

Adaptive Image Sampling using Deep Learning and its Application on X-Ray Fluorescence Image Reconstruction

Qiqin Dai, Henry Chopp, Emeline Pouyet, Oliver Cossairt, Marc Walton,
and Aggelos K. Katsaggelos, *Fellow, IEEE*

Abstract—This paper presents an adaptive image sampling algorithm based on Deep Learning (DL). It consists of an adaptive sampling mask generation network which is jointly trained with an image inpainting network. The sampling rate is controlled by the mask generation network, and a binarization strategy is investigated to make the sampling mask binary. In addition to the image sampling and reconstruction process, we show how it can be extended and used to speed up raster scanning such as the X-Ray fluorescence (XRF) image scanning process. Recently XRF laboratory-based systems have evolved into lightweight and portable instruments thanks to technological advancements in both X-Ray generation and detection. However, the scanning time of an XRF image is usually long due to the long exposure requirements (e.g., $100\mu s - 1ms$ per point). We propose an XRF image inpainting approach to address the long scanning times, thus speeding up the scanning process, while being able to reconstruct a high quality XRF image. The proposed adaptive image sampling algorithm is applied to the RGB image of the scanning target to generate the sampling mask. The XRF scanner is then driven according to the sampling mask to scan a subset of the total image pixels. Finally, we inpaint the scanned XRF image by fusing the RGB image to reconstruct the full scan XRF image. The experiments show that the proposed adaptive sampling algorithm is able to effectively sample the image and achieve a better reconstruction accuracy than that of existing methods.

Index Terms—Adaptive sampling, convolutional neural network, X-Ray fluorescence, inpainting

I. INTRODUCTION

COMPRESSED sensing (CS) has shown that it is possible to acquire and reconstruct natural images under the Nyquist sampling rates [1], [2]. Rather than full image acquisition followed by compression, CS combines sensing and compression into one step, and has the advantages of faster acquisition time, smaller power consumption, and lighter data throughput. Adaptive image sampling is a sub-problem of CS that aims for a sparse representation of signals in the image domain. In this paper, we present a novel adaptive image sampling algorithm based on Deep Learning and show its application to RGB image sampling and recovery. We also applied the proposed adaptive sampling technique to speed

Q. Dai, H. Chopp, O. Cossairt and A. K. Katsaggelos are with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, 60208 USA. E. Pouyet and M. Walton are with the Northwestern University / Art Institute of Chicago Center for Scientific Studies in the Arts (NU-ACCESS), Evanston, IL, 60208 USA.

Manuscript received September 15, 2018.

up the raster scan process of XRF imaging, based on the correlation between RGB and XRF signals.

Irregular sampling techniques have long been studied in the image processing and computer graphics fields to achieve compact representation of images. Such irregular sampling techniques, such as stochastic sampling [3], may have better anti-aliasing performance compared to uniform sampling intervals if frequencies greater than the Nyquist limit are present. Further performance improvement can be obtained if the sampling distribution is not only irregular but also adaptive to the signal itself. The limited samples should be concentrated in parts of the image rich in detail, so as to simulate human vision [4]. Several works have been reported in the literature on adaptive sampling techniques. An early significant work in this direction is made by Eldar *et al.* [5]. A farthest point strategy is proposed which permits progressive and adaptive sampling of an image. Later, Rajesh *et al.* [6] proposed a progressive image sampling technique inspired by the lifting scheme of wavelet generation. A similar method is developed by Demaret *et al.* [7] by utilizing an adaptive thinning algorithm. Ramponi *et al.* [8] developed an irregular sampling method based on a measure of the local sample skewness. Lin *et al.* [9] viewed grey scale images as manifolds with density and sampled them according to the generalized Ricci curvature. Liu *et al.* [10] proposed an adaptive progressive image acquisition algorithm based on kernel construction.

Most of these irregular sampling and adaptive sampling techniques [3], [5]–[10] need their own specific reconstruction algorithm to reconstruct the fully sampled signal. Furthermore, all these sampling techniques are model-based approaches relying on predefined priors, and according to our knowledge, no work has been done on utilizing machine learning techniques to design the adaptive sampling mask.

Inspired by the recent successes of convolutional neural networks (CNNs) [11], [12] in high level computer vision tasks, deep neural networks (DNNs) emerged in addressing low level computer vision tasks as well [13]–[23]. For the task of image inpainting, Pathak *et al.* [21] presented an auto-encoder to perform context-based image inpainting. The inpainting performance is improved by introducing perceptual loss [22] and on-demand learning [23]. Iliadis *et al.* [19] utilized a deep-fully-connected network for video compressive sensing while also learning an optimal binary sampling mask [20]. However, the learned optimal binary sampling mask is not adaptive to the input video signals. According to our knowledge, no work

has been made on generating the adaptive binary sampling mask for the image inpainting problem using deep learning.

With the proposed adaptive sampling algorithm, we can efficiently sample points based on the local structure of the input images. Besides the application of Progressive Image Transmission (PIT) [24], compressed image sampling, image compression, etc., we also show that the proposed adaptive sampling algorithm can speed up many raster scan processes such as XRF imaging. In detail, the binary sampling mask can be obtained using one modality (RGB image for example) of the target, and then applied on the raster scan process for another modality (XRF image for example). A detailed introduction to XRF imaging is in Section II.

The contribution of this paper lies in the following aspects:

- We proposed an effective way to binarize and control the sparseness of the output of a CNN.
- We proposed an efficient network structure to generate the binary sampling mask and showed its advantages over other state-of-the-art adaptive sampling algorithms.
- We proposed an adaptive sampling framework that can be applied to speed up many raster scan processes.
- We proposed a fusion-based image reconstruction algorithm to restore the fully sampled XRF image.
- We experimentally showed that the benefits of the adaptive sampling and the fusion-based inpainting algorithm are additive.

This paper is organized as follows: Section II introduces XRF imaging with adaptive sampling. We illustrate the adaptive sampling mask design in Section III. We describe the XRF image inpainting problem in Section IV. In Section V, we provide the experimental results with both synthetic data and real data to evaluate the effectiveness of the proposed approach. The paper is concluded in Section VI.

II. XRF IMAGING USING ADAPTIVE SAMPLING

During the past few years, XRF laboratory-based systems have evolved to lightweight and portable instruments thanks to technological advancements in both X-Ray generation and detection. Spatially resolved elemental information can be provided by scanning the surface of the sample with a focused or collimated X-ray beam of (sub) millimeter dimensions and analyzing the emitted fluorescence radiation in a nondestructive in-situ fashion entitled Macro X-Ray Fluorescence

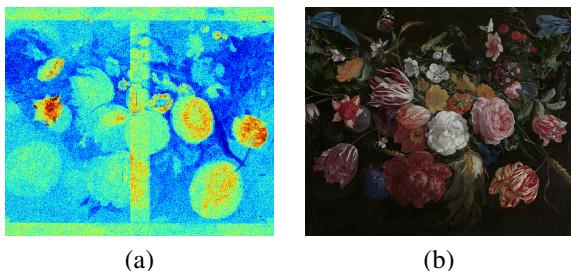


Fig. 1. (a) XRF map showing the distribution of $Pb\ L\eta$ XRF emission line (sum of channel #582 - 602) of the “Bloemen en insecten” (ca 1645), by Jan Davidsz. de Heem, in the collection of Koninklijk Museum voor Schone Kunsten (KMKSA) Antwerp and (b) the HR RGB image.

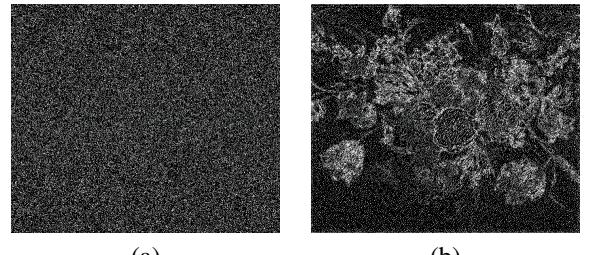


Fig. 2. (a) Random binary sampling mask that skips 80% of pixels and (b) Adaptive binary sampling mask that skips 80% of pixels based on the input RGB images in Fig 1 (b).

(MA-XRF). The new generations of XRF spectrometers are used in the Cultural Heritage field to study the manufacture, provenance, authenticity, etc. of works of art. Because of their fast noninvasive set up, we are able to study large, fragile, and location-inaccessible art objects and archaeological collections. In particular, XRF has been used extensively to investigate historical paintings by capturing the elemental distribution images of their complex layered structure. This method reveals the painting history from the artist creation to restoration processes [25], [26].

As with other imaging techniques, high spatial resolution and high quality spectra are desirable for XRF scanning systems; however, the acquisition time is usually limited, resulting in a compromise between dwell time, spatial resolution, and desired image quality. In the case of scanning large scale mappings, a choice may be made to reduce the dwell time and increase the step size, resulting in noisy XRF spectra and low spatial resolution XRF images.

An example of an XRF scan is shown in Figure 1 (a). Channel #582 – 602 corresponding to the $Pb\ L\eta$ XRF emission line was extracted from a scan of Jan Davidsz. de Heem’s “Bloemen en insecten” painted in 1645 (housed at Koninklijk Museum voor Schone Kunsten (KMKSA) Antwerp). The image is color coded for better visibility. This XRF image was collected by a home-built XRF spectrometer (courtesy of Prof. Koen Janssens) with 2048 channels in spectrum and spatial resolution 680×580 pixels. This scan has a relatively short dwell time, resulting in low Signal-to-Noise Ratio (SNR), yet it still took 18 hours to acquire it. Many other XRF scanners with longer dwell time or slower scanning speed will need a longer acquisition time. Faster scanning speed will be desirable for promoting the popularity of the XRF scanning technique, since the slow acquisition process impedes the use of XRF scanning instruments as high resolution widefield imaging devices. The RGB image of the painting of resolution 680×580 pixels is shown in Figure 1 (b).

