Research Article | Vol. 28, No. 18 / 31 August 2020 / Optics Express 27196
Optics EXPRESS

Rapid super resolution for infrared imagery

NAVOT OZ,[1,2] NIR SOCHEN,[3] OSHRY MARKOVICH,[4] ZIV HALAMISH,[5] LENA SHPIALTER-KAROL,[6] AND IFTACH KLAPP[1,*]

[1]*Department of Sensing, Information and Mechanization Engineering, Institute of Agricultural Engineering, Agricultural Research Organization (ARO),Volcani Center, Rishon LeZion, Israel*
[2]*School of Electrical Engineering, Tel Aviv University, Tel Aviv 69978, Israel*
[3]*Department of Applied Mathematics, Tel Aviv University, Tel Aviv 69978, Israel*
[4]*Rahan Meristem, Kibbutz Rosh Hanikra Western Galilee 22825, Israel*
[5]*Evogene, Gad Feinstein St., Rehovot 7612002, Israel*
[6]*Hazera, Berurim M.P Shikmim 7983700, Israel*
[*]*iftach@volcani.agri.gov.il*

Abstract: Infrared (IR) imagery is used in agriculture for irrigation monitoring and early detection of disease in plants. The common IR cameras in this field typically have low resolution. This work offers a method to obtain the super-resolution of IR images from low-power devices to enhance plant traits. The method is based on deep learning (DL). Most calculations are done in the low-resolution domain. The results of each layer are aggregated together to allow a better flow of information through the network. This work shows that good results can be achieved using depthwise separable convolution with roughly $300K$ multiply-accumulate computations (MACs), while state-of-the-art convolutional neural network-based super-resolution algorithms are performed with around $1500K$ MACs. MTF analysis of the proposed method shows a real $\times 4$ improvement in the spatial resolution of the system, out-preforming the diffraction limit. The method is demonstrated on real agricultural images.

1. Introduction

Infrared (IR) imagery in the $8\mu m - 12\mu m$ atmospheric window is extensively used for agricultural remote-sensing tasks. While a visible light band (VIS) camera measures light reflected from an object, the IR camera measures the object's (plant) temperature associated with thermal radiation. The use of IR imagery in precision agriculture has been well-documented. To name just a few of these uses, [1] showed how to estimate the water status of a grapevine, [2] detected fruit, and [3] measured drought responses of plants. The temperature of the plant is important in deducing information on its well-being.

Once quite costly, today a wide variety of IR cameras are available at affordable prices [4]. While current detector technology allows for a low-cost IR camera, spatial resolution is still limited - two orders of magnitude less than typical VIS imaging. With only a few tens of thousands of pixels in the camera's field of view, there is a trade-off between target coverage and spatial resolution. Since background temperature can vary widely from that of the object of interest, imaging of a small target often results in mixed hot and cold pixels at its edges associated with erroneous temperature estimation.

This work aims to estimate a high-resolution (HR) IR image from a low-resolution (LR) IR image for low-power devices. This process is called super-resolution (SR). Early SR methods were interpolation-based (e.g., nearest-neighbor, bicubic), producing rapid results with low quality, both visually and metrically. Later, advances were made using sparse coding methods (e.g. [5], [6]), image priors (e.g. [7], [8]) and example-based learning (e.g [9], [10]). Recent advances in SR have occurred in the field of machine-learning - specifically deep-learning (DL). The authors of [11] were the first to demonstrate this approach with a convolutional neural network (CNN). As networks grew deeper, the problem of vanishing gradients became more acute, as unraveled

#389926 https://doi.org/10.1364/OE.389926
Journal © 2020 Received 12 Feb 2020; accepted 27 May 2020; published 28 Aug 2020

Optics EXPRESS

by [12]. A solution was to use *skip connections* to allow gradient flow throughout the entire network as in [13]. Later, the notion of *dense skip connections* was proposed by [14], where all of the layers are connected via skip connections. [15] proposed using all of the layer's outputs for the final reconstruction via a *bottleneck layer*.

Recently, extensive work has been made in the optical society on improving the resolution of different imaging systems using DL. To name just a few - [16] applied a CNN to LR slide scanner microscope, [17] upscaled Terahertz images using CNN, [18] used SR on a single-photon camera to increase the signal-to-noise ratio of the outputs, [19] performed SR on an emitting apparatus used for microscopy, [20] used SR to increase the throughput of a lens-free holographic microscope by reducing the number of measurements needed and [21] used a generative adversarial network (GAN) to enhance microscope imagery.

Solutions for SR in IR images were researched in several directions - combining several IR images [22], using prior knowledge from VIS images to enhance the IR image (e.g., via edges, as suggested by [23] and [24]), using iterative regularization [25]. More recent works have used DL. In [26], the authors trained a cascade network - meaning that the SR image is restored in several steps. They first increased the resolution two-fold in each dimension (2×) and subsequently they increased it again four-fold such that the final resolution is eight-fold in each dimension (8×). They demonstrated it on a limited set of examples. Another method was discussed by [27], who offered a DL approach using prior knowledge taken from RGB images. Their method assumed subpixel registration between a pair of VIS and IR images. However, most devices only have either VIS or IR. Even in devices with both channels, accurate subpixel registration is hard to establish with a real commercial-grade apparatus. Low-power approaches for DL are mainly focused on classification and detection. One solution, *MobileNet* by [28], uses *depthwise separable convolution* as suggested by [29] to lower computational complexity. This is explored in Section 2.4.

This work proposes a method for obtaining SR using a single IR image while balancing between the metric quality of the SR image and the low-power requirements posed by the modest hardware of IR cameras. The computational complexity of the proposed solution is considerably lower than that of similar networks while achieving satisfactory results. Today's body of work focuses on improving the metrics of estimation (e.g., peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM)), but pays little attention to run time or power costs. The proposed solution combines both quality and low complexity so that it can be performed on low-power devices. Thus, in this work, a new deep learning SR scheme for IR images is presented. A network that combines the bottleneck layer from [15] with the dense skip connections of [14] is shown to preserve the high-quality performance of a deep network with only a small portion of the required computational power. All calculations are done on the low-resolution space to save on computational costs, and the upscaling is performed using the shuffle block method [30]. Results show that only a handful of skip connections suffice. To further lower computational complexity, depthwise separable convolution [29] is performed, showing good PSNR results as well.

The proposed method is shown to improve the modulation transfer function (MTF) of the imaging system in Section 3.3.

