Probabilistic Approaches to Image Super Resolution

Group Members:
1. Yosua Muliawan - E0674549 / A0228565W
2. Tran Khanh Hung- E0503532 / A0212253W
3. Yong Zhun Hung - E0002807 / A0138922B
4. Prudhvi Kumar Daruvuri - E0680022 / A0229589H
5. Rashi Sharma - E0674476 / A0228492X
6. John Patrick Eala - E0674533 / A0228549R
7. Cheuk Ki Fung - E0674535/A0228551E

1 Introduction

The recovery of high resolution (HR) images from low resolution (LR) images has many practical applications. As an example, automatically inferring HR images from their LR counterparts helps with overcoming limitations of less-capable sensors. This task finds numerous practical applications, ranging from surveillance, medical imaging, and aids in other computer-vision tasks, such as scene recognition and object detection. Deep learning methods currently dominate benchmarks in this area, but are lacking in efficiency and interpretability [3]. As such, we are hoping to contribute along these two main axes by applying Bayesian based methods to the problem.

2 Purpose

We aim to advance image resolution by experimenting using Bayesian techniques and benchmark it against different various image resolution techniques. Methods such as bicubic interpolation and likelihood formulated with a Gaussian distribution were chosen. We believe the definition of the prior model will play a crucial role for estimating HR image quality. We model prior by using pairwise and high-order Markov Random Field (MRF), then evaluate our methods using Peak Signal to Noise Ratio (PSNR) for benchmarking.

3 Problem Statement

Image super resolution (SR) task involves obtaining a high-resolution image from low-resolution image, by upscaling it N times. Single Image Super Resolution in which a higher resolution image is estimated from a lower resolution image is an ill posed problem because the number of unknowns m to be estimated are greater than number of known n. As a result, multiple high-resolution images can be estimated from the same low-resolution image. Use of a priori knowledge can reduce the ill-posedness of this problem. In this project, we address the task of estimating high-resolution images from low-resolution images by the use of Bayesian techniques.

3.1 Formulation

Mathematically the LR image can be represented in terms of convolution of HR image with a low-pass filter and downsampling as:

$$y = DHx + E$$

x: *vector representation of HR Image*

y: *vector representation of LR Image*
H: *matrix for blurring process*
D: *matrix for downsampling the image*
E: *noise term*

We intend to use the above mathematical formulation to estimate HR images by using the following probabilistic model:

$$P(x|y;\theta) = \frac{P(y|x;\theta)P(x)}{P(y)}$$

x: *HR Image*
y: *LR Image*
θ: *Parameters*

4 Literature Review

4.1 Recent Approaches for Image Super Resolution

In recent years, there have been many contributions and a lot of progress in Super resolution of Images using Deep Neural Networks (DNN). The DNN acts as an upscaling function that is trained in a fully supervised manner with LR as input and corresponding HR as output. DNNs learn abstract feature representations in the input image that allow some degree of disambiguation of the fine detail in the HR output.

1. [4] Y Wang et-al, presented a Fully Progressive Approach to Single-Image Super-Resolution method (ProSR) that is progressive both in architecture and training: the network up-samples an image in intermediate steps, while the learning process is organized from easy to hard, as is done in curriculum learning. A form of curriculum learning was used which not only increased the performance for all scales but also reduced the total training time.
2. [5] Chao Dong, Chen Change Loy et-al, proposed a deep learning method for single image super-resolution (SR). This method directly learns an end-to-end mapping between the low/high-resolution images. The mapping is represented as a deep convolutional neural network (CNN) that takes the low-resolution image as the input and outputs the high-resolution one. The deep CNN used in their method has a lightweight structure which demonstrates state-of-the-art restoration quality, and achieves fast speed for practical purposes.
3. [8] Dmitry Ulyanov et. al, proposed the method of deep image prior in order to get high resolution image from low resolution image without training neural networks on a dataset on low- and high-resolution image dataset. They try to bridge the gap between handcrafted priors and pre-trained network based methods. They make use of randomly initialized convnet for each low resolution image and the structure of the convnet is used as the image prior. Like bicubic upsampling, the method used by them does not require prior learning on data and is able to produce high resolution image with cleaner sharper edges at par with state-of-the-art super resolution methods that use ConvNets.

