NANO LETTERS

Letter

pubs.acs.org/NanoLett

Sparsity-Based Super Resolution for SEM Images

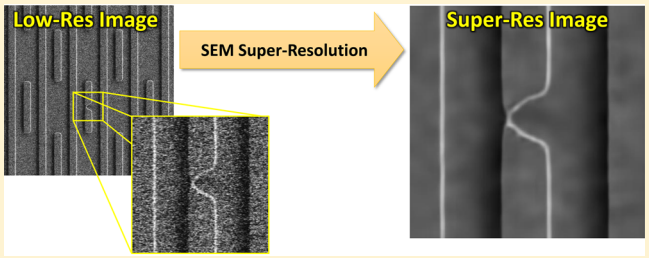
Shahar Tsiper,[†] Or Dicker,[†] Idan Kaizerman,[§] Zeev Zohar,[§] Mordechai Segev,[*,‡] and Yonina C. Eldar[†]

[†]Electrical Engineering Department, [‡]Physics Department and Solid-State Institute, Technion, Haifa 32000, Israel

[§]Applied Materials, 9 Oppenheimer St., Rehovot 76705, Israel

Supporting Information

ABSTRACT: The scanning electron microscope (SEM) is an electron microscope that produces an image of a sample by scanning it with a focused beam of electrons. The electrons interact with the atoms in the sample, which emit secondary electrons that contain information about the surface topography and composition. The sample is scanned by the electron beam point by point, until an image of the surface is formed. Since its invention in 1942, the capabilities of SEMs have become paramount in the discovery and understanding of the nanometer world, and today it is extensively used for both research and in industry. In principle, SEMs can achieve resolution better than one nanometer. However, for many applications, working at subnanometer resolution implies an exceedingly large number of scanning points. For exactly this reason, the SEM diagnostics of microelectronic chips is performed either at high resolution (HR) over a small area or at low resolution (LR) while capturing a larger portion of the chip. Here, we employ sparse coding and dictionary learning to algorithmically enhance low-resolution SEM images of microelectronic chips—up to the level of the HR images acquired by slow SEM scans, while considerably reducing the noise. Our methodology consists of two steps: an offline stage of learning a joint dictionary from a sequence of LR and HR images of the same region in the chip, followed by a fast-online super-resolution step where the resolution of a new LR image is enhanced. We provide several examples with typical chips used in the microelectronics industry, as well as a statistical study on arbitrary images with characteristic structural features. Conceptually, our method works well when the images have similar characteristics, as microelectronics chips do. This work demonstrates that employing sparsity concepts can greatly improve the performance of SEM, thereby considerably increasing the scanning throughput without compromising on analysis quality and resolution.



KEYWORDS: *Super resolution, microscopy, sparsity, dictionary learning, scanning electron microscope (SEM)*

The scanning electron microscope (SEM) is one of the most versatile examination and analysis tools for solid objects at high (subnanometer) resolutions.[1] The SEM works by launching electrons at a specimen by a focused electron beam, and then examining the emission of secondary electrons from the sample. The secondary electrons contain topographic and compositional information about the surface. Their flux is recorded by scanning the sample point by point. A full image is then formed by stitching together the scanning results. Since the first prototypes of SEM microscopes,[2,3] and the first commercial SEM scanner,[4] many improvements have been introduced to modern SEMs. Scanners today have faster acquisition and larger magnification, and they can store images digitally, opening a new path for real-time digital processing. In addition, modern SEMs allow to image the surface of specimens from several perspectives simultaneously (where each perspective translates into a different viewing angle of the sample). This ability of the SEM to generate surface images at subnanometer resolutions has been one of the key factors in the rapid development of nanotechnology. In particular, it had a tremendous impact on the microchip industry, where it aided improving the production process at an exponential rate in accordance with Moore's law. Nowadays, microchip production lines heavily rely on SEM scanners to monitor and analyze the manufacturing process, that is, the wafer is scanned in between various processing steps, detecting defects, and therefore, increasing yield.

Two main parameters control SEM imaging: scanning velocity and resolution. Naturally, fast scanning leads to reduced quality and added noise, while resolution determines the physical size of a pixel. Both parameters impact acquisition time. As microchip dimensions are reduced in accordance with Moore's law, the complexity of the fabrication process increases. This results in a need to scan with ever smaller pixel sizes and over a larger number of locations on the chip, in order to detect and classify defects and monitor the process. Two common modes of SEM operation are scanning a large field of view, employing a fast acquisition time yet low resolution (LR), or slowly scanning a small field of view, achieving a higher resolution (HR) and better image quality. An example of both scanning modes for three different