2. Method

The proposed CNN is presented schematically in Fig. 1. The network is composed of a *LR Block*, a *Shuffle Block* and a final convolution layer. The input to the network is (I_{LR}) a low resolution IR image with dimensions $H \times W \times 1$. The *LR Block* learns the features of the image to extract highly detailed information from the input. A detailed description of the *LR Block* can be seen in Fig. 2. The *LR Block* output dimension is $H \times W \times Ch$ where Ch denotes the number of channels. Following the *LR Block* is the *Shuffle Block*, where the upscaling from I_{LR} to I_{SR} is performed as

Optics EXPRESS

described in [30]. The *Shuffle Block* aims to upscale the features to dimensions of $\alpha H \times \alpha W \times Ch$, where $\alpha > 1$ denotes the SR upscaling. Finally, the convolution layer reorders the features to the required SR image with dimensions $\alpha H \times \alpha W$ with a single channel. The SR image (I_{SR}) is the approximation of the HR image (I_{HR}). The structure of the *LR Block* is presented in Fig. 2. The block is composed of multiple layers. The output of each layer is aggregated via concatenation to the outputs of the previous layers and to the input image. The concatenated matrix goes through a *bottleneck layer* which outputs *Ch* filters. Each *bottleneck layer* convolves all the outputs of its preceding layers. This process is further described in Section 2.2.

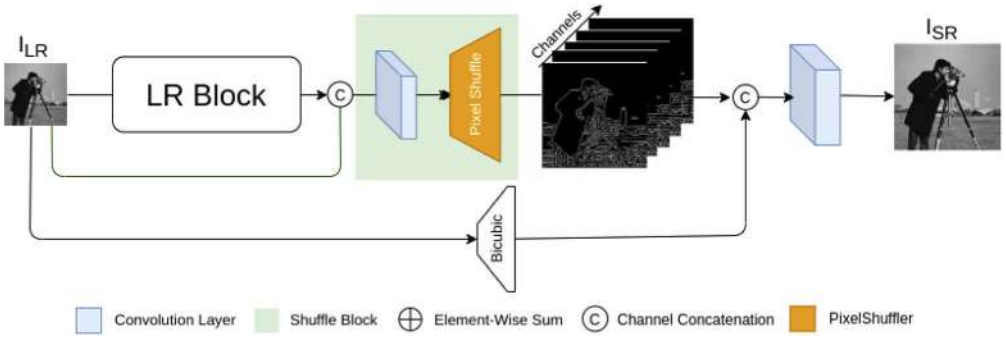


Fig. 1. The proposed method applies the *LR Block* (Fig. 2) to the input. The output of the *LR Block* in concatenated with the input and upsampled by the *Shuffle Block*. The input is interpolated and concatenated to the output of the *Shuffle Block*. The network outputs I_{SR}.

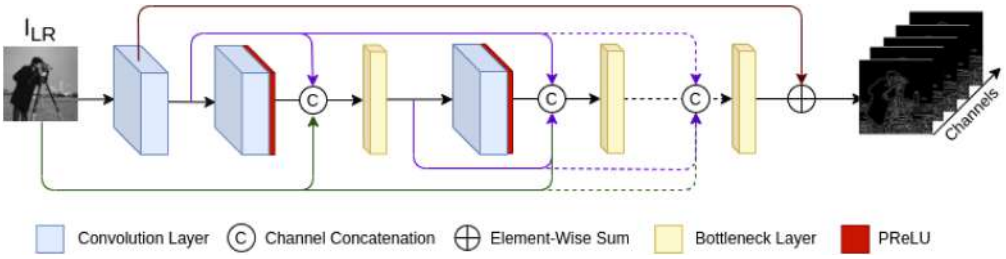


Fig. 2. The *LR Block* decomposes I_{LR} into *Ch* filters. The output of each layer is concatenated with the outputs of all previous layers, along with the input image. The *LR Block* outputs *Ch* filters in the LR domain.

The parametric rectified linear unit (PReLU) proposed by [32] is used as a nonlinear activation function ϕ after each convolution and *bottleneck layer*.

Denoting the *Convolution* between two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ as $(\boldsymbol{A} * \boldsymbol{B})$ and the Concatenation between these matrices as $\{\boldsymbol{A}, \boldsymbol{B}\}$, The network is composed of L layers that can be described as follows:

$$\begin{cases} f^1(\bar{I}_{LR}; \theta_1) = \phi(\theta_1 * \bar{I}_{LR}) \\ f^l(f^{l-1}; \theta_l) = \left\{ f^{l-1},\ \phi\left(\theta_l * S_{l-1}\right) \right\} \quad l \in 2, \ldots, L \end{cases} \tag{1}$$

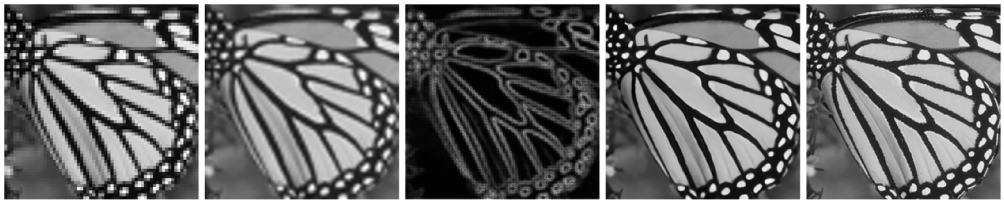
where $\bar{I}_{LR}$ is the normalized low resolution input, θ are learned weights with 3×3 spatial dimensions with *Ch* filters and f^l denotes the output of a layer l. The output of the l bottleneck layer is denoted S_l. The bias term is omitted for brevity.

The network has one initial convolution layer for the input, L convolution layers that are concatenated together and one more final convolution layer for the output. All in all there are $(2+L)$ convolutions and L bottleneck-layers (Fig. 1).

Optics EXPRESS

Depthwise separable convolution layers, as proposed by [29], are used to lower computational cost and are explored in Section 2.4.

The final layer of the network is a convolution with $Ch + 1$ filters as input. The extra channel is a Bicubic interpolation of I_{LR} which is concatenated to the network output before going through the final layer. The concatenation enables the network to learn only the high-frequency difference between I_{LR} and I_{HR}. The final layer outputs a single channel *without* an activation function. The process is illustrated in Fig. 3. Figure 3(a) is a LR image. Figure 3(b) is the bicubic interpolation of the LR image - the low frequencies of LR. Figure 3(c) is the high-frequency information learned by the network pipeline. These interpolation data and the high-frequency data are summed in the final layer, shown in Fig. 3(d). During the training process, the result of this summation is compared to the HR ground-truth image (Fig. 3(e)).



