Current Optics and Photonics
Vol. 2, No. 2, April 2018, pp. 153-164

ISSN: 2508-7266(Print) / ISSN: 2508-7274(Online)
DOI: https://doi.org/10.3807/COPP.2018.2.2.153

A Method for Improving Resolution and Critical Dimension Measurement of an Organic Layer Using Deep Learning Superresolution

Sangyun Kim[1,2] and Heui Jae Pahk[1]*

[1]*School of Mechanical and Aerospace Engineering, Seoul National University, Seoul 08826, Korea*
[2]*SNUPrecision Co., Ltd., Asan 31409, Korea*

(Received February 7, 2018 : revised March 5, 2018 : accepted March 7, 2018)

In semiconductor manufacturing, critical dimensions indicate the features of patterns formed by the semiconductor process. The purpose of measuring critical dimensions is to confirm whether patterns are made as intended. The deposition process for an organic light emitting diode (OLED) forms a luminous organic layer on the thin-film transistor electrode. The position of this organic layer greatly affects the luminescent performance of an OLED. Thus, a system for measuring the position of the organic layer from outside of the vacuum chamber in real-time is desired for monitoring the deposition process. Typically, imaging from large stand-off distances results in low spatial resolution because of diffraction blur, and it is difficult to attain an adequate industrial-level measurement. The proposed method offers a new superresolution single-image using a conversion formula between two different optical systems obtained by a deep learning technique. This formula converts an image measured at long distance and with low-resolution optics into one image as if it were measured with high-resolution optics. The performance of this method is evaluated with various samples in terms of spatial resolution and measurement performance.

Keywords : Critical dimension measurement, Edge detection, Deep learning superresolution
OCIS codes : (100.2980) Image enhancement; (100.6640) Superresolution; (120.3930) Metrological Instrumentation; (150.3045) Industrial optical metrology

I. INTRODUCTION

Next generation displays manufactured with an organic light emitting diode (OLED) have been on the horizon for some time. An OLED utilizes the electroluminescence of organic films. The electroluminescence generates light thru recombination of electrons and holes injected by the electric field. Because an OLED has a self-luminescent property, it can be made without a back light unit and can have a simple structure, a low power consumption, and a high color gamut.

The structure of an OLED, as shown in Fig. 1, consists of glass substrate, thin-film transistor (TFT), organic layer, cathode and anode. When voltage is applied to the OLED, electrons move to the cathode, and holes moves to the anode. At this time, the TFT serves as a switch that controls the flow of electrons that determines the emission from the organic film. Electrons and holes combine at the organic film to form excitons. The energy state of excitons is lowered to the ground state and emits light. Figure 1 shows a case of when a red organic film emits light. The color of the light changes according to the property of the organic film, and the brightness of the light is determined by the amount of current.

The deposition process forms a luminous organic layer on the TFT in a vacuum chamber. This process is standardized by a fine metal mask method for mass production [1]. As shown in Fig. 2(a), the organic source at the solid state is vaporized by the voltage drive. (In general, the organic matter is colorless in the gas state. However, in Fig. 2(a), it is shown in a red color for clarity). A fine metal mask, a thin metal film with numerous holes, is used as a cast

*Corresponding author: hjpahk@snu.ac.kr, ORCID 0000-0002-3838-6928
Color versions of one or more of the figures in this paper are available online.

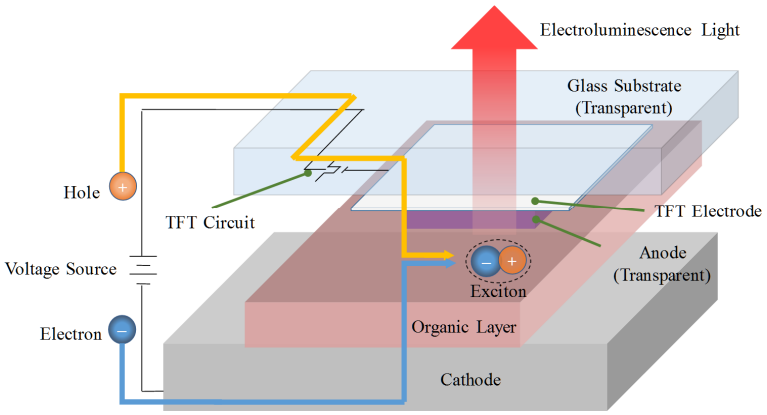

FIG. 1. The structure of OLED is made of glass substrate, TFT circuit, TFT electrode, anode and cathode. The OLED makes the light using electroluminescence by electric field.

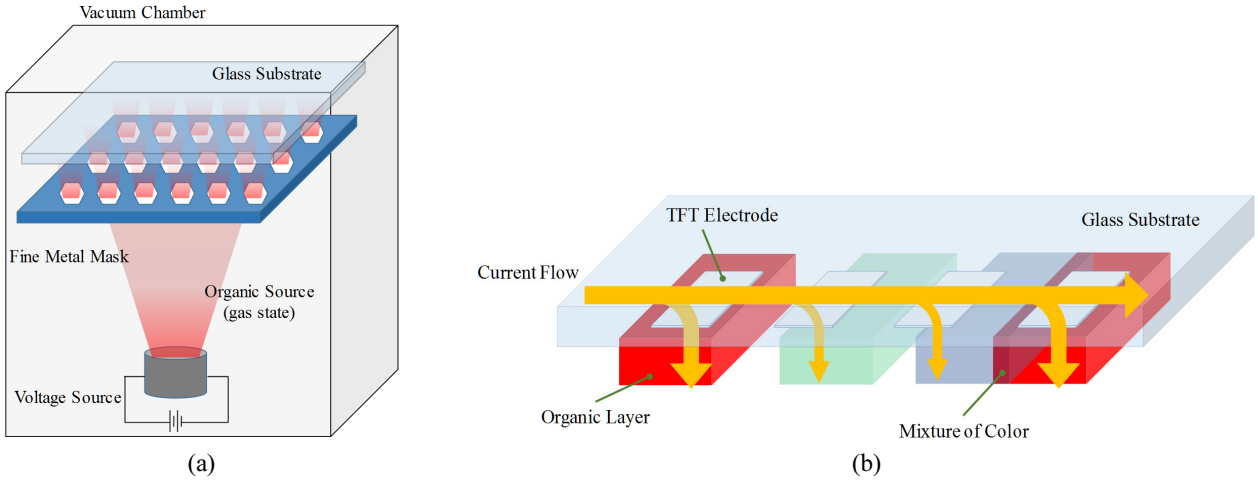

FIG. 2. Organic layers are made on TFT electrodes using a fine metal mask during the deposition process (a). Organic layers of red are formed in position, these emit the light normally. In the case of green and blue, because these are situated at the wrong position, both sharpness of color and luminance efficiency go down.

to shape the organic layer. This is placed in front of the glass substrate to ensure that the vaporized organic material is accurately deposited on the TFT electrode. If a misalignment of the fine metal mask occurs due to vibration, motion error, or deformation in the mask, the organic layers will be mispositioned. Moreover, the amount of contact between the TFT electrode and the organic layer is reduced. Additionally, the amount of current flowing to the organic layer is decreased and the brightness of the color drops. As shown in Fig. 2(b), organic layers for red are deposited at a particular position and emit light normally. However, for green and blue, the amount of current decreases and the emitting performance of these materials deteriorates. Of particular concern, the mixture of color causes the sharpness of the colors to drop. Therefore, the positioning of the organic layer has a critical effect on luminance efficiency. It is an important factor that determines the performance of the OLED.