Received: May 18, 2017
Revised: July 30, 2017
Published: August 14, 2017

ACS Publications © 2017 American Chemical Society

DOI: 10.1021/acs.nanolett.7b02091
Nano Lett. 2017, 17, 5437−5445

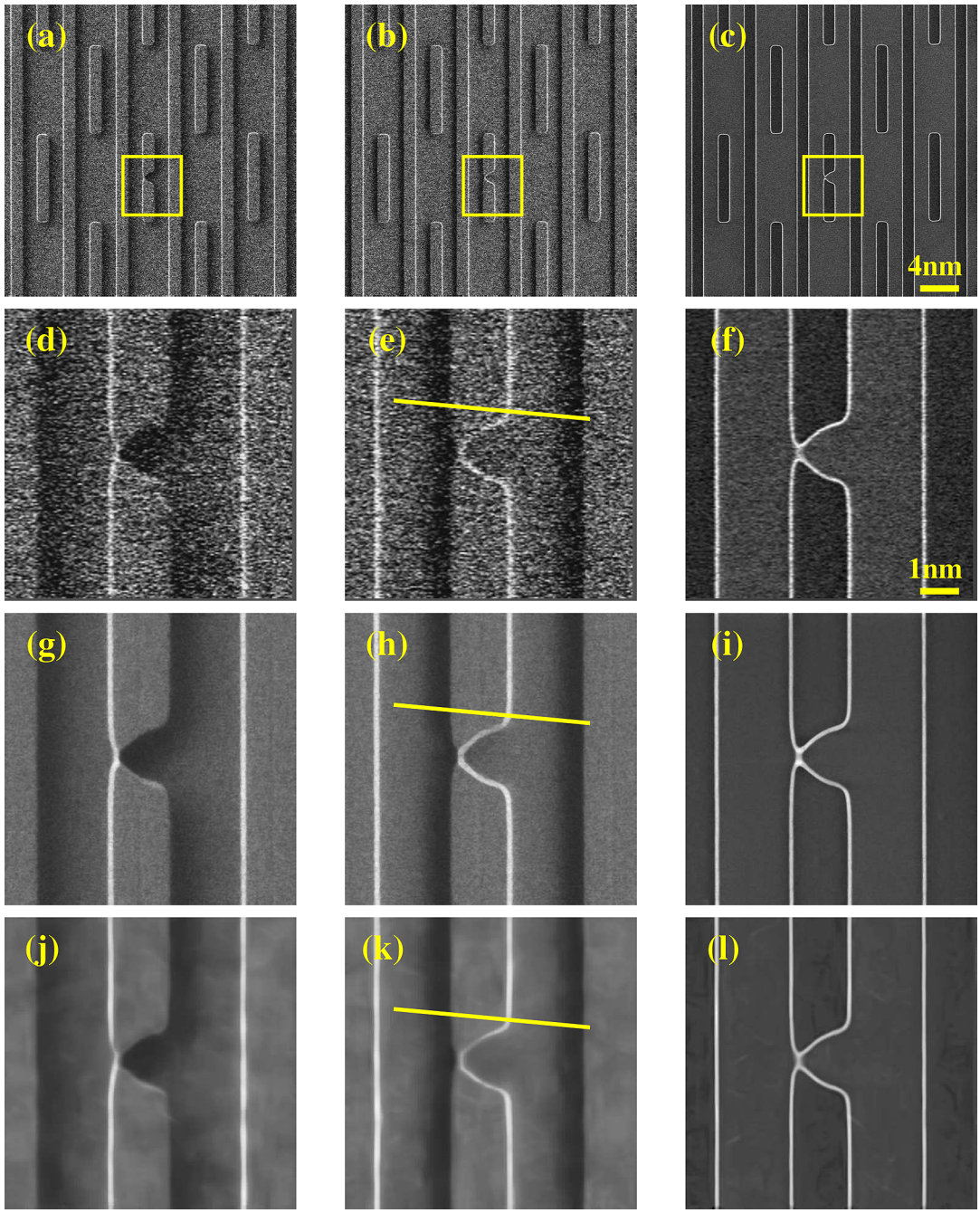


Figure 1. SEM microchip images in low, high, and super resolution. Top row: scanned low resolution (LR) images (20 μm × 20 μm), taken from left (a), right (b), and top (c) perspectives. The yellow square shows the region of interest for enhancement with reduced field-of-view (5 μm × 5 μm). The second row (d–f) shows the corresponding regions of interest from the LR images, enlarged. The third row (g–i) shows the images acquired using the SEM high resolution (HR) mode, scanning the same areas chosen by the yellow squares above. The bottom row (j–l) depicts the enhanced super-resolution (SR) images recovered from their LR counterparts (second row). The yellow lines displayed in (e), (h), and (k) are used for the line cuts analyzed in Figure 3.

perspectives can be seen in Figure 1. HR scans are considered costly since their acquisition time is lengthy and their field of view is small. Scanning the same area in HR mode takes about 2 orders of magnitude longer. On the other hand, high resolution is often a necessity in the semiconductor manufacturing industry, as the ever-increasing demand for faster and more energy efficient microchips is shifting the industry toward smaller structures and more complex manufacturing processes, pushing current SEM technology to its limit. Consequently, there is great interest in scanning large areas quickly, while maintaining high resolution.

These goals, which are clearly conflicting in current SEM technology, inspired our approach of finding a solution that breaks the immediate link between the inherent physical scanning resolution of the SEM and the resolution of the output image. Here, we present an algorithmic method operating on a low-resolution SEM image obtained in a fast scan, to enhance its resolution and image quality up to the level of the HR image obtained in a slow scan using the same system.

DOI: 10.1021/acs.nanolett.7b02091
Nano Lett. 2017, 17, 5437–5445

Our approach consists of two steps: an offline stage of learning a joint dictionary from a sequence of LR and HR images of the same region in the chip, followed by a fast-online super-resolution (SR) step where the resolution of a new LR image is enhanced. The SR process can be simply described as changing the mathematical representation over which the LR image is spanned from a LR basis to a suitable HR one. Both steps are based on concepts of sparsity: the fact that SEM images have a characteristic structure, and hence they can be represented compactly in some mathematical basis that is learned in the first (dictionary learning) step.

The concept of sparsity lies at the heart of our approach and has been used extensively in signal processing, statistics, computer vision, and more. According to sparse approximation theory,[5,6] signals can often be faithfully represented as a linear combination of just a few elements from a given basis, that is, a dictionary. This observation is true for many classes of natural images, and in principle for microchip SEM images, as we demonstrate here. Sparse models underlie many audio and image compression (MP3, JPEG, JPEG2000) methods, as well as many state-of-the-art image processing techniques. In general, assuming a sparse representation allows isolating the important information out of a high-dimensional signal, while removing uninformative noise. In the past decade, the field of compressed sensing (CS)[7−10] has emerged. CS enables the recovery of signals from partial and noisy measurements, by exploiting sparsity. The mathematical framework provided by CS assures that a sparse signal can be exactly recovered from just a small set of linear measurements,[7] given a known dictionary and a known measurement (sensing) system. Relying on these principles, the concepts of sparsity and CS have been used for a variety of applications, ranging from sub-Nyquist sampling[11,12] with applications in radar[13,14] and ultrasound,[15−17] to super-resolution imaging,[18,19] phase retrieval,[20−22] ankylography,[23] mapping the coherence function of light,[24] holography,[25] single pixel camera,[26] ghost imaging,[27] and even quantum state tomography.[28−31] These sparsity-based ideas inspired our current work.

The task of building a dictionary over which a family of signals, sharing the same structural features, is sparsely represented was extensively studied in recent years and is referred to in the literature as dictionary learning[32] (DL). The first step of any DL method is obtaining a large enough training set of measurements from the same family of signals, in our case LR and HR patches of SEM microchip images. Choosing a complete and representative training set allows DL algorithms to construct a compact dictionary which contains only the most basic building blocks that compose the signal family. A correctly designed dictionary can then approximate the structure of signals from within the family well.