(a) Low frequency (b) Bicubic interp. (c) High frequency (d) SR (e) HR

Fig. 3. Illustration of the proposed super-resolution algorithm's summation of low-frequency and high-frequency information. The image was taken from Set5 [31].

Ch is set to 32 channels, L is set to 8 layers. The learned kernel for the filters in the first convolutional layer is set to a 5×5 square to capture local features, while the kernel size for all other convolutional layers is set to 3×3.

2.1. Pre-processing and cost function

I_{LR} single channel represents the object's temperature and its dynamic range is 16 bits. Before to entering the network, the I_{LR} is standardized to the range $[0, 1]$ such that

$$\bar{I}_{LR} = \frac{I_{LR} - \min[I_{LR}]}{(\max[I_{LR}] - \min[I_{LR}])} \tag{2}$$

Network training is done by minimizing the error between a ground-truth HR image and the network's output (SR image). As a cost function, the absolute mean error, i.e. the L_1 norm which is robust to outliers, is applied for the difference between I_{SR} and I_{HR}. It reads

$$\mathcal{L}_{SR}(\theta) = \frac{1}{H \cdot W} \sum_{i=0}^{H} \sum_{j=0}^{W} |I_{SR}^{i,j} - I_{HR}^{i,j}| \tag{3}$$

where H, W are height and width, respectively and θ are the learned weights of the network.

2.2. Bottleneck layer

Bottleneck layers are a 1×1 convolution where the number of output filters is always Ch. This process was described in [33] and used by [34]. The bottleneck layer has several effects. First, it helps mitigate vanishing gradients. Second, the most important features are chosen using the computationally efficient and parameter-conserving bottleneck layer, so operations in other convolution layers are always applied only to Ch channels.

Optics EXPRESS

The mathematical formulation of the bottleneck layer is:

$$S_l = \phi\left(\vartheta_l * \left\{\bar{I}_{LR}, f_1, \ldots, f_l\right\}\right) \tag{4}$$

where ϑ_l denotes the learned weights of the *bottleneck layer* with $l \times Ch$ filters as input and Ch filters as output. ϕ is the nonlinear activation function and f^l the output from the lth convolution layer

2.3. The relation between temperature and pixel intensity

The Stefan-Boltzmann equation formulates the relationship between a surface's temperature and its irradiance. For a typical outdoor temperature (e.g., $280 - 320\ K$), the target and ambient temperatures are similar, such that the change in radiation power in this range can be approximated as linearly dependent on the change of the body temperature relative to the ambient temperature.

$$P(T) = \alpha \cdot \sigma T^4 = \alpha \cdot \sigma(T_0 + \Delta T)^4 \approx \underbrace{\alpha \cdot \sigma(T_0)^4}_{P_0} + 4\alpha \cdot \sigma(T_0)^3 \Delta T \tag{5}$$

where P is the radiant power, T_0 and P_0 are the reference ambient temperature and associated radiance respectively, σ is the Stefan-Boltzmann coefficient, and α is a proportion factor. Equation (5) presents the Taylor expansion around the ambient temperature. Indeed, in a narrow temperature range, the change in radiation is linearly dependent on the change in object temperature ΔT relative to the ambient temperature T_0.

The camera lens concentrates the IR radiation associated with the object temperature on the camera detector. By heating the pixels, the concentrated IR radiation changes the microbolometer resistance, which in turn (approximately) linearly changes the pixel reading. Here, the resulted gray-scale presentation of the scene is assumed to be linearly connected to the image grayscale.

This relation allows training the model on regular VIS images and still achieving satisfactory results, even without fine-tuning of the IR images. Fine-tuning can further enhance performance due to differences in statistics between IR and VIS images, but this issue is not further explored in this work.

2.4. Computational cost

The operations performed in each layer of the network are mainly *dot products*:

$$y = w_0 \cdot x_0 + \cdots + w_n \cdot x_n \tag{6}$$

where $\underline{x}$ and $\underline{w}$ are vectors and y is a scalar. A *multiply-accumulate computation* (MAC) is defined as an operation with a single multiplication and a single addition. This means that in Eq. (6) there are n MAC operations. Note that in terms of *floating point operations* (FLOP) there are $2n - 1$ operations for a dot product.

Let f_l be the feature map of the l'th layer with size $Ch \times H \times W$ where $H \times W$ are the spatial dimensions of the feature map and Ch is the number of channels. For a **convolution layers** with K, C_{in}, C_{out} as the kernel size, number of input and output channels respectively, for each pixel in the feature map a dot-product is taken for a K^2 window across all C_{in} and the process is repeated for C_{out} channels:

$$H \times W \times K^2 \times C_{in} \times C_{out} \tag{7}$$

Meaning that a **bottleneck-layer** where $K = 1$ has:

$$H \times W \times C_{in} \times C_{out} \tag{8}$$

For **depthwise-separable convolution** the calculations for each pixel are done separately for each channel, so only C_{in} times. The resulting number of MACs is a factor of C_{out} less than for a

Optics EXPRESS

convolution layer:

$$H \times W \times K^2 \times C_{in} \tag{9}$$

In the proposed network the first and last layers are always convolution layers, but other layers can be depthwise-separable convolution. Henceforth $C_{in} \equiv C_{out} \equiv Ch$ for brevity. MACs in the initial convolution, final convolution and shuffle block, respectively, are:

$$\#\text{Conv}_{in} = H \times W \times K^2 \times 1 \times Ch \tag{10}$$

$$\#\text{Conv}_{out} = \alpha^2 \times H \times W \times K^2 \times Ch \times 1 \tag{11}$$

$$\#\text{ShuffleBlock} = \alpha^2 \times H \times W \times K^2 \times Ch^2 \tag{12}$$

where α is the upscale factor of the output. The number of MACs for L *convolution* layers with bottlenecks is:

$$\sum_{l=1}^{L} \left[H \times W \times K^2 \times Ch^2\right] + \sum_{l=1}^{L} \left[H \times W \times Ch^2\right] = H \times W \times Ch^2 \times L \times \left(K^2 + 1\right) \tag{13}$$

The number of MACs for L *depthwise-separable convolution* layers with bottlenecks is:

$$\sum_{l=1}^{L} \left[H \times W \times K^2 \times Ch\right] + \sum_{l=1}^{L} \left[H \times W \times Ch^2\right] = H \times W \times Ch^2 \times L \times \left(Ch^{-1}K^2 + 1\right) \tag{14}$$

meaning that factor between the number of MACs performed between the depthwise-separable convolution implementation and the convolution implementation is:

$$\xi = \frac{Ch^{-1}K^2 + 1}{K^2 + 1} \tag{15}$$

with ξ as a reduction factor. A comparison between different networks can be seen in Tables 1 and 2. The contribution of the bias terms and PReLU are neglected for brevity, as each adds C_{out} MACs, which are negligible.

