Singe Image Super Resolution

Application in Image Compression

Interim Report

Author:
Vinay Maniam

CID:
00981045

Supervisor:
Prof. Pier Luigi Dragotti

Date:
January 29, 2018

Contents

1 Introduction

Single Image Super Resolution(SISR) is the process of taking a low resolution(LR) image and upscaling it to a higher resolution(ie increase the number of pixels). This is often described as a highly ill-posed problem, since one LR image can be derived from many high resolution(HR) images. It is used in a wide variety of fields, such as:

Medical Imaging: Devices such as X-rays, MRI, CT scanners and Ultrasound typically produce very low resolution images which require enhancement and filtering before useful information can be read from them.

Surveillance: The presence of security cameras certainly helps significantly in solving crimes, however often times a crime may be committed a great distance away from the closest camera. In this situation, even if the device is high resolution, the fraction of the image containing relevant information will only be represented by a low number of pixels. SISR can then be used to enhance this LR window.

Image Compression: Data traffic across all channels of communication is constantly proliferating. In order to be more efficient in channel bandwidth, images are often compressed during encoding and then decompressed/super resolved at the receiver. Naturally the better SISR algorithms perform, the greater the compression ratio that can be used. Image compression also provides the benefit of economising storage space. Moreover, particularly with videos, compression algorithms will allow for faster loading/download times from the internet, or communication between devices(eg. WhatsApp video message).

The primary focus of this project will be on the Image Compression application of Single Image Super Resolution. Several state of the art SISR approaches are investigated in section 2.2, with comparisons of the relative advantages and disadvantages in section 2.4. The proposed algorithm aims to adapt some of these techniques to improve on an algorithm known as U-FRESH developed by a few members of Imperial College London(see section 2.3.2).

2 Background

2.1 Fundamental Concepts

This section gives a brief description of some generic algorithms in the field of image processing and computer vision which are used as building blocks for several of the advanced methods outlined in section 2.2.

2.1.1 K-means Clustering

K-means Clustering

Inputs: K$\in \mathbb{N}$, set of points $x_1, .., x_n \in \mathbb{R}^m$
Place centroids $c_1, .., c_n$ in random locations in the space
While(stop condition not satisfied) for each x_i
Find the nearest centroid c_j
Assign x_i to the cluster j
For j = 1 to K
c_j = mean of all x_i in the j-th cluster
Stop condition: None of cluster assignments change OR after set # of iterations.

2.1.2 K-nearest Neighbours

K-nearest Neighbours(kNN)

Inputs: Set of points $x_1, .., x_n$ and their corresponding labels $y_1, .., y_n$
For a new point x' with label y':
Find K nearest points based on Euclidean distance.
y' is assigned to the most common label amongst those K points.
Using K nearest neighbours rather than a single nearest neighbour helps prevent overfitting.

2.1.3 Image Quality: PSNR and SSIM

Peak Signal to Noise Ratio(PSNR) is widely used as a quantitative measure of an image's quality. Given a HR image $\mathbf{y^{HR}} \in \mathbb{R}^{m \times n}$ and a SR image $\mathbf{y^{SR}} \in \mathbb{R}^{m \times n}$, the PSNR is calculated as follows:

$$MSE = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} ||y_{ij}^{SR} - y_{ij}^{HR}||_2^2 \tag{1}$$

$$PSNR = 10log_{10}\left(\frac{R^2}{MSE}\right) \tag{2}$$

Where R is the maximum value that a pixel can take(eg. 255 for 8-bit that isn't normalized). Structural Similarity(SSIM) is a metric for measuring the similarity between 2 images. In contrast to PSNR which is based on absolute error between 2 images, SSIM looks at the perceived change in structural information. While still not the perfect measure of reconstruction quality, it factors in more perceptual quality factors than PSNR. The structural similarity between images x and y is given as

$$SSIM = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{3}$$

Where c_1 and c_2 are parameters to stabilize the division when the denominator is very small. SSIM ranges from -1 to 1, with 1 meaning 2 images are identical

2.1.4 Convolutional Neural Networks

Convolutional Neural Networks(CNNs) consist of one or more convolutional layers followed by a standard deep network of fully connected layers. They take advantage of the 2D structure of an input image, through local connections and tied weights, followed by pooling which results in features which translation invariant. They are also easier to train and have significantly fewer parameters than fully connected deep networks.