The DL step is the most important and complex step in our methodology. One of the reasons for the success of our approach, is that we construct dictionaries that contain paired elements of both high and low resolution (the elements are called "atoms" in the context of DL). This specification allows each dictionary atom to naturally map a LR patch to its HR counterpart. In addition, we leverage different perspectives acquired simultaneously by the SEM, as will be explained below.

In our model, each dictionary atom is composed of a LR patch concatenated with a HR twin patch. Choosing a patch size that is too small would result in dictionary atoms that cannot contain the unique inherent structure of the microchip, while specifying a large patch size increases the computational burden. Hence, the patch size should be chosen according to the data at hand (see more details in the Supporting Information). For our data, we use a patch size of 23 × 23 pixels. The next design consideration is to determine the size of the dictionary, namely, the number of atoms in it. In our specific implementation, the number of dictionary elements is defined as 3−5 times the number of coefficients in each atom (which in our case is the patch size $23^2 = 529$). Thus, there is a built-in redundancy in the dictionary, since it has more elements than the intrinsic degrees of freedom. This is needed to ensure the successful description of each microchip patch as a sparse combination of the dictionary atoms.

Once the desired dimensions of the dictionary are set, the next step is to obtain and organize the training data for DL. In our tests, we find that a relatively small set comprising of just several duos of scanned images of both low and high resolution (5 in our case), acquired from the exact same area in the microchip, suffices. From these pairs of LR and HR images we can extract a multitude of patches of appropriate sizes. The patches are simultaneously extracted and stored as part of the training set. The extraction process is randomly performed from the available LR and HR training images, followed by a test for each of the extracted patch duos that ensures it contains a significant part of an actual microchip pattern. The testing prevents us from including patches that contain mostly background noise into the dictionary training set. Further details on the patch selection and pairing process are provided in the Supporting Information. The training set should be much larger than the number of atoms in the dictionary, roughly a hundred times bigger. In our case the dictionary is composed of 2048 atoms, while the training set consists of 250,000 samples. This ensures a thorough learning stage, which yields a robust dictionary, able to faithfully describe many possible microchip features.

The collected training data is the input to the DL algorithm. In what follows, we use a modified version of the K-SVD[32,33] method, yet we emphasize that any DL technique could be adapted to this task. Our algorithm uses an iterative approach to find an approximate solution to the following nonconvex minimization problem:[32]

$$[D, X] = \arg\min_{D',X'} \| T - D'X' \|_F^2 \quad \text{subject to } \| x_j \|_0 < k_0 \qquad \forall\, j \tag{1}$$

Here, $T \in \mathbb{R}^{n \times N_T}$ is a constant matrix containing a training sample in each of its columns, with n pixels in each sample and an overall N_T number of samples. The matrix $D' \in \mathbb{R}^{n \times N_D}$ represents the dictionary where each of its N_D columns is an atom, and $X' \in \mathbb{R}^{N_D \times N_T}$ is a variable sparse matrix, containing the representation coefficients that span T over the dictionary. The zero-norm function $\|\cdot\|_0$ counts the number of nonzero elements in its argument. The constraint $\|x_j\|_0 < k_0$ enforces that every single training sample from T is described by no more than k_0 atoms from the dictionary (further details on the selection of k_0 are provided in the Supporting Information). Solving eq 1 with the K-SVD algorithm yields two output arguments. The first is the dictionary D, containing the atoms that sparsely represent the training set T. Once the dictionary is obtained, the algorithm computes the sparse representation X, such that $T \approx DX$.

