Single Image Text Super Resolution

Suryank Tiwari ($MT19019$) and Anmol Jain ($MT19005$)

***Abstract*— This project attempts the problem of Single Image Super Resolution (SISR) in the context of Single Frame Text Super Resolution. The objective is to derive a higher resolution image from a single low resolution text image frame. This problem was tackled via two approaches - A Bayesian approach and a Generative Adversarial Network (GAN) based approach.**

I. INTRODUCTION

Image Super Resolution is an image resolving power increasing problem to increase the overall capability of an image processing system. Computationally increasing the resolution of an image would result in lower quality images being stored in systems where image or video recording is a constant operation like surveillance equipment. Super zooming of photographs is possible by zooming and increasing the resolution of the scaled sub-portions of the original image. Along with this, astronomical or medical images, where equipment cannot reach or the environment is unstable, a low resolution image can be used to derive crucial deductions.

Super resolution may be done via combining data from multiple sources and images into one for a better quality or by processing on a single image alone. These specified distinctions are categorized into Multi-Image Super Resolution and Single-Image Super Resolution. This project deals with Single Image Super Resolution or SISR in the context of increasing resolving quality of text frames.

Text Frame SISR refers to processing on singular frames of text and not a whole corpus or sentence together. Each frame corresponds a single character of text in low quality which is processed upon to elevate the resolution. This is approachd in two methods - A bayesian network mapping approach and a GAN based approach.

II. METHOD

A. *Dataset*

Kuzushiji Kanji dataset[2] is used for pulling 'frames' or character images. The dataset KMNIST consists of three sub portions: Kuzushiji MNIST, Kuzushiji 49 and Kuzushiji Kanji. The sub-dataset Kuzushiji Kanji was used as it had images in 64x64 resolution as compared to the 28x28 of the rest two parts and standard MNIST. The images present in Kuzushiji Kanji are taken as the high resolution shapes. Low resolution images are generated from these images after processing them with blurring and other data loss techniques mentioned later.

B. *The Bayesian Approach*

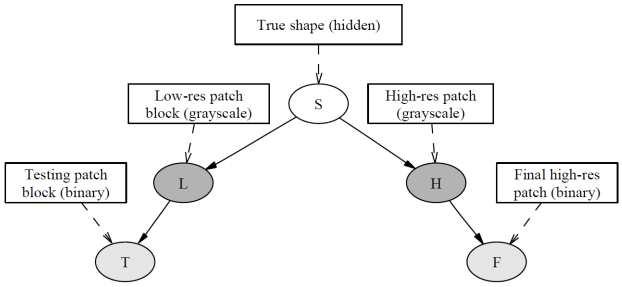

Fig. 1. Bayesian Network Used [1]

*1) **Bayesian Network Framework**:* The Bayesian Network used is depicted in Fig. 1. The nodes in the network each represent an image or its derivative. The Node S is the character image in high resolution that is ideal. It is a hidden node and is not known. L represents a subset of the input image in low resolution and is gray-scale in nature. Corresponding to each L, T is the binary representation of the low resolution image patch. Similarly H is the gray-scale representation of a high resolution patch pertaining to the true shape S and F is its binary representative. Any value greater than 0 is treated as 255 from gray-scale to binary conversion.

*2) **Formulas**:* Once overlapping image patches from all nodes and their respective binary counterparts have been generated, we compute the most likely T given L. That is given by the formula:

$$P(T|L) = \prod_{i=1}^{|L|}(l_i * \delta(1, T_i) + (1 - l_i) * \delta(0, T_i)) \quad (1)$$

$$P(F|H) = \prod_{i=1}^{|F|}(h_i * \delta(1, F_i) + (1 - h_i) * \delta(0, F_i)) \quad (2)$$

Here T, L, F and H are the nodes from Fig. 1. |L| and |F| refer to number of low resolution and high resolution patches that are present for computation. For each low resolution sample $l_i$represents the value of gray-scale at pixel i. $T_i$represents the value of pixel i in the binary low resolution patch. Same naming convention rules are followed for high resolution patches as well.
δ is the Kronecker Delta Function which results 1 when both arguments given to it are equal and 0 otherwise.

$$\hat{P}(F|T) = \frac{1}{N.\hat{P}(T)} * \sum_{i=1}^{n} (P(T|L) * P(F|H))(3)$$

This equation gives the most likely high resolution binary image patch F given a low resolution binary patch T.
The formulas that have guided the Bayesian probabilities are given by Dalley, Freeman and Marks in their work on Text Frame SISR with Bayesian Networks.[1] The super-resolution process that proceeds, and the pairing of patch pairs to compute probabilities is performed differently from their method.

