HUST

ĐẠI HỌC BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

ONE LOVE. ONE FUTURE.

ĐẠI HỌC
BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Image Deblurring on Gopro Dataset

Ngo Duy Dat - 20225480

ONE LOVE. ONE FUTURE.

Table of content

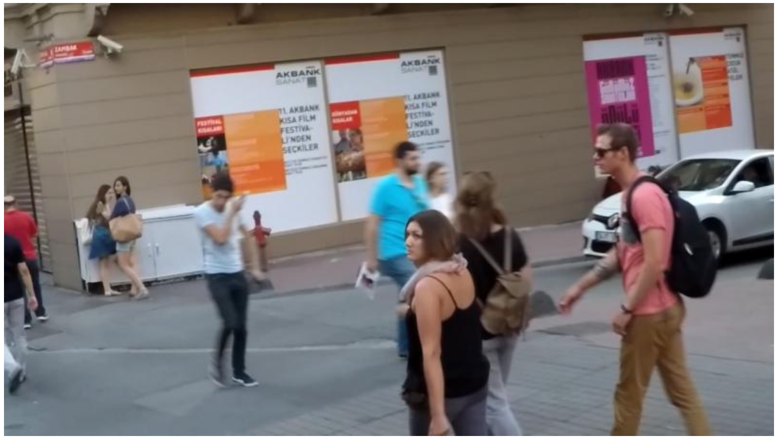
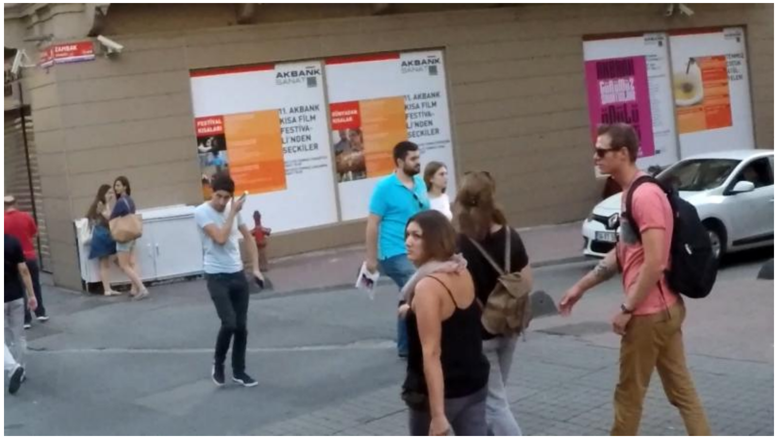
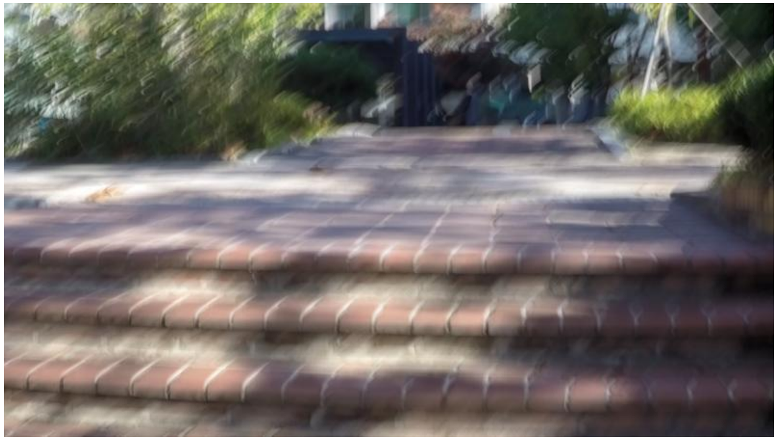
Problem Formulation

- Image deblurring is the task of restoring clear images from those blurred by camera motion or other factors.