2.1.5 JPEG2000

JPEG2000 is an image coding system derived from wavelet theory [17]. Despite being developed almost 20 years ago, it still provides cutting edge performance, particularly at high compression ratios.

2.2 State Of The Art

Multi-image super resolution requires multiple frames of a single scene, each taken from a slightly different vantage point, and leverages them to obtain a single HR image. This is highly impractical in most use cases where only a single image of a particular scene is available, therein creating a demand for single image super resolution(SISR). As mentioned previously, the SISR problem is difficult since a single LR image can map to many HR images.

The most commonly used SISR techniques are interpolation and learning based methods. The most popular interpolation based methods are bicubic and bilinear interpolation, with the former generally giving better results. They are computationally inexpensive [18], but provide poor restoration of high frequency components, and are visually blurry. Conversely, learning based methods have exhibited increasingly spectacular results, particularly in recent years. There are many more degrees of control in these methods, and as we shall soon see the computational cost and SR image quality can vary significantly depending on the method used.

2.2.1 Sparse Representations

Zeyde et al. propose one such approach using sparse representations, further improved with k-SVD and Orthogonal Matching Pursuit(OMP) [23]. Given an HR image y_h, denote a patch at location k as $p_h^k = R_k y_h$. It was shown that if p_h^k can be represented sparsely by some $q_k(p_h^k = A_h q^k)$, then an approximate relationship between low resolution and high resolution images can be derived. This property is incorporated into a dictionary learning algorithm which trains a LR dictionary A_l such that LR patches can be represented sparsely. A corresponding HR dictionary A_h is constructed for HR patches such that it matches the LR one. The training phase also includes patch preprocessing to improve the reconstruction result, and dimensionality reduction to increase training speed.

The testing phase uses A_l to sparse-code pre-processed patches p_l^k using A_l. SR patches are then obtained by multiplying the sparse representations with the HR dictionary($p_h^k = A_h q^k$). Finally, the resulting patches are merged to create the resulting SR image. SISR using sparse coding has shown some improvement in PSNR over interpolation based methods [23]. In comparison to other state of the art learning based techniques, this method has a relatively fast training phase($\approx$ 12 minutes), but is noticeably outperformed in PSNR and structural similarity(SSIM).

2.2.2 A+: Adjusted Anchored Neighbourhood Regression

Timofte et al. describe a highly efficient method which builds on Anchored Neighbourhood Regression(ANR)[19] which also produces state of the art PSNR results [15]. ANR is primarily a strategy to boost runtime speed. It achieves this by precalculating and storing transformations during training so that they can easily be called upon during the testing phase. The patch representation problem is formulated as

$$\min_{\beta} ||y - N_l w||_2^2 + \lambda ||w||^2 \tag{4}$$

Where N_l corresponds to the neighbourhood in the LR space that is chosen to solve this problem, and w is the coefficients/weight matrix. In the case of neighborhood embedding would refer to the K-nearest neighbors(section 2.1.2) of the input patch y and in the case of sparse coding would refer to the LR dictionary, similar to A_l in section 2.2.1. Note that the term $\lambda||\beta||^2$ is the regularisation term which transforms the problem into a ridge regression problem. It is simply there to ensure a closed form solution can be found, and helps prevent overfitting based on the training data. The LR patch y can then be super resolved by projecting it unto the HR space

$$z = N_h(N_l^T N_l + \lambda I)^{-1} N_l^T y = P_j y \tag{5}$$

ANR stores the projection matrices P_j for each dictionary atom d_{lj}. A+ modifies on ANR by considering also the K nearest training samples that lie closest to the dictionary atom matched with the LR input patch y. Equation 4 is then reformulated to give

$$\min_{\beta} ||y - S_l w||_2^2 + \lambda ||w||^2 \tag{6}$$

Where S_l is a matrix containing the aforementioned K nearest training samples. At the time it was published, A+ achieved superior performance to most other SISR techniques in circulation. It was later improved on by Timofte et al. by implementing seven modifications unto it [20]. Three of these techniques in particular were shown to produce significant improvements, which shall be discussed hereafter.