4.2 Bayesian Approaches

The deep-learning approaches discussed above, despite being successful in the image super resolution task, are hard to comprehend because of the black-box nature of the CNNs they use. The following approaches utilize Bayesian modelling techniques and Markov Random Fields (MRF) to perform image-restoration and super resolution. The first approach takes advantage of the generative aspect of MRFs to examine the image priors [1]. It achieves higher performance by utilizing a Gibbs sampler and a general class of MRFs with flexible potentials. Image restoration is done by getting the Bayesian minimum mean squared error and shows promising results with respect to other discriminative approaches.

Another approach is the Generative Bayesian Image Super Resolution with Natural Image Prior [3]. It presents a high-order MRF as the prior for natural images. It also addresses the problem of other Bayesian modeling techniques that only compute for the posterior mode. To obtain the high resolution image, a Markov Chain Monte Carlo-based sampling algorithm is used in computing for the MMSE. This approach benefits from having a flexible prior and computing for the posterior mean which makes it robust to the problems posed by local minima in MAP estimation [3]. Both of these discussed probabilistic approaches present a method that can achieve comparable performance to that of the deep-learning methods while providing efficiency and interpretability.

5 Methodology

5.1 Project pipeline

This project utilizes a comparative approach in analyzing different methods for SR. As such, the proposed project pipeline is as shown below:

1. Training the chosen prior.
2. Establish posterior probabilistic distribution.
3. Estimating HR images from LR images by using different techniques such as maximum a posteriori (MAP) and sampling from Bayesian posterior.
4. Benchmark results with traditional methods (bicubic interpolation)

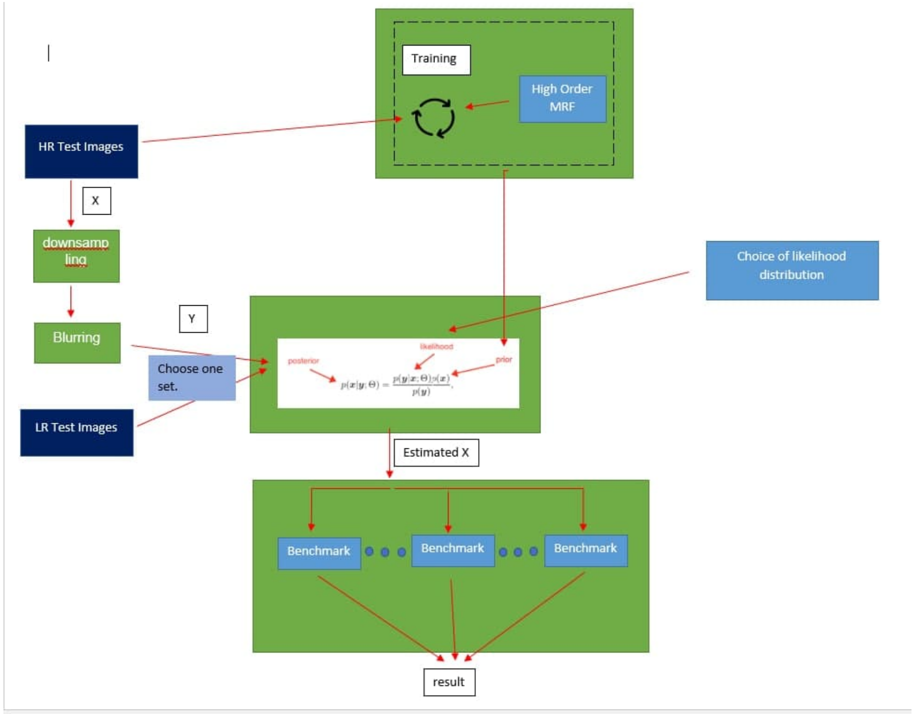
Training
High Order MRF
HR Test Images
X
downsampling
Blurring
Y
Choose one set.
LR Test Images
Choice of likelihood distribution
likelihood
posterior
prior
$p(\boldsymbol{x}|\boldsymbol{y};\Theta) = \frac{p(\boldsymbol{y}|\boldsymbol{x};\Theta)p(\boldsymbol{x})}{p(\boldsymbol{y})}$,
Estimated X
Benchmark
Benchmark
Benchmark
result

