ICASSP 2019:

Gradient Image Super-Resolution for Low-Resolution Image Recognition

Dewan Fahim Noor[1], *Yue Li*[1], *Zhu Li*[1], *Shuvra Bhattacharyya*[2], *George York*[3]

[1]University of Missouri-Kansas City, MO, USA
[2]University of Maryland, College Park, MD, USA
[3]US Air Force Academy, USA

Summary

❑ Problem Statement:

- Low resolution images in a variant of military and surveillance applications
- Recognition of object from a far distance e.g. Counter Unmanned Aircraft System (UAS) applications in DOD use case.
- Low resolution images captured from surveillance cameras
- Super-resolution in pixel domain is not always the right approach as producing better visual quality image might result in losing important features

❑ Solution:

- Developing recognition friendly super-resolution approach
- Proposing a convolutional neural network to perform image SR in Difference-of-Gaussian (DOG) domain for image recognition instead of pixel domain
- Adapting the trained model with SIFT to drive the subsequent key points extraction process

Gradient Image and its Use

- **Gradient image** generally refers to a change in the direction of the intensity or color of an image. In a gradient image, in a certain direction, each pixel finds out the change in intensity of that same point in the original image

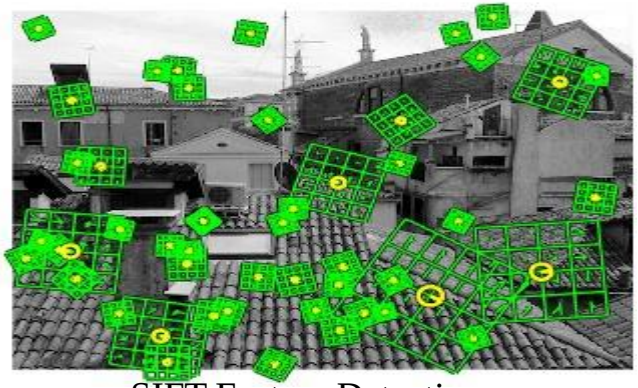
- **Harris Detector** is used to find out the edges and extract corners of the image as well as discovering the infer features of the image

- **Laplacian of Gaussian** is used for blob detection. It detects points that are continuously local maxima or minima with respect to both scale and space

- In **SIFT**, difference of Gaussian (**DoG**) is used for feature detection. From DoG images, maxima and minima are computed to find key points in SIFT detection

Harris Edge Detection

LoG Blob Detection



SIFT Feature Detection

Proposed Method Formulation

- Let , I(x,y) is the original image; G is the Gaussian Kernel,

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{\frac{-(x^2+y^2)}{2\sigma^2}} \quad (1)$$

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (2)$$

L is the function which denotes the scale space of the input image I

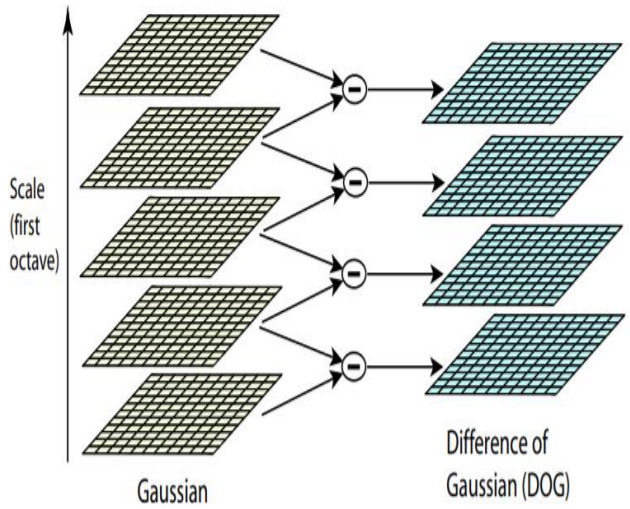


Figure : DoG in SIFT

- Therefore Difference of Gaussian will be:

$$D(x, y, \sigma_1, \sigma_2) = (G_1(x, y, \sigma_1) - G_2(x, y, \sigma_2)) * I(x, y)) \quad (3)$$

$$D(x, y, \sigma_1, \sigma_2) = L_1(x, y, \sigma_1) - L_2(x, y, \sigma_2) \quad (4)$$

- The standard deviation values , σ are 1.24 , 1.54 ,1.94 , 2.45, 3.09 for formulating 4 different DoGs

Proposed Method Formulation

- The loss function E is the MSE loss between the DoG of the super-resolved blurred generated image and the DoG from convolution with original image:

$$E(\hat{D}, D_{original}) = \sum_{i=1}^{n}\sum_{j=1}^{m}(\hat{D}^{ij} - D_{original}^{ij})^2 \qquad (5)$$

Where $\hat{D}$ is the predicted DoG image which is upscaled and $D_{original}$ is the DoG image computed from of the original one convolved with Gaussian filter.

