Learning of Unrolled Optimization with Deep Priors

Weijie Gan
weijie.gan@wustl.edu

Abstract

The report introduces a novel way to solve image restoration(IR) problem. Unrolled optimization means optimization process with fix n iterate epochs. This idea basically considers a IR problem as optimization problem and re-express it as two sub-problems. Both sub-problems then can be combined and seen as a deep-learning network. n iterate epochs can be considered as stack of n same network structure. IR problem could be solved by supervised learning method by training this whole network structure. What' s more, a similar idea called IRCNN is introduced recently. We then compare performance and thinking between the unrolled method and IRCNN method.

1 Introduction

Image restoration (IR) problem has highly practical value in low-level vision application, which basically is aimed to recover clear images from its degraded observed images. Mathematically, let us define observed image as y and clear image as x, the imaging observation can be shown as: $y = Ax + n$, where A represents degradation operation and n is the additive Gaussian noise with zero-mean and σ standard-deviation in general. IR problem is to obtain x from y based on above formulation.

To solve IR problem, one important method is called maximum posterior probability (MAP). If interpreting imaging process with Bayes Method, we get: $P(x|y) = \frac{P(y|x)P(x)}{P(y)}$. In this case, $P(x|y)$ is called posterior probability. $P(y)$ is prior information for observed image and independent from posterior probability. $P(y|x)$ is called likelihood probability, considered in imaging process as $P(n) = P(y - Ax) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-Ax)^2}{2\sigma^2}}$. And $P(x)$ is prior probability, which indicates the estimate of the probability of the estimated clear image before the current image data is observed.

To maximum posterior probability, we need to solve:

$$\begin{aligned} \max\ P(x|y) &= \max \frac{P(y|x)P(x)}{P(y)} = \max\ P(y|x)P(x) \\ &= \max\ \log\left(P(y|x)P(x)\right) \\ &= \max\ \log(P(y|x)) + \log(P(x)) \\ &= \max\ -\frac{1}{2\sqrt{\pi}\sigma}(y - Ax)^2 + \log(P(x)) \\ &= \min\ \frac{1}{2\sqrt{\pi}\sigma}(y - Ax)^2 - \log(P(x)) \end{aligned} \tag{1}$$

More generally, if we express the prior as a function $\Phi(x) = -\log(P(x))$ that is related with x and add a manual coefficient λ to balance value of constant coefficient $\frac{1}{2\sqrt{\pi}\sigma}$ in likelihood probability term, the equation (1) will become:

$$\min\ \frac{1}{2}||y - Ax||_2^2 + \lambda\Phi(x) \tag{2}$$

Different prior function leads to different results in this problem. This idea's also mentioned as image prior method. Common priors would be total variation $\Phi(x) = ||\nabla x||_1$, where ∇ is first-order derivative operation, and sparse restriction, $\Phi(x) = ||Wx||_1$, where W is domain transform operation, like Fourier transform.

With the development of deep-learning(DL) technology, more and more researcher solve IR problem by DL method. Their results also prove advance of DL method. A general objective function for DL method is: $L(\theta, x, y) = ||x - f(\theta, y)||_2$, where f means the Deep network and θ is its parameters that need to be learned. DL methods easily learn complex statistics of natural images, but lack a systematic approach to incorporating prior knowledge, which indicates A matrix in equation (1), of the image formation model.

Thus, what is missing is way to make use of the prior degradation prior knowledge in DL method.

In Steven' paper, he solve this problem by introducing a general framework, that based on LP method and combine with prior information as function(Φ). What' s more, Zhang also indicates a similar idea, IRCNN, in paper [[1]], which is discussed in the class. This class project is based on these two paper. I mainly finish following jobs: (1)I fulfilled the idea in Steven' paper and IRCNN. The code can be found in my GitHub: https://github.com/wjgancn/WashU/tree/master/cse659/proj1.(2)I introduce Steven's idea with details in this report. (3)I compare performances between Steven's idea and IRCNN.