2.2.3 Hierarchical Search, Cascading and Enhanced Prediction

Increasing the size of a learned dictionary typically results in better generalised performance for sparse coding and anchored methods [20]. Naturally this comes as a tradeoff for computational. Alternatively, more regressors of the same size can be used to achieve similar improvements in performance. The run time scales linearly with the number of regressors(O(N)).

Hierarchical search aims to counteract this by reduce the complexity to O($\sqrt{N}$) by using k-means clustering to cluster the N regressors under $\sqrt{N}$ centroids, of which each centroid has $\beta\sqrt{N}$ regressors assigned. The process of finding the appropriate regressor for each input feature can now be reformulated to first finding the most correlated centroid(O($\sqrt{N}$)) and then finding the appropriate regressor within it's cluster(O($\sqrt{N}$)).

SISR inevitably decreases in accuracy as the scale up factor increases. One proposed method of circumventing this problem is by training models on a lower scale up factor, and cascading these models to achieve the desired magnification. This will have little impact on training time, however test runtime will increase linearly with the number of layers.

Enhanced prediction is an especially simple method, however it has been shown to provide significant improvement in PSNR [20]. The idea behind it is that SR images are enhanced by averaging the super resolved estimates on a set of transformed images derived from it. Fundamental transformations used include rotations and reflections. By combining these, we get 8 transformed versions of each LR image, which significantly improves the invariance of the resulting SR image.

These three techniques do not apply specifically to A+, and may be eventually implemented in the final version of this project should they prove useful.

2.2.4 Convolutional Neural Networks on Super Resolution

CNNs have been rapidly growing in popularity in the field of image super resolution in recent years. Dong et al.[3] outline the basic intuition behind this, starting with a sparse coding approach and mapping the stages to components in a CNN. Henceforth this implementation shall be referred to as SRCNN. The main benefit is that unlike sparse dictionary learning approaches, all operations are considered in the optimisation problem. Namely, the LR dictionary, HR dictionary, non-linear mapping and mean subtraction and averaging are all optimised during the training of the CNN [3]. The major drawback to this, and any subsequently discussed deep learning methods is that training time can be incredibly long. Notwithstanding, once a network has been trained it can demonstrate excellent runtime speed.

Dong et al.[4] later outlines some modifications to the SRCNN model which drastically improves training time while not compromising output quality. Unlike SRCNN, which uses bicubic interpolation to make the input to the network match the output in size, Fast SRCNN(FSRCNN) uses the unprocessed LR image as input to its network as in [16]. Further to this, extra layers are added to facilitate shrinking and expanding operations, such that the number of hyperparameters is drastically reduced(from 58k to 17k). The training time on a CPU is quoted at only a few hours, which is a significant improvement over SRCNN which takes several hours even on a GPU.

Kim et al.[12] explore a novel approach using very deep networks($\approx$ 20 layers) with a high learning rate and focused on residual learning(VDSR). The motivation behind this is that the

HR image can be accurately estimated as the sum of a low frequency image(LR) and its high frequency residual components. Moreover, unlike SRCNN, this network can be trained for multiple upscale factors, achieving state of the art results across each scale factor. By leveraging a high learning rate with adjustable gradient clipping, VDSR demonstrates significant improvements in training speed, whilst being protected from vanishing/exploding gradients. The training time on a reasonably powerful GPU is claimed to be 4 hours. Quoted results based on the Berkley Segmentation Dataset 100 show VDSR achieving noticeably superior PSNR and SSIM to SRCNN, in about 5% of the runtime. A similar disparity in image quality is seen between VDSR and A+, with the latter taking roughly 3x as long in testing.