Image inpainting [27]–[29] is the process of recovering missing pixels in images. The XRF images are acquired through a raster scan process. We could therefore speed up the scanning process by skipping pixels and then utilizing an image inpainting technique to reconstruct the missing pixels. If we are to skip 80% of the pixels during acquisition (a 5x speedup), we could use a random sampling mask (shown in Figure 2 (a)), or we could design one utilizing the available RGB image (shown in Figure 2 (b)). The idea of the adaptive

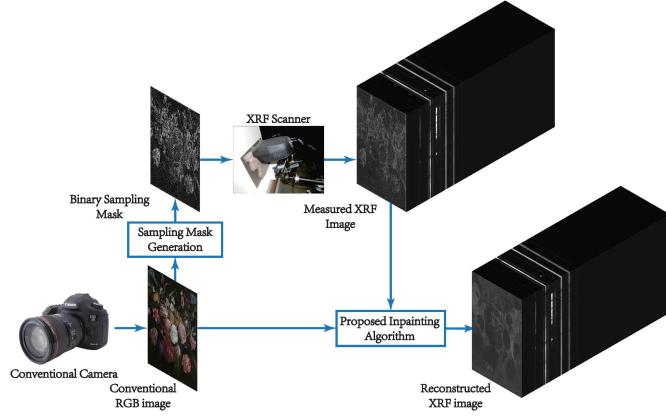


Fig. 3. The proposed pipeline for the XRF image inpainting utilizing an adaptive sampling mask. The binary adaptive sampling mask is generated based on the RGB image of the scan target. Then, the XRF scanner sampled the target object based on the binary sampling mask. Finally, the subsampled XRF image and the RGB image are fused to reconstruct the fully sampled XRF image.

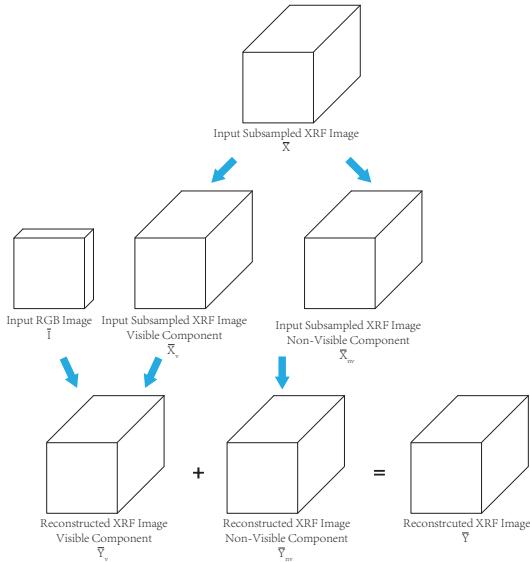


Fig. 4. Proposed pipeline of XRF image inpainting. The visible component of the input subsampled XRF image is fused with the input RGB image to obtain the visible component of the reconstructed XRF image. The non-visible component of the input XRF image is super-resolved to obtain the non-visible component of the reconstructed XRF image. The reconstructed visible and non-visible component of the output XRF image are combined to obtain the final output.

binary sampling mask is based on the assumption that the XRF image is highly correlated with the RGB image. We would like to allocate more pixels to the informative parts of the image, such as high frequency textures, sharp edges, and high contrast details, and spend fewer pixels on the uninformative parts of the image.

With the proposed adaptive sampling algorithm, we propose an image inpainting approach to speed up the acquisition process of the XRF image with the aid of a conventional RGB image, as shown in Figure 3. The proposed XRF image inpainting algorithm can also be applied to spectral images obtained by any other raster scanning processes, such as Scanning Electron Microscope (SEM), Energy Dispersive Spectroscopy (EDS),

and Wavelength Dispersive Spectroscopy (WDS). First, the RGB image of the scanning target is utilized to generate the adaptive sampling mask. Then, the XRF scanner will scan the corresponding pixels according to the binary sampling mask. The speedup in acquisition is achieved since many pixels will be skipped. Finally, the subsampled XRF image is fused with the conventional RGB image to reconstruct the full scan XRF image, utilizing an image inpainting algorithm. For the fusion-based XRF image inpainting algorithm, similarly to our previous super-resolution (SR) approach [30], [31], we model the spectrum of each pixel using a linear mixing model [32], [33]. Because the hidden part of the painting is not visible in the conventional RGB image, but it can be captured in the XRF image [34], there is no direct one-to-one mapping between the visible RGB spectrum and the XRF spectrum. We model the XRF signal as a combination of the visible signal (on the surface) and the non-visible signal (hidden under the surface), as shown in Figure 4. We emphasize that while our framework is general enough to handle separation of visible and hidden layers, it easily handles the case of fully visible layers. To inpaint the visible component XRF signal, we follow an approach similar to the one applied to hyperspectral image SR [35]–[39]. A coupled XRF-RGB dictionary pair is learned to explore the correlation between XRF and RGB signals. The RGB dictionary is then applied to obtain the sparse representation of the RGB input image, resulting in a fully sampled coefficient map. Then the XRF dictionary is applied on the fully sampled coefficient map to reconstruct the XRF image. Differently from those hyperspectral image SR approaches, we experimentally found that for the inpainting problem, a spatially adaptive total variation regularizer [40], [41] is needed to produce a smooth XRF output image. For the non-visible part, we inpaint its missing pixels using a standard total variation regularizer. Finally, the reconstructed visible and non-visible XRF signals are combined to obtain the final XRF reconstruction result. The input subsampled XRF image is not explicitly separated into visible and non-visible parts in advance. Instead, the whole inpainting problem is formulated as an optimization problem. By alternately optimizing over the

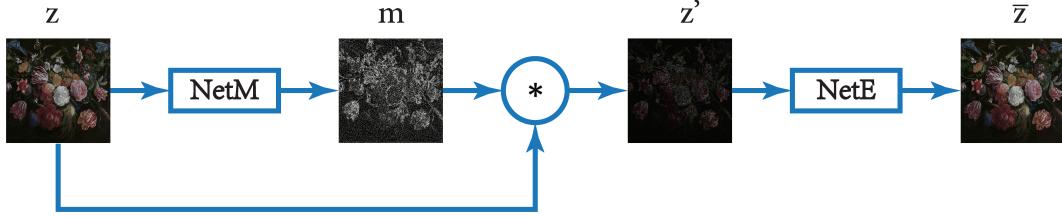


Fig. 5. Pipeline for adaptive sampling mask generation utilizing CNNs.

coupled XRF-RGB dictionary and the visible/non-visible fully sampled coefficient maps, the fidelity of the estimated fully sampled output to both the subsampled XRF and RGB input signals is improved, thus resulting in a better inpainting output. Both synthetic and real experiments show the effectiveness of our proposed method in terms of both reconstruction error and visual quality of the inpainting result compared to other methods [42]–[44].

While there is a large body of work on inpainting conventional RGB images [21]–[23], [27]–[29], [42]–[44], very little work has appeared in the literature on inpainting XRF images [27], and there is no work on fusing a conventional RGB image during the inpainting process. XRF image inpainting poses a particular challenge because the acquired spectrum signal usually has low SNR. In addition, the correlation among spectral channels needs to be preserved for the inpainted pixels. In our previous work on spatial-spectral representation for XRF image super-resolution [31], the input low resolution (LR) XRF image is fused with an HR conventional RGB image to obtain an HR XRF output image. The spatial resolution of the visible component XRF signal is increased by fusing an HR conventional RGB image while the spatial resolution of the non-visible part is increased by using a standard total variation regularizer [40], [41]. Here we propose an XRF image inpainting algorithm by fusing an HR conventional RGB image, which can be regarded as an extension of our previous XRF SR approach.

III. ADAPTIVE SAMPLING MASK GENERATION UTILIZING CONVOLUTIONAL NEURAL NETWORK

In this section, we present our proposed adaptive sampling mask generation using a CNN. In other words, we describe the details of the “Sampling Mask Generation” block in Figure 3. We first formulate the problem of adaptive sampling mask design, followed by the presentation of the overall network architecture consisting of both the inpainting network and the mask generation network.

A. Problem Formulation

As shown in Figure 5, we denote by z an input original image. Our mask generation network $NetM$ produces a binary sampling mask $m = NetM(z, c)$, where $c \in [0, 1]$ is the predefined sampling percentage. The entries of m are equal to 1 for the sampled and 0 otherwise. The corrupted image z' is obtained by

$$z' = z \odot m = z \odot NetM(z, c), \quad (1)$$

where \odot is the element-wise product operation. The reconstructed image \bar{z} is obtained by the inpainting network $NetE$,

$$\bar{z} = NetE(z') = NetE(z \odot NetM(z, c)). \quad (2)$$

The overall pipeline is shown in Figure 5. We could regard the whole pipeline (Equation 2) as one network with input z and output \bar{z} and perform an end to end training. If we simultaneously optimize the mask generation network $NetM$ and the inpainting network $NetE$ according to the following loss function,

$$\mathcal{L}(z) = \|z - \bar{z}\|_2 = \|z - NetE(z \odot NetM(z, c))\|_2, \quad (3)$$

$NetM$ will perform an optimized adaptive sampling strategy according to the input image, and $NetE$ will perform optimized image inpainting. After the mask has been generated by the network $NetM$, we can replace the inpainting network $NetE$ with other image inpainting algorithms. The detailed network architecture of $NetE$ and $NetM$ are discussed in the following two subsections III-B and III-C, respectively.

B. Deep Learning Network Architecture for Inpainting Network

The network architecture in [23] is used for the inpainting network, as shown in Figure 6. The network is an encoder-decoder pipeline. The encoder takes a corrupted image z' of size 64×64 as input and encodes it in the latent feature space. The decoder takes the feature representation and outputs the restored image $\bar{z} = NetE(z')$. The encoder and decoder are connected through a channel-wise fully-connected layer. For the encoder, four convolutional layers are utilized. A batch normalization layer [45] is placed after each convolutional layer to accelerate the training speed and stabilize the learning process. The Leaky Rectified Linear Unit (LeakyReLU) activation [46], [47] is used in all layers in the encoder.

The convolutional layers in the encoder only connect all the feature maps together, but there are no direct connections among different locations within each specific feature map. Fully-connected layers are then applied to handle this information propagation. To reduce the number of parameters in the fully connected layers, a channel-wise fully-connected layer is used to connect the encoder and decoder, as in [21]. The channel-wise fully connected layer is designed to only propagate information within activations of each feature map. This significantly reduces the number of parameters in the network and accelerates the training process.