2 Background & Related Work

Basically, one of the most effective ways to solve MAP problem in equation (2) is the variable splitting technique, which means splitting the original problem as two or three sub-problem then solve them iteratively, such as ISTA, FISTA, HQS and ADMM.

Take HQS as solution in our problem. By introducing an extra auxiliary variable z, the equation (2) can be re-express as:

$$\min \frac{1}{2}||y - Ax||_2^2 + \lambda\Phi(z) \text{ s.t } z = x \tag{3}$$

Its Lagrange function will be:

$$L(x, u) = \frac{1}{2}||y - Ax||_2^2 + \lambda\Phi(z) + \frac{u}{2}||z - x||_2^2 \tag{4}$$

where u is a manually constant. This problem can be solved with HQS via following iterative steps:

$$x_{k+1} = \arg\min_x ||y - Ax||_2^2 + u||z_k - x||_2^2 \tag{5a}$$

$$z_{k+1} = \arg\min_z \frac{u}{2}||z - x_{k+1}||_2^2 + \lambda\Phi(z) \tag{5b}$$

The first equation (5a) is a simple least-square problem and we can find its close-form solution:

$$x_{k+1} = (A^TA + uI)^{-1}(A^Ty + uz_k) \tag{6}$$

Importantly, the second equation (5b) can be seen as the a similar problem with our original problem but set $A = I$, which indicates a image denoising problem. Image denoising problem is one of the IR problem. DL method have great performance in solving this problem and denoising problem is not related with prior matrix A. So, instead of solving the equation (5b), we can construct a denoising neutral network with x_{k+1} as input:

$$z_{k+1} = f(\theta, x_{k+1}) \tag{7}$$

where f is deep network and θ are its parameters. This is the basic idea that combine DL method with MAP method.

Actually, what Zhang proposed in his/her paper [1] used the above thinking. S/He make network function f as 7-layers deep neutral network and train it for common image denoising problem by residual loss function. After obtaining this network, s/he make use of it to general IR problem by solving equation (5a) and (5b) iteratively. It separates the training of network and objective IR problem, which reminds us if we can combine them together?

As mentioned above, in general variable splitting technique, we need to solve two equations iteratively and then stop until the algorithm reaches a certain condition, like change of value x between two step. The reason we can not combine Zhang' s two step is because we can not estimate the iterative epoch before solving IR problem.

However, in some cases, we can suppose and fix the iterative step n, and stop the algorithm after n iterates. It is called Unrolled Optimization Method [2].

By this suppose, we can further show n iterative epoch as a n-layers neutral network, where each layer contain both (5a) and (5b) equation. The part in figure (1) with yellow rectangle shows this idea.

Since equation (5b) is already expressed as a denoising network, as shown as equation (7), what we need to do is to embed equation (5a) into exist denoising network and repeat this structure n times. With this idea, we can then train this whole n-layers network and successfully relate two equations in IR problem. What remains in our method is how to express the equation (6) as a part of neutral network, which will be introduced in the next part.

3 Method

In Zhang' s method, different IR problem share the same denoising network. But in unrolled method, the network is highly related with matrix A, which means different IR problem is corresponding with different neutral network.

Since equation (7) is a deep network for all IR problem, to specific the concrete structure of our neutral network, we only need to figure out the process of solving x-sub-problem shown as equation (6).

In this project, we limit the IR problem as image deblur problem. Under this condition, the equation (2) becomes:

$$\min \frac{1}{2}||y - k * x||_2^2 + \lambda\Phi(x) \tag{8}$$

where k is blur kernel and $*$ means convolution operation. Then, the equation (5a) will become:

$$x_{k+1} = \arg\min_x ||y - k * x||_2^2 + u||z_k - x||_2^2 \tag{9}$$

With the aim of computation efficiency, we convert this problem into Fourier domain as:

$$X_{k+1} = \arg\min_X ||Y - KX||_2^2 + u||Z_k - X||_2^2 \tag{10}$$

where X, Y, K and Z are Fourier transform of x, y, k and z. To set the derivation of above formulation as zero:

$$\begin{aligned} &2K^T(KX - Y) + 2u(X - Z_k) = 0 \\ &(2u + K^TK)X = K^TY + 2uZ_k \\ &X = (2u + ||K||_2^2)^{-1}(K^TY + 2uZ_k) \end{aligned} \tag{11}$$