DOI: 10.1021/acs.nanolett.7b02091
Nano Lett. 2017, 17, 5437−5445

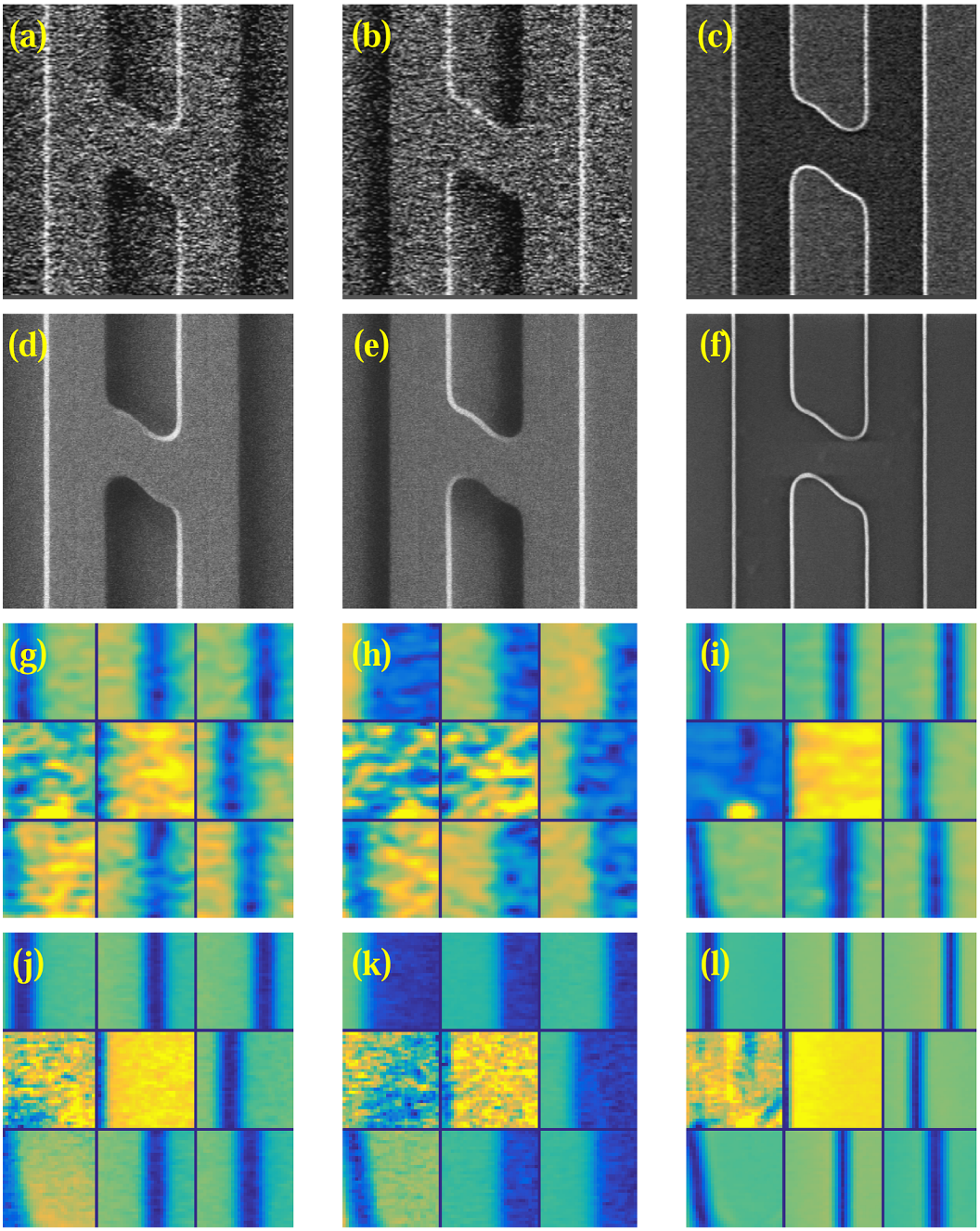


Figure 2. Dictionary atoms mapping. The top and second rows (a–f) show the LR and HR SEM images in the three perspectives, used to assemble the dictionary training set. The two bottom rows display subsets of nine ordered and paired atoms from each trained dictionary. The atoms in the LR trained dictionaries (g–i) are clearly mapped to the HR atoms in (j–l). This mapping connects every LR feature to a HR one and acts as a key enabler for the super resolution methodology.

We adapt eq 1 to our goals, by expanding the training set and dictionaries into concatenated matching pairs of low and high resolution, namely, solving

$$\left[\begin{pmatrix} D_l \\ D_h \end{pmatrix}, X\right] = \underset{D_l', D_h', X'}{\arg\min} \left\| \begin{pmatrix} T_l \\ T_h \end{pmatrix} - \begin{pmatrix} D_l' \\ D_h' \end{pmatrix} X' \right\|_F^2$$
$$\text{subject to } \| x_j' \|_0 < k_0 \qquad \forall\, j \tag{2}$$

From here on, the subindex h denotes an HR term, and the subindex l refers to a LR term. The training set $T = \begin{pmatrix} T_h \\ T_l \end{pmatrix} \in \mathbb{R}^{2n \times N_T}$, used as input for the DL process in eq 2, consists of concatenated and paired LR and HR elements, thus ensuring that the trained dictionaries $D_l \in \mathbb{R}^{n \times N_D}$ and $D_h \in \mathbb{R}^{n \times N_D}$ are fully synchronized and paired as well. An example for the pairing established between the atoms within each dictionary and an image duo used in the creation of the training set are shown in Figure 2 and will be further explained below. The accurate mapping between the dictionaries is an essential requirement that paves the way towards the next step, the actual SR reconstruction.

DOI: 10.1021/acs.nanolett.7b02091
Nano Lett. 2017, 17, 5437–5445

After a dictionary is learned, it is used to perform SR. This stage consists of taking a new LR image acquired by the SEM (but was not part of the training set used to construct the dictionary) and enhancing its resolution by relying on the feature mapping established between low and high-resolution samples in the DL stage.

The SR reconstruction process requires no prior information, apart from the dictionary, constructed in the learning stage. This process begins by taking the LR image and interpolating it to the HR grid as seen in Figure 1, using standard interpolation techniques (see details in the Supporting Information). Although the interpolated LR image resides on the same grid as the HR image, its spectral content has an inherent cutoff frequency, determined by the ratio between low and high resolutions. Next, the interpolated LR image is divided into an ordered set of patches. Each LR patch is stored in a matrix $Y_l \in \mathbb{R}^{n\times N}$, where n is the number of pixels in each patch, and N is the overall number of patches that compose the full LR image. After decomposing the image, we solve the following minimization problem for obtaining the sparse representation X^* over the dictionary D_l:

$$X^* = \arg\min_X \| Y_l - D_l X \|_F^2 \quad \text{subject to } \| x_j \|_0 < k_0 \qquad \forall\, j \tag{3}$$

The information encompassed in X^* is now used to span the microchip at a higher resolution by using it in the overcomplete basis expansion of the HR dictionary via

$$\hat{Y}_h = D_h X^* \tag{4}$$

This step has low computational complexity, since both the dictionary size is small and only a few of the coefficients in each column of X^* are different from zero. After calculating the matrix $\hat{Y}_h$, the patches contained in it are stitched together to form the final SR image.

Example sets of LR, HR, and SR images are displayed in Figure 1. As shown there, the SR images are sharper and much cleaner, as a result of completing the missing spectral information and removal of the unwanted noise. Enhancing the resolution of SEM images is essentially extrapolating the unknown high-frequencies missing from the LR images, while removing much of the noise introduced during the acquisition. That is, what we demonstrate here is sparsity-based bandwidth extrapolation, based on dictionary learning.

In our experiments, we solve eq 3 using a fast greedy algorithm, specifically a fast variation of orthogonal matching pursuit (OMP),[34,35] although any other sparse recovery technique can be used as well. The OMP algorithm detects, for each patch from the LR image, the most correlated dictionary atoms from within the LR dictionary D_l. It then finds the best representation for the patch using just those few selected atoms. Here, usually only 3−5 atoms are needed to faithfully describe the information contained in each patch. The OMP algorithm is fast by nature since it is a greedy algorithm: in each step OMP adds a single element to every column of X, which is the single best atom to sparsely describe a given patch. If the difference between the sparse combination of the already found atoms and the original patch does not meet a predetermined tolerance (dependent on the noise level of the LR image), another atom is added to minimize the difference. Consequently, the number of nonzero elements in each column of X is incrementally growing, one by one in each iteration of the OMP, until either the tolerance is achieved, or the limit of k_0 nonzero terms is reached. For computational efficiency, we solve eq 3 by using the BatchOMP[33] extension, which solves for all the columns of X in parallel. The sparse representation stored in the matrix X^* now contains all of the intrinsic information on the scanned sample and allows spanning it at higher resolution, while removing much of the noise.