Figure 1: Pipeline

5.2 Likelihood

The noise term can usually be represented by a Gaussian distribution. By assuming that the samples of the noise term are independent and identically distributed (i.i.d) and the standard deviation is σ, we can model the noise term as follows:

$$E \sim N(0, \sigma^2)$$

We can then combine this with our initial formulation to obtain the following:

$$p(y|x) = N(y; DHx, \sigma^2)$$

Hence, our likelihood model is a Gaussian distribution with μ = y where y is obtained by downsampling and blurring the original HR image and $\sigma = \sigma$ from the noise term distribution. Note that the variance term σ^2 here is unobserved, we can utilise a hyperprior to model it as a random variable for estimation.

5.3 Prior

This experiment tests out three different priors for the SR model:

Firstly, the field of experts (high-order markov random fields) which advances from pairwise markov random field, is used to model prior of an image by multiplying potential clique values of all experts together. Whereby, each experts is modeled by a Gaussian Scale Mixture (GSM)

In the second test case, a Deep Image prior is used. It utilizes unlearned priors. It implements a variant of encoder and decoder model and uses neural network architecture as the prior for the SR model.

In the third test case, a Gaussian prior is learnt over HR images and then used for obtaining the posterior distribution.

5.3.1 Field of experts (FoE)

Formulation FoE is a framework for learning generic image priors [1]. Using MRFs fitted with potentials over a large area of pixels, and are trained on natural images. Traditionally, pairwise MRFs have been proven to not capture the statistics of natural images well. Hence, FoE counters this by predicting parameters of ‘experts‘ over a large area of image (clique).

$$f(x_{(k)}) = f_{PoE}(x_{(k)}; \Theta) = \prod_{l=1}^{L} \phi((k_l * x)_c; \Theta_l)$$

FoE parameters are shared between all maximal cliques. Hence the amount of parameters needed for representation can be kept to a minimum. As an end result, we will obtain the weights of each ‘expert‘ as well as learned weights of the filters.

There have been comparisons of FOE to convolution neural networks, due to their similarity. However, FoE differs mainly with the computing of the partition function of the FoE formulation is usually intractable. As a convention, FoE is normally written in its log form

$$\begin{aligned} P_{FoE}(X; \Theta) &= \tfrac{1}{Z(\Theta)} \textstyle\prod_{c \in \zeta} \prod_{l=1}^{L} \phi((k_l * x)_c; \Theta_l) \\ &= \tfrac{1}{Z(\Theta)} exp\{\textstyle\sum_{c \in \zeta} \sum_{l=1}^{L} \psi((k_l * x)_c; \Theta_l)\} \\ &= \tfrac{1}{Z(\Theta)} exp\{-E_{FoE}(x; \Theta)\} \end{aligned}$$

$$\psi((k_l * x)_c; \Theta_l) = log\phi((k_l * x)_c; \Theta_l)$$

Training FoE Training of FoE is achieved through a process called Contrastive Divergence (CD). First proposed by Hinton in 2002[9], this method of training finds extensive usage in the training of Restricted Botlzmann Machines. The basic idea is to minimize the Kullback-Leibler divergence between the model distribution and the data distribution. Training is achieved by performing gradient descent on the log likelihood. Hence, the change of weights in each iterations is formulated as below

$$\partial\theta_i = \eta \left[\left\langle \frac{\partial E_{FOE}}{\partial \theta_i} \right\rangle_p - \left\langle \frac{\partial E_{FOE}}{\partial \theta_i} \right\rangle_X \right] \qquad (1)$$

However, in practice, even an efficient MCMC algorithm would not converge shortly. Hence training of FOE in practice only utilizes a fixed number of iterations of MCMC, which is denoted as p^j. The equation below reflects the process described above.

$$\partial\theta_i = \eta\left[\left\langle \frac{\partial E_{FOE}}{\partial\theta_i}\right\rangle_{p^j} - \left\langle \frac{\partial E_{FOE}}{\partial\theta_i}\right\rangle_{p^o}\right] \quad (2)$$

Posterior sampling from FoE We implemented FoE based on (2). And such the algorithm used to estimate the final HR is as follows.