Gopro dataset

- The dataset consists of pairs of a realistic blurry image and the corresponding ground truth sharp image that are obtained by a high-speed camera.
- 3214 pairs : 2103 for training and 1111 for tests

Gopro dataset

Blur	Sharp
	
	

Traditional Methods

Traditional methods

Some common traditional image deblurring methods

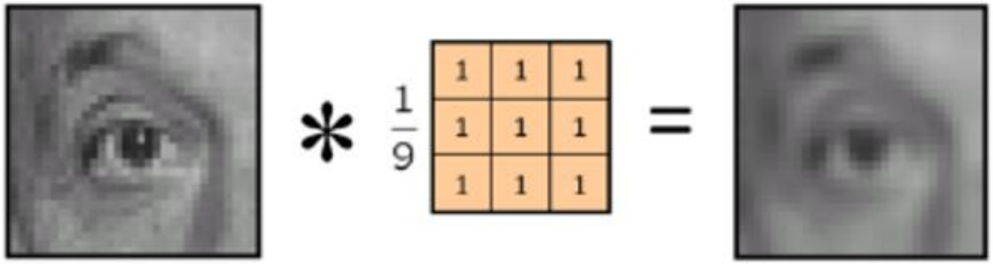
- **Wiener Filtering:** balances deblurring and noise suppression using a statistical model of the signal and noise.
- **Lucy-Richardson Deconvolution:** An iterative algorithm based on maximum likelihood estimation, assuming Poisson noise and known blur kernel.

Blur Kernel

- A blur kernel is a small matrix that represents how a single point spread out of “blurs” in an image.
- Traditional methods model image blurring explicitly as:

$$y = x * k + n$$

where y: blurred image
x: sharp image
k: blur kernel
n: noise



Original

Blur (with a mean filter)