The formulation in eqs 3 and 4 considers just a single perspective, acquired by the SEM in the LR mode, to recover the SR image. We can further improve performance by exploiting the fact that all different perspectives are acquired simultaneously from the exact same area of the scanned specimen. For generalizing the step of DL to multiple perspectives, we would like to train dictionaries for both low and high resolution for every available perspective (three perspectives in our case), for a total of six paired dictionaries. To produce these dictionaries, paired training samples are simultaneously extracted for all the perspectives and resolutions, generating the training sets $T^i = \begin{pmatrix} T_l^i \\ T_h^i \end{pmatrix}$, for each ith perspective (i = 1, 2, 3). It is imperative that the training samples are extracted from the same respective areas for all of the perspectives, so that all of the sets are matched, leading to paired dictionaries. The DL process is performed again using the K-SVD method, but now solving a bigger concatenated DL problem that takes into account all of the available information simultaneously:

$$\left[\begin{pmatrix} D^1 \\ D^2 \\ D^3 \end{pmatrix}, X\right] = \arg\min_{D^{1'},D^{2'},D^{3'},X'} \left\| \begin{pmatrix} T^1 \\ T^2 \\ T^3 \end{pmatrix} - \begin{pmatrix} D^{1'} \\ D^{2'} \\ D^{3'} \end{pmatrix} X' \right\|_F^2 \quad \text{subject to } \| x_j' \|_0 < k_0 \qquad \forall\, j \tag{5}$$

Here $D^i = \begin{pmatrix} D_h^i \\ D_l^i \end{pmatrix}$ are the high and low-resolution dictionaries for the ith persepective. Equation 5 guarantees that the dictionaries for all perspectives and resolutions are paired, as seen in Figure 2.

The sparse pursuit and reconstruction steps are also generalized to account for the information shared between the different perspectives. By modifying eq 3 the sparse representation that spans all perspectives is obtained at once:

$$X^* = \arg\min_X \left\| \begin{pmatrix} Y_l^1 \\ Y_l^2 \\ Y_l^3 \end{pmatrix} - \begin{pmatrix} D_l^1 \\ D_l^2 \\ D_l^3 \end{pmatrix} X \right\|_F^2 \quad \text{subject to } \| x_j \|_0 < k_0 \qquad \forall\, j \tag{6}$$

where now $X^* \in \mathbb{R}^{3N_D\times 3N}$. The reconstruction relies on the combined sparse representation X^*, via

$$\begin{pmatrix} \hat{Y}_h^1 \\ \hat{Y}_h^2 \\ \hat{Y}_h^3 \end{pmatrix} = \begin{pmatrix} D_h^1 \\ D_h^2 \\ D_h^3 \end{pmatrix} X^* \tag{7}$$

DOI: 10.1021/acs.nanolett.7b02091
Nano Lett. 2017, 17, 5437−5445

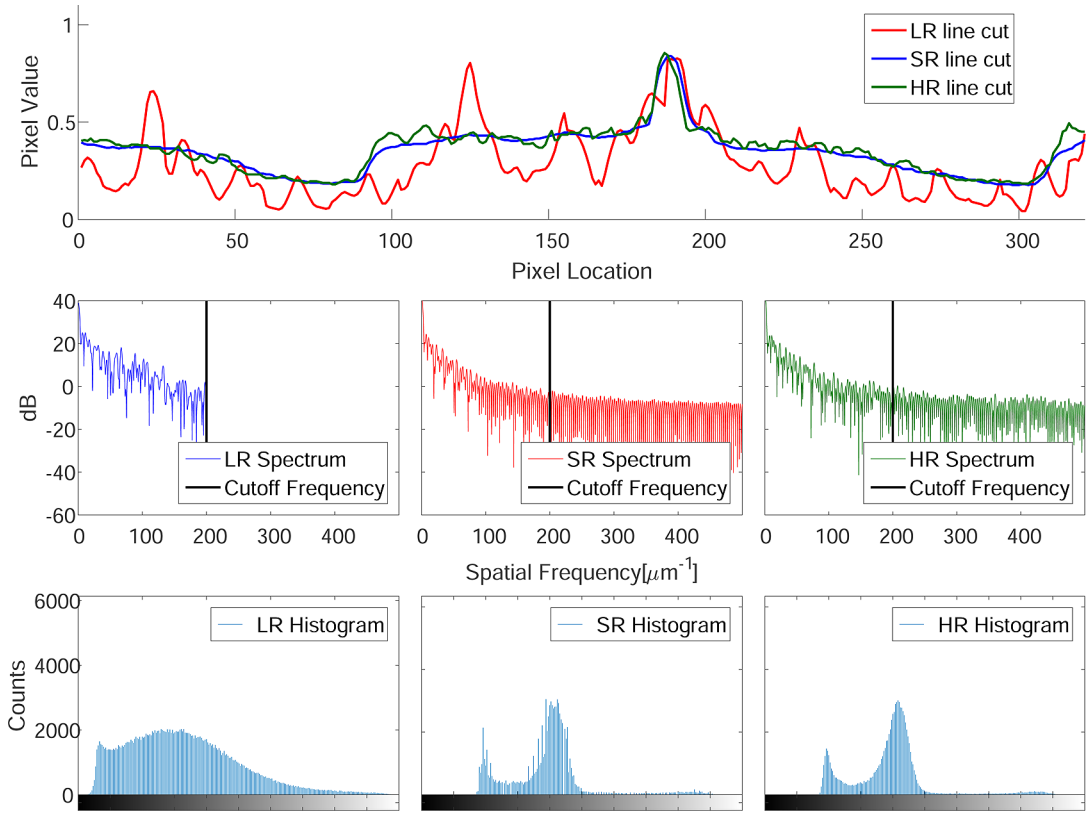


Figure 3. Performance analysis. The top row shows the values of the yellow line cuts from Figure 1, for the three images (LR, SR, and HR). The SR line cut is an almost perfect reconstruction of the HR line, except for considerable reduction of noise. The LR cut contains severe errors (e.g., the peaks at pixels 25 and 125) that are removed by our method, since the algorithm uses the shared information from the different perspectives simultaneously to counter gross errors inflicted by high noise. The middle row displays the 1D discrete Fourier transform of the line cuts, with the maximal spatial frequency of the LR image marked by a vertical bold line, emphasizing that 60% of the spectral content of the HR image is extrapolated. The bottom row shows the histograms of the whole respective images from which the line cuts are taken, highlighting that the statistical nature of the SR image is well approximated. The horizontal axis depicts the gray intensity values, from black to white, of the pixels in the image, while the vertical axis displays the number of pixels in each intensity bin (divided into 512 gray level bins).