Table 1. Results of different Cucumber datasets for an upscale factor of $\alpha = 4$.

	Method	Temperature Mean Error [C°]	PSNR [dB]	SSIM	kMACs
A	**Cucumber in the greenhouse** - 180 IR Images				
1	32Ch, 8L, Convolution	0.16	34.15	0.945	409
2	32Ch, 8L, Depthwise	0.18	33.12	0.937	333
3	16Ch, 16L, Convolution	0.18	33.00	0.937	158
4	SRCNN [11]	0.18	32.05	0.943	1746
5	VDSR [13]	0.16	34.13	0.944	11814
6	SRDenseNet [14]	0.20	32.16	0.930	12228
7	Bicubic	0.68	26.68	0.927	-
B	**Cucumber in the field** - 556 IR Images				
1	32Ch, 8L, Convolution	0.17	32.95	0.936	409
2	32Ch, 8L, Depthwise	0.19	31.93	0.928	333
3	16Ch, 16L, Convolution	0.19	31.82	0.927	158
4	SRCNN [11]	0.18	31.75	0.937	1746
5	VDSR [13]	0.17	33.05	0.935	11814
6	SRDenseNet [14]	0.20	31.22	0.922	12228
7	Bicubic	0.98	22.80	0.908	-

Optics EXPRESS

Table 2. Results of different Banana datasets for an upscale factor of $\alpha = 4$.

	Method	Temperature Mean Error [C°]	PSNR [dB]	SSIM	kMACs
C	**Banana Leaves in the greenhouse** - 6523 IR Images				
1	32Ch, 8L, Convolution	0.32	30.12	0.925	409
2	32Ch, 8L, Depthwise	0.35	29.23	0.915	333
3	16Ch, 16L, Convolution	0.36	29.17	0.913	158
4	SRCNN [11]	0.32	30.93	0.926	1746
5	VDSR [13]	0.32	30.26	0.927	11814
6	SRDenseNet [14]	0.44	27.94	0.889	12228
7	Bicubic	1.66	19.55	0.826	-
D	**Banana Leaves in the field** - 2371 IR Images				
1	32Ch, 8L, Convolution	0.16	35.10	0.943	409
2	32Ch, 8L, Depthwise	0.17	34.19	0.936	333
3	16Ch, 16L, Convolution	0.17	34.07	0.935	158
4	SRCNN [11]	0.17	33.27	0.943	1746
5	VDSR [13]	0.16	35.00	0.940	11814
6	SRDenseNet [14]	0.18	33.43	0.931	12228
7	Bicubic	1.03	24.02	0.899	-

3. Results

The proposed method was evaluated on a database composed of 9,630 outdoor IR images of four crops - cucumbers and banana leaves in the field and in a greenhouse—where performances were compared in terms of restoration temperature value, PSNR and MACs against other previously suggested state-of-the-art SR networks. Tables 1 and 2 presents the average results for the four crops - (A) Cucumbers in greenhouse, (B) Cucumbers in the field, (C) Banana leaves in greenhouse. (D) Banana leaves in the field. For convenience, the table is separated into four parts (A-D). Each with seven rows. Row 1-3 present different implementations of the proposed network. Rows 4-7 present the performances of three previously suggested SR networks: SRCNN [11], SRDenseNet [14], VDSR [13] and bicubic interpolation. The proposed network outperformed SRCNN [11], SRDenseNet [14] and bicubic interpolation, both in restoration quality and with lower MACs. Comparison to VDSR [13] shows a 28× improvement in computational requirements, with only a negligible 0.1_{dB} reduction in PSNR.

Figures 5 and 6 show results for 2× and 4× SR respectively, and a comparison between I_{LR}, bicubic, I_{SR} and VDSR [13]. Observing the figures, the proposed method indeed appears to be at the same level as the VDSR, with a significantly lower computation effort. Both methods performed better than bicubic interpolation in both appearance and metrics. In Fig. 4, we present a zoomed-in replica of Fig. 5(E) (cucumber in the greenhouse). The proposed method appears much better than VDSR [13]. This advantage will be further discussed in Section 4.

All results were obtained on a desktop computer equipped with an i7 processor.

3.1. Training details

The network was implemented using Pytorch [35]. The mini-batch size was set to 16. Each image was cropped randomly to 192×192 to create I_{HR} and then downscaled with a bicubic kernel by 2× or 4× to create I_{LR}. The training dataset was augmented with horizontal flips and 90° rotations. All image processing was done using python *PIL* image library.

All network trainable weights were initialized via the method proposed by [32], with a scaling factor of 0.1 as proposed by [36]. The network was optimized using gradient descent with Adam