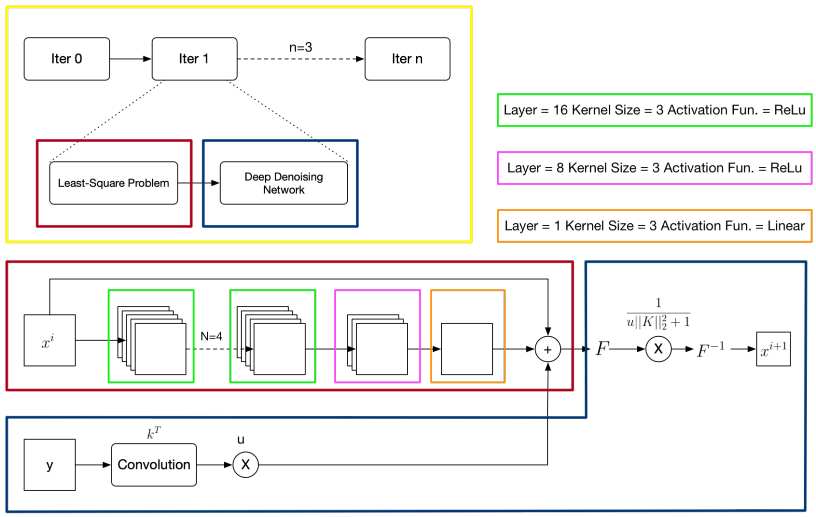

Figure 1: The Yellow Rectangle part shows the basic idea of Unrolled Optimization Method contains Solving Least-Square Problem part (Red Rectangle) and Deep Denoising part(Blue Rectangle). Beside the Yellow Rectangle, the figure also offer details structure of Unrolled Optimization Method in image deblur problem, with same color corresponding with same information

Let F as Fourier transform operation and F^{-1} as converted operation, then $K^T Y = F(k^T * y)$ and above formulation can be also seen as:

$$\begin{aligned} F(x_{k+1}) &= (2u + ||K||_2^2)^{-1}(F(k^T * y) + 2uF(z_k)) \\ F(x_{k+1}) &= (2u + ||K||_2^2)^{-1}F((k^T * y) + 2uz_k) \\ x_{k+1} &= F^{-1}((2u + ||K||_2^2)^{-1}F((k^T * y) + 2uz_k)) \end{aligned} \tag{12}$$

The equation can be expressed as a part of neutral network, shown as Blue rectangle part in figure (1).

4 Experimental Results

In this part, I show experiment result from both proposed unrolled deep network and IRCNN.

As mentioned above, the unrolled deep network is trained for image deblur problem. The concrete structure is shown as Red Rectangle and Blue Rectangle part in figure (1). I also implement the IRCNN method, with deep denoising network shown as figure (2). Those network are trained with mean-square error loss function and Adam optimizer.

In data preparation phase, I use both train, validation and test data from BSDS300[3]. I treat is as ground-truth image. In order to obtaining input images, I firstly blur

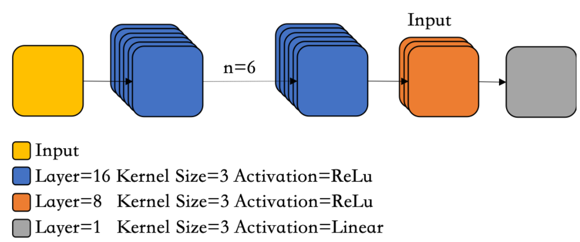

Figure 2: Deep Denoising Network in IRCNN

these image, with blur kernel shown as follow:

$$k = \frac{1}{256}\begin{bmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix}$$

Then, I add Gaussian distribution noise with mean as 0 and two kind of standard deviation, 0.01 as low noise and 0.03 high noise. So, I have two kind of input data. I do the same experiments with those two different noise level data. To estimate the quality of reconstructed images, we use both value of PSNR and SSIM.