The three different perspectives are taken at the exact same time and location, so that the resulting measurements have the same sparse representation X^* over different dictionaries, as implied by eq 6. Joint processing improves the signal-to-noise ratio by increasing the number of equations. In addition, since noise from different perspectives is uncorrelated, errors resulting from outliers in a single perspective are mitigated, as will be further discussed below, and illustrated in Figure 3.

Our experimental data is produced by an Applied Materials SEMVision G6 Semiconductor Defect Review (DR) tool. The resolution of the SEMVision G6 is 1.4 nm, the best resolution of any DR tool available in the industry to date. The images are of various semiconductor microchip samples whose features range between sizes of 28 and 10 nm. Respectively, the field of view of all LR images is between 20 and 2 μm, while that of the HR images is 5 to 0.5 μm. HR scanning takes significantly longer than LR scanning, because the scan is performed over a significantly finer grid, involving more measurements for imaging the same area. In addition, for producing a high-quality HR scan, each grid point is scanned multiple times, and the result is averaged to increase SNR. Furthermore, in the SEM setup we use, the scanning beam width and noise statistics are considered the same for LR and HR scans.

Typical results of recovering HR SEM images from their LR versions are displayed in Figure 1. The top row depicts the entire field of view (20 μm × 20 μm) of low-resolution scans acquired simultaneously from three perspectives of the same microchip. The second, third, and fourth rows depict a magnified region (×2.5) of a narrower field of view (5 μm × 5 μm), marked by the yellow square in the top row. The goal is to enhance the resolution of these LR images to the level of their respective HR images, while removing the noise.

We next test our methodology on real SEM data. The first stage is to construct six dictionaries, one for each perspective and resolution, using the sparsity-based methodology described by eq 5. The dictionaries are constructed from a training set of just five images, acquired from the same SEM at three perspectives. Examples for images used in the DL stage are depicted in the top rows of Figure 2. Importantly, the images used for the training stage are different than those we use later for enhancing resolution, although they are acquired in the same SEM and from the same perspectives as the data whose resolution we enhance. A small subset of ordered atoms from each dictionary is displayed in the two bottom rows of Figure 2, where the pairing (LR and HR) between the atoms of different dictionaries is clearly visible, emphasizing the established mapping between the elements of the three pairs of dictionaries.

After completing the training stage, we apply our method to LR images newly acquired by the SEM and enhance their resolution by first applying eq 6 and then eq 7. Figure 1 shows an example of one LR image at the three perspectives (second

DOI: 10.1021/acs.nanolett.7b02091
Nano Lett. 2017, 17, 5437−5445

row) and their HR counterparts (third row). The respective SR images, obtained by our technique, are shown in the bottom row. When comparing the enhanced SR images (bottom row) to the LR images (second row), it is clear that the overall sharpness of the contours and features is considerably improved, while the noise is greatly reduced. Comparing the SR (bottom row) to the HR images (third row) reveals that the details and edges are accurately reconstructed for all three perspectives.

It is important to examine the performance of our technique in a quantitative manner. To this end, we examine cross sections (line cuts) of the LR, HR, and SR images, taken along the yellow lines shown in the middle column of Figure 1. The line cuts are shown in Figure 3. The top panel displays the values of the pixels along the cross-section line for each image. Clearly, the SR cross section (red) is very similar to the HR one, preserving all its features, while removing small variances that occur due to noise in the HR images. That is, apart from accurately reconstructing the HR images, the SR line demonstrates lower noise variance even when compared to the HR images. This fact is not surprising, since the sparsity constraint we enforce during the SR enhancement fundamentally removes noise in a highly efficient manner. Notably, even large differences, between the LR line and HR one, are mitigated almost entirely from the SR cross section. For example, the large peak near pixel 125 in the LR cross-section is absent in the SR recovery, and indeed that peak is absent in the HR image, exemplifying the effectiveness of our technique. In other words, our reconstruction can remove false information that appears in the LR images, even when this information contains peaks of values twice as large as the true (HR) values. This feature has critical impact on the SR reconstruction: without it—the SR image would often contain errors. This is a result of the fact that the recovery process considers the mutual information from all perspectives at once. Therefore, a systematic error present in a single perspective of the LR image is removed by taking into account valid information from the other perspectives.

The spectral contents of the cross-section lines are presented in the middle row of Figure 3, showing that the high spatial frequencies present in the HR image are extrapolated in the SR image, well beyond the cutoff frequency present in the LR spectrum (marked by a solid vertical line). This shows that the algorithm extrapolates 60 percent of the spectral content of the HR image, effectively expanding the spatial spectrum of the LR image by a factor of two and a half. The bottom row of Figure 3 shows the histograms of LR and HR images taken from the same area and compares them to the histogram of the SR image, reconstructed from the LR image. Clearly, the histograms of the LR and HR images have very different statistical nature; nonetheless, the histogram of the SR image is very similar to the HR image.

To further assess the performance of our algorithm, we compare the SR enhanced images to the HR ones, noting that the SR images used for the fidelity test are reconstructed from new LR images. We use a data set of 30 images for the fidelity test, containing various patterns and structures. Unfortunately, the specific test images cannot be displayed here, due to confidentiality commitments to their respective owners, as these are electronic chips used by commercial companies.

What we show here (Figure 4) is the outcome of the fidelity tests, which is meant to assess the success of our methodology. To that end, we note that both the LR images and their HR counterparts were not included in the training set used to construct the dictionary. Rather, the paired images are new, obtained in the same machine under conditions similar to those used to obtain the training set. For evaluating our methodology, we consider the HR images as the "ground truth", although they do contain some inherent measurement noise, which naturally impedes our comparison. The evaluation is performed using the PSNR metric, calculated as