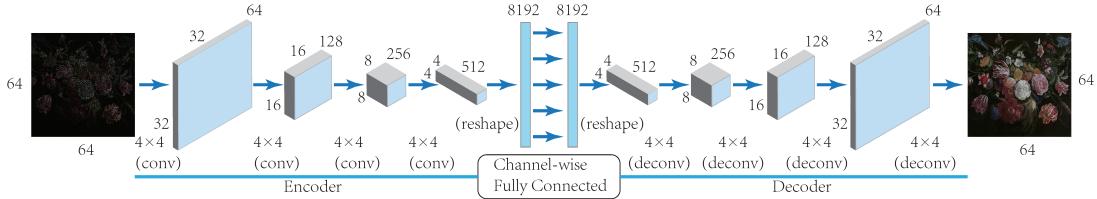


Fig. 6. Network architecture for the image inpainting network (*NetE*). The inpainting framework is an autoencoder style network with the encoder and decoder connected by a channel-wise fully-connected layer.

The decoder consists of four deconvolutional layers [48]–[50], each of which is followed by a ReLU activation except for the output layer. The tanh function is used in the output layer to restrict the pixel range of the output image. The series of up-convolutions and nonlinearities perform a nonlinear weighted upsampling of the feature produced by the encoder and generates an inpainted image of the target size (64×64).

C. Deep Learning Network Architecture for the Mask Generation Network

According to our knowledge, no prior work has been reported on generating the adaptive binary sampling mask utilizing CNNs. The desired mask generation network *NetM* should satisfy the following criteria:

- The output image should have the same spatial resolution as the input image.
- The network architecture should be fully convolutional to handle arbitrary input sizes.
- The output image should be binary.
- The output image should have a certain percentage c of 1's.

Inspired by [51], a network architecture with residual blocks is applied here, as shown in Figure 7. The network *NetM* consists of B residual blocks with identical layout. Following the method used in [52], we use two convolutional layers with small 3×3 kernels and 64 feature maps followed by batch-normalization layers [45] and ParametricReLU [53] as the activation function. The network *NetM* is fully convolutional to handle an arbitrary input size. To keep the spatial dimensions at each layer the same, the images are padded with zeros. At the end of each residual block is a final elementwise sum layer, followed by a Sigmoid activation layer. Let us denote by L_{ij} the $(i, j)^{th}$ element of L , which is the output of the Sigmoid activation layer, and mapped to the range $[0, 1]$. The Mean Adjustment layer F is defined as

$$D_{ij} = F(L_{ij}) = \frac{c}{\bar{L}} \times L_{ij}, \quad (4)$$

where D_{ij} is the $(i, j)^{th}$ element of matrix $D \in \mathbf{R}^2$, which is the output of the Mean Adjustment layer, and \bar{L} is the mean of L . Then $D_{ij} \in [0, 1]$ and the mean value of D , denoted by \bar{D} , will be equal to c . Finally, the Bernoulli distribution $Ber()$ is applied to binarize the values of D ; that is,

$$B_{ij} = Ber(D_{ij}) = \begin{cases} 1, & p = D_{ij} \\ 0, & p = 1 - D_{ij} \end{cases}. \quad (5)$$

Notice that

$$\bar{B} \approx \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N E(B_{ij}) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N D_{ij} = c, \quad (6)$$

where N^2 is the total number of pixels of L and $E(B_{ij})$ is the expected value of B_{ij} . Therefore, B is binary matrix with mean value equal to c , implying that it has c percent of 1's.

Since applying the function $Ber(F(\cdot))$ on the input L will make the output of the network be binary and have c percent of 1's, we then make it the last Binarization layer activation function. Notice that function $Ber(D)$ is not continuous and its derivatives do not exist, making the back propagation optimization during training impractical. We use its expected value D to approximate it during training and apply the original function $Ber(D)$ during testing.

IV. SPATIAL-SPECTRAL REPRESENTATION FOR X-RAY FLUORESCENCE IMAGE INPAINTING

In this section, we propose the XRF image inpainting algorithm by fusing it with a conventional RGB image, detailing the “Proposed Inpainting Algorithm” block in Figure 3. The proposed fusion style inpainting approach has similarities with our previous fusion style SR approach [31]. We first formulate the XRF image inpainting problem, then demonstrate our proposed solution to this inpainting problem.

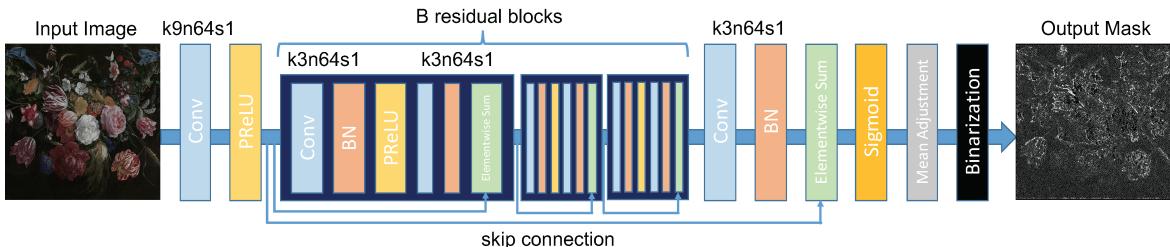


Fig. 7. Network architecture for the mask generation network (*NetM*). Residual blocks with skip connections are used in the binary sampling mask network.

A. Problem Formulation

As shown in Figure 4, we are seeking the estimation of a reconstructed XRF image $\bar{Y} \in \mathbb{R}^{W \times H \times B}$ that is fully sampled, with W , H , and B the image width, height, and number of spectral bands, respectively. We have two inputs: a subsampled XRF image $\bar{X} \in \mathbb{R}^{W \times H \times B}$ with the known binary sampling mask $S \in \mathbb{R}^{W \times H}$ ($\bar{X}(i, j, :) = 0$ if not sampled, i.e., corresponding to $\bar{S}(i, j) = 0$), and a conventional RGB image $\bar{I} \in \mathbb{R}^{W \times H \times b}$ with the same spatial resolution as the target XRF image \bar{Y} , but a small number (equal to 3) of spectral bands, i.e., $b \ll B$. Note that the RGB image is fully sampled, and therefore the primary goal of the reconstruction algorithm is to transfer image information from the RGB image to regions of the XRF where no samples are acquired. The input subsampled XRF image \bar{X} can be separated into two parts: the visible component $\bar{X}_v \in \mathbb{R}^{W \times H \times B}$ and the non-visible component $\bar{X}_{nv} \in \mathbb{R}^{W \times H \times B}$, with the same binary sampling mask \bar{S} as \bar{X} . We propose to estimate the fully sampled visible component $\bar{Y} \in \mathbb{R}^{W \times H \times B}$ by fusing the conventional RGB image \bar{I} with the visible component of the input subsampled XRF image \bar{X}_v , and the fully sampled non-visible component $\bar{Y}_{nv} \in \mathbb{R}^{W \times H \times B}$ by using standard total variation inpainting methods. Note that while our general framework is capable of handling the more difficult problem of separating visible and non-visible components in the XRF image, it also gracefully handles the problem when no underpainting is present.

To simplify notation, the image cubes are written as matrices, i.e., all pixels of an image are concatenated, such that every column of the matrix corresponds to the spectral response at a given pixel, and every row corresponds to a lexicographically ordered spectral band. Those unsampled pixels are skipped in this matrix representation. Accordingly, the image cubes are written as $Y \in \mathbb{R}^{B \times N_h}$, $X \in \mathbb{R}^{B \times N_s}$, $I \in \mathbb{R}^{b \times N_h}$, $X_v \in \mathbb{R}^{B \times N_s}$, $X_{nv} \in \mathbb{R}^{B \times N_s}$, $Y_v \in \mathbb{R}^{B \times N_h}$, $Y_{nv} \in \mathbb{R}^{B \times N_h}$, where $N_h = W \times H$ and $N_s = W \times H \times c$ is the number of sampled XRF pixels. We therefore have

$$X = X_v + X_{nv}, \quad (7)$$

$$Y = Y_v + Y_{nv}, \quad (8)$$

according to the visible/non-visible component separation models as shown in Figure 4.

Let us denote by $y_v \in \mathbb{R}^B$ and $y_{nv} \in \mathbb{R}^B$ the one-dimensional spectra at a given spatial location of \bar{Y}_v and \bar{Y}_{nv} respectively. That is, a column of Y_v and Y_{nv} is represented according to the linear mixing model [54], [55] described as

$$y_v = \sum_{j=1}^M d_{v,j}^{xrf} \alpha_{v,j}, \quad Y_v = D_v^{xrf} A_v, \quad (9)$$

$$y_{nv} = \sum_{j=1}^M d_{nv,j}^{xrf} \alpha_{nv,j}, \quad Y_{nv} = D_{nv}^{xrf} A_{nv}, \quad (10)$$

where $d_{v,j}^{xrf}$ and $d_{nv,j}^{xrf}$ are column vectors representing respectively the endmembers for the visible and non-

visible components, M is the total number of endmembers, $D_v^{xrf} \equiv [d_{v,1}^{xrf}, d_{v,2}^{xrf}, \dots, d_{v,M}^{xrf}] \in \mathbb{R}^{B \times M}$, $D_{nv}^{xrf} \equiv [d_{nv,1}^{xrf}, d_{nv,2}^{xrf}, \dots, d_{nv,M}^{xrf}] \in \mathbb{R}^{B \times M}$, and $\alpha_{v,j}$ and $\alpha_{nv,j}$ are the corresponding per-pixel abundances. Equation 8 holds for a specific column y_v of matrix Y_v (say the k^{th} column). We take the corresponding $\alpha_{v,j, j=1,\dots,M}$ and stack them into an $M \times 1$ column vector. This vector then becomes the k^{th} column of the matrix $A_v \in \mathbb{R}^{M \times N_h}$. In a similar manner, we construct matrix $A_{nv} \in \mathbb{R}^{M \times N_h}$. The endmembers D_v^{xrf} and D_{nv}^{xrf} act as basis dictionaries representing Y_v and Y_{nv} in a lower-dimensional space \mathbb{R}^M , with $\text{rank}\{Y_v\} \leq M$, and $\text{rank}\{Y_{nv}\} \leq M$.