Algorithm 1: (Generative Bayesian Super-Resolution.)

```
1: Input: LR observation y, zooming factor r, number of samples M, burn in step length N
2: Initialize: set up D and H according to r and set the initial HR estimation as $\tilde{x}_0 = H^T D^T y$
3: For i = 1, 2, $\cdots$ ,M, do
    • Sample $\tilde{z}_i \sim p(z|\tilde{x}_{i-1}, y, D, H)$ via:
        $$p\big(z_{cl}|x, y\big) \propto p(z_{cl}) \cdot \mathcal{N}((k_l * x)_c; 0, \frac{\eta_l^2}{s_{z_{cl}}})$$
    • Sample $\tilde{\tau}_i \sim p(\tau|\tilde{x}_{i-1}, \tilde{z}_i, y, D, H)$ via:
        $$\begin{aligned} p(\tau|x, y, z) &\propto \mathcal{N}(y; DHx, \tau^{-1}I)\mathcal{G}(\tau; a^0, b^0) \\ &\propto \tau^{\frac{n}{2}} exp\big(-\tau\frac{\|y - DHx\|_2^2}{2}\big) \cdot \tau^{a^0-1} exp(-\tau b^0) \\ &\propto \mathcal{G}\big(\tau; \frac{n}{2} + a^0, \frac{\|y - DHx\|_2^2}{2} + b^0\big) \end{aligned}$$
    • Sample $\tilde{x}_i \sim p(x|\tilde{z}_i, \tilde{\tau}_i, y, D, H)$ :
        – Solve for $\tilde{u}_i$ via:
            $$W_z u = \tau H^T D^T y$$
        – Solve for $\tilde{v}_i$ via:
            $$\begin{aligned} r &\sim \mathcal{N}(0, I) \\ K^T Z K v &= K^T \sqrt{Z} r \end{aligned}$$
        – Obtain a sample via:
            $$\tilde{x}_i = \tilde{u} + \tilde{v}$$
4: End
5: Perform SR image estimation via:
    $$\hat{x} = \frac{1}{M - N} \sum_{i=N+1}^{M} \tilde{x}_i$$
6: Output: SR estimation $\hat{x}$
```

For more detail on the exact implementation of the algorithm, please refer to the original paper [2].

5.3.2 Deep Image Prior

In deep image priors the authors[8] try to bridge the gap between handcrafted priors and priors learned from data. This is done by the use of Maximum posteriori to estimate the value of hidden values by the use of data available. Using the terminology mentioned before the equations can be written in the form of:

$$P(x|y) = \frac{p(y|x)p(x)}{p(x)} \propto p(y|x)p(x)$$

$$x = \underset{x}{\text{argmax}}\ p(x|y) = \underset{x}{\text{argmin}}\ [-log\ p(y|x) - log\ p(x)]$$

The approach used in the paper to minimize the above equation involves constructing a function g with random weights θ whose output can be mapped to the high-resolution image. The random weights are corrected to the correct weights who's output from a different space can be mapped to image x by using gradient descent until convergence. Thus, the new equations to be optimised becomes:

$$\underset{g(\theta)}{\text{argmin}}[-log\ p(y|g(\theta)) - log\ p(g(\theta))]$$

As g is subjective $g : \theta \mapsto x$ (if at least one θ maps to image x) then this optimization problem is equivalent, and thus have same solutions. In practice, g changes depending on how the optimization method searches the image space. The function g can be treated as a hyper parameter and tuned. $g(\theta)$ acts as a prior which helps in selecting a good mapping which gives a desired output image and prevents from getting the wrong images.

Instead of optimizing the sum of two components, only the first term is optimised. The equation can be represented as:

$$g(\theta) = f_\theta(y)$$

$$\underset{f_\theta(y)}{\text{argmin}} - \Bigg(\log P\Big(y|f_{\theta(y)}\Big)\Bigg) = \underset{f_\theta(y)}{\text{argmin}}\, E\Bigg(f_\theta(y); y\Bigg)$$

$$\theta^{t+1} = \theta^t + \tfrac{\alpha\delta(E(f_\theta(y);y))}{\delta\theta}$$

and finally,

$x = f^*_\theta(y)$ where θ^* is the optimal θ.