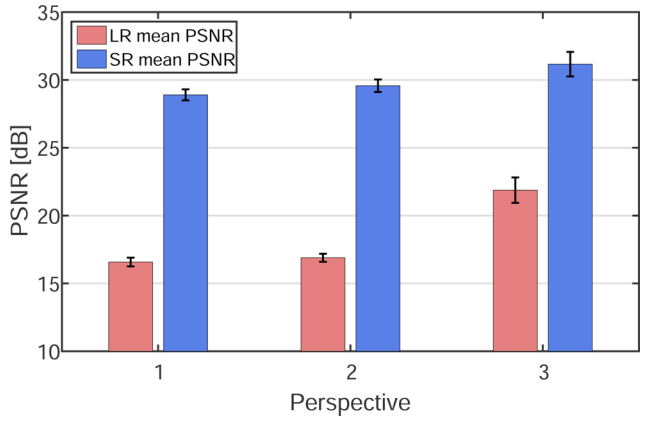


Figure 4. Algorithm fidelity testing. The SR algorithm is tested on 30 images that were not part of the training set used for constructing the dictionary. All the LR and HR data is taken from real SEM experiments (not simulated data). The performance is calculated using the peak signal-to-noise ratio (PSNR), using eq 8, and the standard deviation of the PSNR is marked by the error bar. Both the LR and SR images are compared with the HR images acting as reference. A notable improvement of 10−15 dB is achieved when comparing the SR images to their LR counterparts. Furthermore, the results show that performance is consistent across all perspectives throughout the image set, indicating the robustness of our methodology.

$$\mathrm{PSNR}(\widehat{Y}_h, Y_h) = -10 \log_{10}\left(\frac{\| \hat{Y}_h - Y_h \|_F^2}{nN}\right) \tag{8}$$

The performance scores of the SR enhanced set of images and their reference HR versions are displayed in Figure 4, as blue bars. For comparison, the PSNR values are calculated between the LR and HR images, and their average is shown for each perspective by a red bar. An improvement of 10−15 dB is consistently gained across all images and perspectives tested. The values achieved of approximately 30 dB, indicate very good reconstruction quality, especially since the reference HR images suffer from considerable measurement noise as well.

Before closing, we comment briefly on other SR approaches. Classical resolution enhancement methods[36−38] relied on the prior acquisition of several LR images, each with subpixel shifts from one another, to generate one SR image. Modern methods, as well as our methodology, construct an SR image from just a single LR sample. While our solution is inspired by techniques[39−42] that combine learned dictionaries with sparse representations, nowadays, there are many alternative methods for single-image SR. Among these are approaches that incorporate deep neural networks,[43] internal patch recurrence,[44,45] super-resolution forests,[46,47] and more, that were shown to produce successful results for natural images. Here, our work demonstrates that tailoring a modern signal processing technique to the SEM setup can improve performance to a great extent. It is plausible that combining concepts

DOI: 10.1021/acs.nanolett.7b02091
Nano Lett. 2017, 17, 5437−5445

from other modern SR methods, as those mentioned above, may further improve the results for SEM imaging.

In summary, this work is part of an ongoing attempt to incorporate cutting-edge technologies from the domain of signal processing into the realm of applied physics and optics in particular. We proposed a simple and fast method that produces an enhanced HR image from a single noisy LR SEM scan. Our method achieves unprecedented results in terms of reconstruction quality of SEM images while operating under high noise levels and achieving up to ×4 resolution enhancement in our tests (see Supporting Information for an example). The fast-online reconstruction step accurately recovers spectral information well beyond the cutoff frequency of the image. Our approach relies on two main foundations. First, the exploitation of the unique structural properties of the scanned samples, specifically finding the sparse representation of the samples over a redundant mathematical basis, learned offline from available training data. ***We stress that the trained dictionaries assume no prior knowledge on the samples***. Second, information from several acquired perspectives is exploited to achieve good reconstruction results, even when gross errors in the measured data are present. We show that previously lost high frequency spectral components of the LR images are extrapolated and reconstructed and that consistent results are achieved throughout our testing set, demonstrating robustness of our method.

At the core of our methodology we rely on a single assumption and one necessary condition. The assumption is that the images are sparse over some mathematical representation. This assumption holds in a variety of image modalities, including natural images observed at any scale, as well as manmade fabricated samples. The necessary condition is having an appropriate and representative training set, to train the dictionaries. The solution proposed can easily be adopted to other modalities with only minor adjustments. We believe that our method can impact electron microscopy imaging, aiding both researchers and industry alike to faithfully produce high-resolution images from low-resolution scans.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.nanolett.7b02091.

> A more detailed overview of the SR algorithm and mathematical details on both the registration and interpolation processes is described. The parameters we use in our method are discussed, followed by further information on the sparse pursuit procedure and the overall computational complexity. Finally, an additional SR example with a magnification ratio of ×4 is presented, with explanatory figures that provide intuition on the dictionary learning process (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: msegev@technion.ac.il.

ORCID

Mordechai Segev: 0000-0002-9421-2148

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research was carried out in the context of the MAGNET program "METRO450". M.S. gratefully acknowledges the support of a PoC (Proof of Concept) grant from the European Research Council. In addition, this work was supported by the Ministry of Science, by the ISF I-CORE joint research center of the Technion and the Weizmann Institute, Israel, and by the European Union's Horizon 2020 research and innovation program under grant agreement no. 646804-ERC-COG-BNYQ.

■ REFERENCES

(1) Goldstein, J.; Newbury, D.; Joy, D.; Lyman, C.; Echlin, P.; Lifshin, E.; Sawyer, L.; Michael, J. R. *Scanning electron microscopy and X-ray microanalysis: a text for biologists, materials scientists, and geologists*; Springer Science & Business Media, 2012.

(2) von Ardenne, M. *Z. Phys.* **1938**, *109* (9−10), 553−572.

(3) Zworykin, V. K. *Sci. Am.* **1942**, *167* (3), 111−113.

(4) Pease, R. F. W.; Nixon, W. C. *J. Sci. Instrum.* **1965**, *42*, 2−7.

(5) Eldar, Y. C. *Sampling Theory: Beyond Bandlimited Systems*; Cambridge University Press, 2015.

(6) Elad, M. *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*; Springer, 2010.

(7) Candes, E.; Romberg, J.; Tao, T. *Commun. pure Appl. Math.* **2006**, *59* (8), 1207−1223.

(8) Candès, E. J.; Romberg, J.; Tao, T. *IEEE Trans. Inf. Theory* **2006**, *52* (2), 489−509.

(9) Donoho, D. L. *IEEE Trans. Inf. Theory* **2006**, *52* (4), 1289−1306.

(10) Eldar, Y. C.; Kutyniok, G. *Compressed Sensing, Theory and Applications*; Cambridge University Press, 2012.

