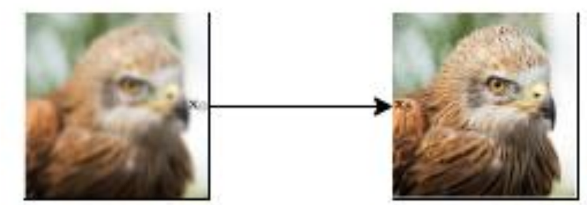
Deep Unfolding Network for Image Super-Resolution

Kai Zhang Luc Van Gool Radu Timofte Computer Vision Lab, ETH Zurich, Switzerland

Review by Abhishek Soni, Roll No. 180100004

1. Problem Statement

Single image super-resolution is the process of recovering the natural and sharp detailed high resolution (HR) counterpart from a low-resolution (LR) image. It is an underdetermined inverse problem, for which no unique solution is there. This problem can be typically resolved by limiting the solution space along with providing strong prior information. This



has a wide range of applications for real-world problems, such as enhancing the image visual quality on high-definition displays and improving the performance of other high-level vision tasks.

2. Previous Work

Previous works was based on Model based methods(interpolation) and Learning based methods (Neural networks). Both has its own limitations.

2.1 Degradation model

Previous work in this domain assumed it as the problem of image interpolation. Their degradation model assumes the Low-Resolution image is directly down sampled from the High-Resolution image without blurring.

2.2 Flexible Single Image Super Resolution methods

Despite CNN-based SISR methods got impressive results in handling bicubic degradation, using them with other more practical degradation models is quite cumbersome. Considering practical aspects, designing a flexible super-resolver that can handle considering factors like *scale factor, blur kernel* and *noise level*.

Some methods tackled bicubic degradation with different scale factors via a single model, like as **LapSR** with progressive up sampling, **MDSR** with scales-specific branches, **Meta-SR** with meta-upscale module. But these are limited to Gaussian blur kernels. The **deep plug and play** model, is most flexible CNN which handled the blur kernels, scale factors and noise levels, but suffered from high computational burden.

2.3 Deep unfolding image restoration

Compared to plain learning-based methods, deep unfolding methods are interpretable and can fuse the degradation constraint into the learning model. Most Learning based models have drawbacks as: (i) The solution of the prior subproblem without using a deep CNN is not powerful enough for good performance. (ii) The convergence may be hindered as the data subproblem is not solved by a closed-form solution. (iii) The model is trained end to end as opposed to the stage-wise and fine-tuning method.

No such deep unfolding SISR exists which can simultaneously handle the classical degradation problem.

3. Motivation

Taking advantage of end-to-end training, Learning based Single Image Super Resolution (SISR) methods are getting attention due to superior effectiveness and efficiency as compared to model based methods.

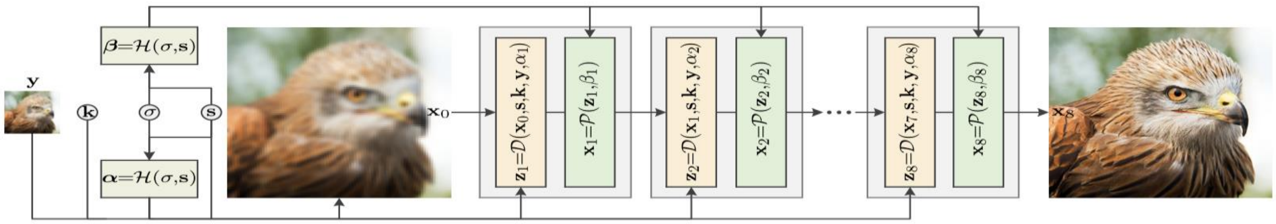
But these learning based methods, lack flexibility as compared to model-based methods which can handle the SISR problem with different scale factors, blur kernels and noise levels under a unified MAP (maximum a posteriori) framework. There's still not enough work is done on *a single end-to-end trained deep model* to obtain back High resolution image from Low resolution image.

To address this problem, an end-to-end trainable unfolding network which combines both learning based methods and model-based methods can be proposed.