The training plots are shown as figure (3). From training process, we notices the training process is stable and there is no over-fitting in this network, since both training data and validation data have same trend in performance.

Image results are introduced in figure(4).

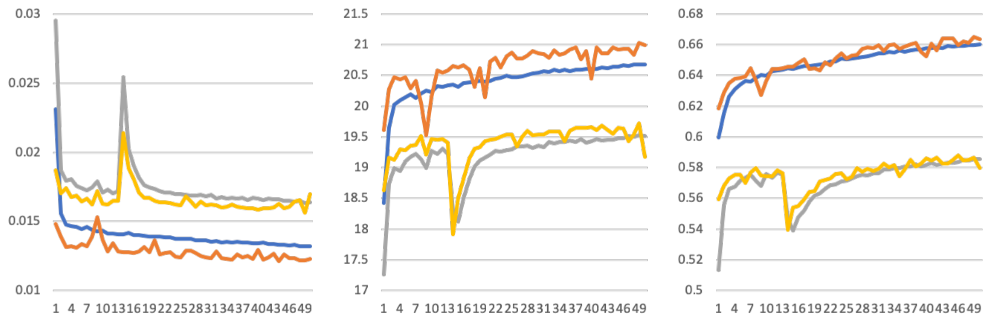

Figure 3: Training Plot. From Left to Right is value of loss function, PSNR and SSIM. For all three plots, the x-axis is iterate epochs. The BLUE line indicates training data in low noise. The ORANGE line means validating data in low noise. The GRAY line is training data in high noise. The YELLOW line shows validating data in high noise.

	PSNR	SSIM
Proposed(Low Noise)	22.2605	0.7250
Proposed(High Noise)	20.5896	0.6588
IRCNN(Low Noise)	22.7452	0.7133
IRCNN(High Noise)	16.7155	0.5904

Table 1: Average PSNR and SSIM in test data

From image results, we found that both proposed method and IRCNN have good performance in low level noise data. And the reconstructed images from proposed method are more smooth, such as the face in Human image and river surface in River image. Due to that, the PSNR value of IRCNN in low noise is better than proposed method. However, we find that proposed method have better performance in high noise. Though PSNR value is different, the reconstructed images from proposed method in both low noise and high noise look similar. And the IRCNN can not recover image from high noise condition, where images are lack of fidelity and no enough details, such as trees in Building image and human in River image.

Including the three images shown in figure (4), I test two method in totally 70 images within dataset. I compute the average value of PSNR and SSIM. The result is shown as table (1). From these results, we find that proposed method is better than IRCNN in SSIM value and PSNR in high noise, but not PSNR in the low noise.

5 Conclusion

The basic idea of both unrolled method and IRCNN is to treat the MAP optimization problem as two sub-problem using HQS method. Then, in IRCNN, one of sub-problem can be solved by a trained network. The main different between unrolled method and IRCNN is that unrolled method combine deep learning network, which is used to solve one of sub-problem, with the other sub-problem and train the whole structure together, which means on the other word, it consider imaging process equation in network-training phase. Despite that, IRCNN train the network first, solve two sub-problem iteratively and does not consider the degrade process in network training.

From experiment results, IRCNN have better performance in low noise data while it fails in high level noise. And unrolled method have great performance in both low and high noise. Especially, unrolled method have a better SSIM value than IRCNN.

References

[1] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. Learning Deep CNN Denoiser Prior for Image Restoration. *arXiv:1704.03264 [cs]*, April 2017. arXiv: 1704.03264.

[2] Steven Diamond, Vincent Sitzmann, Felix Heide, and Gordon Wetzstein. Unrolled Optimization with Deep Priors. *arXiv:1705.08041 [cs]*, May 2017. arXiv: 1705.08041.

[3] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics. In *Proc. 8th Int'l Conf. Computer Vision*, volume 2, pages 416–423, July 2001.

