The University of Hong Kong
COMP 4801 Final Year Project
Detail Project Plan
September 30, 2018
Topic: Face Image Super-Resolution using Deep Learning

Supervisor: Kenneth K.Y. Wong
Team Name: DeepStrive
Team Members: Lo Tsz Fung (3035270953)
Tsang Suet Ying Florence (3035272341)
Wong Man Yu (3035274313)
# Contents

1. Introduction 2
2. Objectives 3
3. Scope 4
4. Related Works 5
5. Methodology 7
6. Schedule 10
7. Project management and work allocation 12

References 13

# Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN</td>
<td>Convolutional Neural Network</td>
</tr>
<tr>
<td>GAN</td>
<td>Generative Adversarial Nets</td>
</tr>
<tr>
<td>GPU</td>
<td>Graphical Processing Unit</td>
</tr>
<tr>
<td>SR</td>
<td>Super-resolution</td>
</tr>
<tr>
<td>SISR</td>
<td>Single-Image Super-resolution</td>
</tr>
<tr>
<td>MISR</td>
<td>Multiple-Image Super-resolution</td>
</tr>
<tr>
<td>LR</td>
<td>Low-resolution</td>
</tr>
<tr>
<td>HR</td>
<td>High-resolution</td>
</tr>
<tr>
<td>MSE</td>
<td>Minimum Squared Error</td>
</tr>
<tr>
<td>PNSR</td>
<td>Peak Noise to Signal Ratio</td>
</tr>
</tbody>
</table>
1. Introduction

High-resolution human face images are important in various fields. Particularly, in security and surveillance fields, high-resolution face images are required for face recognition or detection purpose [1]. The multimedia industry also desires high-resolution face images since they provide better audience perception and satisfaction [2].

In reality, resolution of face images may not necessarily be as high as desired. These may sometimes be limited by hardware of imaging devices [2], or the fact that the image is sometimes out of focus. Image resolution may also be limited by the distance between the human subject and the imaging device, such that the face image taken is small [2]. Therefore, it is highly desirable and valuable if high-resolution face image could be reconstructed from low-resolution image using super-resolution techniques. The latest advancement in deep learning technology opens possibility for more plausible image super-resolution by learning from realistic facial features.

There are several rooms for improvements on existing solutions of super-resolution by deep learning and algorithmic models and algorithmic models. In terms of:

1) Accuracy and perceptual quality of outputs
2) Computational resources required, such as time and memory usage

Moreover, existing solutions for super-resolution are usually more generic, without specializing on a category of images. We believe that studying face image super-resolution could bring new value to the field of image super-resolution.
2. Objectives

This project aims at implementing a deep neural network for super-resolution with a narrower input scope, which is face images. By experimenting various deep learning algorithms and improving existing algorithms, we expect that our model will produce photo-realistic outputs while reducing computation power and being able to handle multi-scale super-resolution. Also, the output image is expected to be a high-resolution face image of the same person.
3. Scope

Image super-resolution falls into two categories - multiple-image super-resolution (MISR) and single-image super-resolution (SISR). MISR refers to the fusing of information from multiple low-resolution images to produce a high-resolution face image [3]. On the other hand, the input for SISR is one single low-resolution image, from which the resolution is enhanced by previous learning of the relationship between the low-resolution examples of the same category (e.g. face images) and their high-resolution counterparts [4].

Our project will mainly focus on single-image super-resolution for face images. We believe that SISR is a more valuable and challenging problem as the outcome is more dependent on the capability of the algorithm to reconstruct image information missing in the low-resolution single image source. Additionally, data for SISR is more readily available. The training and testing data for SISR are independent face images, while that for MISR are sets of images of the same face, which the source is scarcer. A rich data source will ease our data collection process and will provide a better ground for training and experimenting on the deep learning model.

Furthermore, our project will mainly focus on frontal face images which provides the most facial features. Limiting the scope to only frontal face images should also make the deep learning model more specialized and mature in the task.
4. Related Works

Traditional approaches of super-resolution generate high-resolution (HR) images from low-resolution (LR) images without the use of deep learning. Interpolation is one of the basic approaches in which the interpolated pixels are calculated from neighbouring LR pixel values [5]. Several other SISR algorithms focused on the reconstruction of sharp edges in HR images, or the utilization of image statistics or exemplar image patches for generating HR images [5]. However, in these models, high frequency details of the images are lost.

Recently, deep learning techniques have been used extensively in SISR. This enables the reconstruction of more realistic results and finer details. Dong et al. [6] introduced a three-layer SRCNN, which extracts features from the input image, performs a non-linear mapping, and reconstructs an HR image.