*3) **Super Resolution Process**:* Overlapping sub-image patches are generated in a deterministic manner from the data set mentioned. Binary image patches are then created from gray-scale image thus obtained and a mapping of most likely pairs is created between them. After computing the most likely high resolution binary image patch F with shape (m,m) from a low resolution image patch, of shape (n, n) found in the image that is being super-resolved, the high resolution patch is padded with already existing values in the super-resolved image to match the size of the low resolution patch and added to it with a mean. This leads to the image being added with high resolution patches over time and constructing an image with it. A more complete model might select the location of replacement of the high resolution patch within the low resolution patch with a higher degree of success, in this experiment it was replaced at the middle to account for higher probability of empty frames around the corners.

C. The Generative Adversarial Network Approach

Super-resolution via GAN applies a deep network which is a combination of a Generator model and a Discriminator model. Generator model generates images from scratch with the use of probability distribution, Discriminator checks whether those images are correct or not. In a way discriminator helps generator to reach correctness and make more realistic data. Both models are learning at the same time, and once Generator is trained it has enough knowledge to generate new samples which share very similar properties.

Process :

- Down-sampling(factor-4) is done on high resolution images.
- After this all down-sampled images are passed into generator model, which up-samples images to give newly generated images.
- Now this new generated images are passed into discriminator model and back-propagate to refine our models.
- Residual blocks: 16 residual blocks are used in Generator, which were useful in building network for faster training.
- PixelShuffler x2: This is use for feature up-scaling.
- PRelu(Parameterized Relu): It generates parameters that can be learned to learn negative part coefficient.

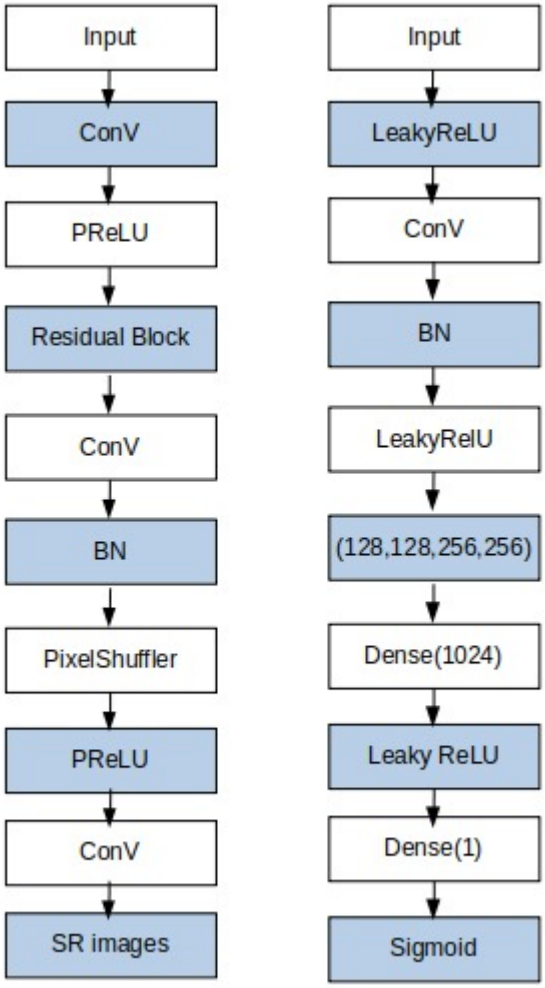

Fig. 2. a: Generator-Network b: Discriminator-Network

III. Experiments and Results

The image patches F, H, T and L can be seen in fig 3. The most likely mappings between low res grayscale and low res binary can be seen in fig 4. Bayesian SISR and GAN based SISR results are in Fig 5 and Fig 6 respectively. Loss generated on experimenting with GAN based model Experimentation between flipping the 0s to 1s and vice versa and the switch between flooring and ceiling the values while normalizing was perfomed for bayesian approach. No significant improvement in results was seen.

Loss HR - 0.0023 , Loss LR - 0.048, Loss GAN -0.078"

IV. Observations

Clearly with Bayesian SISR there is some loss of information. This is suspectedly because most of the high resolution patches had white boxes around them and the 255 ranges overshoot the otherwise intricate corners.