The Lanczos filter is used for downsampling while training the neural network because it is differentiable and hence loss can be calculated easily. The neural network is assigned random weights initially and y i.e. the LR image is used as the input and no output is used instead the update of the weights is calculated by using the above equation.

The network used has the shape of an autoencoder with skip connections and upsampling method involving upsampling using nearest neighbours with deconv layers in order to prevent checkerboard effect.

5.3.3 Gaussian Image Prior

We trained a Gaussian prior over High-Resolution images of size 128x128 using NUTS sampler and zoom factor of 2. After learning the prior, we used the likelihood equation and the Gaussian prior over x to obtain the posterior. A down sampling kernel was trained over high resolution and low-resolution images. While sampling from the posterior this down sampling kernel was used in replacement of D and H. The HR image was then obtained form the posterior using MAP estimate and taking the mean of samples obtained from Metropolis sampling.

5.4 Datasets

Natural images of different categories might have different underlying distributions, which often results in varying performance of models across different datasets. We originally tried to provide comparisons of approaches to several different datasets. However, due to time constraints, we were unable to extend our scope to different datasets. Hence, we decided to employ the DIV2k dataset, which is a standard dataset that has been extensively used in SR. [3]

Name	Number of Images	Resolution	Image Categories
Div2k	1000	1972 x 1437	Environment, nature

Table 1: Dataset we use to train and test our models

5.5 Benchmarks

Image super resolution is inherently an ill-posed problem. For a given LR image, there might be several qualitatively acceptable HR solutions. Two main benchmarks for SR models are PSNR (Peak Signal to Noise Ratios) and SSIM (Structural Similarity Index). However, we are only including the PSNR benchmark in the scope of this project.

Name	Formal definition
(Peak-Signal-to-Noise-Ratio) PSNR	$PSNR = 10\log_{10}(\frac{L^2}{\frac{1}{N}\sum_{i=1}^{N}(I(i)-\hat{I}(i))^2})$

Table 2: Benchmark we use to evaluate our models

6 Results

We used PSNR metric to compute the results for Deep image Prior and Gaussian Prior. The DIV2K dataset contained both HR and LR images and the results observed were as follows:

	Starfish Image (Zoom Factor 2)		Crab Image (Zoom Factor 4)	
PSNR	Gaussian	Bicubic	Deep Image	Bicubic
	21.16	20.62	26.53	26.314

Table 3: Results

6.1 Gaussian Prior

In our experiments with Gaussian image prior, a grayscale starfish image of size 64x64 with a zoom factor of 2 was used. Both MAP and sampling were used to obtain our results. The MAP estimate gave a PSNR of 19.6 which was poorer than that of the Bicubic interpolation results. The upscaled image obtained using the mean of sampling gave a PSNR of 21.16 which was better than the Bicubic interpolation results which was 20.62. Gaussian priors do not perform as well on natural images as heavier tail distributions due to the nature of natural image statistics so average performances were observed on using Gaussian prior.

6.2 Deep Image Prior

The crab image was converted to grayscale and the size of the LR image was reduced to 28X28. The image was then input to the deep neural net and zoomed up to a factor of 4. After training the upsampled image gave slightly better results than the bicubic interpolation method as mentioned in the above table. The image upsampled by the neural net was able to contain more edge information as compared to the image upsampled by the bicubic interpolation as can be seen in the images below.