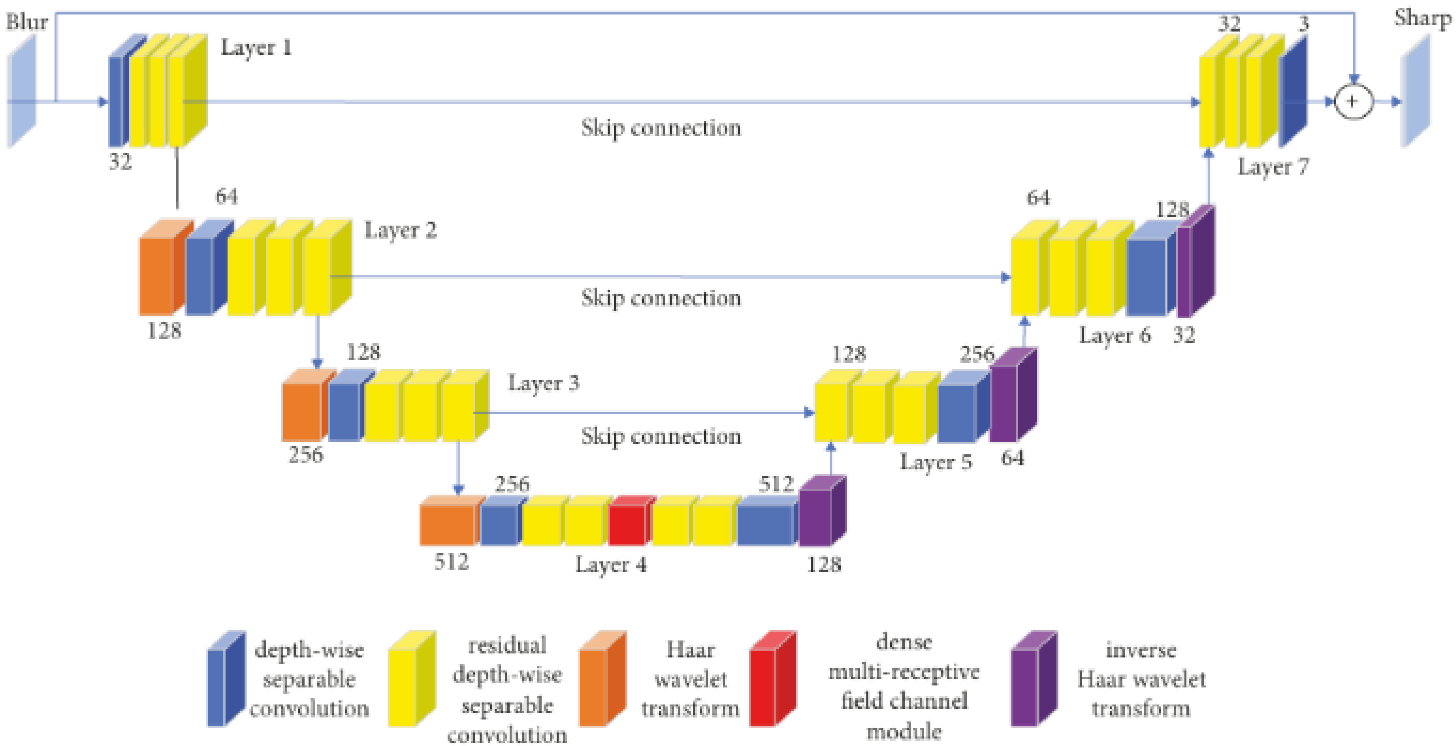
Problem with traditional methods

- Traditional methods heavily depends on known blur kernel to reverse the process of blurring image.
- This is not suitable for the task of deblurring on Gopro dataset:
  - GoPro images suffer from complex, spatially varying blur due to fast camera motion.
  - In GoPro images, different parts of the image are blurred differently.

Improved Unet model

Overall Architecture

A U-shaped architecture with an encoder, a bottleneck, and a decoder:



Haar wavelet transform

Haar wavelet transform divide image into 4 subbands to achieve different image frequency information.

a	b	a	b
c	d	c	d
a	b	a	b
c	d	c	d

$$x1 = \begin{bmatrix} a & a \\ a & a \end{bmatrix},$$

$$x2 = \begin{bmatrix} b & b \\ b & b \end{bmatrix},$$

$$x3 = \begin{bmatrix} c & c \\ c & c \end{bmatrix},$$

$$x4 = \begin{bmatrix} d & d \\ d & d \end{bmatrix},$$

$$I_{HL} = \frac{(-x1 - x2 + x3 + x4)}{2},$$

$$I_{HL} = \frac{(-x1 + x2 - x3 + x4)}{2},$$

$$I_{HH} = \frac{(x1 - x2 - x3 + x4)}{2},$$

$$I_{LL} = \frac{(x1 + x2 + x3 + x4)}{2}.$$

Inverse Haar wavelet transform

Inverse Haar wavelet transform recover the original matrix by calculating x1, x2, x3, x4.