The visible X_v and non-visible X_{nv} components of the input subsampled XRF image are spatially subsampled versions of Y_v and Y_{nv} respectively; that is

$$X_v = Y_v S = D_v^{xrf} A_v S, \quad (11)$$

$$X_{nv} = Y_{nv} S = D_{nv}^{xrf} A_{nv} S, \quad (12)$$

where $S \in \mathbb{R}^{N_h \times N_s}$ is the subsampling operator that describes the spatial degradation from the fully sampled XRF image to the subsampled XRF image.

Similarly, the input RGB image I can be described by the linear mixing model [54], [55],

$$I = D^{rgb} A_v, \quad (13)$$

where $D^{rgb} \in \mathbb{R}^{b \times M}$ is the RGB dictionary. Notice that the same abundance matrix A_v is used in Equations 9 and 11. This is because the visible component of the scanning object is captured by both the XRF and the conventional RGB images. Matrix A_v encompasses the spectral correlation between the visible component of the XRF and the RGB images.

B. Proposed Solution

To solve the XRF image inpainting problem, we need to estimate A_v , A_{nv} , D^{rgb} , D_v^{xrf} and D_{nv}^{xrf} simultaneously. Utilizing Equations 7, 11, 12, and 13, we formulate the following constrained least-squares problem:

$$\begin{aligned} & \min_{A_v, A_{nv}, D^{rgb}, D_v^{xrf}, D_{nv}^{xrf}} \|X - D_v^{xrf} A_v S - D_{nv}^{xrf} A_{nv} S\|_F^2 \\ & \quad + \gamma \|\nabla_I(D_v^{xrf} A_v)\|_F^2 + \lambda \|\nabla(D_{nv}^{xrf} A_{nv})\|_F^2 \\ & \quad + \|I - D^{rgb} A_v\|_F^2 \end{aligned} \quad (14a)$$

$$\text{s.t. } 0 \leq D_v^{xrf} \leq 1, \forall i, j \quad (14b)$$

$$0 \leq D_{nv}^{xrf} \leq 1, \forall i, j \quad (14c)$$

$$0 \leq D_{ij}^{rgb} \leq 1, \forall i, j \quad (14d)$$

$$A_{v,ij} \geq 0, \forall i, j \quad (14e)$$

$$A_{nv,ij} \geq 0, \forall i, j \quad (14f)$$

$$\mathbf{1}^T (A_v + A_{nv}) = \mathbf{1}^T, \quad (14g)$$

$$\|A_v + A_{nv}\|_0 \leq s, \quad (14h)$$

with $\|\cdot\|_F$ denoting the Frobenius norm and $\|\cdot\|_0$ the ℓ_0 norm, i.e., the number of nonzero elements of the given matrix.

$D_{v,ij}^{xrf}$, $D_{nv,ij}^{xrf}$, D_{ij}^{rgb} , $A_{v,ij}$, and $A_{nv,ij}$ are the (i,j) elements of matrices D_v^{xrf} , D_{nv}^{xrf} , D^{rgb} , A_v , and A_{nv} respectively, and $\mathbf{1}^T$ denotes a row vector of 1's compatible with the dimensions of A_v and A_{nv} . Equations 14b, 14c, and 14d enforce the nonnegative, bounded spectrum constraints on endmembers, Equations 14e and 14f enforce the nonnegative constraints on abundances, and Equation 14g enforces the constraint that the visible component abundances and non-visible component abundances for every pixel sum up to one. These physically grounded constraints from [37] are shown to be effective in our previous work [31], by making full use of the fact that the XRF endmembers are XRF spectra of individual materials, and the abundances are proportions of those endmembers.

The first term in Equation 14a represents a measure of the fidelity to the subsampled XRF data X , the second term is the total variation (TV) regularizer of the visible component, the third term is the TV regularizer of the non-visible component, and the last term is the fidelity to the observed RGB image I . The TV regularizer of the visible component $\nabla_I(D_v^{xrf} A_v)$ is defined as

$$\begin{aligned} & \|\nabla_I(D_v^{xrf} A_v)\|_F^2 \\ &= \sum_{i=1}^{H-1} \sum_{j=1}^{W-1} w_{i,j}^{down} \|D_v^{xrf} \bar{A}_v(i,j,:) - D_v^{xrf} \bar{A}_v(i+1,j,:)\|_2^2 \\ &\quad + w_{i,j}^{right} \|D_v^{xrf} \bar{A}_v(i,j,:) - D_v^{xrf} \bar{A}_v(i,j+1,:)\|_2^2 \\ &= \|D_v^{xrf} A_v P(I)\|_F^2, \end{aligned} \quad (15)$$

where $\bar{A}_v \in \mathbb{R}^{W \times H \times M}$ is the 3D volume version of A_v and $\bar{A}_v(i,j,:) \in \mathbb{R}^M$ is the non-visible component abundance of pixel (i,j) . $w_{i,j}^{down}$ and $w_{i,j}^{right}$ are the adaptive TV weights in the vertical and horizontal directions respectively; that is,

$$w_{i,j}^{down} = e^{-\alpha \|\bar{I}(i,j,:) - \bar{I}(i+1,j,:)\|_2^2}, \quad (16)$$

$$w_{i,j}^{right} = e^{-\alpha \|\bar{I}(i,j,:) - \bar{I}(i,j+1,:)\|_2^2}, \quad (17)$$

where $\bar{I}(i,j,:)$ is the RGB image pixel at position (i,j) . $P(I) \in \mathbb{R}^{N_h \times ((W-1)(H-1))}$ in Equation 15 is the adaptive horizontal/vertical first order difference operator determined by the input RGB image I according to Equations 16 and 17. Equations 16 and 17 indicate that the TV regularizer of the visible component adapts to the dense set of information that is available in the conventional RGB image \bar{I} . When the difference between two adjacent RGB pixels is small, a strong spatial smoothness constraint is placed on their corresponding XRF pixels, and vice versa. This adaptive TV regularizer is one of the main differences between this fusion-based XRF image inpainting algorithm and our previous fusion-based XRF image SR algorithm [31]. We found out that such TV regularizer on the visible component is essential for the inpainting problem, otherwise the inpainting results are not satisfactory. For the SR approach, we do not need such a TV regularizer on the visible component. The SR degradation model assumes that the LR measured XRF image is a weighted sum of all the pixels in the target HR XRF image, so there is an implicit spatial smoothness constraint imposed by the LR XRF

image. However, for the XRF image inpainting problem, we subsample the XRF image to obtain the measurement so that many pixels are not sampled at all, making the reconstruction more difficult than for the SR problem. Also, the mapping from RGB to XRF pixels is one to many, meaning that utilizing the RGB image cannot guarantee a spatially smooth XRF reconstruction.

The optimization in Equation 14 is non-convex and difficult to carry out if we are to optimize over all the parameters A_v , A_{nv} , D^{rgb} , D_v^{xrf} , and D_{nv}^{xrf} directly. We found it effective to alternately optimize over these parameters. Also, because Equation 14 is highly non-convex, a good initialization is needed. Let $Y^{(0)} \in \mathbb{R}^{B \times N_h}$ be the initialization of Y . Such initialization can be obtained by utilizing some standard image inpainting algorithms [42], [44] to inpaint the subsampled XRF image channel by channel. Then, the coupled dictionary learning technique in [56], [57] can be utilized to initialize D^{rgb} and D_v^{xrf} by

$$\begin{aligned} \min_{D^{rgb}, D_v^{xrf}} \quad & \|I - D^{rgb} A_v\|_F^2 + \|Y^{(0)} - D_v^{xrf} A_v\|_F^2 \\ & + \beta \sum_{k=1}^{N_l} \|A_v(:,k)\|_1, \\ \text{s.t.} \quad & \|D^{rgb}(:,k)\|_2 \leq 1, \forall k, \\ & \|D_v^{xrf}(:,k)\|_2 \leq 1, \forall k, \end{aligned} \quad (18)$$

where $\|\cdot\|_1$ is the ℓ_1 vector norm, parameter β controls the sparseness of the coefficients in A_v , and $A_v(:,k)$, $D^{rgb}(:,k)$, and $D_v^{xrf}(:,k)$ denote the k^{th} column of matrices A_v , D^{rgb} , and D_v^{xrf} respectively. Details of the optimization can be found in [56], [57]. D^{rgb} and D_v^{xrf} are initialized using Equation 18 and D_{nv}^{xrf} is initialized to be equal to D_v^{xrf} . A_v is initialized by Equation 18 as well, while A_{nv} is set equal to zero at initialization. Note that this formulation allows our approach to naturally handle the problem when no hidden layers are present. The iterative optimization algorithms we deploy are similar to the ones used in our previous work [31], and they are not repeated here.

Once the optimization problem in Equation 14 is solved according to Equations 8, 9, and 10, the reconstructed fully sampled XRF output image Y can be computed by

$$Y = Y_v + Y_{nv} = D_v^{xrf} A_v + D_{nv}^{xrf} A_{nv}. \quad (19)$$

V. EXPERIMENTAL RESULTS

In this section, we will first demonstrate the advantages of our developed adaptive sampling mask when applied to RGB image sampling and reconstruction. We will then demonstrate the performance of the proposed fusion-based XRF inpainting algorithm.

A. Adaptive Sampling Mask for RGB Image Sampling and Reconstruction

For the RGB image sampling and reconstruction task, we show the benefits of our developed adaptive sampling mask, over the use of a random sampling mask, as well as other adaptive sampling algorithms, such as Adaptive Irregular

Sampling (AIrS) proposed in [8] and Kernel-based Adaptive Sampling (KbAS) proposed in [10].

1) *Datasets*: To train our proposed adaptive sampling mask generation CNN in Section III, the ImageNet [58] database without any of the accompanying labels is used. We randomly select 1,000,000 images for the training set, 100 for the validation set, and 100 for the testing set. All images are selected randomly among all categories to capture as diverse image structures as possible and are cropped to have spatial resolution 64×64 pixel.

2) *Implementation Details*: Our proposed adaptive sampling mask generation CNN (Section III) is implemented in PyTorch¹. ADAM [60] is applied as the stochastic gradient descent solver for optimization. We use the same hyperparameters suggested in [23] and batch size equal to 128 during training. 400 epochs are applied during the training process. We test 3 sampling rates: $c = 5\%$, $c = 10\%$, and $c = 20\%$ for the sampling percentage parameter. A 5%, 10%, and 20% sampling percentage roughly speed up the raster scanning procedure by a factor of 20, 10, and 5, respectively.