GAN based approach outdid the Bayesian approach clearly. Results from the GAN based approach are better. While still not comparable to synthetic rendering but

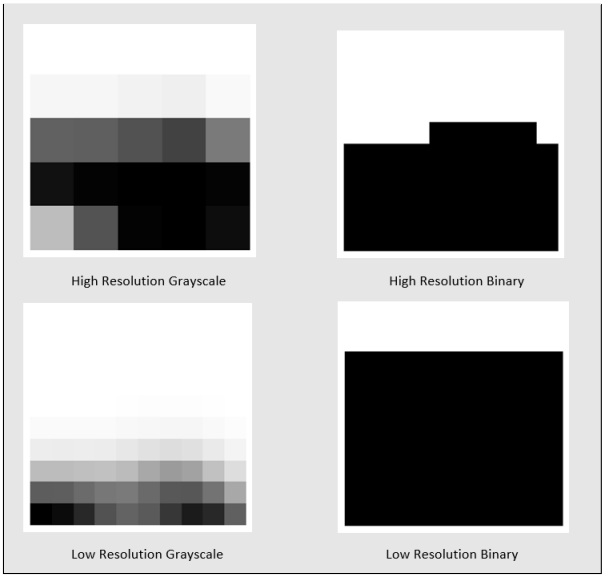

Fig. 3. Image patches

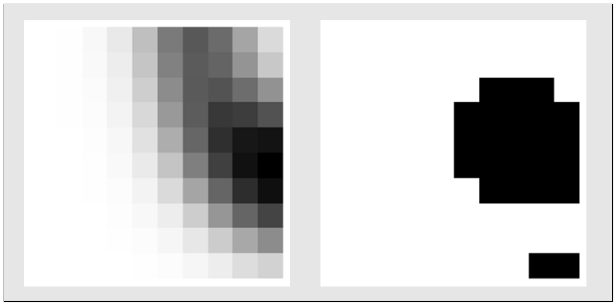

Fig. 4. Likely mappings

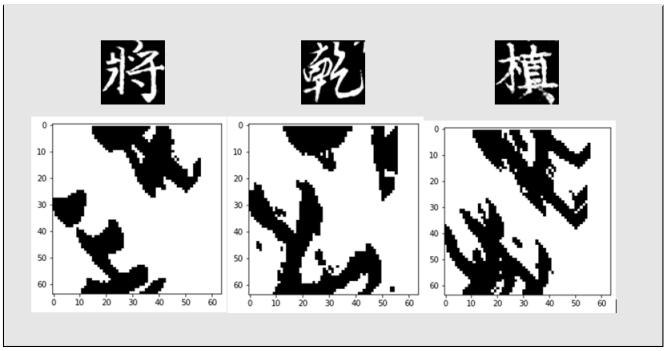

Fig. 5. Bayesian SISR results

Fig. 6. GAN based SISR results, a ratio comparison

Bayesian Approach and GAN based SISR were successfully implemented.

Fig. 7. GAN based SISR result depiction

V. Individual Contributions

After project inception, we read up on research papers and medium articles to obtain information on how we can proceed with achieving satisfactory results for this problem. We worked together in the ideation phase to set up a distinct route for the project along with our future plans. The writing of project proposals, presentations and this final report has been a joint effort. We tackled the SISR problem with two different approaches and each team member took one approach. While we coded on separate solutions to the same end result, we always discussed the problems we faced and the approaches we should make in our respective problems. Each team member is aware of the intricacies of their own problem and simultaneously has a good grasp on the whole project as a whole.

References

[1] Dalley, Gerald & Freeman, B. & Marks, J.. (2004). Single-frame text super-resolution: a Bayesian approach. 5. 3295 - 3298 Vol. 5. 10.1109/ICIP.2004.1421818.

[2] Clanuwat, Tarin, et al. “Deep Learning for Classical Japanese Literature.” ArXiv.org, 3 Dec. 2018, arxiv.org/abs/1812.01718.

[3] C. Ledig et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 105-114, doi: 10.1109/CVPR.2017.19.

[4] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[5] Johnson, J., Alahi, A., Fei-Fei, L. (2016). Perceptual losses for real-time style transfer and super-resolution. In European Conference on Computer Vision.

[6] Projects doing the same things:
https://github.com/MathiasGruber/SRGAN-Keras
https://github.com/eriklindernoren/Keras-GAN/tree/master/srgan
https://github.com/tensorlayer/srgan
https://github.com/deepak112/Keras-SRGAN