A similar residual learning approach is adopted by Jia et al.[10], which introduces a single model, multi-scale super resolution network(MSSR).

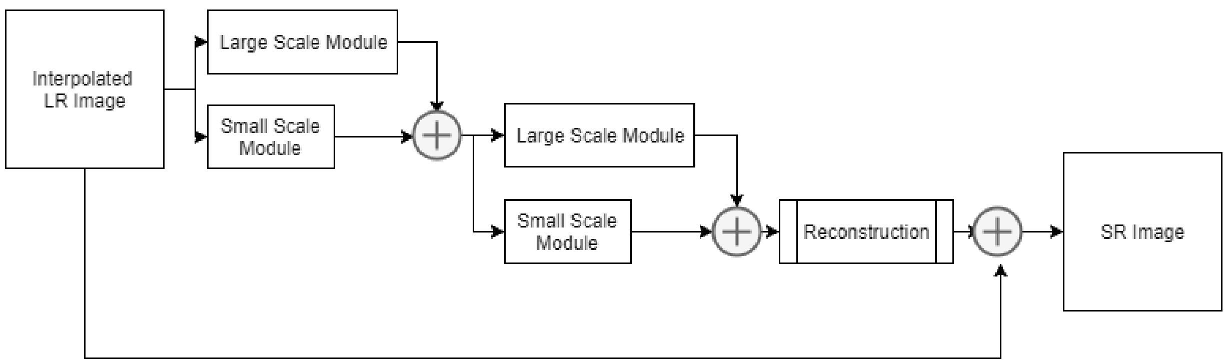

Figure 1: Network architecture of MSSR

Each module in figure 1 is a set of cascaded convolutional layers, with the large scale module being a deeper network for higher scale factors, while the small scale module is more lightweight and efficient, focusing on low scale factors. In this example three different scales can be obtained through a single training session of the network, by choosing a different module at each of the 2 stages shown. Training time is significantly reduced by sharing parameters between "parallel" modules. Additionally, Kim et. al augment the data, similar to the enhanced prediction technique outlined in [20]. MSSR was shown to produce state of the art results, outperforming VDSR in both speed and image quality.

2.2.5 Image Quality Metrics

Peak signal to noise ratio(PSNR) has widely been used as the primary metric for assessing an image's quality. As it is inversely related to MSE, it is perfectly valid to assume that this generally implies superior accuracy of an SR image. However, it has been shown that PSNR does not perfectly correlate with subjective image quality [7][8][5]. One such experiment outlining an example of this is detailed hereafter.

2.2.6 Generative Adversarial Networks(GAN)

In 2014, a novel approach for estimating generative models via an adversarial approach has been invented [6]. The framework of a GAN comprises of two networks, competing with each other in a Minimax game. More specifically as it pertains to image super resolution, the generator network's goal is to create SR images from LR inputs that are indistinguishable to the discriminator network. Concurrently, the discriminator network is trained to maximize the probability of correctly discriminating between the SR image and the true HR image. These networks have since seen groundbreaking success in the field of photorealism [13] with the primary drawback being training time, however batch normalization can be used to counteract the internal covariate shift and diminish this problem[9].

Ledig et al. elaborate on one such implementation by starting with a deep residual network(ResNet) which has been optimized for PSNR, and uses this to initialize the generator network. This is done to prevent the adversarial network from reaching a trivial local minimum. Secondly, the perceptual loss function is adapted from Johnson et al.[11] and modified to

$$L^{SR} = \underbrace{L_X^{SR}}_{\text{content(VGG) loss}} + 10^{-3} \underbrace{L_{Gen}^{SR}}_{\text{adversarial loss}} \tag{7}$$