Measurement can also evaluate the quality level of the deposition process. Figure 3(a) shows the definition of pixel position accuracy (PPA), which is the distance between the center of an organic layer and that of the TFT. By measuring PPA, it is possible to evaluate deposition accuracy and to control the deposition process. A smaller PPA value indicates higher accuracy and higher stability in the deposition process.

The most difficult challenge for accurate PPA measurement is the degradation of image resolution due to a long working distance (WD). Working distance means the space between an object and the optical system. In order to eliminate impurities as much as possible in the vacuum chamber, the optical system has to be located outside of the vacuum chamber and images are taken through a viewport. This means that the WD has to be at between 180~190 mm, as show in Fig. 3(b).

In Eqs. (1) and (2), having a large WD means that the spatial resolution loss is caused by the diffraction light. At the same time, measurement accuracy is poor.

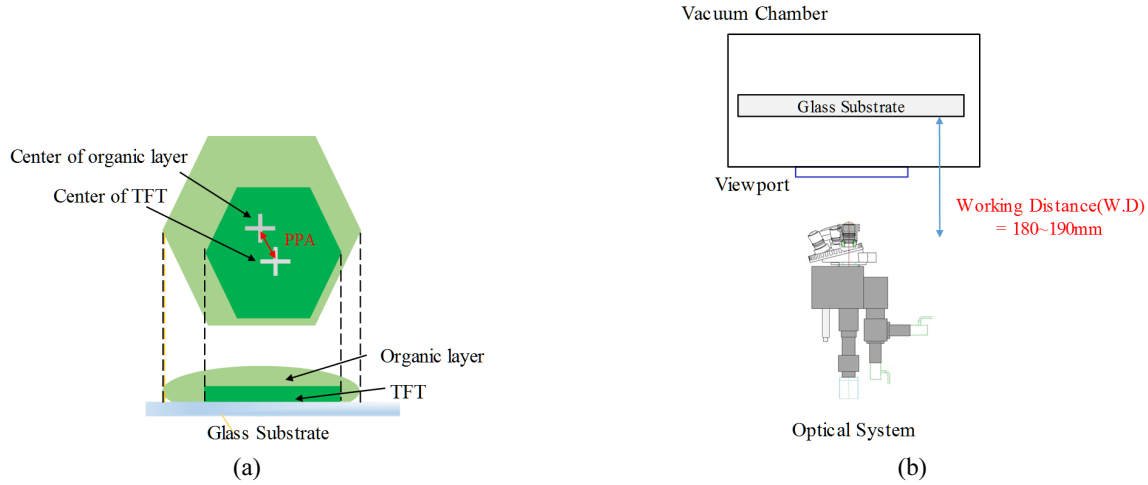

FIG. 3. Definition of pixel position accuracy (PPA) is deviation between center of organic layer and the one of TFT (a). The optics for PPA measurement have to be located outside of the chamber and have the long working distance (WD) (b).

$$NA(object) = n\sin(\theta) \tag{1}$$

where θ is the objective angular aperture, n is the refractive index of the medium between the objective and the specimen.

$$r(Airy) = \frac{1.22\lambda}{2NA(object)} \tag{2}$$

$r(Airy)$ is the airy radius, and λ is the wavelength of the illuminating light.

A number of techniques to improve resolution have been proposed. Most methods in the industrial field are primarily reconstruction-based superresolution approaches. In these methods, a high-resolution (HR) image is made from multiple low-resolution (LR) images. LR images serve as the subpixel data of the reconstructed HR image. To obtain LR images, additional structures are needed, such as dozens of camera arrays [2-4], various illumination angles [5-7], shifting of the actuator for camera aperture [8] or the camera sensor [9]. The approach also requires acquisition time for the multiple images, which are sensitive to external noise and limit real-time application.

Recently, deep learning methods have gained considerable attention for single image superresolution [10-13], restoration [14, 15], deconvolution [16], and denoising [17, 18]. Deep learning for rich details in an image is based on a set of prior examples with LR images and corresponding HR images [19]. This method, called example-based superresolution, uses model mapping from LR to HR images that are trained in advance in order to obtain a high-quality image. In 2014, C. Dong *et al.* introduced a deep learning technique into superresolution for the first time, using convolution neural networks composed of three layers [10]. Fast superresolution convolutional neural networks (FSRCNN) [11] use the deconvolution layer instead of bicubic interpolation, aiming for faster and better performance. Very deep superresolution (VDSR) [12] includes 20 layers and uses residual learning, finding that deep networks improved image quality. There have been attempts to work toward both efficiency and robustness by compressing image data [13, 14] and producing a simpler model architecture [15]. Because the deep learning method is based on single image, it does not need multiple images, eliminating the concerns regarding capture time and additional structure.

This paper proposes a new superresolution method that utilizes the relationship between two different optics expressed by the neural network. This method trains the neural networks for a conversion formula, which is used to convert the LR image measured at the long distance and low-resolution optics into a HR image as if the measurements were done with high-resolution optics. For the first time, this novel single-image superresolution method uses a deep learning technique for resolution enhancement and industrial-level measurement.

II. EXPERIMENTAL SETUP

Figure 4(a) shows the measuring optics configuration. This system consisted of a 10 × objective lens for magnifying the sample, an ultraviolet (UV) laser diode at a 405 nm wavelength to obtain images of the organic layers, a white light emitting diode (LED) to capture TFT image, a CCD camera with a sensor size of 5.5 by 5.5 square microns, and a UV cut filter to block UV light reflected by the sample. Optics specifications were: 195 mm WD, 0.12 numerical aperture (NA), and 0.55-μm per pixel of spatial resolution.

Compared to a conventional optical microscope, the optics had two distinct characteristics. First, it had a dramatically long WD, and thus, can be located outside of a vacuum chamber. However, the resolution was too low to achieve appropriate accuracy. As a comparison, a 10 × CFI Plan Nikon objective lens was frequently used for industrial measurements, with specification of 16.0 mm WD and 0.3