Optics EXPRESS

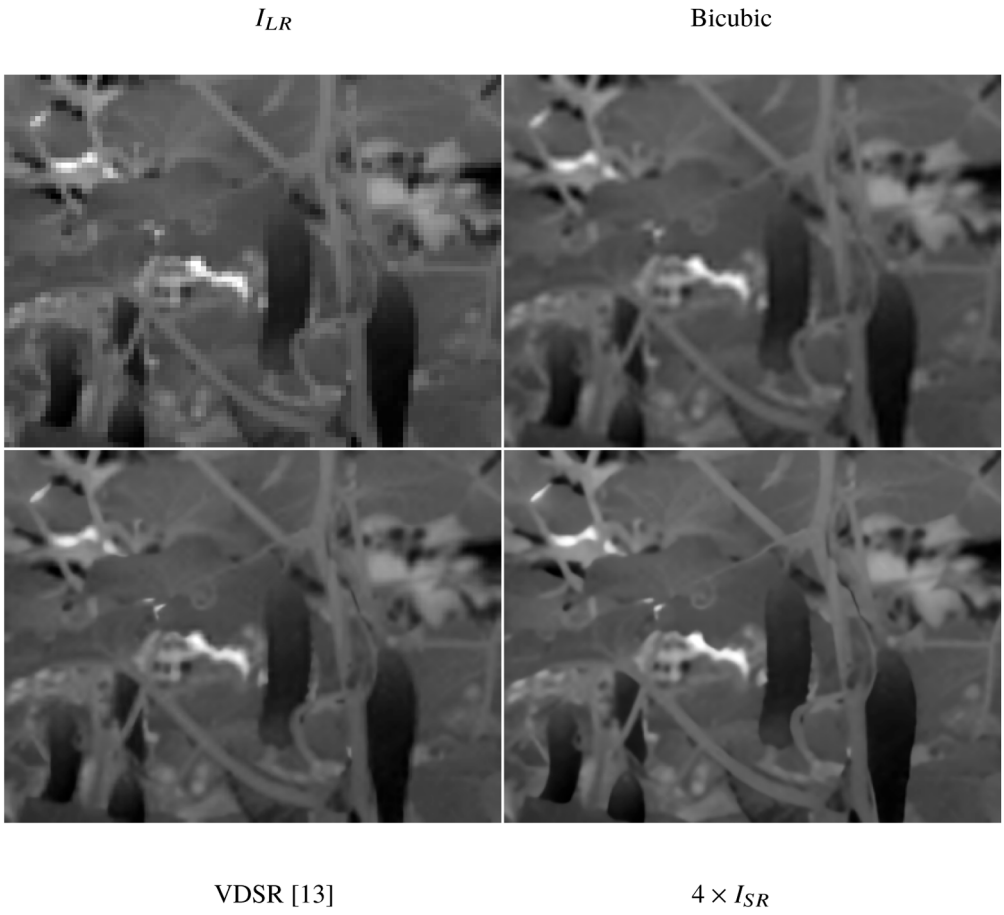


Fig. 4. Examples of 4x SR results. Typical examples taken from five different datasets are presented one below the other. From left to right: LR input, bicubic interpolation, VDSR [13] restoration and the results from the proposed method.

[37] with $\beta_1 = 0.9, \beta_2 = 0.999$ and the initial learning rate set to $5 \cdot 10^{-4}$. The learning rate was halved at 10^4 and 10^5 iterations. The training was run for $3 \cdot 10^5$ iterations.

The training was done using NVIDIA 2080ti GPU. Each permutation of the network was trained for $300k$ iterations.

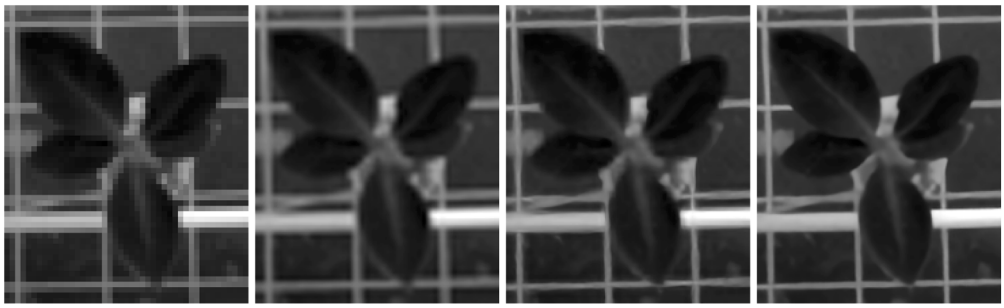
3.2. Data

The training was done on the DIV2K [38] and Flicker2K [39] datasets. The images in these datasets have a resolution of 2k, so each image contains fine details. To obtain LR images, the training set was down-sampled using bicubic interpolation. The training was done on the Y channel in the YCbCr color representation scheme, because of the proportionality between temperature and pixel intensity (Section 2.3).

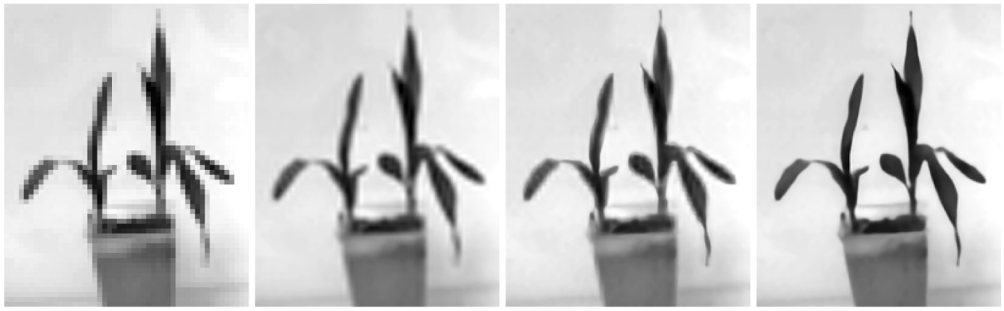
The training results were evaluated on Set5 [31] and Set14 [40]. The metrics used were PSNR and SSIM. Both metrics were calculated between the SR image (I_{SR}) and the HR image (I_{HR}) using `compare_psnr()` and `compare_ssim()` from the `skimage` library in Python. The borders of the images were each cropped by 10 pixels to neglect border effects.

Aside from these training and testing sets, several test sets of different plants were gathered using a Therma-App TH Infra Red camera at midday.

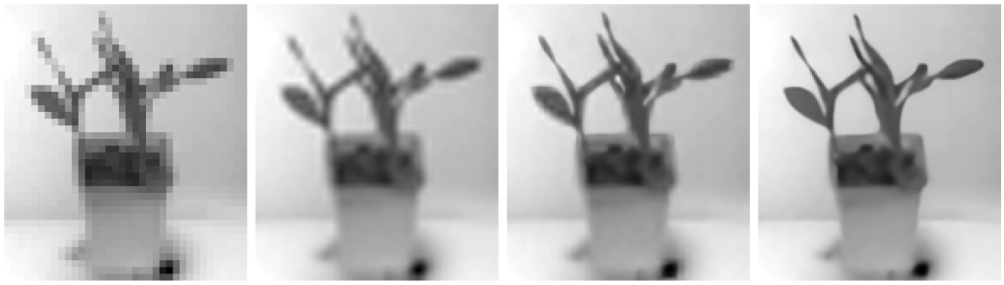
Optics EXPRESS



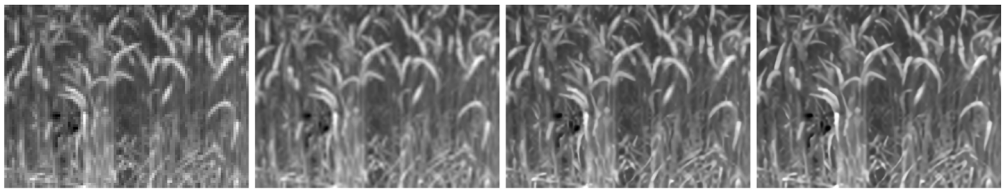
(a) Banana plant in Greenhouse.