Where the VGG loss is defined as the euclidean distance between the feature representations of a SR image and the ground truth HR image. The adversarial loss, or generative loss is given by the equation below

$$L_{Gen}^{SR} = -\sum_{i=1}^{N} log\big[D_{\theta_D}(G_{\theta_G}(I^{LR}))\big] \tag{8}$$

$D_{\theta_D}(G_{\theta_G}(I^{LR}))$ in equation 8 is the probability calculated by the discriminator that the super resolved image $G_{\theta_G}(I^{LR})$ output from the generator is in fact a true HR image. Intuitively, a high probability here implies that the generator has likely convinced the discriminator that the SR image is a true HR image, suggesting a high degree of photorealism. Here $log\big[D_{\theta_D}(G_{\theta_G}(I^{LR}))\big]$ is minimised rather than $log\big[D_{\theta_D}(G_{\theta_G}(I^{LR}))\big]$ for better gradient behaviour [6].

Ledig et al. evaluate performance under three criteria: PSNR, Structural Similarity(SSIM) and Mean Opinion Score(MOS). The latter is a human based metric, where a sample population is asked to rate images on a scale of 1 to 5. The main objective of this form of testing was to demonstrate the disparity between high PSNR scores and true photorealism. Results showed that while the ResNet produced state of the art PSNR and SSIM benchmarks and the GAN noticeably lower, the sample population heavily favoured the images up scaled by the latter.

2.3 Related Work

2.3.1 FRESH

FRESH is a Finite Rate of Innovation(FRI) based SISR technique which leverages multi-resolution analysis of wavelet theory [21]. It begins by modeling each line of an image as a piecewise smooth signal, which can be decomposed into a piecewise polynomial and a globally smooth function. The former is reconstructed using FRI theory to estimate the missing wavelet coefficients. Wei and Dragotti, 2016 examine the 1D problem of increasing the resolution of a signal using filter banks. This procedure is then extended to 2D signals, thereby allowing it to be applied to images. The globally smooth function is reconstructed using linear reconstruction.

2.3.2 U-FRESH

The main limitation imposed on the performance of FRESH stems from the assumption that image lines can be modeled as piecewise smooth, which is not highly accurate [2]. This gives rise to U-FRESH, which builds off of FRESH and adds a local regression learning approach similar to [1][22] but with certain key modifications:

1. Unlike other SISR approaches which start with bicubic interpolation of the LR image, U-FRESH uses FRESH enhanced LR images which already contain a large amount of high frequency detail.

2. A different method of selecting the appropriate regression dictionary is used.

The training phase is done on LR-HR patch pairs, where the corresponding LR patch mean is subtracted from each patch pair. Patches of low variance are removed as they contribute little to the learning process, and patch pairs of low correlation are discarded as these were found to degrade the training process. The optimisation problem is similar to other dictionary learning approaches previously discussed [15][23], and can be summarized in three stages. Firstly, K-means clustering(section 2.1.1) is used to obtain K centroids for the training set. Each centroid is assigned a cluster of J training examples using k-nearest neighbours(section 2.1.2). Lastly, the linear regression matrix for each cluster is obtained by minimizing the squared error, with added regularization, gives rise to the closed form expression

$$W_i = H_i L_i^T (L_i L_i^T + \lambda I)^{-1} \tag{9}$$

Where W_i is the regression matrix for the i-th cluster, H_i and L_i respectively are the HR and LR feature matrices associated with cluster i, and λ and I are the regularization parameter and identity matrix. U-FRESH also uses Wavelet Back Projections and Ensemble Learning(on rotated images) to obtain better performance. Results show U-FRESH produces state of the art image quality, both in PSNR and SSIM, when compared to the current standard.

U-FRESH is also applicable in image compression, with the proposed framework[2] illustrated in figure 2.