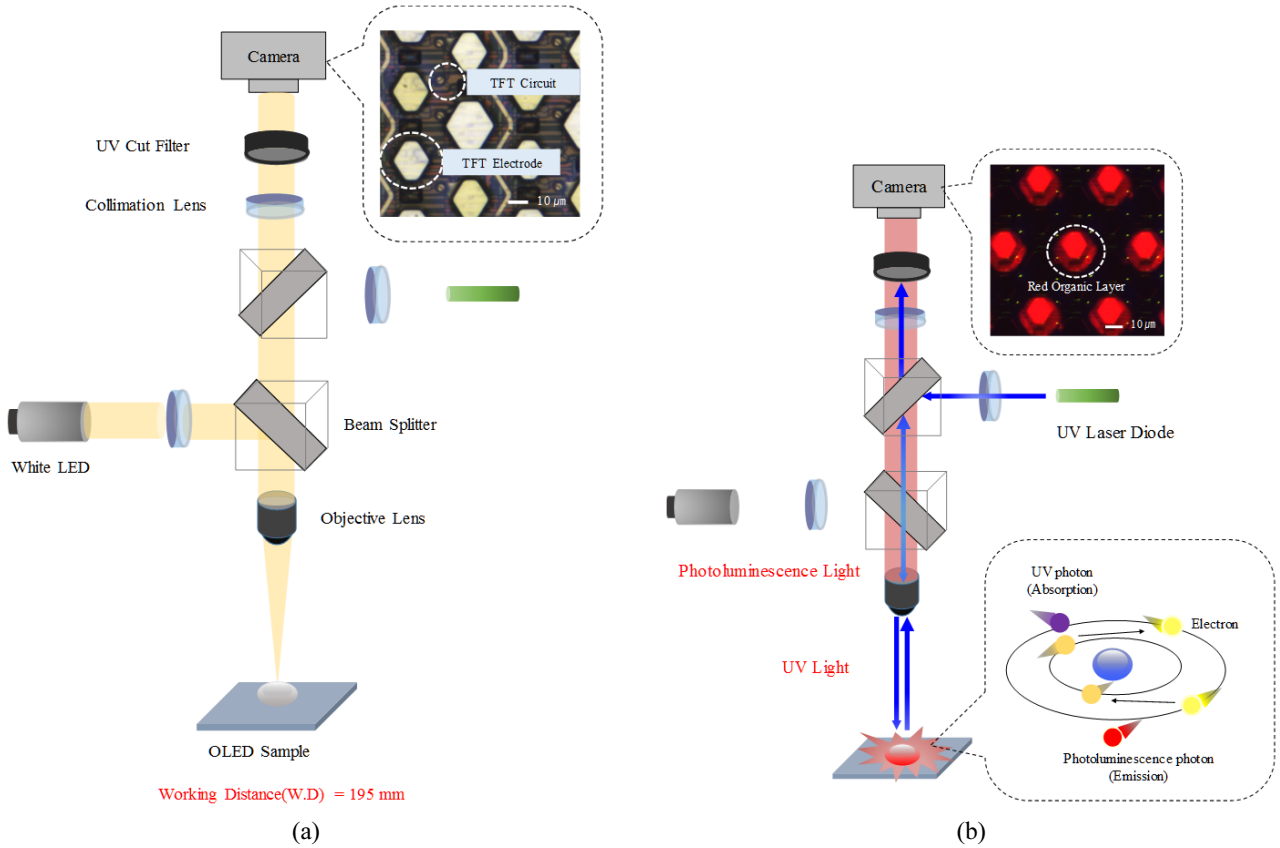

FIG. 4. The measuring optics had very long working distance (WD) in order to capture the image outside of the vacuum chamber. The image of TFT can be obtained by white light illumination (a). Organic layers that were transparent can be measured by UV light using photo-luminescence. The UV light was used for an energy source, not for curing. Because UV light was bluish, the reflected light from the sample should be prevented from entering the camera using the UV cut filter.

of NA.

Another unique feature of this optics setup was the inclusion of two different illumination sources. The first source was a white LED to illuminate the image for TFT. Figure 4(a) shows a real image of hexagonal TFT electrodes and TFT circuits illuminated by the white light. (Note: as the anode was made of indium tin oxide, which was a transparent electrode, it can't be seen in the image). The image shows that electric circuits and electrodes existed on the glass substrate to flow current to each of the TFT electrodes. Because the organic layers were transparent, it was impossible to see organic layers using white light.

Figure 4(b) shows the second light source which was a UV laser diode. This source was not used for curing, but as an energy source in order to generate photo-luminescence light for the organic layer. Electrons orbiting the nucleus at a ground-state energy level absorbed energy from the UV photons and revolved at an excited state. However, these were very unstable, and the emitted photo-luminescence photons occurred in a very short amount of time. The position of the transparent organic film can be confirmed by acquiring the emitted photo-luminescence light with the camera. The shape of the organic layer made by the deposition process had blurrier edges than the TFT made by photolithography (Images of Figs. 4(a) and 4(b) were obtained by a color camera for research purposes).

Because the laser diode for the UV had a long coherence length, it can provide energy to the organic layer over a long distance. UV light with a wavelength of 405 nm was bluish and can be mixed with the emission of the organic film. Therefore, as shown in Fig. 4(b), the UV light reflected from the sample should be blocked from entering the camera using the UV cut filter.

The training optical system in Fig. 5 was manufactured to capture HR images used as ground truth images when training the network. This optics was composed of a 20 × objective lens of the CFI Plan from Nikon, a 405 nm UV LED, and a piezoelectric transducer (PZT). The WD of this system was 8.7 mm and the NA was 0.46.

The measurement optics used a laser diode with a long coherent length to supply the energy to the organic film at a distance. Because the working distance of the training optics was shorter than for the measurement optics, this system used LEDs with short coherent lengths but with more stable power compared to the laser diode. The PZT moved the objective lens along the z direction to obtain

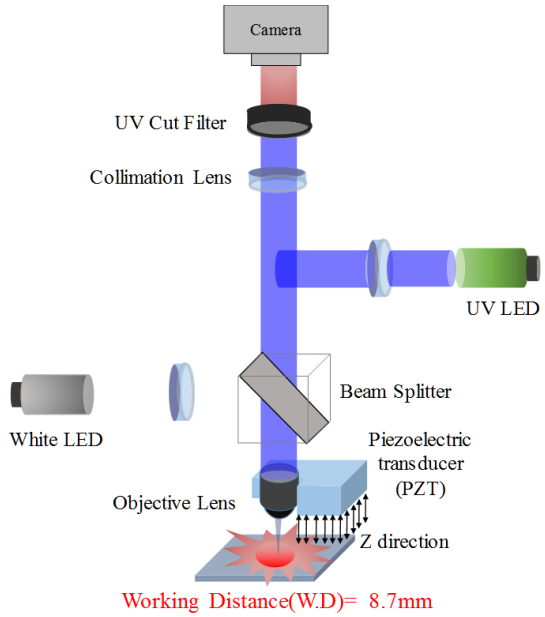

FIG. 5. The configuration of training optics with short working distance (WD) and high numerical aperture (NA). This optics used UV LED for stable power and piezoelectric transducer for capturing a clearer image.

the best focus and a clear image. It had an accuracy of 0.1 nm and scanned at 15 μm to acquire an image at 0.075 μm intervals and found the optimal focus position. Other configuration and measurement principles were the same as for the measurement optics.

III. METHOD

Figure 6(a) shows a flow chart of the overall measurement procedure. The procedure was comprised of two steps: a training process and a measurement process. The training process provided the conversion formula between the measurement optics and the training optics using a deep learning technique. The objects used for the training data sets were measured in each optical system. The image obtained from the measurement optics was set as the LR image and the one taken by the training optics was set as the HR image. Because the sample was measured by changing the optical system, the position of the sample in the LR and HR images must inevitably vary. With a scale-invariant-feature transform [20], both images were corrected for the translational and rotational deviations from each other. This process was repeated for various samples to obtain the LR-HR training sets. It was possible to obtain the conversion formula by training the deep learning network with these data sets.

The measurement process calculated the PPA value with a superresolution image and edge detection. First, the image of a test sample was captured thru the measurement optics and was converted to a superresolution image using the conversion formula obtained by the training process. The next step was to calculate the center of the TFT and organic layers with edge detection. Edge detection senses the intensity change of pixels and separates the object from the background by finding a boundary line. Finally, the PPA values can be calculated by these center coordinates.