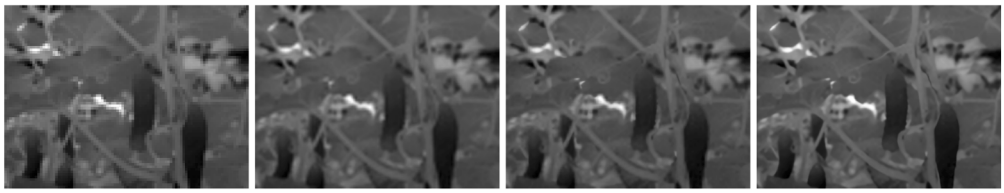
(b) Banana plant in Greenhouse - 50cm.



(c) Banana plant in Greenhouse - 90cm.



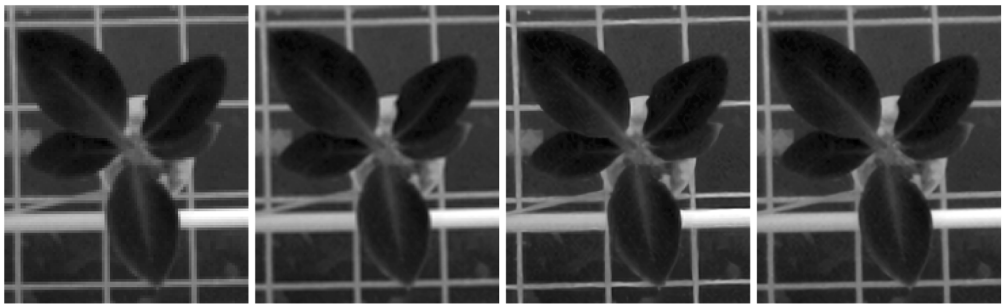
(d) Wheat in a field at midday.



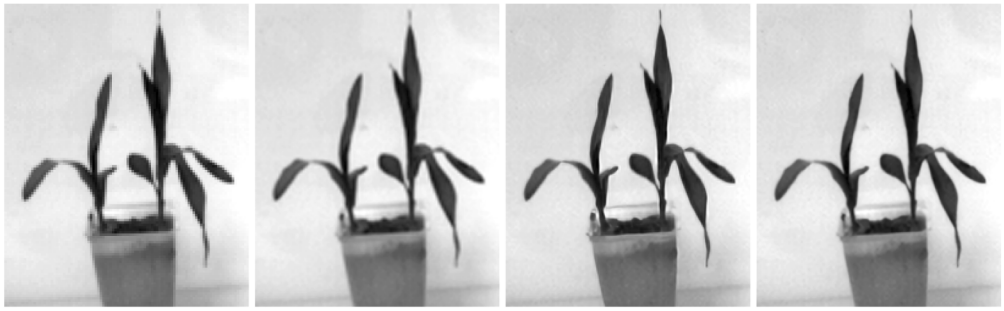
(e) Cucumber in Greenhouse.

Fig. 5. Examples of 4× SR results. Typical examples taken from five different datasets are presented one below the other. From left to right: LR input, bicubic interpolation, VDSR [13] restoration and the results from the proposed method.

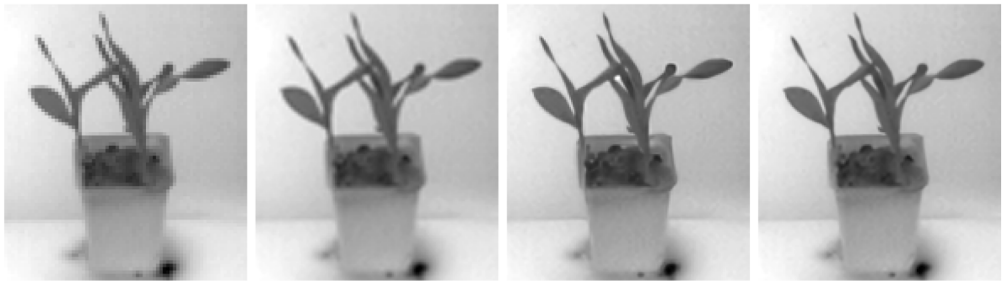
Optics EXPRESS



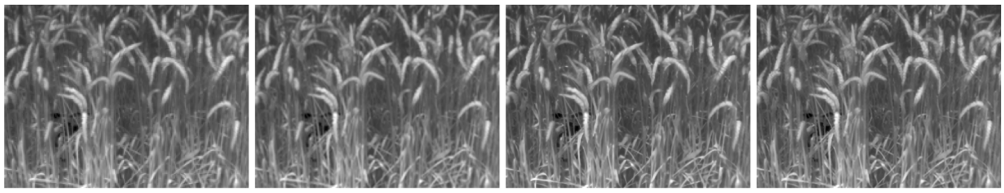
(a) Banana plant in Greenhouse.



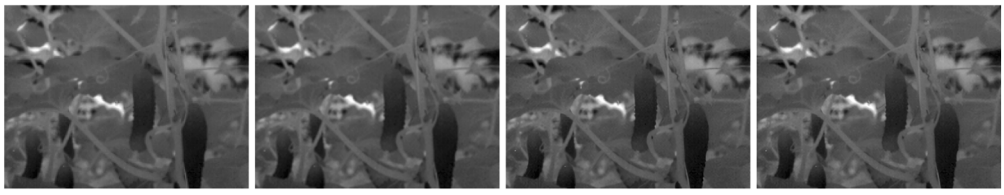
(b) Banana plant in Greenhouse - 50cm.



(c) Banana plant in Greenhouse - 90cm.



(d) Wheat in a field at midday.



(e) Cucumber in Greenhouse.

Fig. 6. Examples of 2× SR results. Typical examples taken from five different datasets are presented one below the other. From left to right: LR input, bicubic interpolation, VDSR [13] restoration, and the results from the proposed method.

Optics EXPRESS

3.3. MTF comparison

Following the above, the improvement in the MTF [41] is examined. To measure the MTF, an experiment was conducted to evaluate the Line Spread Function (LSF) of the system [42]. First, a step response was measured by imaging an aluminum sheet heated to 50°, which is held in the open room's air. The image was taken with `ThermApp TH` camera from a distance of $L = 310_{mm}$, which resulted in a sampling resolution of 0.40_{mm} in the object plane. The LSF was evaluated by a derivation of the edge between the aluminum and the air. The MTF was evaluated from the Fourier transform of the LSF. Three additional MTF curves were evaluated for comparison. Figure 7 shows the MTF curves evaluated from the LR image, the 4× bicubic interpolation, 4× SR image and the diffracted limited MTF. The edge width is composed of a few hundreds of pixels; thus, each of the first three plots is an average over an ensemble of many edge points.