Some studies aimed at improving the model efficiency [7, 8] while some others focused on the model accuracy. Kim et al. [9] made use of a recursive layer to make the network deeper such that it is capable to enlarge the receptive field, which provides more contextual information for learning. In Kim et al. [10], a very deep network that comprises 20 layers brought about notable gain in the accuracy. Although deeper networks can be more powerful, they are more difficult to be trained due to the vanishing gradient problem [11]. Residual learning, gradient clipping and skip connections are some popular techniques to mitigate this issue [10, 12, 13, 14, 15, 16].

Among these models, Mean Squared Error (MSE) and Peak Signal-to-Noise Ratio (PSNR) are the common choices of the loss function and evaluation metrics respectively. A common issue is that although minimizing MSE tends to give a higher PSNR, it does not necessarily reflect the perceptual quality of the output HR image since MSE often overly smoothens the fine details in the images [17].

Ledig et al. [17] proposed a generative adversarial network (GAN) and a perceptual loss that combines adversarial loss and content loss. Despite that the PSNR stays the same, they demonstrated finer HR output images.

More recently, to deal with the instability of GANs, Dahl et al. [18] extended the pixelCNNs and developed a new probabilistic model for SISR, which consists of a prior network and a conditioning network. Using a log-likelihood function for training, the prior network can potentially learn the prior information of a particular type of images such as face images. Additionally, a study involving human observers was conducted to evaluate the performance of the model in generating realistic images.
On the other hand, the model would be more useful if it is able to support different scaling factors from the input image to the output image. Some multi-scale models achieve this by enhancing the image resolution along the network and at the same time producing intermediate outputs of various resolutions [10, 13, 15].

It is also worth noting that some models uses bicubic-upsampled images as input and train the mapping function between the inputs and HR images [6, 9, 10], while other models take LR images as input and directly learn the upscaling filters [8, 15, 19] inside the network.

All the above related works provide insights for us on formulating our project strategy. It will also guide us in exploring which technologies or approaches could be re-used or improved.
5. Methodology

This section highlights the major tasks for this project. While image collection and implementation of neural network for super-resolution are necessary, a few other plausible directions related to face image super-resolution will also be considered.

![Project Stages](image)

**Stage 1 - Image collection and preprocessing**

Input data is a key element of our project as it affects the performance and scope of our model. We are using two sources of face image data - an online face image data source called Helen dataset [20], and collection of face images through Google Custom Search API [21]. To standardize the data, we have also created a pipeline using Python scripts. It detects and crops faces in the image [22], then resizes the cropped image to a desired resolution. We will also manually filter images which matches our criteria: 1) Frontal human face images 2) Photo images instead of drawings / cartoons.

**Stage 2 - Results replication and analysis**

Replication of results from research papers is the first and important step of our deep learning research. This is because research papers usually present the best examples to illustrate the performance of the algorithm. Replication of results allows us to visualize the performance of the algorithms for our own input face images. This helps us discover the limitations and potential areas of improvement of the algorithms. The other benefit of replicating results is to get ourselves familiar with the tools and environment to be used, such as two Python deep learning libraries - Tensorflow and PyTorch.

We would also like to carry out some further analysis after replication of results. A convolutional neural network has several layers where each layer is trying to learn different aspects about the
data. We could try to visualize and understand more on how each different layer contributes to the result, beyond what was described in research papers. We could also adjust the number of hidden layers or adjust various parameters of the model and observe how they affect the result. Another area we could take note of is the amount of computer resources, such as processing speed and memory usage, required by different models.

By the end of this stage, we hope that our understanding of different models will have been improved, and the limitations of various models will have been discovered. This provides us a ground to determine the directions for later stages.

**Stage 3 - Model creation, evaluation and improvement cycle**

In this stage, we will create our own deep learning model. From the analysis in the previous stage, we will have already gained insights on the strengths and weaknesses of different models. In general, a deep learning model consists of the following components:

1) Objective functions. They are mathematical functions describing what the model is trying to optimize and the aim for the output.

2) Neural network architectures. A neural network consists of layers, where each layer is responsible for learning different features (e.g. facial features) about the input data.

We intend to build a new model by incorporating the approaches in various existing models and adding some new ideas based on our own knowledge in computer vision and deep learning. Possible directions include incorporating face images prior into our model (i.e. allowing the model to learn the prior information of the specific category of images - face image), employing appropriate neural network architectures from previous studies, or modifying the objective function to be trained.

After creating a deep learning model, we will evaluate it using the metrics as described in the literature review - Minimum Squared Error and Peak Noise to Signal Ratio. We will also compare the performance of our model with the models from research papers. We might also explore the possibility of conducting an experiment to appraise the perceptual quality of our output.