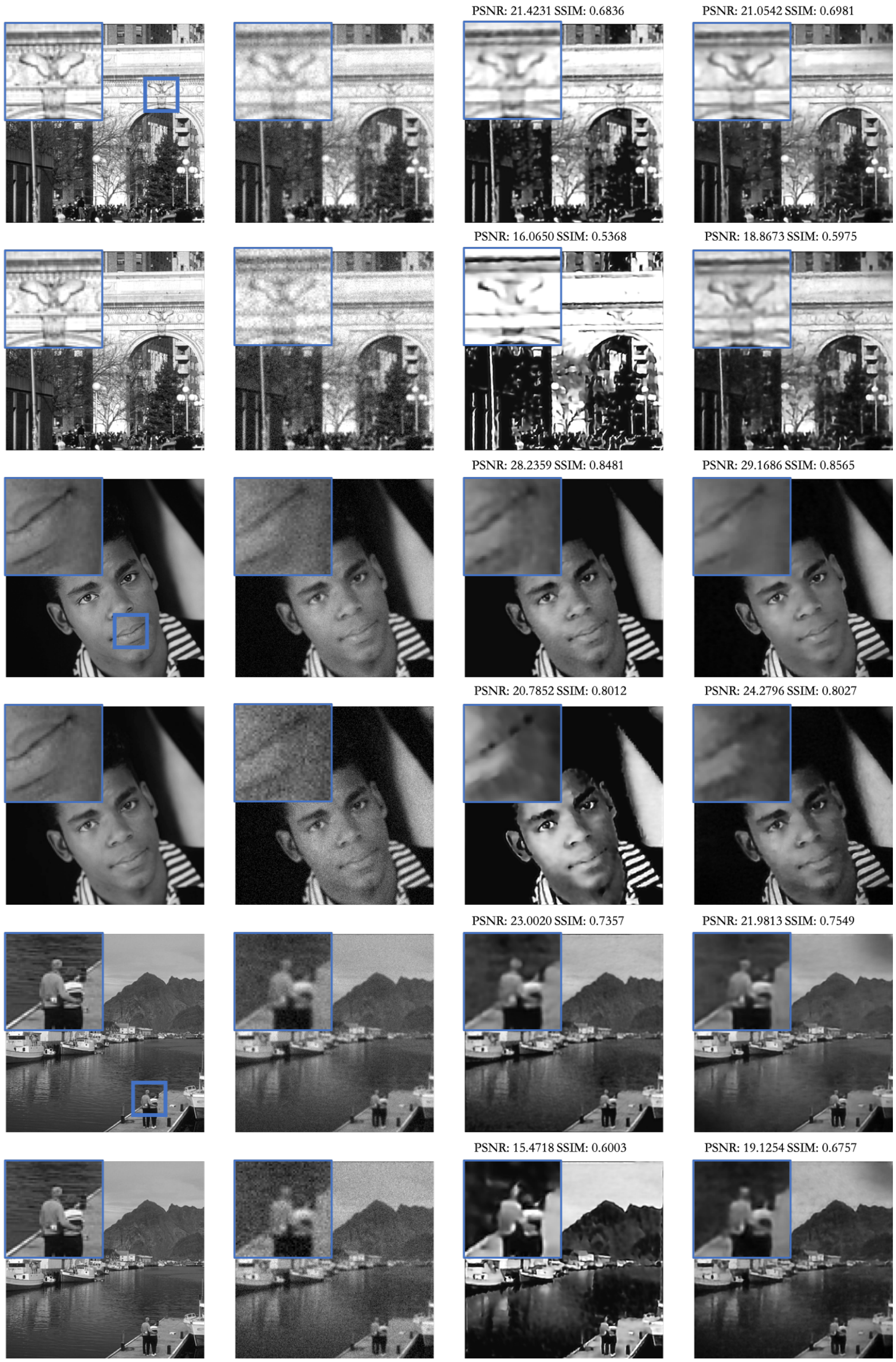

Figure 4: Three experiment images. In each image, the UP four images is computed under low noise with standard derivation as 0.001, while high noise with standard derivation as 0.003 in the DOWN four images. In up or down image, from left to right, images are ground-truth, noise image, results from IRCNN and results from proposed method(Unrolled Optimization Method). From up to bottom, the images are called as Building, Human and River.