The diffracted limited MTF was evaluated analytically. The camera is characterized by $F\# = 1.1$. The diffraction-limited MTF [41] of the imaging system was calculated for a circular aperture with a cutoff frequency of $\xi_{cutoff} = \frac{f}{F\#\lambda \cdot L} = 3.87 \frac{Cycles}{mm}$ with $\lambda = 10_{\mu m}$ which is the middle of the camera's sampled spectrum. The diffraction-limited MTF has the form [43]:

$$MTF\left(\frac{\xi}{\xi_{cutoff}}\right) = \frac{2}{\pi}\left[\arccos\frac{\xi}{\xi_{cutoff}} - \frac{\xi}{\xi_{cutoff}}\left(1 - \frac{\xi}{\xi_{cutoff}}^2\right)\right]^{\frac{1}{2}} \tag{16}$$

Due to a practical limit of imaging systems that stems from noise, a 5% contrast is taken as a minimal requirement for the minimal required contrast [44].

Observing Fig. 7, while the Bicubic interpolation improves the native cutoff frequency of the LR image, it's contrast magnitude over the extended range is poor, hardly exceeding 5% until a virtual cutoff at $4.5 \frac{Cycles}{mm} - 0.22_{mm}$ sampling resolution. The SR falls to 5% contrast in a sampling frequency of more than $9.5 \frac{Cycles}{mm}$ which is equivalent to 0.10_{mm} sampling resolution - i.e a true 4× improvement of the original 0.40_{mm} sampling resolution. Furthermore, the method is shown to significantly improve the diffraction-limit of the imaging system.

Optics EXPRESS

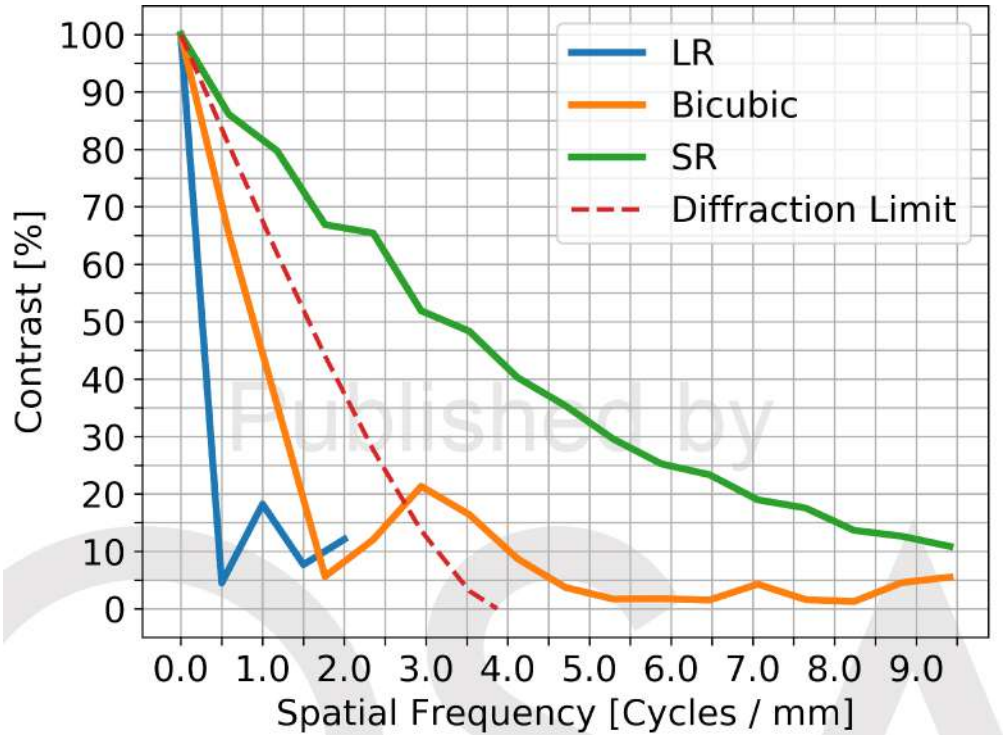


Fig. 7. MTF curves of the LR image, the effective MTF of the ×4 bicubic interpolation, the effective MTF of the 4× SR and the diffracted limited MTF of the lens. The y-axis is the MTF contrast given in %. The x-axis is the spatial frequency [cycle/mm] given in object space coordinate

4. Summary and discussion

This work was aimed at finding a suitable solution to a real-world problem - enhance plant traits resolution in low-power IR cameras. The model developed in this paper can be used with low-power devices under field-conditions, making it suitable for agricultural and environmental uses.

As seen in Tables 1 and 2, the restoration metrics of the proposed method were on par with state-of-the-art methods in terms of PSNR, SSIM and temperature estimation, while requiring 4 to 30 times less MACs.

As for the appearance of the restoration (Fig. 5 and 6), the model produces visually pleasing results, supported by the enlarged comparison between I_{LR}, bicubic interpolation, I_{SR} and VDSR [13] shown in Fig. 4. The results of the proposed method are sharper and look better than the other results, including VDSR. This is believed to be due to the propagation of features from all layers throughout the network using *bottleneck layers*. Moreover, VDSR is trained using the minimization of the L2 norm, which improves the PSNR but tends to produce blurry results.

The MTF comparison experiment shown in Section 3.3 shows a true 4× improvement in the sampling resolution compared to the LR image. Furthermore, the proposed method results in a significant advantage in the diffracted limited results and bicubic interpolation.

Thus, the method offered in this work provides a suitable solution in both quality and complexity.

Optics EXPRESS

Funding

Israeli Innovation authority; Ministry of Agriculture and Rural Development (20-12-0030).

Acknowledgments

The authors would like to thank the Israeli Ministry of Agriculture's Kandel Program for funding this research under grant no. 20-12-0030 and the Israeli Innovation authority Phenomics consortium. we would like to personally thank our partners Dr. Victor Alchanatis and Dr. Yafit Cohen of Volcani center for our joint work at the Kandel program and To Dr. Ilya Leizerson of Elbit corporation for our joint work at the Phenomics consortium.

Disclosures

The authors declare no conflicts of interest.