$$x1 = \frac{(I_{LL} - I_{HL} - I_{LH} + I_{HH})}{4},$$

$$x2 = \frac{(I_{LL} - I_{HL} + I_{LH} - I_{HH})}{4},$$

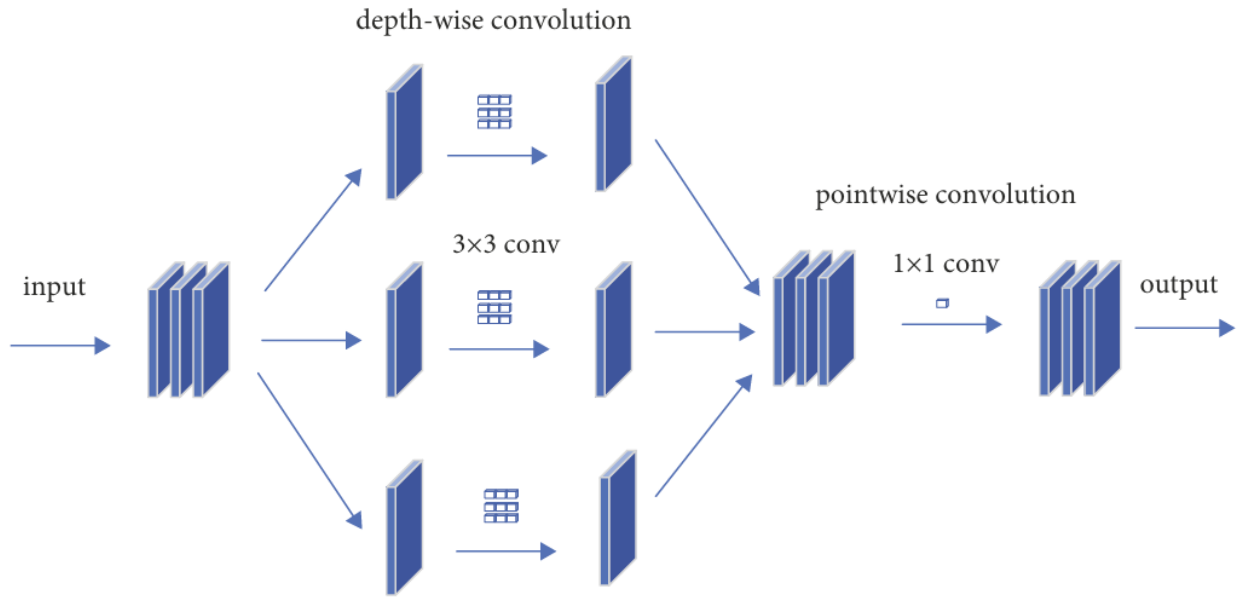
$$x3 = \frac{(I_{LL} + I_{HL} - I_{LH} - I_{HH})}{4},$$

$$x4 = \frac{(I_{LL} + I_{HL} + I_{LH} + I_{HH})}{4}.$$

Depth-wise seperable convolution (DSC)

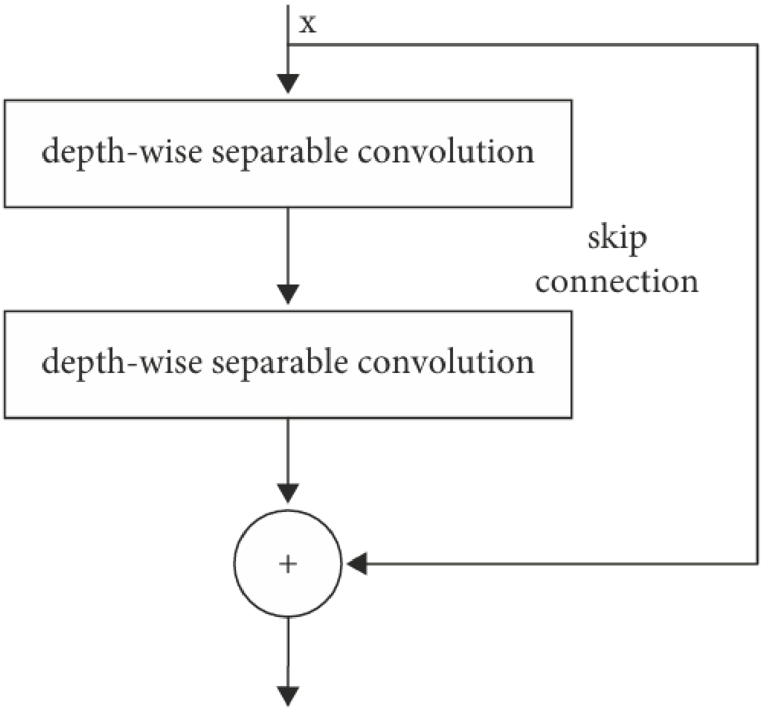
Depth-wise convolution block use depth-wise convolution instead of standard convolution, which :