This stage would be an iterative process. Through all these evaluations, we hope to gain insights on how to further improve our model. After that, we will create another version of the model and repeat this cycle, until we come up with a reasonably satisfactory model.
Stage 4 - Optional Features

After building a satisfactory model for face image super-resolution, we might extend our scope to relevant tasks if time allows. Such tasks might include:

- Allowing customization of input image resolution and upscaling ratio.
- Super-resolution of non-frontal face images
- Implement MISR - face image super-resolution using multiple LR images of the same face
- Implement a pipeline for facial landmark detection
- Adding visual effects to upsampled face images
- Optimization on training and prediction speed.
- Super-resolving face images in multiple video frames. This might be achieved by using a special class of neural network called Recurrent Neural Network.
6. Schedule

<table>
<thead>
<tr>
<th>Task</th>
<th>Estimated start time</th>
<th>Estimated completion time</th>
<th>Completion status</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stage 1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial self-study on deep learning</td>
<td>Aug 20, 2018</td>
<td>Oct 15, 2018</td>
<td>80%</td>
</tr>
<tr>
<td>Image collection and pre-processing</td>
<td>Sep 1, 2018</td>
<td>Sep 30, 2018</td>
<td>100%</td>
</tr>
<tr>
<td>Literature Review</td>
<td>Sep 1, 2018</td>
<td>Sep 30, 2018</td>
<td>100%</td>
</tr>
<tr>
<td>Phase 1 deliverable: Detailed project plan</td>
<td>Sep 1, 2018</td>
<td>Sep 30, 2018</td>
<td>100%</td>
</tr>
<tr>
<td>Phase 1 deliverable: Project webpage</td>
<td>Sep 1, 2018</td>
<td>Sep 30, 2018</td>
<td>100%</td>
</tr>
<tr>
<td><strong>Stage 2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Get familiar with tools and environment</td>
<td>Sep 15, 2018</td>
<td>Oct 15, 2018</td>
<td>50%</td>
</tr>
<tr>
<td>Replication of results from research papers</td>
<td>Oct 1, 2018</td>
<td>Oct 31, 2018</td>
<td>0%</td>
</tr>
<tr>
<td><strong>Stage 3</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model creation, evaluation and implementation</td>
<td>Nov 1, 2018</td>
<td>Jan 6, 2018</td>
<td>0%</td>
</tr>
<tr>
<td>• Deliverable: Deep learning model – initial version</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First presentation</td>
<td>Jan 7, 2019</td>
<td>Jan 11, 2019</td>
<td>0%</td>
</tr>
<tr>
<td>Phase 2 deliverable: Interim Report</td>
<td>Dec 24, 2018</td>
<td>Jan 20, 2019</td>
<td>0%</td>
</tr>
<tr>
<td>Enhance deep learning model</td>
<td>Jan 21, 2019</td>
<td>Feb 28, 2019</td>
<td>0%</td>
</tr>
<tr>
<td>• Deliverable: deep learning model – enhanced version</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Stage 4</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Implement optional features</td>
<td>Mar 1, 2019</td>
<td>Apr 7, 2019</td>
<td>0%</td>
</tr>
<tr>
<td>• Deliverable: deep learning model – extended version</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deliverable</td>
<td>Start Date</td>
<td>End Date</td>
<td>Progress</td>
</tr>
<tr>
<td>------------------------------------------------------</td>
<td>-------------</td>
<td>------------</td>
<td>----------</td>
</tr>
<tr>
<td>Phase 3 deliverable: Implementation and Final Report</td>
<td>Mar 18, 2019</td>
<td>Apr 14, 2019</td>
<td>0%</td>
</tr>
<tr>
<td>Final presentation</td>
<td>Apr 15, 2019</td>
<td>Apr 19, 2019</td>
<td>0%</td>
</tr>
<tr>
<td>Project exhibition</td>
<td>Apr 15, 2019</td>
<td>Apr 29, 2019</td>
<td>0%</td>
</tr>
<tr>
<td>Project competition (if selected)</td>
<td>Apr 30, 2019</td>
<td>May 29, 2019</td>
<td>0%</td>
</tr>
</tbody>
</table>
7. Project management and work allocation

We will adopt the agile methodology in this project. This project is highly exploratory. It would better to determine the tasks for the next step after a review on the preview step. Furthermore, we will also utilize a Kanban dashboard to visualize the work in progress and foster collaboration.

At stage 2 of the project, each person could be responsible for replicating the algorithms from different research papers. After that, we would communicate closely on what have been discovered from this stage.

At stage 3, we believe that allowing every member to contribute whenever they have ideas for certain parts would be more efficient than dictating each team member’s responsible activities. Thus, the exact task for each member will be refined during stage 3.

At stage 4, which is after a satisfactory face image super resolution model is built, the team would proceed with implementing the optional features. We intend to have each member working separately on 1 or 2 optional features of his / her interest.
References


