component generation and enhancement. In *IJCAI*, pages 4537–4543, 2017.

[27] Q. Cao, L. Lin, Y. Shi, X. Liang, and G. Li. Attention-aware face hallucination via deep reinforcement learning. In *CVPR*, pages 1656–1664, 2017.

[28] Yu Chen, Ying Tai, Xiaoming Liu, Chunhua Shen, and Jian Yang. Fsrnet: End-to-end learning face super-resolution with facial priors. In *CVPR*, pages 1–8, 2008.

[29] X. Yu and F. Porikli. Imagining the unimaginable faces by deconvolutional networks. *IEEE Trans. Image Process.*, 27(6):2747–2761, June 2018.

[30] Laetitia Loncan, Luis B De Almeida, José M Bioucas-Dias, Xavier Briottet, Jocelyn Chanussot, Nicolas Dobigeon, Sophie Fabre, Wenzhi Liao, Giorgio A Licciardi, Miguel Simoes, et al. Hyperspectral pansharpening: A review. *IEEE Geoscience and remote sensing magazine*, 3(3):27–46, 2015.

[31] Naoto Yokoya, Claas Grohnfeldt, and Jocelyn Chanussot. Hyperspectral and multispectral data fusion: A comparative review of the recent literature. *IEEE Geoscience and Remote Sensing Magazine*, 5(2):29–56, 2017.

[32] Naoto Yokoya, Takehisa Yairi, and Akira Iwasaki. Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 50(2):528–537, 2011.

[33] Weisheng Dong, Fazuo Fu, Guangming Shi, Xun Cao, Jinjian Wu, Guangyu Li, and Xin Li. Hyperspectral image super-resolution via non-negative structured sparse representation. *IEEE Trans. Image Process.*, 25(5):2337–2352, 2016.

[34] Qi Xie, Minghao Zhou, Qian Zhao, Deyu Meng, Wangmeng Zuo, and Zongben Xu. Multispectral and hyperspectral image fusion by ms/hs fusion net. In *CVPR*, pages 1585–1594, 2019.

[35] R. A. Borsoi, T. Imbiriba, and J. C. M. Bermudez. Super-resolution for hyperspectral and multispectral image fusion accounting for seasonal spectral variability. *IEEE Trans. Image Process.*, 29:116–127, 2020.

[36] Huijuan Huang, Anthony G Christodoulou, and Weidong Sun. Super-resolution hyperspectral imaging with unknown blurring by low-rank and group-sparse modeling. In *ICIP*, pages 2155–2159. IEEE, 2014.

[37] Jie Li, Qiangqiang Yuan, Huanfeng Shen, Xiangchao Meng, and Liangpei Zhang. Hyperspectral image super-resolution by spectral mixture analysis and spatial–spectral group sparsity. *IEEE Geoscience and Remote Sensing Letters*, 13(9):1250–1254, 2016.

[38] Yuan Yuan, Xiangtao Zheng, and Xiaoqiang Lu. Hyperspectral image superresolution by transfer learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(5):1963–1974, 2017.

[39] Weiying Xie, Xiuping Jia, Yunsong Li, and Jie Lei. Hyperspectral image super-resolution using deep feature matrix factorization. *IEEE Trans. Geosci. Remote Sens.*, 57(8):6055–6067, 2019.

[40] Shaohui Mei, Xin Yuan, Jingyu Ji, Yifan Zhang, Shuai Wan, and Qian Du. Hyperspectral image spatial super-resolution via 3d full convolutional neural network. *Remote Sensing*, 9(11):1139, 2017.

[41] Qiang Li, Qi Wang, and Xuelong Li. Mixed 2d/3d convolutional network for hyperspectral image super-resolution. *Remote Sensing*, 12(10):1660, 2020.

[42] Yong Li, Lei Zhang, Chen Dingl, Wei Wei, and Yanning Zhang. Single hyperspectral image super-resolution with grouped deep recursive residual network. In *BigMM*, pages 1–4. IEEE, 2018.

[43] J. Jiang, H. Sun, X. Liu, and J. Ma. Learning spatial-spectral prior for super-resolution of hyperspectral imagery. *IEEE Trans. Computat. Imag.*, 6:1082–1096, 2020.

[44] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *CVPR*, pages 9446–9454, 2018.

[45] O. Sidorov and J. Y. Hardeberg. Deep hyperspectral prior: Single-image denoising, inpainting, super-resolution. In *ICCVW*, pages 3844–3851, 2019.

[46] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *CVPR*, pages 624–632, 2017.

[47] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016.

[48] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[49] X. Ma, J. Zhang, and C. Qi. Hallucinating face by position-patch. *Pattern Recogn.*, 43(6):2224–2236, 2010.

[50] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *CVPRW*, pages 136–144, 2017.

[51] Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang. Second-order attention network for single image super-resolution. In *CVPR*, pages 11065–11074, 2019.

[52] W. Di, L. Zhang, D. Zhang, and Q. Pan. Studies on hyperspectral face recognition in visible spectrum with feature band selection. *IEEE Trans. Syst., Man, and Cybern. A: Sys. and Humans*, 40(6):1354–1361, 2010.

[53] Louis J. Denes, Peter Metes, and Yanxi Liu. Hyperspectral face database. Technical Report CMU-RI-TR-02-25, Carnegie Mellon University, Pittsburgh, PA, September 2002.

[54] Muhammad Uzair, Arif Mahmood, and Ajmal Mian. Hyperspectral face recognition with spatiospectral information fusion and pls regression. *IEEE Trans. Image Process.*, 24(3):1127–1137, 2015.

[55] Roberta H Yuhas, Alexander FH Goetz, and Joe W Boardman. Discrimination among semi-arid landscape endmembers using the spectral angle mapper (sam) algorithm. In *JPL, Summaries of the Third Annual JPL Airborne Geoscience Workshop. Volume 1: AVIRIS Workshop, pp. 147-149*, 1992.

[56] Lucien Wald. *Data fusion: definitions and architectures: fusion of images of different spatial resolutions*. Presses des MINES, 2002.

[57] Zhou Wang, Alan C Bovik, Hamid R Sheikh, Eero P Simoncelli, et al. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.*, 13(4):600–612, 2004.

[58] Chenxi Liu, Barret Zoph, Maxim Neumann, Jonathon Shlens, Wei Hua, Li-Jia Li, Li Fei-Fei, Alan Yuille, Jonathan Huang, and Kevin Murphy. Progressive neural architecture search. In *ECCV*, pages 19–34, 2018.