For training, we first initialize the inpainting network $NetE$ according to [23]. A random sampling mask with sampling rate c is utilized to sample the input RGB image. The mask generation network $NetM$ is initialized randomly. We then train the whole network architecture in Figure 5. The learning rate of the mask generation network $NetM$ is set equal to 0.0002 during training. The learning rate of the inpainting network $NetE$ is set to 0, i.e., we fix $NetE$ when training $NetM$. We did not optimize $NetM$ and $NetE$ simultaneously; although the best reconstruction of z would have been obtained, the two networks would have been dependent on each other. Notice that the channel-wise fully connected layer in $NetE$ (Figure 6) is able to learn a high-level feature mapping, making $NetE$ able to perform semantic image inpainting. However, we would like to utilize other image inpainting algorithms other than $NetE$ to make the adaptive sampling mask be as general and applicable to as many image inpainting algorithms as possible. By fixing $NetE$, which is pre-trained by random sampling masks, $NetM$ is constrained to be optimized for the general image inpainting problem instead of the semantic image inpainting problem. An adaptive sampling mask is thus trained suitable for general image inpainting problem. Note that, since $NetE$ architecture is differentiable, it allows us to train the $NetM$ architecture to optimize the mask, while taking into account both the RGB image content, and the reconstruction algorithm.

3) *Performance on ImageNet Testing Images*: To compare the performance of our adaptive sampling mask ($NetM$) with the random, AIrS [8], and KbAS [10] sampling masks, we apply all samplings to corrupt those 100 testing images from the ImageNet database. Three sampling rates $c = 0.05$, $c = 0.1$ and $c = 0.2$ are tested for the sampling methods. For image inpainting algorithms, $NetE$ Inpainting [23], Harmonic Inpainting [59], Mumford-Shah (Mum-Sh) Inpainting [42], and BPFA inpainting [44] are used to reconstruct the fully sampled RGB images.

The average PSNRs in dB over all 100 test images are shown in Table I. First, we observe that under all three sampling rates, the proposed $NetM$ mask outperforms the random, AIrS, and KbAS sampling masks consistently over all inpainting reconstruction algorithms in terms of PSNR, showing the effectiveness of our proposed adaptive sampling mask generation network. Furthermore, we find that the proposed $NetM$ is significantly faster than AIrS and KbAS for sampling mask generation. Finally, it can be concluded that the smaller the sampling rate, the larger the advantage of our proposed algorithm compared to other sampling algorithms. This implies that our proposed $NetM$ is able to handle challenging sampling tasks (small sampling rates) and have better reconstruction accuracy under various reconstruction algorithms.

The visual quality comparison of the adaptive sampling and random sampling masks is shown in Figure 8. Two test images are picked from the testing set of 100 images in total. Under sampling rate $c = 0.2$, the advantages of the proposed $NetM$ mask over the random, AIrS, and KbAS sampling masks can be observed by comparing the reconstruction results of the same inpainting algorithm over different sampling strategies. For the test image #39, the $NetM$ mask is able to capture the white dots in the red hat, resulting in accurate reconstruction results of those white dots. KbAS misses one white dot in the image. AIrS is able to sample all the white dots in the image, however it fails to capture the structure of these white dots. For test image #91, compared to random sampling mask, the proposed $NetM$ mask samples the contour structure of the bird, resulting in its better reconstruction. When compared to AIrS and KbAS masks, the proposed $NetM$ mask samples the whole image more evenly, resulting in fewer artifacts in the inpainted images. The improved performance of the adaptive sampling mask over other sampling masks is consistent over all inpainting reconstruction algorithms we tested.

4) *Performance on Painting Images*: We also tested our proposed adaptive sampling algorithm on painting images at sampling rate $c = 0.1$. As shown in Figures 9 (a), the RGB image of the painting “Bloemen en insecten” is tested. It has spatial resolution 580×680 pixels. Random, AIrS, KbAS, and $NetM$ sampling masks are generated as shown in Figures 9(b), (f), (j), and (n) with the corresponding computation time. Harmonic Inpainting [59], Mumford-Shah Inpainting [42], and BPFA [44] algorithms are utilized to reconstruct the sampled RGB images, and the reconstruction results are shown in Figures 9 (c)-(e), (g)-(i), (k)-(m), and (o)-(q) with the corresponding PSNR values. By comparing the rows of different sampling masks, it can be concluded that our proposed $NetM$ mask outperforms other sampling masks in terms of both visual quality of the reconstructed images and the PSNR values. Notice that $NetM$ is significantly faster than AIrS and KbAS in computation speed. Although KbAS samples densely on the foreground, it still misses many details due to the complexity of the flower structure. $NetE$ Inpainting [23] is not utilized in this experiment since it is trained to inpaint RGB images with spatial resolution 64×64 pixels. The network structure shown in Figure 6 is not fully convolutional, as there is the channel-wise fully connected

¹<https://github.com/usstdqq/deep-adaptive-sampling-mask>

	c=0.05					c=0.1					c=0.2				
	NetE	Harmonic	Mum-Sh	BPFA	time(s)	NetE	Harmonic	Mum-Sh	BPFA	time(s)	NetE	Harmonic	Mum-Sh	BPFA	time(s)
Random	19.72	19.49	20.09	16.32	-	21.56	21.45	21.78	20.47	-	23.61	23.43	23.80	23.47	-
AIrS	18.25	17.21	19.63	13.41	0.22	19.88	19.60	21.80	17.26	0.23	23.91	24.67	25.41	22.77	0.24
KbAS	19.58	18.78	21.61	14.99	23.37	21.97	22.28	24.00	19.73	50.87	24.24	25.41	26.58	23.83	104.55
<i>NetM</i>	22.05	21.71	22.37	17.06	0.02	23.51	24.07	24.76	22.54	0.02	25.81	26.14	26.93	25.96	0.02

TABLE I

SAMPLING AND INPAINTING RESULTS ON IMAGENET TEST IMAGES. RANDOM, AIRS [8], KBAS [10], AND PROPOSED *NetM* SAMPLING STRATEGIES ARE COMPARED IN PSNR UNDER *NetE* [23], HARMONIC [59], MUM-SH [42], AND BPFA [44] INPAINTING ALGORITHMS. COMPUTATION TIME OF DIFFERENT SAMPLING STRATEGIES IS ALSO COMPARED. BEST RESULTS ARE SHOWN IN BOLD. THE RESULTS SHOWN ARE AVERAGED OVER A SET OF 100 TEST IMAGES.

layer in the middle.

B. Adaptive Sampling Mask for X-Ray Fluorescence Image Inpainting

In the previous section (Section V-A), we demonstrated the effectiveness of our proposed *NetM* sampling mask on the RGB image sampling and inpainting problem. To further evaluate the effectiveness of the *NetM* sampling mask and evaluate the performance of our proposed fusion-based inpainting algorithm (Section IV), we have performed experiments on XRF images. The basic parameters of the proposed reconstruction method are set as follows: the number of atoms in the dictionaries D^{rgb} , D_{nv}^{xrf} and D_v^{xrf} is $M = 200$; $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 1.01$, which only affects the speed of convergence; parameter λ and γ in Equation 14 are set equal to 0.1; parameter α in Equation 16 and Equation 17 is set to 16. The optional constraint in Equation 14h is not applied here.

1) *Error Metrics*: The root mean squared error (RMSE), the peak-signal-to-noise ratio (PSNR), and the spectral angle mapper (SAM, [61]) between the estimated fully sampled XRF image Y and the ground truth image Y^{gt} are used as the error metrics.

2) *Comparison Methods*: According to our knowledge, no work has been reported on solving the XRF (or Hyperspectral) image inpainting problem by fusing a conventional RGB image. So, we can only compare our results with traditional image inpainting algorithms such as Harmonic Inpainting [59], Mumford-Shah Inpainting [42], and BPFA inpainting [44]. Harmonic Inpainting and Mumford-Shah Inpainting methods are for image inpainting, so we have to inpaint the XRF image channel by channel. BPFA inpainting [44] is able to inpaint multiple channels simultaneously. For the sampling mask comparison, we still compare with random, AIrS [8], and KbAS [10] sampling masks.

3) *Real Experiment*: For this experiment, real data were collected by a home-built X-ray fluorescence spectrometer (courtesy of Prof. Koen Janssens), with 2048 channels in spectrum. Studies from the XRF image scanned from Jan Davidsz. de Heem's "Bloemen en insecten" (ca 1645), in the collection of Koninklijk Museum voor Schone Kunsten (KMKSA) Antwerp, are presented here. We utilize the super-resolved XRF image in our previous work [31] as the ground truth. The ground truth XRF image has dimensions $680 \times 580 \times 2048$. We first extract 20 regions of interest (ROI) spectrally and work on them, to decrease the spectral dimension from 2048 to 20.

We decrease the spectral dimension so as to compare with other inpainting algorithms, e.g., [42], [59], which therefore reconstruct the subsampled XRF image channel by channel and large spectral dimensions will make the computational time very long. The sampling ratio c is set to be 0.05, 0.1, and 0.2. Different sampling strategies are applied and analyzed. Various inpainting methods are subsequently applied to reconstruct those subsampled XRF images.

As shown in Table II, our proposed fusion-based inpainting algorithm with the proposed adaptive sampling mask provides the closest reconstruction to the ground truth XRF image compared to all other methods. Our proposed algorithm utilizes as guidance a conventional fully sampled and high contrast RGB image (Figure 9 (a)), resulting in better inpainting performance. By comparing the difference between results by "Mum-Sh" and results by "Proposed" under various sampling rates, it can be concluded that the benefit gained by our proposed fusion-based inpainting is large when the adaptive sampling masks are applied. For example, at sampling rate $c = 0.2$, there is a 0.78 dB improvement in PSNR by applying our proposed fusion-based inpainting when a random sampling mask is applied, while there is a 6.71 dB , 7.94 dB , and 7.33 dB improvement in PSNR when an AIrS, KbAS, and *NetM* sampling mask is applied, respectively. This is because the adaptive sampling masks sampled the corresponding visible component of the XRF image efficiently and the fusion inpainting propagated the measured XRF pixels properly. Furthermore, the *NetM* sampling mask provides the best XRF image reconstruction compared to other sampling masks.