3.1. Training Process

We denote our data set S, which was used to train the network

$$S=\{(X_n, Y_n), n=1, \ldots, N\} \tag{3}$$

where

$$X_n=\{x_j^{(n)}, j=1, \ldots, |X_n|\} \tag{4}$$

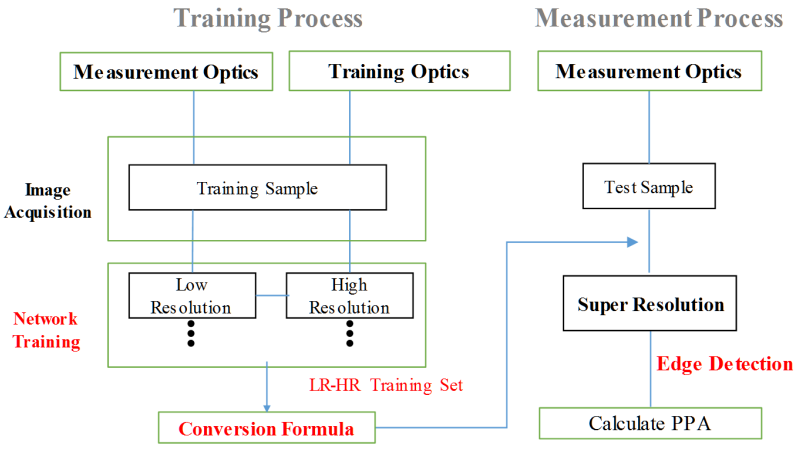

FIG. 6. Flow chart of PPA measurement with training and measurement process. The training process was making conversion formula between two different optics using LR-HR set of training samples. The measure process calculated the PPA value. The image of test sample was converted to superresolution image using conversion formula. Next step was to calculate the center of TFT and organic layers with edge detection and the PPA value can be calculated.

$$Y_n = \{y_j^{(n)}, j = 1, \ldots, |Y_n|\} \tag{5}$$

X_n denotes the low-resolution images captured thru the measurement optics and Y_n denotes the high-resolution images, which correspond to the object image measured in the training optics. n is the number of training samples.

Our goal is to learn a model F precisely to predict the values of output Y_n.

$$\widehat{Y_n} = F(X_n) \tag{6}$$

where $\widehat{Y_n}$ is its estimate of output Y_n (We drop the n subscript for brevity).

Our model F, as shown in Fig. 7, was comprised of a multiple layer structure and can be expressed as the composition of layer function f.

$$F(x) = (f_l \circ f_{l-1} \circ \cdots \circ f_1)(x) \tag{7}$$

where

$$f_k(x) = \max(w_k * x + b_k) \tag{8}$$

k represents the order of layer, l is the total number of layers, and x is the input data. The operator $\circ$ denotes the function of the composition, the operator $*$ denotes a convolution, and $\max(0, \cdot)$ corresponds to a ReLU. w_k and b_k are a 3-by-3 convolution filter parameter and bias term for the $k-th$ layer, respectively. These network parameters were trained in order to minimize the loss function using a stochastic gradient descent with the standard back-propagation. The loss function was L_1 distance expressed as

$$L_1(W) = \frac{1}{n}\sum_{n=1}^{N}\sum_{j=1}^{|Y_n|}\left|y_j^{(n)} - \widehat{y_j^{(n)}}\right| \tag{9}$$

Rivenson *et al.* [21] implemented a convolution neural network for statistical transformation between a low-resolution and a high-resolution microscopic image, enhancing its spatial resolution, a large field of view and depth of field. Our proposed method applied the same concept in order to improve measurement accuracy and the spatial resolution degraded by diffraction blur. Our papers have a very similar methodology. However, there are two differences in the deep neural network architecture. The first is depth of layers. Rivenson *et al.* [21] composed a network of 10 layers (5 blocks × 2 convolution layers), which was a relatively shallow network in the field of deep learning superresolution. For example, most networks that competed in the New Trends in Image Restoration and Enhancement workshop and challenge on image superresolution (NTIRE) 2017 [19] had more than 20 layers, and the 1st prize network of this contest consisted of 36 layers). We assume that the reason behind employing a smaller shallow layer might have been to settle convergence for various biological samples, while accepting a small loss in performance. In comparison, our samples were OLEDs and these had repeated patterns. Additionally, the shape deviations among samples tended to be small. Therefore, we can construct our network to be deeper, 20 layers, achieving stable convergence and high superresolution performance at the same time. The second difference was convergence speed. The number of epochs till convergence by Rivenson *et al.* was 460 (lung tissue sample) or 630 (breast tissue sample). Our method was only 150 epochs and this resulted from a simple loss function. Rivenson *et al.* exploited the combination of Mean Square Error and the L2 norm of the image gradient and focused on the stability of the convergence. We employed a very simple L1 loss function and high value for the learning rate in order to speed up the training process. This enabled convergence of the training process such that it was stable by virtue of similar structures in the samples, repetitive patterns in the image, and a carefully initialization of kernels and bias terms.

Various previous deep learning superresolution techniques [10-15], including VDSR, basically used one image to create an LR-HR set. The original image was reduced and enlarged as a scale factor. Because the compressed image underwent a bicubic interpolation process, the resolution of the image was degraded, and it became the LR image. The original image was then used as the HR image in order to

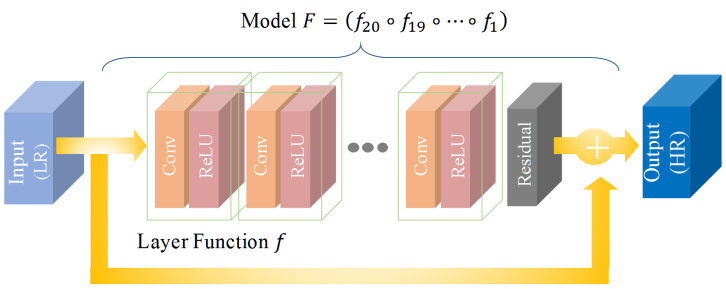

FIG. 7. Our goal is to learn a model F and this is expressed as composition of layer function f. This model is based on a deep convolution neural network. This network was made of 20 layers. Each layer has a convolution layer composed of 3-by-3 convolution filters and a ReLu layer. The number of filters are 64. The last layer is a residual image, subtracted result between input and output image. Most values of residual image are likely to be zero or small.