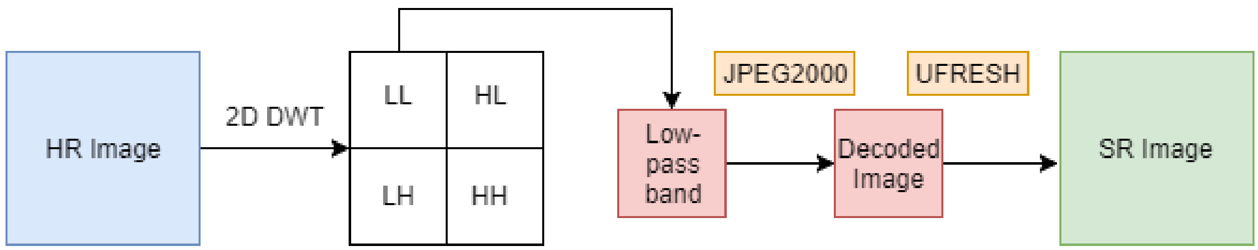

Figure 2: U-FRESH applied to image compression

2.4 Summary of Current Approaches

The aforementioned algorithms can be loosely classified under the categories of dictionary learning and deep network based approaches. Dictionary learning approaches(Sparse Coding, U-FRESH, A+) have fast training times, and generally offer reasonably high SR image quality when compared to Bicubic Interpolation. On the other hand, deep network methodologies(MSSR, SRGAN) offer superior image quality at the cost of drastically increased training time. Methods such as batch normalization and residual learning using high learning rates have been shown to effectively reduce this problem, however they still take orders of magnitude longer to train than other learning based approaches(hours vs minutes). It has also been shown that the function the SISR method is aiming to minimize may require reformulation, as PSNR/MSE does not perfectly correlate with image quality(section 2.2.5).

3 Evaluation

3.1 Safety, Legality and Ethics

This project is entirely software based, consequently no human labour would be required to sustain the finished product as opposed to more hardware based projects which may have ethical qualms with bad practices in the way raw materials are acquired. Moreover, SISR cannot be used in an impactful way in military application, or other violent purposes. To this extent it can be concluded that there are no major ethical concerns. Similarly, no external hardware is concerned, there are no safety precautions to be taken. From a legal point of view as it pertains to patent infringement, this project primarily focuses on further developing algorithms derived by the supervisor, P.L.D, and his research group. Any external methodologies adapted will be open source and appropriately referenced in any formal literature that should come out of this project.

3.2 Criteria and Benchmarks

The proposed algorithm will be compared with other state of the art approaches across the following criteria:

1. Image quality(PSNR, SSIM, MOS*) - Naturally high SR image quality is the primary objective of this project.

2. Training time - In order for the end product to be practical it ought not to have extremely long training times, especially if the algorithm requires retraining for different types of input images.

3. Runtime speed(After training) - Typically in the order of magnitude of seconds/milliseconds, so this is mostly important for real time implementations of the algorithm, as well as applications in video super resolution.

Mean Opinion Scoring(MOS) relies on the availability of a sufficiently large sample population(>25 people[13][11]) to rate image quality. As such, this will only be used as a benchmark provided sufficient test users are available. Of the several approaches to SISR have been discussed in section 2.2, the proposed algorithm will be compared against the following:

1. Bicubic Interpolation

2. A+: Adjusted Anchored Neighbourhood Regression

3. Multiscale SRCNN(MSSR)

4. Generative Adversarial SR Network(SRGAN)

5. Current version of U-FRESH

In addition, it will be compared against both JPEG2000 and the current version of U-FRESH with respect to image compression.

3.3 Training and Test Datasets

The training and test images will be sourced from the Berkley Segmentation Dataset. This is a large dataset of natural images that is widely used as the standard in image processing and computer vision research [14].

3.4 Improvements to U-FRESH

The literature on SISR has inspired many potential ways of improving the existing algorithm. One such technique that particularly stands out is reducing the scale and cascading multiple layers of U-FRESH in a manner that will achieve similar improvements as outlined in [20]. Through successive stages of upscaling and refinement, the overall ambiguity, and therefore reconstruction error, is expected to decrease as compared to doing everything in one step.