4. Current Method

The proposed method focuses on non-blind SISR which requires some quantities beforehand, ie the Low-Resolution image, blur kernel and noise level to be available beforehand.

By unfolding the MAP inference via a half-quadratic splitting algorithm, a fixed number of iterations consisting of alternately solving a data subproblem and a prior subproblem can be obtained and by neural modules the two problems then can be solved resulting in an end-to-end trainable, iterative network. So just via single model, the devised current network has the flexibility of model-based methods to super-resolve blurry, noisy images for different scale factors.



4.1 Degradation model

This paper adopts the data-driven method to solve the following kernel estimation problem by minimizing the reconstruction error over a large HR/bicubic-LR pairs {(x, y)}. The downscaling operation selects the upper-left pixel for each distinct patch.

4.2 Unfolding optimization

The High-Resolution image estimation is done by minimizing the following energy function:

$$E_\mu(\mathbf{x}, \mathbf{z}) = \frac{1}{2\sigma^2}\|\mathbf{y} - (\mathbf{z} \otimes \mathbf{k})\downarrow_{\mathbf{s}}\|^2 + \lambda\Phi(\mathbf{x}) + \frac{\mu}{2}\|\mathbf{z} - \mathbf{x}\|^2$$

4.3 Deep unfolding network

After defining the unfolding optimization, we proceed to the next step of designing the unfolding super-resolution network (***USRNet***)

4.3.1 Data module D

Data Module aims to find a clearer High Resolution image which minimizes a weighted combination of two terms data term: (i) $||y - (z \otimes k) \downarrow_s||^2$ and (ii) $||z - x_{k\text{-}1}||^2$ the quadratic regularization term with trade-off hyper-parameter α_k.
$z_k = D(x_{k-1}, s, k, y, \alpha_k)$

4.3.2 Prior module P

The prior module receives the Z_k inputs, and along with a denoiser with noise level β_k it aims to obtain a cleaner HR image x_k.

$$x_k = P(z_k, \beta_k)$$

4.3.3 Hyper-parameter module H

The hyper-parameter module controls the outputs of the data module and prior module by tweaking the parameters α, β

$$[\alpha, \beta] = H(\sigma, s)$$

The hyper-parameter *H* module architecture consists of three fully connected layers where *ReLU* activation for the first two layers and *Softplus* activation for the last layer.

5. Implementation

The devised model is tested along with other models on the widely-used color ***BSD68*** dataset which basically consists of 68 images with tiny structures and fine textures and thus is challenging to improve the quantitative metrics, such as PSNR to quantitatively evaluate different methods. To avoid this, only 12 representative kernels were considered which includes 4 isotropic Gaussian kernels with different widths, 4 anisotropic Gaussian kernels from and 4 motion blur kernels.

5.1 Comparison With Other Models

a. Current model (***USRNet***) outperforms significantly as compared with others on different scale factors, blur kernels and noise levels with much fewer iterations it has at least an average PSNR gain of 1dB over IRCNN with 30 iterations due to the end-to-end training.
b. Although **RCAN** model achieved good performance on the bicubic degradation but it performed very bad when the degradation deviates from bicubic.
c. ***ZSSR*** performs well on both isotropic and anisotropic Gaussian blur kernels for small scale factors but not on motion blur kernel and large-scale factors.
d. ***IKC*** does not generalize well to anisotropic Gaussian kernels and motion kernels.

Conclusions

The devised method shows good generalizability (in Fig. below) by to decoupling of the data term and the prior term. The most notable achievement of this devised network is that this single model can handle the classical degradation model.

Zoomed LR (×3) USRNet

So basically this proposed network have two modules for HR estimation clearer and cleaner and one to control their behavior. As a result, both degradation and prior constrain can be imposed on the solution by this single model.

Limitations

This method focuses only on non-blind SISR which assumes the Low-Resolution image, blur kernel and noise level are known beforehand, which limits the usability of this model.

Scope for Improvements

Non-blind Single Image Super Resolution can be an intermediate step towards solving blind SISR, which can help a revolutionary development in the field of image restoration.