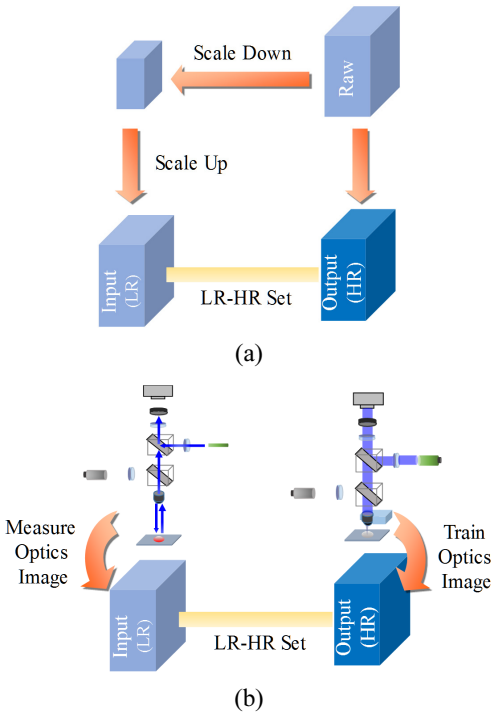

FIG. 8. Conventional deep learning makes LR-HR set with one image (a), this paper composes LR-HR set with two images obtained from each optical systems (b).

train a network (Fig. 8(a)). Generally, a high-quality image dataset such as ImageNet, BSD300, BSD500, Set5, Set14, and DIV2K had been exploited. The quality of trained network output cannot surpass the one of raw images. In this paper, however, all images of a single specimen acquired from two optical systems were then used as LR-HR image sets (Fig. 8(b)). Configuring an image set in this way provided several benefits, including a higher resolution, which exceeds that of the original image. Therefore, the superior performance came from a new way of preparing training data sets, not a neural network architecture. For example, the NA of training optics was approximately four times that of the measurement optics, and its magnification from the objective lens was twice that of the measurement optics. Theoretically, if the learning was done perfectly, the image of the measurement optics can be converted into a super-resolved image whose resolution would be four times better and the size doubled. This was a surprising result, given that resolution usually went down as image size increased. In contrast, the super-resolved image of a conventional method might have a resolution that was less than that of the original image.

The network, as shown in Fig. 7, was based on a VDSR structure [12] and was composed of 20 layers. Each layer had 64 convolutional neurons with a 3-by-3 convolution filter. The last layer was a residual image, which was the subtracted result between an input and output image. Most values of the residual image were likely to be zero or small. The proposed network was trained by a stochastic gradient descent. Weight decay parameters and momentum were set to 0.0001 and 0.9, respectively. The learning rate was set from 0.005 to 0.0001, and it was reduced gradually at each epoch. The number of epochs was 150, and the patch was a 32×32 grid of the image. The network was implemented using the MatConvNet toolbox (beta.24) in a MATLAB 2014a environment.

The purpose of this paper was to obtain a high-resolution image by developing a conversion equation between optical systems using a deep learning network. We noted that the goal of our study was not to find a network structure with the best performance in a given dataset, as was the goal of many papers on deep learning.

3.2. Measurement Process

The measurement process in Fig. 8 calculated the PPA value with superresolution images and edge detection. First, the image of a measurement object is captured thru a measurement optical system (Fig. 9(a)). This image was converted to a superresolution image using the pre-learned network in a training process (Fig. 9(b)). Next, calculate the center of the TFT and organic layer with edge detection. The edge detection senses the intensity change in pixels and separates the object from the background by finding a boundary line. The edge detection image shows the boundary for the TFT and the organic layer (Fig. 9(c)). The orange cross mark and blue cross mark show the center coordinates of the TFT and of the organic film, respectively (Fig. 9(d)). The distance between center coordinates indicates the PPA value. Therefore, the accuracy of the edge detection is directly related to the PPA value. If one pixel in the edge detection is miscalculated, the error is accumulated by 0.275 μm, which is the size per pixel. Figure 9(e) illustrates the area within the yellow circle in Fig. 9(c). One small rectangle corresponds to one pixel of the image. The white, gray, black rectangles are the intensity of the organic layer, intermediate, and background, respectively. Blue rectangles mean a pixel-level edge and the red line corresponds to a subpixel level edge. To obtain a higher level of accuracy, it is necessary to have edge detection at a sub-pixel level.

A pixel level of accuracy, and finding the pixels of an image corresponding to an edge, can be obtained by the Sobel, Canny, Laplacian of Gaussian (LoG) operator. After edge detection with pixel-level accuracy is complete, the intensity around pixels corresponding to an edge are assumed as a continuous function. Then, a sub-pixel level of accuracy can be accomplished by finding the greatest intensity value change in one pixel again. This paper exploited pixel-level edge detection thru LoG operator, which was expressed as

$$\nabla^2 G(x, y, \sigma) = \left(\frac{x^2 + y^2 - 2\sigma^2}{\sigma^4} \right) e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{10}$$

where $G(x, y, \sigma)$ represents a two-dimensional Gaussian function.

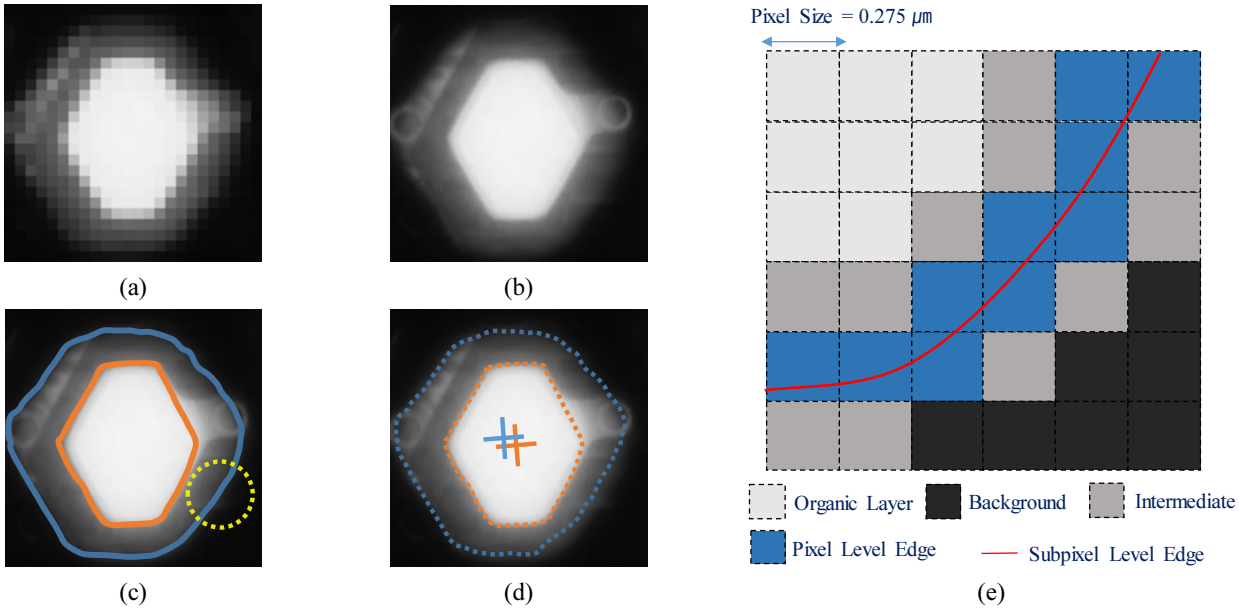

FIG. 9. The measure process calculated PPA value with superresolution image and edge detection. The image of test sample (a), superresolution image with the trained neural network (b), edge detection image (c), the center coordinate of TFT and organic layer (d) and difference between pixel level and subpixel level edge detection (e) show the process of measure process.