Further to this, a hierarchical search method as described in section 2.2.3 can be implemented to increase training speed, and therefore facilitating a larger training set. Moreover, the current technique used to reduce invariance by rotating the image by multiples of 90 degrees outlined in [2] can be improved by also including the reflected/flipped image along with its rotations, effectively doubling the training set.

The existing algorithm will also be analysed for any areas of weakness that can be improved upon. For example, due the paradigm that PSNR is not perfectly correlated with perceptual quality [13], the minimisation problem may require reformulation to accommodate this.

4 Planning

Due to an asymmetric choice of modules betweek Autumn and Spring terms, the project was planned with milestones that mostly fall in Spring and Summer terms.

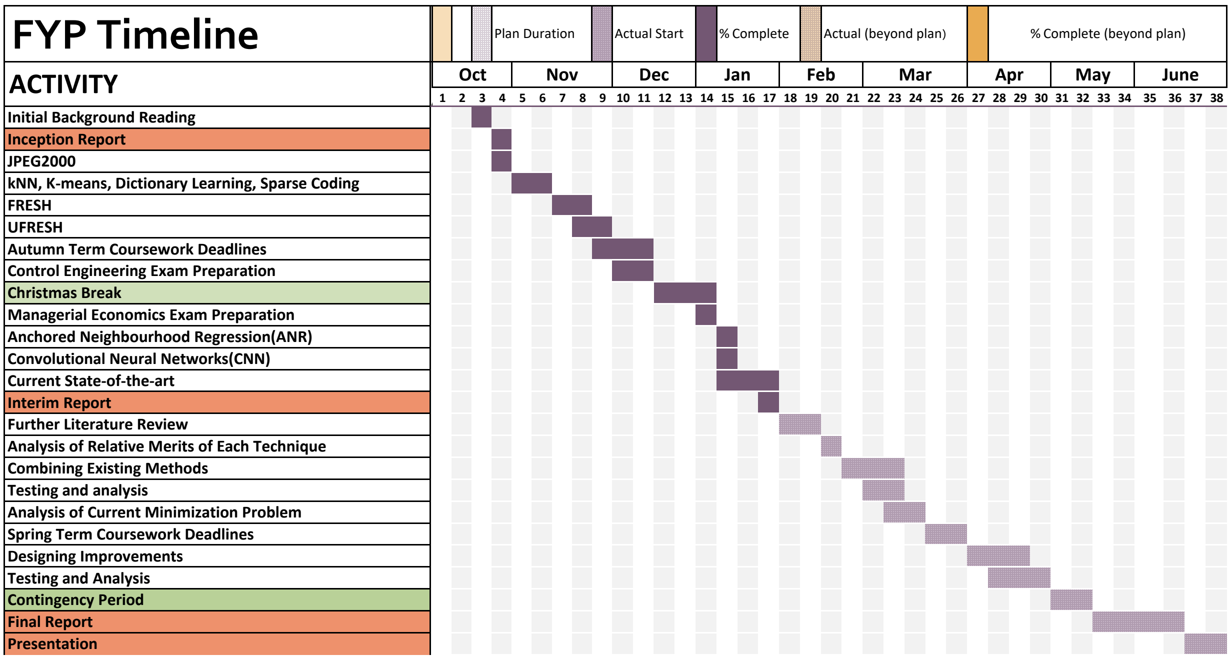

Figure 3: Project timeline and predicted milestones

As this project aims at further developing a cutting edge algorithm, several weeks have been allocated expressly for reviewing relevant literature in the field. Notwithstanding, research will be conducted in the background throughout the duration of the project, beyond what is shown in figure 3. The activities highlighted in red are dates in the project which cannot be shifted, that is formal deadlines. Those highlighted in green are buffer periods for contingency purposes. The final contingency period is only allocated 2 weeks, however it is expected that writeup of the final report and presentation should take fewer than 6 weeks, thereby extending this period. Objectives have been set in a staggered pipeline, such that tasks with dependencies can be carried out as and when the relevant prerequisites are completed. For example, "designing improvements" and "testing and analysis" have significant overlap since tests will be carried out immediately following each design change.