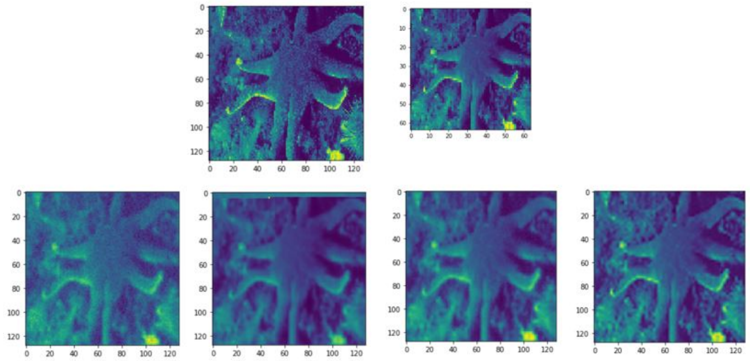

Figure 2: Results for gaussian image prior, following the order from left to right and top to bottom: HR Image , LR image, Sample from posterior sampling, MAP estimate, Mean of sampling, Bicubic interpolation

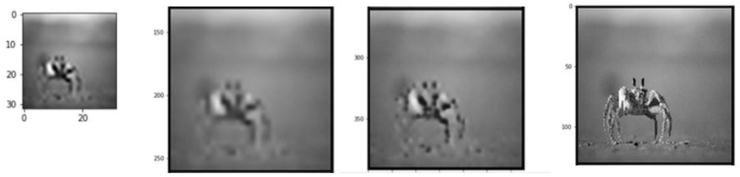

Figure 3: Results for deep image prior, following the order from left to right:LR image, Bicubic interpolation upscaled image, Deep image prior upscaled image, original HR image

7 Conclusions

Ultimately, our group were only able to complete the training and posterior prediction using the Deep Image Prior as well as the Gaussian Image Prior. The FoE prior was unexpectedly complex to implement within the time scope of this project. Although we were able to complete the training and prediction using FoE prior, the results we obtained were not congruent with the ones cited in the paper (the images were only noise) due to issues with our code.

Regardless, we were still able to achieve results that are comparable and even better than the standard bicubic interpolation from the other two priors.

Through this project, we were able to get hands-on experience working with Bayesian-based models, applied our knowledge from the course and even explored beyond the scope of CS5340. While we were unable to achieve one of our main goals, we were glad to have gotten two working models for SR using Bayesian approaches.

7.1 Limitations of our project

Within the timeline of the project, we were able to achieve several working models for SR. However, we could not fully take advantage of our results by analysing specific parts of our models, and producing better insights into the intermediary parts of our model.

Furthermore, given that our models outperformed the bicubic interpolation method for the dataset we used, it would be better if these results are consistent with other datasets as well.

7.2 Future work

Given more time, we would like to debug and properly implement the FoE model. Furthermore, we would like to extend predictions of our models beyond the DIV2k datasets to two other main datasets, the ‘91 Images‘ and random samples of the ImageNet dataset. Last but not least, we would like to compare our models across other benchmarking methods such as SSIM to provide a more comprehensive evaluation of our approaches.

8 References

[1]Schmidt, U., Gao, Q., & Roth, S. (2010, June). A generative perspective on MRFs in low-level vision. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (pp. 1751-1758). IEEE.

[2]Zhang, H., Zhang, Y., Li, H., & Huang, T. S. (2012). Generative Bayesian image super resolution with natural image prior. IEEE Transactions on Image processing, 21(9), 4054-4067.

[3]Wang, Z., Chen, J., & Hoi, S. C. (2020). Deep learning for image super-resolution: A survey. IEEE transactions on pattern analysis and machine intelligence.

[4]Y. Wang, F. Perazzi, B. McWilliams, A. Sorkine-Hornung, O. Sorkine-Hornung and C. Schroers, ”A Fully Progressive Approach to Single-Image Super-Resolution,” 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 2018, pp. 977-97709, doi: 10.1109/CVPRW.2018.00131.

[5] C. Dong, C. C. Loy, K. He and X. Tang, ”Image Super-Resolution Using Deep Convolutional Networks,” in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 2, pp. 295-307, 1 Feb. 2016, doi: 10.1109/TPAMI.2015.2439281.

[6]J. Kim, J. K. Lee and K. M. Lee, ”Accurate Image Super-Resolution Using Very Deep Convolutional Networks,” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 1646-1654, doi: 10.1109/CVPR.2016.182.

[7] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In ICLR, 2015.

[8] Dmitry Ulyanov et al. Deep Image Prior. arXiv:1711.10925, 2017

[9] G. Hinton. Training products of experts by minimizing contrastive divergence. In Aistats (Vol. 10, pp. 33-40)