Here, in order to shorten the computation time, the one-dimensional convolution operation in the row direction and the column direction were successively performed to replace the two-dimensional convolution operation. So,

$$\begin{aligned}\nabla^2 G(x, y, \sigma) &= h_{12}(x, y) + h_{21}(x, y) \\ &= h_1(x)h_2(y) + h_2(x)h_1(y)\end{aligned} \tag{11}$$

where

$$h_1(\xi) = \sqrt{K}\left(1 - \frac{\xi^2}{\sigma^2}\right)e^{-\frac{\xi^2}{2\sigma^2}} \tag{12}$$

$$h_2(\xi) = \sqrt{K}e^{-\frac{\xi^2}{2\sigma^2}} \tag{13}$$

K represents scale factor.

The continuous function of intensity around an edge in subpixel edge detection was assumed as the facet model. Using the notation of [22, 23], let the discrete integer index R and C be an index satisfying the condition, $r \in R$, $c \in C$. The facet model was expressed as

$$\begin{aligned}I_{LoG}(r, c) = {} & k_0 + k_1 r + k_2 c + k_3 r^2 + k_4 rc \\ & + k_5 c^2 + k_6 r^2 c + k_7 rc^2 + k_8 r^2 c^2\end{aligned} \tag{14}$$

where k_n is the polynomial coefficient. If we assume $I_{LoG}(r, c)$ was a continuous function, it can be expressed as

$$I_{LoG}(r, c) = \sum_{n=1}^{N} a_n P_n(r, c) \tag{15}$$

where

$$a_n = \frac{\sum_r \sum_c P_n(r, c) d(r, c)}{\sum_i \sum_j P_n^2(i, j)} \tag{16}$$

$P_n(r, c)$ represents a Chebyshev discrete orthogonal polynomial set, and $d(r, c)$ represents a data value be observed. The image intensity $I_{LoG}(r, c)$ can be obtained using the facet model technique. The zero-crossing operator is applied to the reconstructed image intensity $I_{LoG}(r, c)$ e, so that the edge points can finally be detected.

IV. RESULTS

4.1. Resolution Enhancement Analysis

The proposed method was validated with peak signal-to-noise ratio (PSNR), structural similarity index (SSIM) compared to conventional deep learning networks [10, 12, 14], multiple image based interpolation [9], and unsharp mask [24]. The deep learning network created a superresolution image using a measurement optics image as an LR image. The multiple image based interpolation produced 9 images when moving the sample with micrometer as a one-third distance of pixel resolution along the X, Y direction. Each of the images was reconstructed in order to create a superresolution image. The unsharp mask method was used

as an image sharpening method to enhance local contrast based on edge detection.

To prepare LR-HR sets for training, various samples were used, such as those for an OLED mobile phone and OLED TV. Various patterns in these samples were illuminated with UV light to obtain organic film images, and a white light LED was used to obtain TFT images. In the training process, both TFT and organic layer images were used together. There were 150 training sets. In contrast, test images were made by measuring sample 1~6 with the measurement optics. There were 40 test images.

Figures 10 and 11 show the proposed method compared to other top performing methods. HR images of the training optics were used as the ground truth. In Fig. 10, only our method reconstructs TFT lines in the middle and at the corners. Similarly, in Fig. 11, the boundary lines of the TFT circuits are clean and vivid in our method, whereas these lines are severely blurred or distorted in the other methods. Table 1 provides a summary of the quantitative evaluation results for the test samples. The proposed method outperforms all previous methods in these samples.

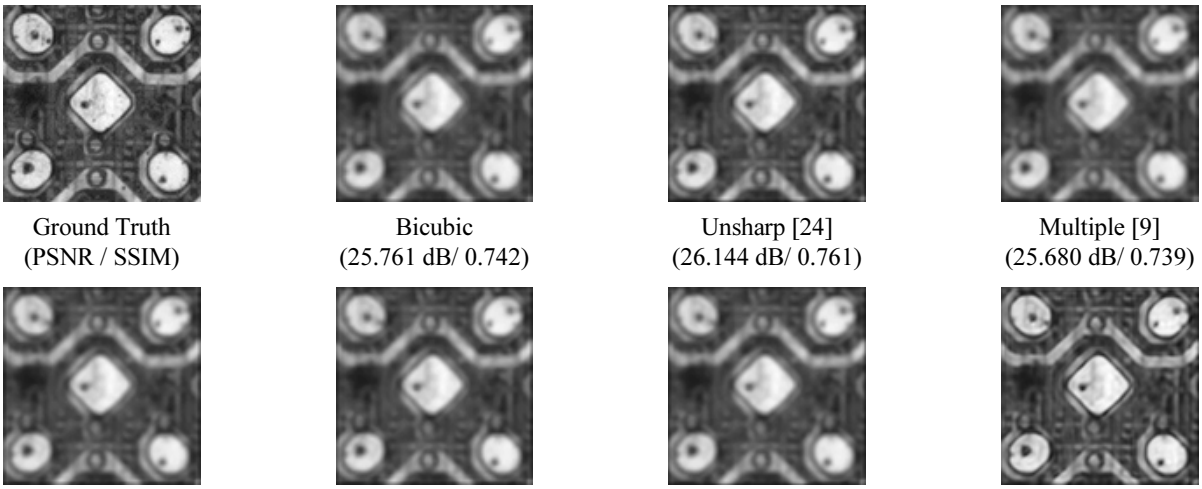

FIG. 10. Qualitative comparison of proposed method with other works on ×2 superresolution about Sample 1. Proposed method recovers the sharp edge line of TFT.

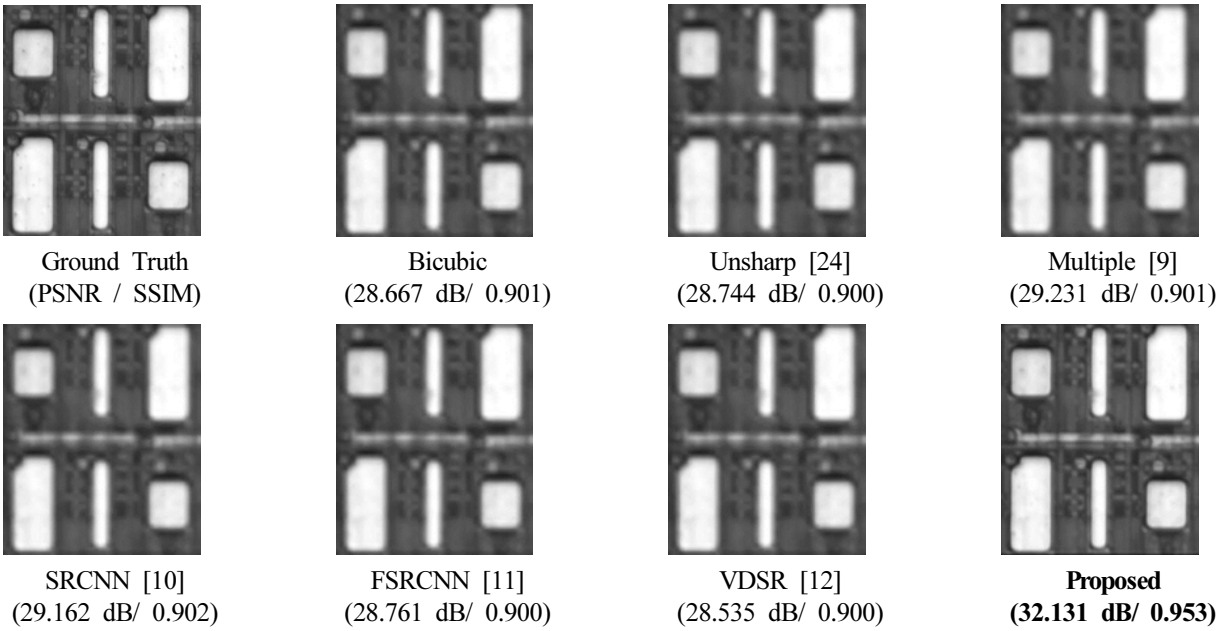

FIG. 11. Qualitative comparison of proposed method with other works on ×2 superresolution about Sample 4. Proposed method shows boundary lines of TFT circuits are clearly visible.