- The gradient descent of the loss function will be:

$$\frac{\delta E}{\delta \hat{D}} = \frac{\delta(\sum_{i=1}^{n}\sum_{j=1}^{m}(\hat{D}^{ij} - D_{original}^{ij})^2)}{\delta \hat{D}} \qquad (6)$$

$$\frac{\delta E}{\delta \hat{D}} = 2\sum_{i=1}^{n}\sum_{j=1}^{m}(\hat{D}^{ij} - (\frac{1}{2\pi{\sigma_1}^2}P - \frac{1}{2\pi{\sigma_2}^2}Q)) \\ (1 - (\frac{1}{2\pi{\sigma_1}^2}\frac{\delta P}{\delta \hat{D}} - \frac{1}{2\pi{\sigma_2}^2}\frac{\delta Q}{\delta \hat{D}})) \qquad (7)$$

$$P = e^{\frac{-(x_i^2+y_j^2)}{2{\sigma_1}^2}} * I(x_i, y_j), Q = e^{\frac{-(x_i^2+y_j^2)}{2{\sigma_2}^2}} * I(x_i, y_j) \qquad (8)$$

- The simplified loss function can be written as MSE between Gaussian blurred images and computing DoG images separately.

$$E(\hat{L}, L_{original}) = \sum_{i=1}^{n}\sum_{j=1}^{m}(\hat{L}^{ij} - L_{original}^{ij})^2 \qquad (9)$$

Network Implementation

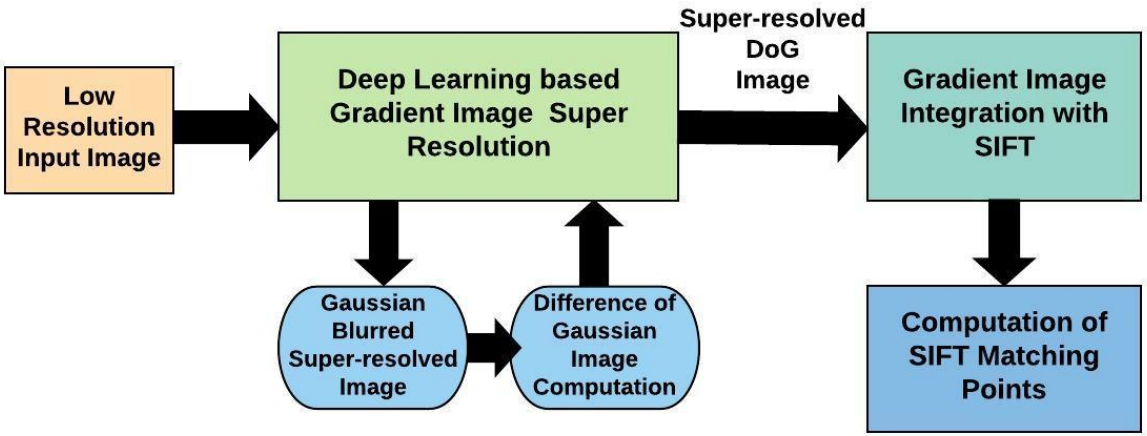


Figure: Proposed Network Architecture

- Low Resolution input images will be passed through a deep learning based gradient image super resolution stage. There are five SR networks for the purpose
- Each SR network produces a super-resolved Gaussian blurred image with different σ values [σ ={1.24 , 1.54 ,1.94 , 2.45, 3.09}]
- Four Gradient images (DoG image) are computed from five Gaussian Blurred images
- Four Gradient images are integrated to SIFT method for the computation of key matching points