- Reduce computational cost
- Allow multi-scale feature extraction



Residual depth-wise seperable convolution

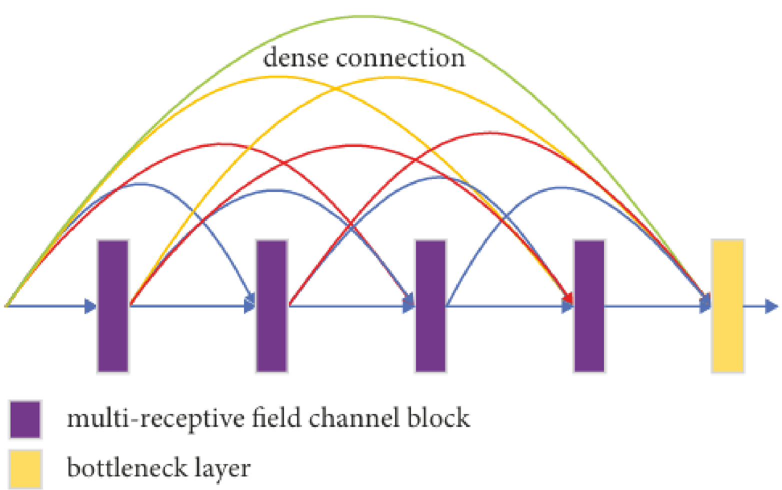
The residual DSC block consists 2 DSC blocks and addition of skip connections, which solves the problem of gradient descent.



Dense multi-receptive field channel(DMRFC) module

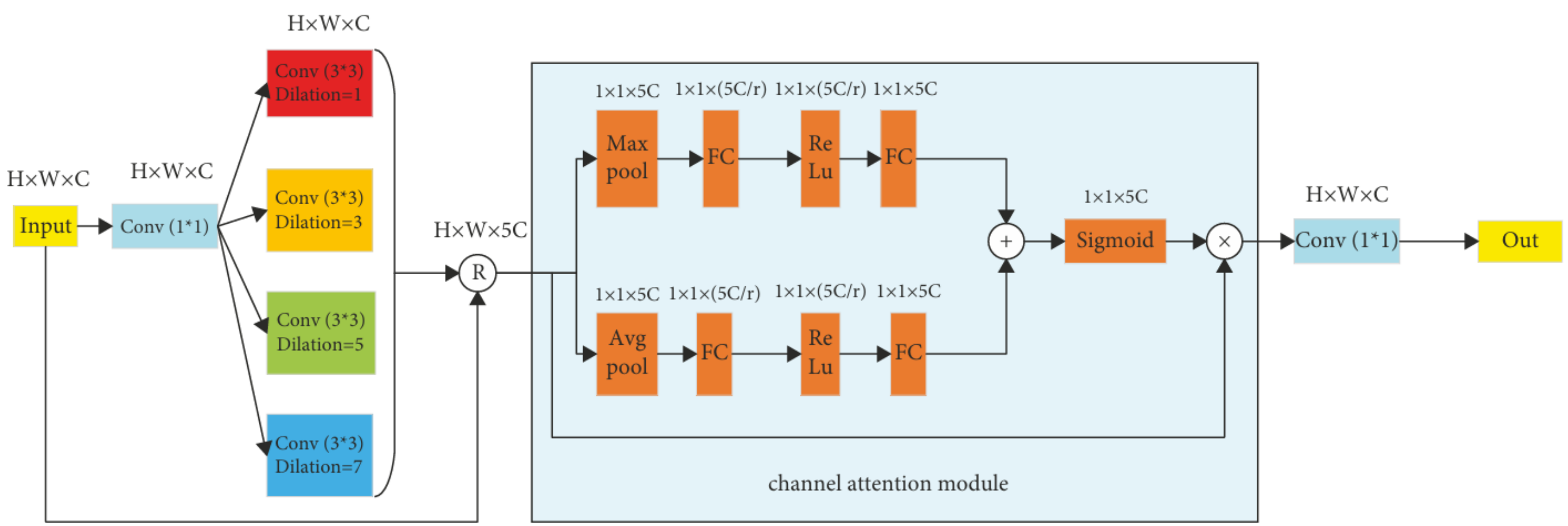
The DMRFC module is composed of four multi-receptive field channel(MRFC) blocks and a bottleneck layer:

- MRFC blocks: obtain deep sematic information from image
- Bottleneck layer: reduce number of feature for computational efficiency