The iteration process of our proposed fusion inpainting algorithm when $c = 0.2$ is shown in Figure 10. Notice that at the initial iterations of our proposed fusion inpainting algorithm, the RMSE is higher than the Mumford-Shah inpainting algorithm. This is because we decompose the inpainting result of Mumford-Shah inpainting algorithm by sparse representation, according to Equation 18. Due to the complexity of the "Bloemen en insecten" data, we lose some accuracy during the first few iterations. However, with more iterations, the RMSE of both random sampling and adaptive sampling decreases and becomes smaller than the RMSE of Mumford-Shah inpainting algorithm. Our proposed *NetM* sampling mask provides the smallest RMSE for both Mumford-Shah initialization and the proposed fusion inpainting algorithm, showing the effectiveness of *NetM* in the task of XRF image sampling and reconstruction.

When $c = 0.2$, the visual quality of the different in-

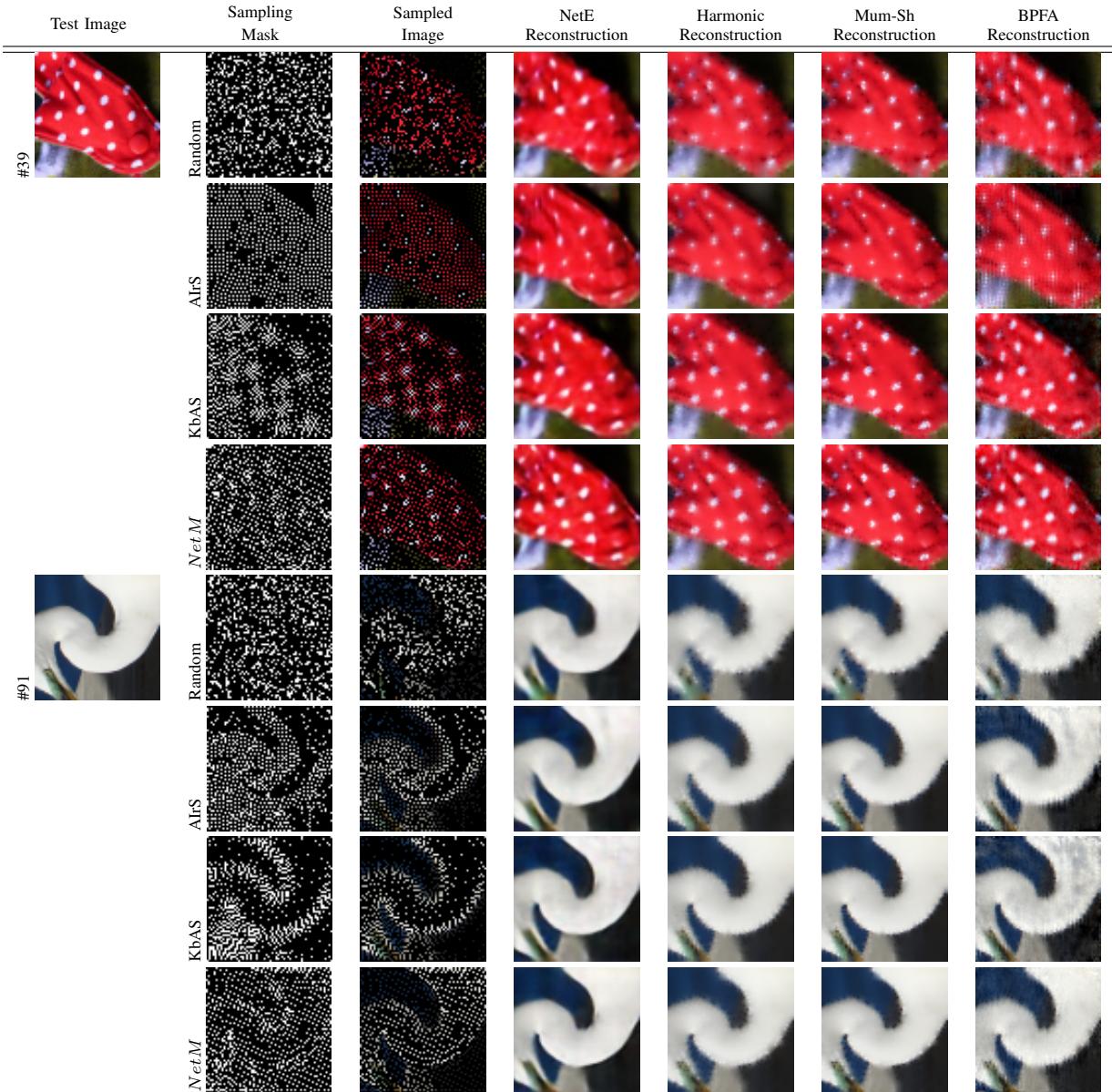


Fig. 8. Visual Comparison of the reconstructed images using random, AIRs, KbAS, and *NetM* sampling masks at sampling rate $c = 0.2$. The first column is the input test image and the second column is the sampling mask, either random, AIRs, KbAS, or *NetM*, the third column is the sampled image obtained by the sampling mask, and the rest of the columns are the reconstruction results of *NetE* Inpainting [23], Harmonic Inpainting [59], Mumford-Shah Inpainting [42], and BPFA inpainting [44] respectively.

painting algorithms and sampling strategies on channel #16, corresponding to the $Pb\ L\gamma$ XRF emission line, is compared in Figure 11. The same random, AIRs, KbAS, and *NetM* sampling masks as the sampling masks in Figures 9 (b), (f), (j), and (n), are applied here. Notice that the two long rectangles in the ground truth XRF image in Figure 11 (a) are the stretcher bars under the canvas, which is not visible on the RGB image. The reconstruction results of Harmonic, Mum-Sh, BPFA, and Proposed using AIRs, KbAS, and *NetM* sampling masks are sharper than those using the random sampling mask. This is because the majority of the XRF signal in the “Bloemen en insecten” data correlates to the RGB signal, and the sampling masks, which are designed for the RGB image, would also be suitable for the visible component of the XRF signal. Among the AIRs, KbAS, and *NetM* sampling masks, *NetM*

sampling mask outperforms others in different reconstruction algorithms, both quantitatively (Table II) and qualitatively (Figure 11). The proposed fusion inpainting algorithm further improves the contrast and resolves more fine details in (q). When compared to the ground truth image (a), we can conclude that those resolved details have high fidelity to the ground truth image (a). The reconstructed stretcher bars in all cases are blurry compared to other objects because they do not utilize any information from the RGB image, both in sampling and inpainting.

Our proposed adaptive sampling algorithm *NetM* is not designed and trained for the task of XRF image sampling and reconstruction. However, *NetM* still outperforms random sampling, as well as two other adaptive sampling algorithms, under various XRF reconstruction algorithms and under differ-

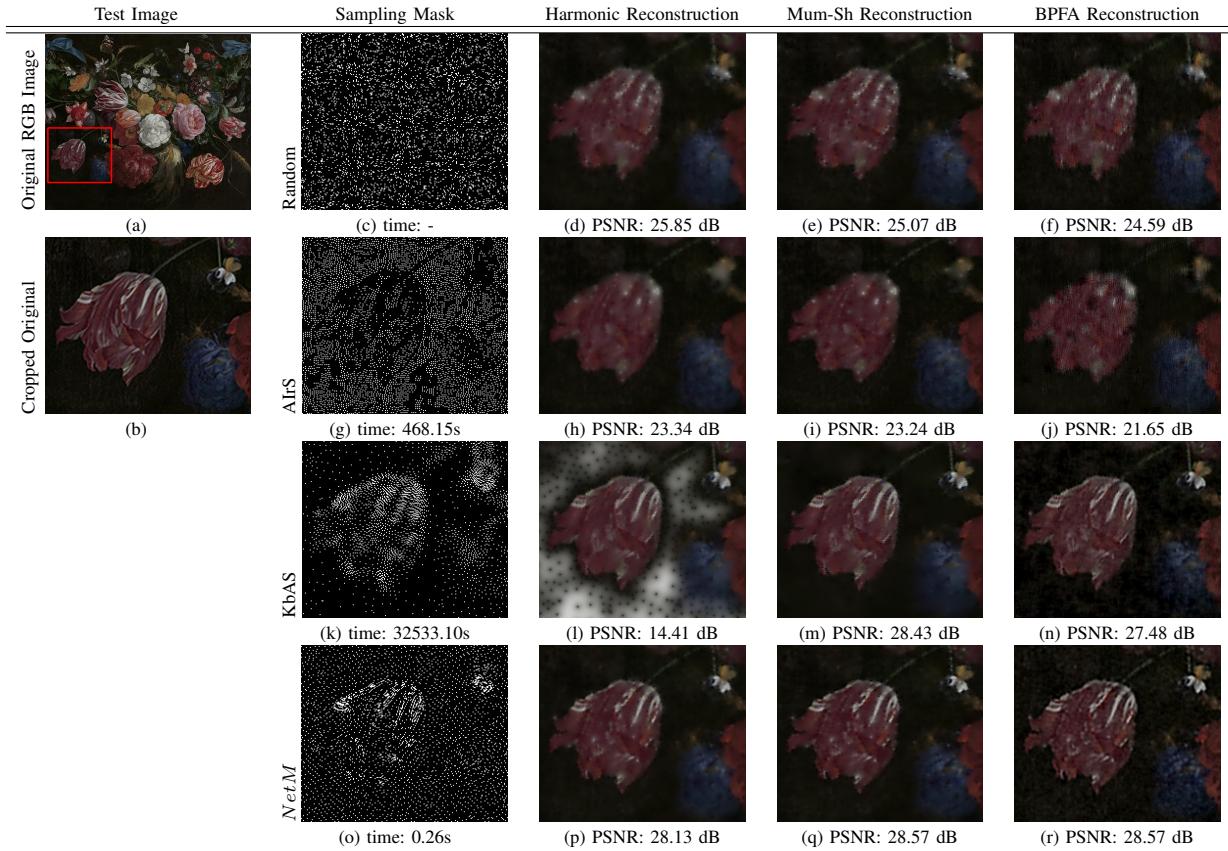


Fig. 9. Visualization of sampling and inpainting result of the “Bloemen en insecten” painting. (a) original RGB image with red bounding box. (b) region inside the bounding box of (a) for visualization purposes. (c), (g), (k), and (o) random, AIsS, KbAS, and *NetM* sampling masks respectively. (d), (h), (l), and (p) reconstruction results of each sampling mask using Harmonic algorithms. (e), (i), (m), and (q) reconstruction results of each sampling mask using Mumford-Shah algorithm. (f), (j), (n), and (r) reconstruction results of each sampling mask using BPFA algorithm. Computation time of each sampling mask and PSNR of the entirety of each reconstructed images are also shown.

some other imaging tasks.