TABLE 1. Average PSNR(dB)/SSIM for scale factor ×2 on datasets for Sample 1~6

Test sample	Bicubic	Unsharp	Multiple	SRCNN	FSRCNN	VDSR	**Proposed**
Sample 1	27.273/0.742	27.236/0.741	27.509/0.742	27.494/0.741	27.282/0.740	27.043/0.741	**32.483/0.779**
Sample 2	30.105/0.910	29.805/0.910	30.557/0.912	30.142/0.916	29.917/0.913	29.661/0.913	**31.904/0.941**
Sample 3	27.730/0.907	27.038/0.902	28.481/0.910	27.684/0.910	27.190/0.907	26.928/0.909	**28.884/0.938**
Sample 4	26.985/0.843	26.497/0.844	27.684/0.846	26.951/0.847	26.575/0.843	26.207/0.844	**29.116/0.900**
Sample 5	26.205/0.882	29.968/0.877	30.496/0.889	30.433/0.901	29.960/0.897	29.658/0.896	**32.299/0.929**
Sample 6	27.730/0.907	27.038/0.902	28.481/0.910	27.684/0.910	27.190/0.907	26.928/0.909	**28.884/0.938**

4.2. Modulation Transfer Function Analysis

We conducted modulation transfer function (MTF) analysis in order to quantify the effect of our neural network on the spatial frequencies of the output image. We applied the trained neural networks on a resolution target (USAF Resolving Power Test Target 1951), which was imaged using the measurement optics. The theoretical Abbe limit of measurement optics was 2.29 μm (436.36 line pair/mm), assuming an illumination wavelength of 550 nm. MTF was evaluated by calculating the contrast for different elements of the resolution target [21]. We calculated the contrast value with element's local maximum value I_{max}, and minimum value I_{min} in Eq. (17).

$$Contrast = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \tag{17}$$

The maximum value and minimum value can be obtained from the intensity profile of average cross-section of the resolution target element. We set the contrast to 0 in case there is no difference between the maximum value and the minimum value, assuming that the MTF was not preserved. Based on this experimental analysis, MTFs for the input image and the output image of the deep neural network are compared to each other. As shown in Fig. 12, the output image of network shows an increase in modulation contrast for a significant portion of the spatial frequency spectrum at high frequencies.

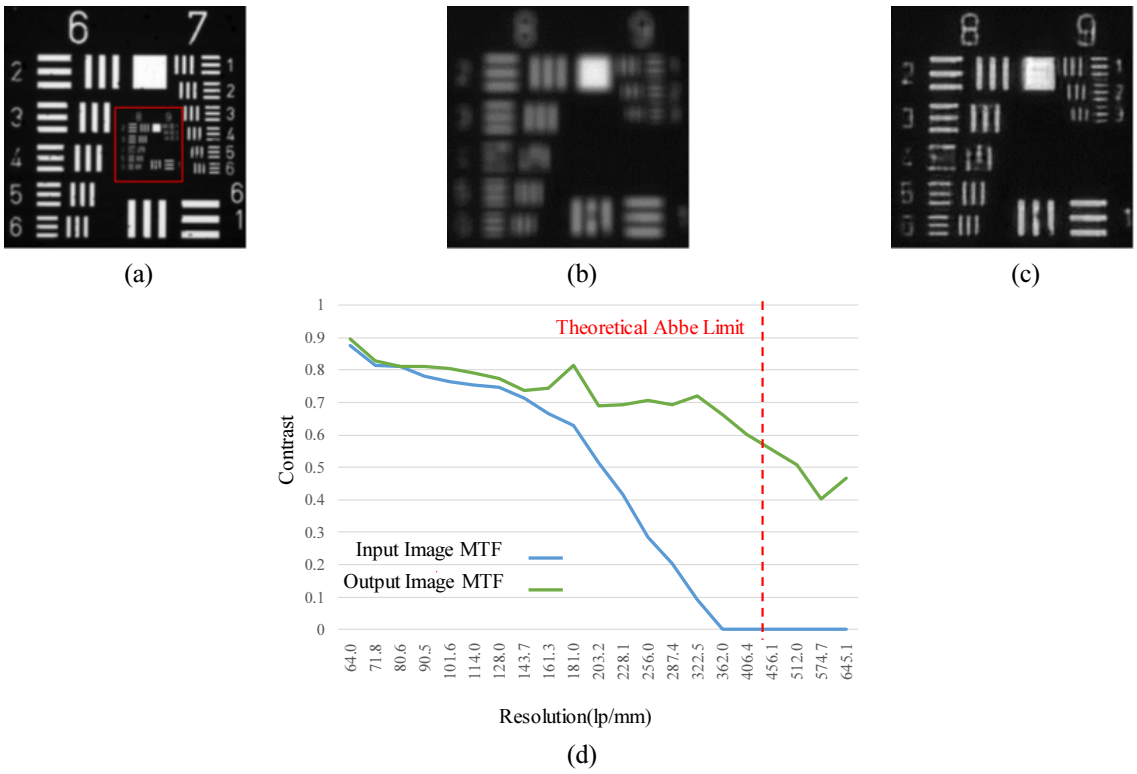

FIG. 12. Modulation transfer function (MTF) comparison for the input image and the output image of trained deep neural network. The US Air Force (USAF) resolution target illuminated with a numerical aperture of 0.12, leading to theoretical Abbe limit of 2.29 μm. Input image acquired with a measurement optics (a), zoom in on the red highlighted region (b), output image of the deep neural network (c), MTF calculated from the input and output images of the deep neural network (d).

4.3. Measurement Accuracy Analysis

Measurement accuracy represents the difference of PPA values between data to be measured and true value. A high measurement accuracy means having the small error between two values. If the resolution is high, it is possible to extract the correct edge, resulting in accuracy of a high order. Table 2 shows the measurement accuracy of various samples. These values are the average of deviation between PPA values obtained by each method and those of the HR image. The proposed method has the smallest errors on these samples. Because the present pixel resolution is 0.275 μm, it can be said that the accuracy of the proposed method is 1/4~1/30 pixel level. Also, the relationship between image enhancement and measurement performance can also be verified. As the values of PSNR and SSIM increase, the measurement accuracy tends to be higher, but does not exactly coincide. This is because PSNR and SSIM are evaluated for the improvement of overall images, but the measurement accuracy is affected only by how much the edge has higher resolution. Further studies are underway to improve the resolution of the edge area preferentially. This also has the effect of compressing data, which can also save time on network training.