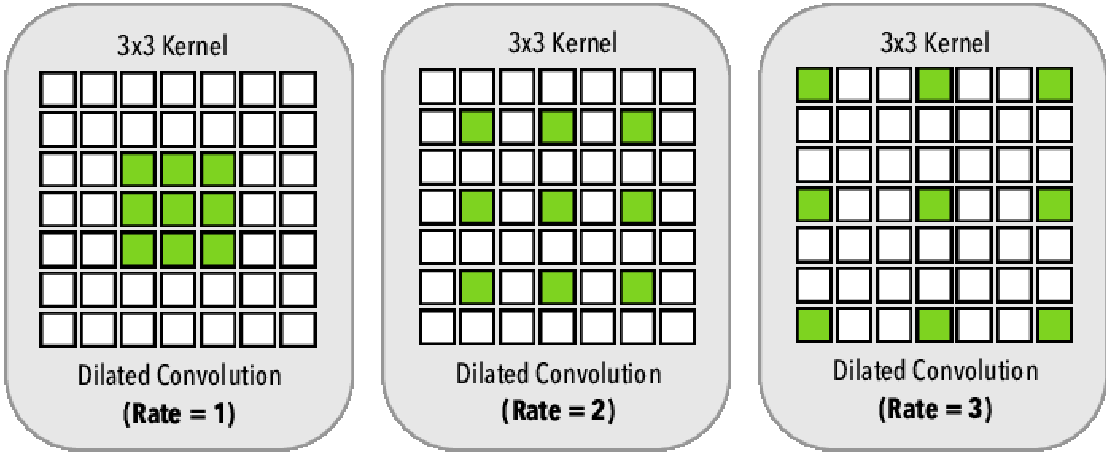
Multi-receptive field channel (MRFC) block

- Extract features in 4 paralleled branches with different dilation rate, allow simultaneously perceive features at multiple scales.
- Channel attention module capture most important feature and amplify them



Multi-receptive field channel (MRFC) block

- Different dilation rate allows capture features at multi-scale, which is essential for image deblurring.



Training Setup

Loss function

- Structural similarity index (SSIM) is a metrix to measure the similarity between 2 images

$$\mathrm{SSIM}(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)}.$$

- The loss function used here is the combination of SSIM loss and MSE loss, w1 is a hyperparameter set at 0.001.

$$L_{\mathrm{SSIM}} = 1 - \mathrm{SSIM}\,(R, S),$$

$$L_{\mathrm{total}} = L_{\mathrm{MSE}} + w_1 L_{\mathrm{SSIM}}.$$

Training detail

- The model is trained over 1000 epochs
- The learning rate initially is 1e-4, is gradually decreased until 1e-5
- The optimizer is Adam
- The model is trained on Kaggle

Result

Peak Signal-to-Noise Ratio (PSNR)

- PSNR measures the quality of a reconstructed image by comparing it to its original, perfect version

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{\mathrm{MAX}^2}{\mathrm{RMSE}^2}\right),$$

Where:

- MAX is the highest pixel value of the deblurred image
- RMSE is the root mean squared error between deblurred image and the sharp image

Result

- We evaluate the model using 2 metrics: PSNR and SSIM

	Our model	Deblur-GAN V2	Restormer
PSNR ↑	28.85	29.55	32.92
SSIM ↑	0.8688	0.932	0.961

Result


Blurred Image
Ground Truth Sharp
Predicted Deblurred
Blurred Image
Ground Truth Sharp
Predicted Deblurred
Blurred Image
Ground Truth Sharp
Predicted Deblurred

Conclusion

Conclusion

- This project has successfully implemented an Unet-based model for image deblurring task on GoPro Dataset.
- The model has achieved the result of 28.85 for PSNR and 0.8688 for SSIM.
- However, the deblurred images generated still lack sharp details and struggle with human faces.
- Further work:
  - Further training for more epochs, optimizing hyperparameters
  - Leverage transformer-based models to achieve state-of-the-art result

HUST

hust.edu.vn fb.com/dhbkhn

THANK YOU !