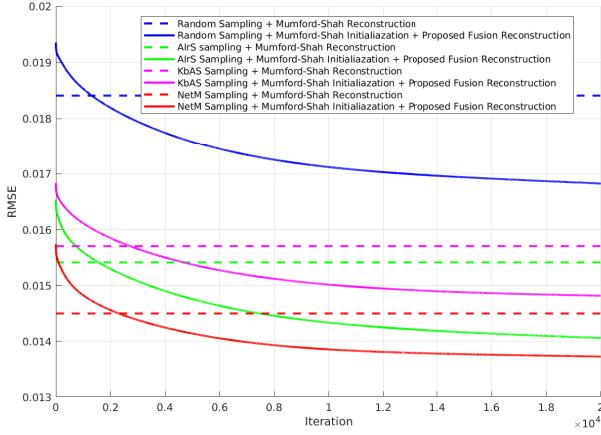


Fig. 10. RMSE versus number of iterations for our proposed fusion inpainting algorithm on the “Bloemen en insecten” data when $c = 0.2$. The Mumford-Shah inpainting algorithm is utilized as initialization of our proposed algorithm. Different sampling masks are compared.

ent sampling rates. This illustrates the effectiveness of *NetM* in extracting image information with limited sampling budget. This also demonstrates that *NetM* can generalize well into

VI. CONCLUSION

In this paper, we presented a novel adaptive sampling mask generation algorithm based on CNNs and a novel XRF image inpainting framework based on fusing a conventional RGB image. For the adaptive sampling mask generation, we trained the mask generation network *NetM* along with the inpainting network *NetE* to obtain an optimal binary sampling mask based on the input RGB image. For the fusion-based XRF image inpainting algorithm, the XRF spectrum of each pixel is represented by an endmember dictionary, as well as the RGB spectrum. The input subsampled XRF image is decomposed into visible and non-visible components, making it possible to find the nonlinear mapping from the RGB to the XRF spectrum. A total variation regularizer is applied to both visible and non-visible components to ensure the spatial smoothness of the reconstructed XRF image. The reconstructed fully sampled visible XRF component and the fully sampled non-visible XRF component are combined to obtain the final fully sampled XRF image. Experiments show the effectiveness of our proposed network *NetM* in both RGB and XRF image sampling and reconstruction tasks. Higher reconstruction accuracy is achieved in both RGB and XRF image sampling, and

	c=0.05											
	Harmonic			Mum-Sh			BPFA			Proposed		
	RMSE	PSNR	SAM									
Random	0.0428	27.37	5.06	0.0319	29.93	3.85	0.1141	22.57	8.27	0.0300	30.45	3.67
AIrS	0.1224	18.24	8.20	0.0378	28.45	4.33	0.1446	17.80	13.49	0.0355	37.03	4.14
KbAS	0.2374	12.49	12.62	0.0309	30.19	3.95	0.1791	15.08	19.51	0.0297	39.25	3.85
<i>NetM</i>	<i>0.0333</i>	29.56	4.22	<i>0.0279</i>	<i>31.09</i>	<i>3.43</i>	<i>0.0910</i>	24.22	7.21	0.0261	39.54	3.27
	c=0.1											
	Harmonic			Mum-Sh			BPFA			Proposed		
	RMSE	PSNR	SAM									
Random	0.0260	31.70	3.10	0.0247	32.14	2.84	0.0369	28.71	3.27	0.0229	32.80	2.66
AIrS	0.0408	27.78	3.77	0.0284	30.92	3.04	0.0703	23.12	4.30	0.0260	38.54	2.84
KbAS	0.1473	16.63	7.78	0.0231	32.74	2.83	0.1268	18.06	9.00	0.0220	41.59	2.73
<i>NetM</i>	<i>0.0227</i>	<i>32.87</i>	<i>2.84</i>	<i>0.0211</i>	<i>33.53</i>	<i>2.50</i>	<i>0.0353</i>	<i>29.66</i>	<i>3.61</i>	0.0195	41.87	2.38
	c=0.2											
	Harmonic			Mum-Sh			BPFA			Proposed		
	RMSE	PSNR	SAM									
Random	0.0195	34.19	2.18	0.0184	34.70	1.92	0.0176	35.29	2.01	0.0168	35.48	1.79
AIrS	0.0185	34.66	1.94	0.0154	36.24	1.59	0.0336	29.50	2.48	0.0141	42.95	1.49
KbAS	0.0518	25.71	3.35	0.0157	36.08	1.78	0.0683	23.39	3.21	0.0148	44.02	1.72
<i>NetM</i>	<i>0.0160</i>	<i>35.90</i>	<i>1.86</i>	<i>0.0145</i>	<i>36.77</i>	<i>1.54</i>	<i>0.0151</i>	<i>36.70</i>	<i>1.80</i>	0.0137	44.10	1.48

TABLE II

EXPERIMENTAL RESULTS ON THE “BLOEMEN EN INSECTEN” DATA COMPARING DIFFERENT INPAINTING METHODS, UNDER RANDOM, AIrS, KBAS, AND *NetM* SAMPLING STRATEGIES, DISCUSSED IN SECTION V-B2 IN TERMS OF RMSE, PSNR, AND SAM. BEST SAMPLING MASKS UNDER THE EACH RECONSTRUCTION ALGORITHM ARE SHOWN IN ITALIC. BEST RECONSTRUCTION RESULTS ARE SHOWN IN BOLD.

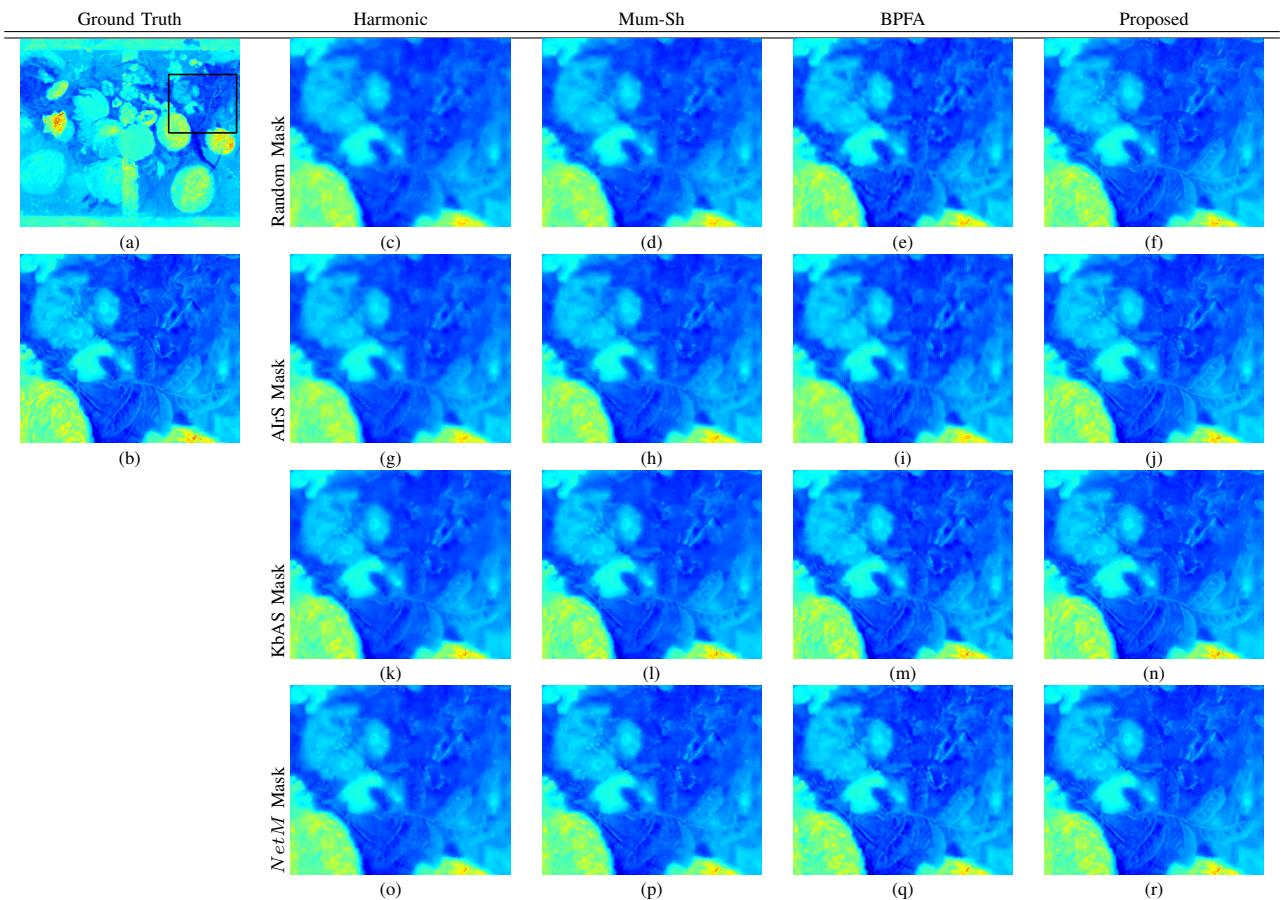


Fig. 11. Visualization of inpainting results on the “Bloemen en insecten” data when $c = 0.2$. Channel #16 related to the $Pb\ L\eta$ XRF emission line is selected. (a) ground truth XRF image with black bounding box. (b) region inside the bounding box of (a), shown for visual comparison purposes. The sampling masks for random, AIrS, KbAS, and *NetM* are the same as the sampling masks in Figures 9 (b), (f), (j), and (n) respectively. (c)-(r) reconstruction results of different inpainting algorithms for different sampled XRF image within the same region of (a) as (b).

also the computation time of mask generation is significantly smaller than other adaptive sampling methods.