(11) Mishali, M.; Eldar, Y. C. *IEEE Signal Process. Mag.* **2011**, *28* (6), 98−124.

(12) Mishali, M.; Eldar, Y. C. *IEEE J. Sel. Top. Signal Process.* **2010**, *4* (2), 375−391.

(13) Bar-ilan, O.; Eldar, Y. C. *IEEE Trans. Signal Process.* **2014**, *62* (7), 1796−1811.

(14) Baransky, E.; Itzhak, G.; Wagner, N.; Shmuel, I.; Shoshan, E.; Eldar, Y. C. *IEEE Trans. Aerosp. Electron. Syst.* **2014**, *50* (2), 809.

(15) Tur, R.; Eldar, Y. C.; Friedman, Z. *IEEE Trans. Signal Process.* **2011**, *59* (4), 1827−1842.

(16) Wagner, N.; Eldar, Y. C.; Friedman, Z. *IEEE Trans. Signal Process.* **2012**, *60* (9), 4643−4657.

(17) Chernyakova, T.; Eldar, Y. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **2014**, *61* (8), 1252−1267.

(18) Gazit, S.; Szameit, A.; Eldar, Y. C.; Segev, M. *Opt. Express* **2009**, *17* (26), 23920−23946.

(19) Szameit, A.; Shechtman, Y.; Osherovich, E.; Bullkich, E.; Sidorenko, P.; Dana, H.; Steiner, S.; Kley, E. B.; Gazit, S.; Cohen-Hyams, T.; Shoham, S.; Zibulevsky, M.; Yavneh, I.; Eldar, Y. C.; Cohen, O.; Segev, M. *Nat. Mater.* **2012**, *11* (5), 455−459.

(20) Shechtman, Y.; Eldar, Y. C.; Szameit, A.; Segev, M. *Opt. Express* **2011**, *19* (16), 14807−14822.

(21) Sidorenko, P.; Kfir, O.; Shechtman, Y.; Fleischer, A.; Eldar, Y. C. *Nat. Commun.* **2015**, *6*, 8209.

(22) Shechtman, Y.; Eldar, Y. C.; Cohen, O.; Chapman, H. N.; Miao, J.; Segev, M. *IEEE Signal Process. Mag.* **2015**, *32* (3), 87−109.

(23) Mutzafi, M.; Shechtman, Y.; Eldar, Y. C.; Cohen, O.; Segev, M. *Nat. Commun.* **2015**, *6*, 7950.

(24) Tian, L.; Lee, J.; Oh, S. B.; Barbastathis, G. *Opt. Express* **2012**, *20* (8), 8296.

(25) Rivenson, Y.; Stern, A.; Javidi, B. *Appl. Opt.* **2013**, *52* (1), A423.

(26) Duarte, M. F.; Davenport, M. A.; Takhar, D.; Laska, J. N.; Sun, T. S. T.; Kelly, K. F.; Baraniuk, R. G. *IEEE Signal Process. Mag.* **2008**, *25*, 83−91.

(27) Katz, O.; Bromberg, Y.; Silberberg, Y. *Appl. Phys. Lett.* **2009**, *95* (13), 131110.

DOI: 10.1021/acs.nanolett.7b02091
Nano Lett. 2017, 17, 5437−5445

(28) Mirhosseini, M.; Magaña-Loaiza, O. S.; Hashemi, S. M.; Boyd, R. W. *Phys. Rev. Lett.* **2014**, *113* (9), 090402.
(29) Howland, G. A.; Schneeloch, J.; Lum, D. J.; Howell, J. C. *Phys. Rev. Lett.* **2014**, *112* (25), 253602.
(30) Oren, D.; Shechtman, Y.; Mutzafi, M.; Eldar, Y. C. *Optica* **2016**, *3* (3), 226.
(31) Oren, D.; Mutzafi, M.; Eldar, Y. C.; Segev, M. *Optica* **2017**, *4*, 993−999.
(32) Aharon, M.; Elad, M.; Bruckstein, A. *IEEE Trans. signal Process.* **2006**, *54* (11), 4311−4322.
(33) Rubinstein, R.; Zibulevsky, M.; Elad, M. *CS Technol.* **2008**, *40* (8), 1−14.
(34) Pati, Y. C. C.; Rezaiifar, R.; Krishnaprasad, P. S. S. In *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*,1993; pp 1−5.
(35) Davis, G.; Mallat, S. G.; Avellaneda, M. *Constr. Approx.* **1997**, *13* (1), 57−98.
(36) Irani, M.; Peleg, S. In *IEEE Proceedings, 10th International Conference on Pattern Recognition*, 1990; Vol. *2*, pp 115−120.
(37) Borman, S.; Stevenson, R. L. *Midwest Symp. Circuits Syst.* **1998**, 374−378.
(38) Park, S. C.; Park, M. K.; Kang, M. G. *IEEE Signal Process. Mag.* **2003**, *20* (3), 21−36.
(39) Yang, J.; Wright, J.; Huang, T. S.; Ma, Y. *IEEE Trans. Image Process.* **2010**, *19* (11), 2861−2873.
(40) Zeyde, R.; Protter, M.; Elad, M. *Technol. - Comput. Sci. Dep. - Technol. Rep.* **2010**, *1* (1), 1−10.
(41) Timofte, R.; Smet, V.; De Gool, L.; Van Leuven, K. U. In *Proceedings of the IEEE International Conference on Computer Vision*, 2013; pp 1920−1927.
(42) Timofte, R.; de Smet, V.; van Gool, L. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **2015**, *9006*, 111−126.
(43) Dong, C.; Loy, C. C.; He, K.; Tang, X. *arXiv*:1501.00092v2 **2015**, 1−14.
(44) Glasner, D.; Bagon, S.; Irani, M. *Proc. IEEE Int. Conf. Comput. Vis.* **2009**, No. Iccv, 349−356.
(45) Michaeli, T.; Irani, M. In *European Conference on Computer Vision*; Springer, 2014; pp 783−798.
(46) Schulter, S.; Leistner, C.; Bischof, H. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 2015, 07-12-June; pp 3791−3799.
(47) Salvador, J.; Perez-Pellitero, E. *Proc. IEEE Int. Conf. Comput. Vis.* 2016, 11-18-Dec (1); pp 325−333.

DOI: 10.1021/acs.nanolett.7b02091
Nano Lett. 2017, 17, 5437−5445