5 References

References

[1] P. V. B. A. Choudhury. Boosting performance and speed of single-image super resolution based on a partitioned linear regression. In *2016 IEEE International Conference on Image Processing*, pages 1419–1423, 2016.

[2] X. Deng, J. Huang, M. Liu, and P. Dragotti. U-fresh: An fri-based single image super resolution algorithm and an application in image compression. under review.

[3] C. Dong, C. C. Loy, K. He, and X. Tang. Learning a deep convolutional network for image super-resolution. In D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, editors, *Computer Vision – ECCV 2014*, pages 184–199, Cham, 2014. Springer International Publishing.

[4] C. Dong, C. C. Loy, and X. Tang. Accelerating the super-resolution convolutional neural network. *CoRR*, abs/1608.00367, 2016.

[5] Ferwerda and James. Three varieties of realism in computer graphics. 5007, 03 2003.

[6] I. J. Goodfellow, J. PougetAbadie, M. Mirza, B. Xu, D. WardeFarley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. *CoRR*, abs/1603.08155, 2016.

[7] P. Hanhart, P. Korshunov, and T. Ebrahimi. Benchmarking of quality metrics on ultra-high definition video sequences. In *2013 18th International Conference on Digital Signal Processing (DSP)*, pages 1–8, July 2013.

[8] Q. Huynh-Thu and M. Ghanbari. Scope of validity of psnr in image/video quality assessment. 44:800 – 801, 02 2008.

[9] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *CoRR*, abs/1502.03167, 2015.

[10] X. Jia, X. Xu, B. Cai, and K. Guo. Single image super-resolution using multi-scale convolutional neural network. *CoRR*, abs/1705.05084, 2017.

[11] J. Johnson, A. Alahi, and F. Li. Perceptual losses for real-time style transfer and super-resolution. *CoRR*, abs/1603.08155, 2016.

[12] J. Kim, J. K. Lee, and K. M. Lee. Accurate image super-resolution using very deep convolutional networks. *CoRR*, abs/1511.04587, 2015.

[13] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. P. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. *CoRR*, abs/1609.04802, 2016.

[14] 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.

[15] L. V. G. Radu Timofte, Vincent De Smet. A+: Adjusted anchored neighborhood regression for fast super-resolution. In *Asian Conference on Computer Vision*, pages 111–126, 2014.

[16] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. *CoRR*, abs/1609.05158, 2016.

[17] D. S. Taubman and M. W. Marcellin. Jpeg2000: standard for interactive imaging. *Proceedings of the IEEE*, 90(8):1336–1357, Aug 2002.

[18] P. Thevenaz, T. Blu, and M. Unser. Interpolation revisited [medical images application]. *IEEE Transactions on Medical Imaging*, 19(7):739–758, July 2000.

[19] R. Timofte, V. De, and L. V. Gool. Anchored neighborhood regression for fast example-based super-resolution. In *2013 IEEE International Conference on Computer Vision*, pages 1920–1927, Dec 2013.

[20] R. Timofte, R. Rothe, and L. V. Gool. Seven ways to improve example-based single image super resolution. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1865–1873, June 2016.

[21] X. Wei and P. Dragotti. Fresh—fri-based single-image super-resolution algorithm. *IEEE Transactions on Image Processing*, 25(8):3723–3735, Aug 2016.

[22] C. Y. Yang and M. H. Yang. Fast direct super-resolution by simple functions. In *2013 IEEE International Conference on Computer Vision*, pages 561–568, Dec 2013.

[23] R. Zeyde, M. Elad, and M. Protter. On single image scale-up using sparse-representations. In J.-D. Boissonnat, P. Chenin, A. Cohen, C. Gout, T. Lyche, M.-L. Mazure, and L. Schumaker, editors, *Curves and Surfaces*, pages 711–730, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[24] K. Zhang, X. Gao, D. Tao, and X. Li. Multi-scale dictionary for single image super-resolution. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1114–1121, June 2012.