REFERENCES

- [1] E. J. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency informa-

- tion,” *IEEE Transactions on information theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [2] D. L. Donoho, “Compressed sensing,” *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [3] R. L. Cook, “Stochastic sampling in computer graphics,” *ACM Transactions on Graphics (TOG)*, vol. 5, no. 1, pp. 51–72, 1986.
- [4] M. Soumekh, “Multiresolution dynamic image representation with uniform and foveal spiral scan data,” *IEEE transactions on image processing*, vol. 7, no. 11, pp. 1627–1635, 1998.
- [5] Y. Eldar, M. Lindenbaum, M. Porat, and Y. Y. Zeevi, “The farthest point strategy for progressive image sampling,” *IEEE Transactions on Image Processing*, vol. 6, no. 9, pp. 1305–1315, 1997.
- [6] S. Rajesh, K. Sandeep, and R. Mittal, “A fast progressive image sampling using lifting scheme and non-uniform b-splines,” in *Industrial Electronics, 2007. ISIE 2007. IEEE International Symposium on*. IEEE, 2007, pp. 1645–1650.
- [7] L. Demaret, N. Dyn, and A. Iske, “Image compression by linear splines over adaptive triangulations,” *Signal Processing*, vol. 86, no. 7, pp. 1604–1616, 2006.
- [8] G. Ramponi and S. Carrato, “An adaptive irregular sampling algorithm and its application to image coding,” *Image and Vision Computing*, vol. 19, no. 7, pp. 451–460, 2001.
- [9] A. S. Lin, B. Z. Luo, C. J. Zhang, and D. E. Saucan, “Generalized ricci curvature based sampling and reconstruction of images,” in *Signal Processing Conference (EUSIPCO), 2015 23rd European*. IEEE, 2015, pp. 604–608.
- [10] J. Liu, C. Bouganis, and P. Y. Cheung, “Kernel-based adaptive image sampling,” in *Computer Vision Theory and Applications (VISAPP), 2014 International Conference on*, vol. 1. IEEE, 2014, pp. 25–32.
- [11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
- [12] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.
- [13] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, “Extracting and composing robust features with denoising autoencoders,” in *Proceedings of the 25th international conference on Machine learning*. ACM, 2008, pp. 1096–1103.
- [14] H. C. Burger, C. J. Schuler, and S. Harmeling, “Image denoising: Can plain neural networks compete with bm3d?” in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2392–2399.
- [15] C. Dong, C. C. Loy, K. He, and X. Tang, “Learning a deep convolutional network for image super-resolution,” in *European Conference on Computer Vision*. Springer, 2014, pp. 184–199.
- [16] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, “Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1874–1883.
- [17] P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid, “Deepflow: Large displacement optical flow with deep matching,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1385–1392.
- [18] A. Kappeler, S. Yoo, Q. Dai, and A. K. Katsaggelos, “Video super-resolution with convolutional neural networks,” *IEEE Transactions on Computational Imaging*, vol. 2, no. 2, pp. 109–122, 2016.
- [19] M. Iliadis, L. Spinoulas, and A. K. Katsaggelos, “Deep fully-connected networks for video compressive sensing,” *Digital Signal Processing*, vol. 72, pp. 9–18, 2018.
- [20] ——, “Deepbinarymask: Learning a binary mask for video compressive sensing,” *arXiv preprint arXiv:1607.03343*, 2016.
- [21] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, “Context encoders: Feature learning by inpainting,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2536–2544.
- [22] R. Yeh, C. Chen, T. Y. Lim, M. Hasegawa-Johnson, and M. N. Do, “Semantic image inpainting with perceptual and contextual losses,” *arXiv preprint arXiv:1607.07539*, 2016.
- [23] R. Gao and K. Grauman, “From one-trick ponies to all-rounders: On-demand learning for image restoration,” *arXiv preprint arXiv:1612.01380*, 2016.
- [24] K.-H. Tzou, “Progressive image transmission: a review and comparison of techniques,” *Optical engineering*, vol. 26, no. 7, p. 267581, 1987.
- [25] M. Alfeld, J. V. Pedroso, M. van Eikema Hommes, G. Van der Snickt, G. Tauber, J. Blaas, M. Haschke, K. Erler, J. Dik, and K. Janssens, “A mobile instrument for in situ scanning macro-xrf investigation of historical paintings,” *Journal of Analytical Atomic Spectrometry*, vol. 28, no. 5, pp. 760–767, 2013.
- [26] A. Anitha, A. Brasoveanu, M. Duarte, S. Hughes, I. Daubechies, J. Dik, K. Janssens, and M. Alfeld, “Restoration of x-ray fluorescence images of hidden paintings,” *Signal Processing*, vol. 93, no. 3, pp. 592–604, 2013.
- [27] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, “Image inpainting,” in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 2000, pp. 417–424.
- [28] A. Criminisi, P. Pérez, and K. Toyama, “Region filling and object removal by exemplar-based image inpainting,” *IEEE Transactions on image processing*, vol. 13, no. 9, pp. 1200–1212, 2004.
- [29] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher, “Simultaneous structure and texture image inpainting,” *IEEE transactions on image processing*, vol. 12, no. 8, pp. 882–889, 2003.
- [30] Q. Dai, E. Pouyet, O. Cossairt, M. Walton, F. Casadio, and A. Katsaggelos, “X-ray fluorescence image super-resolution using dictionary learning,” in *Image, Video, and Multidimensional Signal Processing Workshop (IVMSP), 2016 IEEE 12th*. IEEE, 2016, pp. 1–5.
- [31] Q. Dai, E. Pouyet, O. Cossairt, M. Walton, and A. Katsaggelos, “Spatial-spectral representation for x-ray fluorescence image super-resolution,” *IEEE Transactions on Computational Imaging*.
- [32] C. M. Pieters and P. A. Englert, *Remote geochemical analysis, elemental and mineralogical composition*, 1993, vol. 1.
- [33] D. Manolakis, C. Siracusa, and G. Shaw, “Hyperspectral subpixel target detection using the linear mixing model,” *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 39, no. 7, pp. 1392–1409, 2001.
- [34] M. Alfeld, W. De Nolf, S. Cagno, K. Appel, D. P. Siddons, A. Kuczewski, K. Janssens, J. Dik, K. Trentelman, M. Walton *et al.*, “Revealing hidden paint layers in oil paintings by means of scanning macro-xrf: a mock-up study based on rembrandt’s an old man in military costume,” *Journal of Analytical Atomic Spectrometry*, vol. 28, no. 1, pp. 40–51, 2013.
- [35] R. Kawakami, J. Wright, Y.-W. Tai, Y. Matsushita, M. Ben-Ezra, and K. Ikeuchi, “High-resolution hyperspectral imaging via matrix factorization,” in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 2329–2336.
- [36] N. Akhtar, F. Shafait, and A. Mian, “Sparse spatio-spectral representation for hyperspectral image super-resolution,” in *Computer Vision–ECCV 2014*. Springer, 2014, pp. 63–78.
- [37] C. Lanaras, E. Baltsavias, and K. Schindler, “Hyperspectral super-resolution by coupled spectral unmixing,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 3586–3594.
- [38] W. Dong, F. Fu, G. Shi, X. Cao, J. Wu, G. Li, and X. Li, “Hyperspectral image super-resolution via non-negative structured sparse representation,” 2016.
- [39] R. C. Hardie, M. T. Eismann, and G. L. Wilson, “Map estimation for hyperspectral image resolution enhancement using an auxiliary sensor,” *Image Processing, IEEE Transactions on*, vol. 13, no. 9, pp. 1174–1184, 2004.
- [40] S. D. Babacan, R. Molina, and A. K. Katsaggelos, “Total variation super resolution using a variational approach,” in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. IEEE, 2008, pp. 641–644.
- [41] A. Marquina and S. J. Osher, “Image super-resolution by tv-regularization and bregman iteration,” *Journal of Scientific Computing*, vol. 37, no. 3, pp. 367–382, 2008.
- [42] S. Esedoglu and J. Shen, “Digital inpainting based on the mumford–shah–euler image model,” *European Journal of Applied Mathematics*, vol. 13, no. 04, pp. 353–370, 2002.
- [43] J. Shen and T. F. Chan, “Mathematical models for local nontexture inpaintings,” *SIAM Journal on Applied Mathematics*, vol. 62, no. 3, pp. 1019–1043, 2002.
- [44] M. Zhou, H. Chen, J. Paisley, L. Ren, L. Li, Z. Xing, D. Dunson, G. Sapiro, and L. Carin, “Nonparametric bayesian dictionary learning for analysis of noisy and incomplete images,” *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 130–144, 2012.
- [45] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” *arXiv preprint arXiv:1502.03167*, 2015.
- [46] A. L. Maas, A. Y. Hannun, and A. Y. Ng, “Rectifier nonlinearities improve neural network acoustic models,” in *Proc. ICML*, vol. 30, no. 1, 2013.

- [47] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," *arXiv preprint arXiv:1505.00853*, 2015.
- [48] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.
- [49] A. Dosovitskiy, J. Tobias Springenberg, and T. Brox, "Learning to generate chairs with convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1538–1546.
- [50] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European conference on computer vision*. Springer, 2014, pp. 818–833.
- [51] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European Conference on Computer Vision*. Springer, 2016, pp. 694–711.
- [52] G. Sam and W. Michael. (2016, February) Training and investigating residual nets. [Online]. Available: <http://torch.ch/blog/2016/02/04/resnets.html>
- [53] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.
- [54] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 5, no. 2, pp. 354–379, 2012.
- [55] N. Keshava and J. F. Mustard, "Spectral unmixing," *Signal Processing Magazine, IEEE*, vol. 19, no. 1, pp. 44–57, 2002.
- [56] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.
- [57] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *Image Processing, IEEE Transactions on*, vol. 19, no. 11, pp. 2861–2873, 2010.
- [58] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 248–255.
- [59] T. F. Chan and J. Shen, "Nontexture inpainting by curvature-driven diffusions," *Journal of Visual Communication and Image Representation*, vol. 12, no. 4, pp. 436–449, 2001.
- [60] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [61] R. H. Yuhas, A. F. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (sam) algorithm," 1992.