References

1. M. Möller, V. Alchanatis, Y. Cohen, M. Meron, J. Tsipris, A. Naor, V. Ostrovsky, M. Sprintsin, and S. Cohen, "Use of thermal and visible imagery for estimating crop water status of irrigated grapevine*," J. Exp. Bot. **58**(4), 827–838 (2006).
2. D. Bulanon, T. Burks, and V. Alchanatis, "Image fusion of visible and thermal images for fruit detection," Biosystems Eng. **103**(1), 12–22 (2009).
3. B. Berger, B. Parent, and M. Tester, "High-throughput shoot imaging to study drought responses," J. Exp. Bot. **61**(13), 3519–3528 (2010).
4. R. Bhan, R. Saxena, C. Jalwania, and S. Lomash, "Uncooled infrared microbolometer arrays and their characterisation techniques," Defence Sci. J. **59**(6), 580–589 (2009).
5. J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," *Proc IEEE Comput Vis Pattern Recognit* (2008).
6. J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," IEEE Trans. on Image Process. **19**(11), 2861–2873 (2010).
7. K. In Kim and Y. Kwon, "Single-image super-resolution using sparse regression and natural image prior," IEEE Trans. Pattern Anal. Mach. Intell. **32**(6), 1127–1133 (2010).
8. S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," IEEE Trans. on Image Process. **13**(10), 1327–1344 (2004).
9. D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *ICCV*, (2009).
10. W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," IEEE Comput. Grap. Appl. **22**(2), 56–65 (2002).
11. C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," IEEE Trans. Pattern Anal. Mach. Intell. **38**(2), 295–307 (2016).
12. Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," IEEE Trans. Neural Netw. **5**(2), 157–166 (1994).
13. J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR Oral)*, (2016).
14. T. Tong, G. Li, X. Liu, and Q. Gao, "Image super-resolution using dense skip connections," in *2017 IEEE International Conference on Computer Vision (ICCV)*, (2017), pp. 4809–4817.
15. J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* pp. 1637–1645 (2016).
16. L. Mukherjee, A. Keikhosravi, D. Bui, and K. W. Eliceiri, "Convolutional neural networks for whole slide image superresolution," Biomed. Opt. Express **9**(11), 5368–5386 (2018).
17. Z. Long, T. Wang, C. You, Z. Yang, K. Wang, and J. Liu, "Terahertz image super-resolution based on a deep convolutional neural network," Appl. Opt. **58**(10), 2731–2735 (2019).
18. Z. Niu, J. Shi, L. Sun, Y. Zhu, J. Fan, and G. Zeng, "Photon-limited face image super-resolution based on deep learning," Opt. Express **26**(18), 22773–22782 (2018).
19. E. Nehme, L. E. Weiss, T. Michaeli, and Y. Shechtman, "Deep-storm: super-resolution single-molecule microscopy by deep learning," Optica **5**(4), 458–464 (2018).
20. Z. Luo, A. Yurt, R. Stahl, A. Lambrechts, V. Reumers, D. Braeken, and L. Lagae, "Pixel super-resolution for lens-free holographic microscopy using deep learning neural networks," Opt. Express **27**(10), 13581–13595 (2019).
21. H. Zhang, C. Fang, X. Xie, Y. Yang, W. Mei, D. Jin, and P. Fei, "High-throughput, high-resolution deep learning microscopy based on registration-free generative adversarial network," Biomed. Opt. Express **10**(3), 1044–1063 (2019).

Optics EXPRESS

22. S. Chikamatsu, T. Nakaya, M. Kouda, N. Kuroki, T. Hirose, and M. Numa, "Super-resolution technique for thermography with dual-camera system," in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*, (2010), pp. 1895–1898.
23. A. Zomet and S. Peleg, "Multi-sensor super-resolution," in *Proceedings of the Sixth IEEE Workshop on Applications of Computer Vision*, (IEEE Computer Society, Washington, DC, USA, 2002), WACV '02, p. 27.
24. K. Choi, C. Kim, M. Kang, and J. B. Ra, "Resolution improvement of infrared images using visible image information," IEEE Signal Process. Lett. **18**(10), 611–614 (2011).
25. S. Dai, H. Xiang, Z. Du, and J. Liu, "Adaptive regularization of infrared image super-resolution reconstruction," in *Fifth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, (2014), pp. 1–4.
26. Z. He, S. Tang, J. Yang, Y. Cao, M. Y. Yang, and Y. Cao, "Cascaded deep networks with multiple receptive fields for infrared image super-resolution," IEEE Trans. Circuits Syst. Video Technol. **29**(8), 2310–2322 (2019).
27. T. Y. Han, Y. J. Kim, and B. C. Song, "Convolutional neural network-based infrared image super resolution under low light environment," in *2017 25th European Signal Processing Conference (EUSIPCO)*, (2017), pp. 803–807.
28. A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," (2017).
29. F. Chollet, "Xception: Deep learning with depthwise separable convolutions," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* pp. 1800–1807 (2017).
30. 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," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* pp. 1874–1883 (2016).
31. M. Bevilacqua, A. Roumy, C. Guillemot, and M. line Alberi Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding,".
32. K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," IEEE Intl. Conf. Comput. Vis. (ICCV 2015) **1502**, 1026–1034 (2015).
33. C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)* (Springer-Verlag, Berlin, Heidelberg, 2006).
34. E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," IEEE Trans. Pattern Anal. Mach. Intell. **39**(4), 3431–3440 (2017).
35. A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in PyTorch," in *NIPS Autodiff Workshop*, (2017).
36. X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, "Esrgan: Enhanced super-resolution generative adversarial networks," in *The European Conference on Computer Vision Workshops (ECCVW)*, (2018).
37. Kingma, P. Diederik, and J. Ba, "Adam: A method for stochastic optimization," (2014). Cite arxiv:1412.6980Comment: Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015.
38. E. Agustsson and R. Timofte, "Ntire 2017 challenge on single image super-resolution: Dataset and study," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, (2017).
39. R. Timofte, E. Agustsson, L. Van Gool, M.-H. Yang, L. Zhang, and B. Lim, *et al.*, "Ntire 2017 challenge on single image super-resolution: Methods and results," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, (2017).
40. R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Curves and Surfaces*, J.-D. Boissonnat, P. Chenin, A. Cohen, C. Gout, T. Lyche, M.-L. Mazure, and L. Schumaker, eds. (Springemr Berlin Heidelberg, Berlin, Heidelberg, 2012), pp. 711–730.
41. J. W. Goodman, *Introduction to Fourier optics* (Roberts and Company Publishers, 2017), 4th ed.
42. N. S. Kopeika, *A system engineering approach to imaging* (SPIE Press, Bellingham, WA, 1998).
43. G. D. Boreman, *Modulation Transfer Function in Optical and Electro-Optical Systems* (SPIE Press, Bellingham, Washington, 2001).
44. I. Klapp, A. Solodar, and I. Abdulhalim, "Tunable extended depth of field using a liquid crystal annular spatial filter," Opt. Lett. **39**(6), 1414 (2014).