Network Implementation

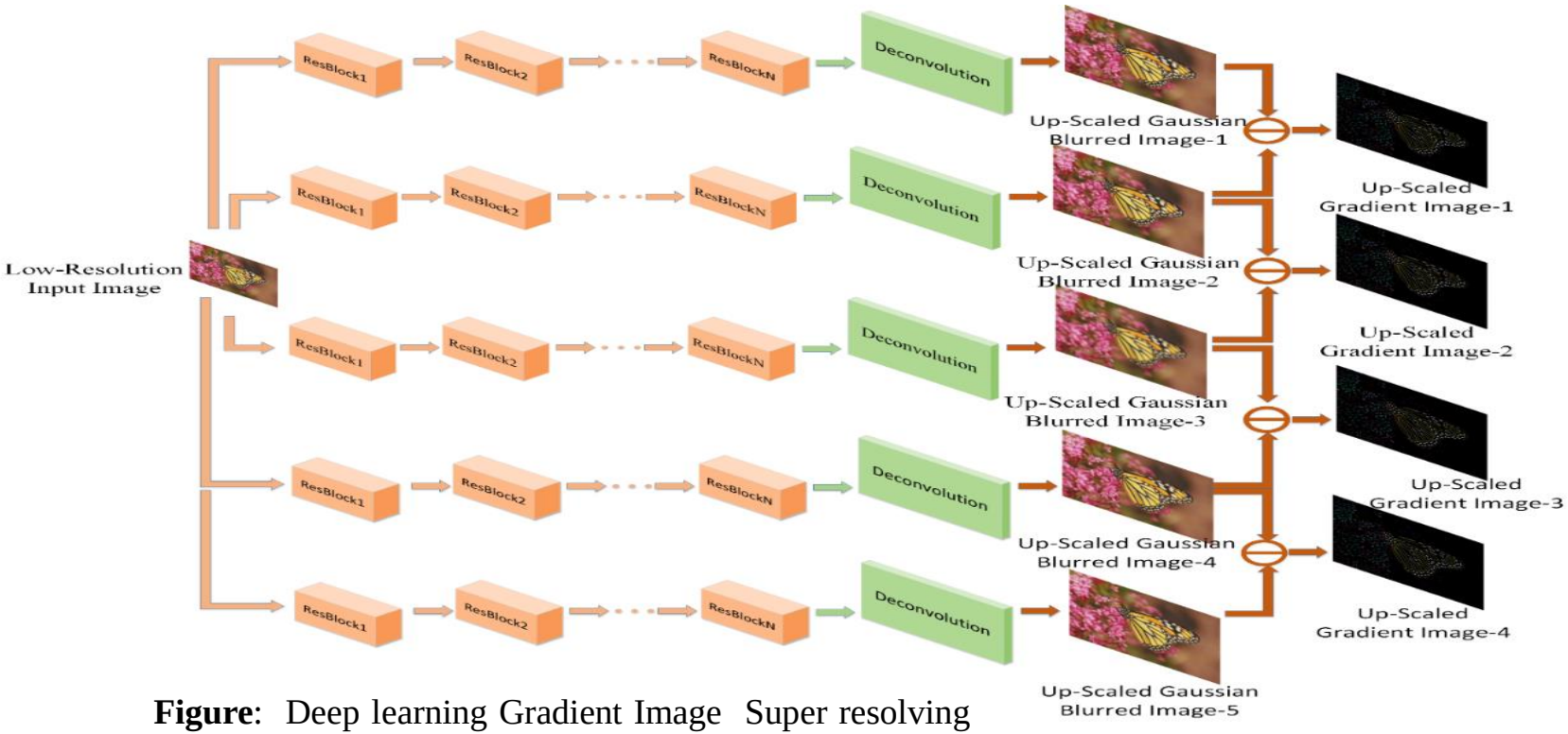


Figure: Deep learning Gradient Image Super resolving network to compute upscaled gradient image

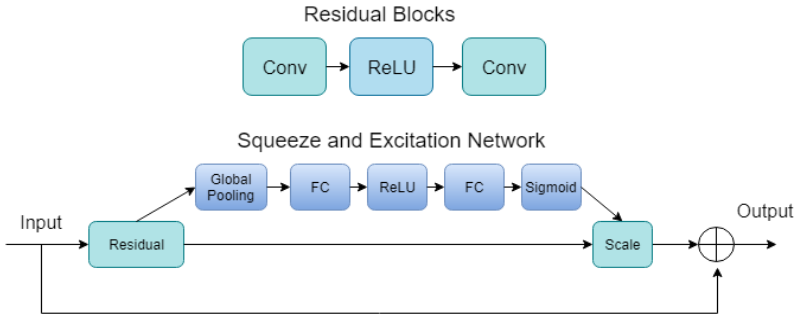


Figure: Residual Blocks

- Filter kernel size of 3X3 with 64 number of features
- Deconvolutional Layer is used to upscale.