V. CONCLUSION

This paper presented a novel method that demonstrates the enhancement of both spatial resolution in long distance imaging and measurement-level of an organic layer for an in-line deposition process by utilizing a deep learning technique. Existing industry superresolution techniques have reconstructed an HR image with multiple LR images acquired by a camera array, various illumination angles, and by the shifting of aperture and sensor. However, the proposed method obtained an HR image with only one LR image through the trained neural network. This network was able to convert an image measured in the LR optical system into an image as if it were measured in a HR optical system. This method enables real-time measurement that is suitable for an in-line process. The experimental results show that the proposed method had the best performance of PSNR, SSIM and measurement accuracy compared to conventional single image superresolution methods on various samples.

ACKNOWLEDGEMENT

This work was supported in part by Institute of Engineering Research, Institute of Advanced Machines and Design at Seoul National University. The work also supported by SNU Precision Company.

REFERENCES

1. B. Geffroy, P. L. Roy, and C. Prat, "Organic light-emitting diode (OLED) technology: Materials, devices and display technologies," Polym. Int. **55**, 572-582 (2006).
2. K. Venkataraman, D. Lelescu, J. Duparr´e, A. McMahon, G. Molina, P. Chatterjee, R. Mullis, and S. Nayar, "Picam: An ultra-thin high performance monolithic camera array," ACM Trans. Graph. **32**, 166 (2013).
3. G. Carles, J. Downing, and A. R. Harvey, "Super-resolution imaging using a camera array," Opt. Lett. **39**, 1889-1892 (2014).
4. J. Holloway, Y. Wu, M. K. Sharma, O. Cossairt, and A. Veeraraghavan, "SAVI: Synthetic apertures for long-range, subdiffraction-limited visible imaging using fourier ptychography," Sci. Adv. **3**, e1602564 (2017).
5. G. Zheng, R. Horstmeyer, and C. Yang, "Wide-field, high-resolution fourier ptychographic microscopy," Nat. Photonics, **7**, 739-745 (2013).
6. S. Dong, Z. Bian, and R. Shiradkar, "Sparsely sampled fourier ptychography," Opt. Express **22**, 5455-5464 (2014).
7. K. Guo, S. Dong, P. Nanda, and G. Zheng, "Optimization of sampling pattern and the design of fourier ptychographic illuminator," Opt. Express **23**, 6171-6180 (2015).
8. S. Dong, R. Horstmeyer, R. Shiradkar, K. Guo, X. Ou, Z. Bian, H. Xin, and G. Zheng, "Aperture-scanning fourier ptychography for 3d refocusing and super-resolution," Opt. Express **22**, 13586-13599 (2014).
9. N. T. Doan, J. H. Moon, T. W. Kim, and H. J. Pahk, "A fast image enhancement technique using a new scanning path for critical dimension measurement of glass panels," Int. J. Precis. Eng. Man. **13**, 2109-2114 (2012).
10. 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* (ECCV) (Chem, 6-12 Sept. 2014), pp. 184-199.
11. C. Dong, C. C. Loy, K. He, and X. Tang, "Accelerating the superresolution convolutional neural network," in *European*

TABLE 2. Average measurement accuracy of PPA value for Sample 1~6 [unit: μm]

Test sample	Bicubic	Unsharp	Multiple	SRCNN	FSRCNN	VDSR	**Proposed**
Sample 1	0.341	0.131	0.098	0.117	0.120	0.116	**0.022**
Sample 2	0.151	0.089	0.151	0.093	0.192	0.116	**0.062**
Sample 3	0.134	0.061	0.032	0.075	0.092	0.076	**0.010**
Sample 4	0.117	0.080	0.117	0.087	0.108	0.114	**0.043**
Sample 5	0.067	0.067	0.085	0.033	0.044	0.063	**0.015**
Sample 6	0.100	0.126	0.068	0.138	0.137	0.137	**0.065**

Conference on Computer Vision (ECCV), (Amsterdam, 11-14, Oct. 2016), pp. 391-407.
12. J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2016), pp. 1646-1654.
13. W. Bae and J. Yoo, "Beyond deep residual learning for image restoration: Persistent homology-guided manifold simplification," https://arxiv.org/abs/1611.06345.
14. Y. S. Han, J. Yoo, and J. C. Ye, "Deep residual learning for compressed sensing CT reconstruction via persistent homology analysis," https://arxiv.org/abs/1611.06391.
15. B. Lim, S. H. Son, H. W. Kim, S. J. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshop* (IEEE, 2017), pp. 136-144.
16. L. Xu, J. Ren, C. Liu, and J. Jia, "Deep convolutional neural network for image deconvolution," in *Proc. IEEE Conference on Advances in Neural Information Processing Systems* (IEEE, 2014), pp. 1790-1798.
17. J. Xie, L. Xu, and E. Che, "Image denoising and inpainting with deep neural networks", in *Proc. IEEE Conference on Advances in Neural Information Processing Systems* (IEEE, 2012), pp. 341-349.
18. K. Zhang, W. Zuo, Y. Chen. D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising," IEEE Trans. Image Process. **26**, 3142-3155 (2017).
19. R. Timofte, E. Agustsson, L. V. Gool, M.-H. Yang, L. Zhang, B. Lim, S. Son, H. Kim, S. Nah, K. M. Lee, X. Wang, Y. Tian, K. Yu, Y. Zhang, S. Wu, C. Dong, L. Lin, Y. Qiao, C. C. Loy, W. Bae, J. Yoo, Y. Han, J. C. Ye, J.-S. Choi, M. Kim, Y. Fan, J. Yu, W. Han, D. Liu, H. Yu, Z. Wang, H. Shi, X. Wang, T. S. Huang, Y. Chen, K. Zhang, W. Zuo, Z. Tang, L. Luo, S. Li, M. Fu, L. Cao, W. Heng, G. Bui, T. Le, Y. Duan, D. Tao, R. Wang, X. Lin, J. Pang, J. Xu, Y. Zhao, X. Xu, J. Pan, D. Sun, Y. Zhang, X. Song, Y. Dai, X. Qin, X.-P. Huynh, T. Guo, H. S. Mousavi, T. H. Vu, V. Monga, C. Cruz, K. Egiazarian, V. Katkovnik, R. Mehta, A. K. Jain, A. Agarwalla, C. V. S. Praveen, R. Zhou, H. Wen, C. Zhu, Z. Xia, Z. Wang, and Q. Guo, "NTIRE 2017 challenge on single image super-resolution: Method and results," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops* (IEEE, 2017), pp. 1110-1121.
20. D. G. Lowe, "Distinctive image features from scale-invariant keypoints," Int. J. Comput. Vis. **60**, 91-110 (2004).
21. Y. Rivenson, Z. Göröcs, H. Günaydin, Y. Zhang, H. Wang, and A. Ozcan, "Deep learning microscopy," Optica **4**, 1437-1443 (2017).
22. J. Lee, Y. Kim, S. Kim, I. Lee, and H. Pahk, "Real-time application of critical dimension measurement of TFT-LCD pattern using a newly proposed 2D image-processing algorithm," Opt. Lasers Eng. **46**, 558-569 (2008).
23. R. M. Haralick, "Digital step edge from zero crossing of second directional derivatives", IEEE Trans. Pattern Anal. Mach. Intell. **6**, 58-68 (1984).
24. S. McHugh, *Digital photography tutorials* (2005).