Experimental Dataset

Training Dataset:

1. CVPR DIV_2k dataset with 800 images is used for training.
2. They are first downsampled by 2 /4 times
3. Cropped patch size:32X32.
4. Total input data 300k

Test Dataset:

1. MPEG CDVS Full dataset.
2. MPEG CDVS is a comprehensive collection of images of various objects which consists of 186k labeled images of CDs and book covers, paintings, video frames, buildings and common objects
3. 200 matching pairs from each category were chosen
4. They are first downsampled by 2 /4 times

Figure: CDVS Dataset

Results

MPEG CDVS Full dataset results:

Table 1: Average number of SIFT matching points for 200 matching image pairs from each category

Category	Upscaling Factor	Avg. no. of matching SIFT points for the original image	Avg. no. of matching SIFT points using proposed method	Avg. no. of matching SIFT points using EDSR	Avg. no. of matching SIFT points using bi-cubic interpolation
Building	2	125.8	130.4	116.3	112.4
Building	4	125.8	115.4	105.6	100.4
Graphics	2	101.6	102.8	94.5	92.8
Graphics	4	101.6	90.4	86.7	85.4
Objects	2	115.3	118.5	106.9	102.6
Objects	4	115.3	108.8	99.1	96.2
Painting	2	114.4	120.5	105.9	100.7
Painting	4	114.4	109.8	101.5	96.1
Video	2	94.3	94.4	87.2	85.2
Video	4	94.3	85.5	80.1	79.2

Results

PSNR Comparison:

Table 2: PSNR(in dB) comparison of DoG Images for 2X upscaling for CDVS full dataset

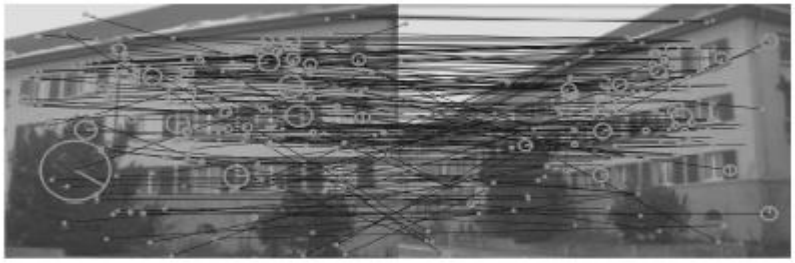
DoG(σ_1, σ_2)	Proposed method	EDSR	Bi-cubic
σ_1=1.24, σ_2=1.54	33.30	31.24	30.2
σ_1=1.54, σ_2=1.94	37.60	35.58	34.75
σ_1=1.94, σ_2=2.45	44.75	42.48	41.5
σ_1=2.45, σ_2=3.09	48.38	46.12	45.55

Table 3: PSNR(in dB) comparison of DoG Images for 4X upscaling for CDVS full dataset

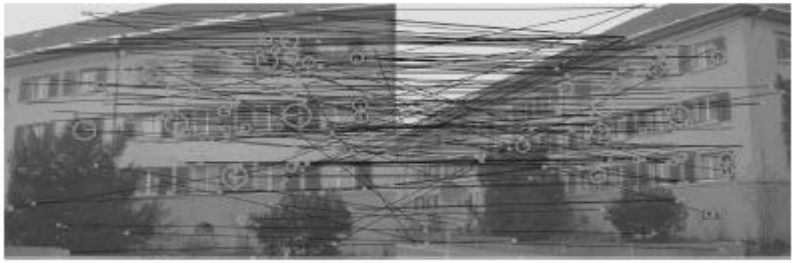
DoG(σ_1, σ_2)	Proposed method	EDSR	Bi-cubic
σ_1=1.24, σ_2=1.54	31.20	29.15	28.65
σ_1=1.54, σ_2=1.94	35.68	33.53	33.05
σ_1=1.94, σ_2=2.45	40.6	38.15	37.68
σ_1=2.45, σ_2=3.09	45.9	43.60	43.16

Comparative Results for SIFT matching points

(a) SIFT matching points for original image (102 points)



(b) SIFT matching points using proposed method (112 points)



(c) SIFT matching points using EDSR (100 points)

(d) SIFT matching points using bicubic interpolation(96 points)

Figure: SIFT Matching Points Comparison for a sample matching image pair with 2x upscaling

Conclusion

- we developed a novel gradient image super-resolution solution that opens up more degree of freedom (DoF) in the SR network design by allowing scale space adaptation in both network architecture and depth.

- Simulation results demonstrated that the SR performance in both gradient image quality and subsequent machine vision tasks like key point repeatability are improved compared with the state of art solutions in pixel domain super-resolution.

- In the future, we will develop task-specific deep neural network integration with triplet loss and softmax loss networks to drive better task level

End

Q &A