Department of Electrical and Electronic Engineering

Text based Image Super-Resolution: Benchmark and GUI

Jongyoon (John), Kim
Dimitris Agrafiotis

1. Introduction

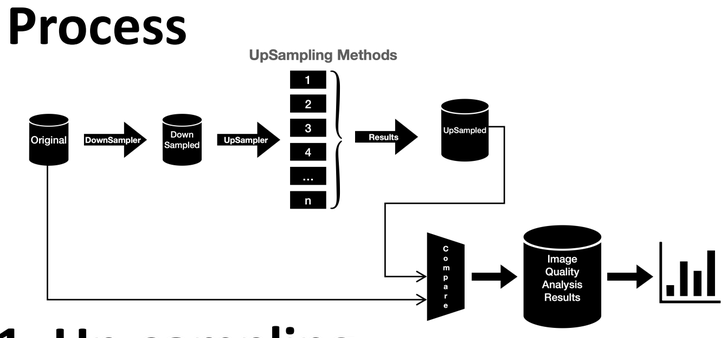
Reading texts on the low resolution images is difficult. Text based images with low resolution typically occur on old cameras, such as CCTV. The basic solution is to replace old CCTV cameras with new cameras, but this requires an increased budgets. Another solution to this problem is to use "Single-Image Super-Resolution (SISR)" technique. This resolution enhancing technique is the focus of this project.

In this project, some SISR methods that have already been implemented will be analysed – focusing on those specifically used for text based images.

2. Aim/Objectives

- This project aims to analyse which SISR method performs best on text based images.
- 13 SISR methods in 3 different categories will be analysed.
- 4 image datasets will be used for analysis.
- SISR methods' performance will be analysed by 4 Objective "Image Quality Analysis(IQA)" metrics.
- Finally, by comparing IQA results on 4 different image sets, optimal method for each image set will be found.

3. Process



4-1. Up-sampling Method/Model Selection

- **Interpolation**

  Interpolation methods are usually used as a baseline to compare other methods.[1]

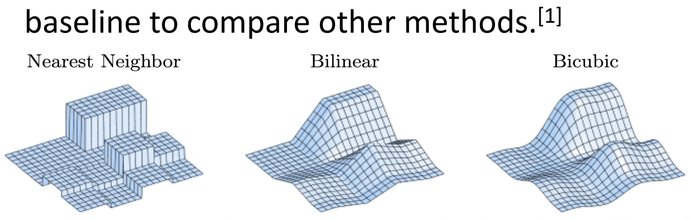


Fig 1. Illustration of interpolation methods. Adapted from [1]

- **Edge-Preserving**

  Used to up-sample texts as detecting and preserving edges of the character are important.

  - **iCBI (interactive Curvature Based Interpolation):** Modified NEDI method which introduces second-order derivatives of image brightness for interpolation [2]
  - **iNEDI (improved New Edge Directed Interpolation):** Improved NEDI method which solves several problems of NEDI such as the estimation method of the edge component. [3]

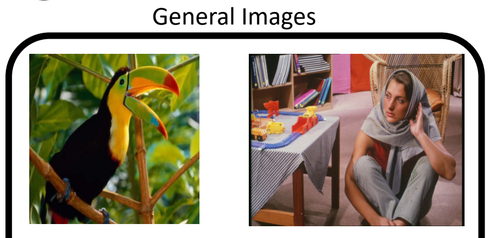
- **Deep-Learning**

  The deep-learning method is a widely used state-of-the-art technique which has dramatically improved the performance of SISR in recent years. [4]

  - **VDSR:** 20 layers using high learning rate with residuals of image for training the model. [5]
  - **FSRCNN:** Fast SRCNN by introducing a deconvolution layer on the last layer of the network, reducing the dimension of the input feature and applying smaller filter but more mapping layers. [7]
  - **ESPCN:** Convolutional neural network with complex filters for each feature map. [6]
  - **EDSR:** Unnecessary modules from ordinary ResNet architecture are removed and a residual scaling function is implemented for stable training of this model. [8]
  - **DRCNN:** Up to 16 recursive layers applied with recursive supervision and skip-connection to handle training stage problems. [9]
  - **LapSRN:** Laplacian Pyramid Super-Resolution Network. This method does not use the pre-interpolation process, hence reducing the complexity. [10]

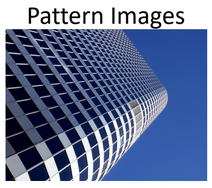
4-2. Image Set Selection (Ground-Truth Images)

General Images



Set5 [11] Set14 [12]

Pattern Images



Urban100 [13]

Text Images

KAIST Scene Text100 [14], [15]

5-1. Image Quality Analysis (IQA)

(objective)

$$RMSE = \sqrt{MSE} = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[Original - Upsampled]^2$$

Similar to PSNR, it is often used as an image quality metric to measure overall difference between original image and the up-sampled image. (Smaller value correlates to better quality)

$$PSNR = 20\log_{10} MAX_{Original} - 10\log_{10} MSE$$

The most popular image quality metric.[18] (Higher value correlates to better quality)

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

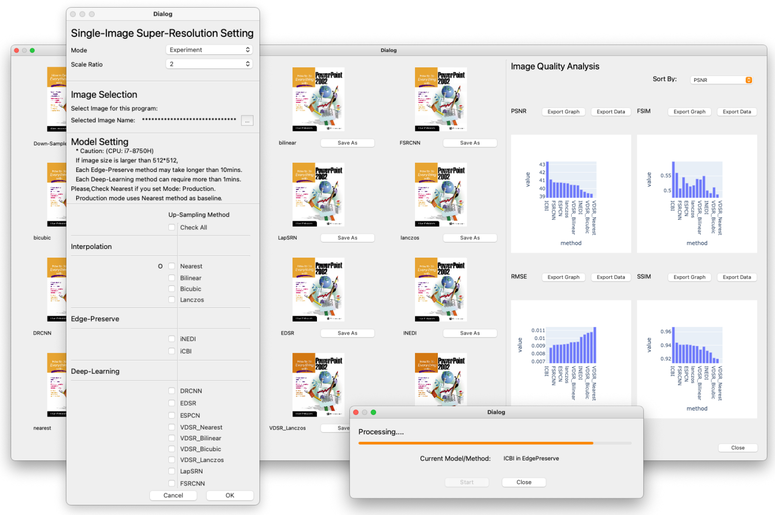
This method assesses the reduction in structural information.[16] (Higher value correlates to better quality)

$$FSIM_c = \frac{\Sigma_\Omega S_{PC}\cdot S_G(\mathrm{x})\cdot\left[S_I(\mathrm{x})\cdot S_Q(\mathrm{x})\right]^\lambda \cdot PC_m(\mathrm{x})}{\Sigma_\Omega PC_m(\mathrm{x})}$$

This method uses low-level features to imitate Human Visual Perception.[17] (Higher value correlates to better quality)

5-2. Process Implementation

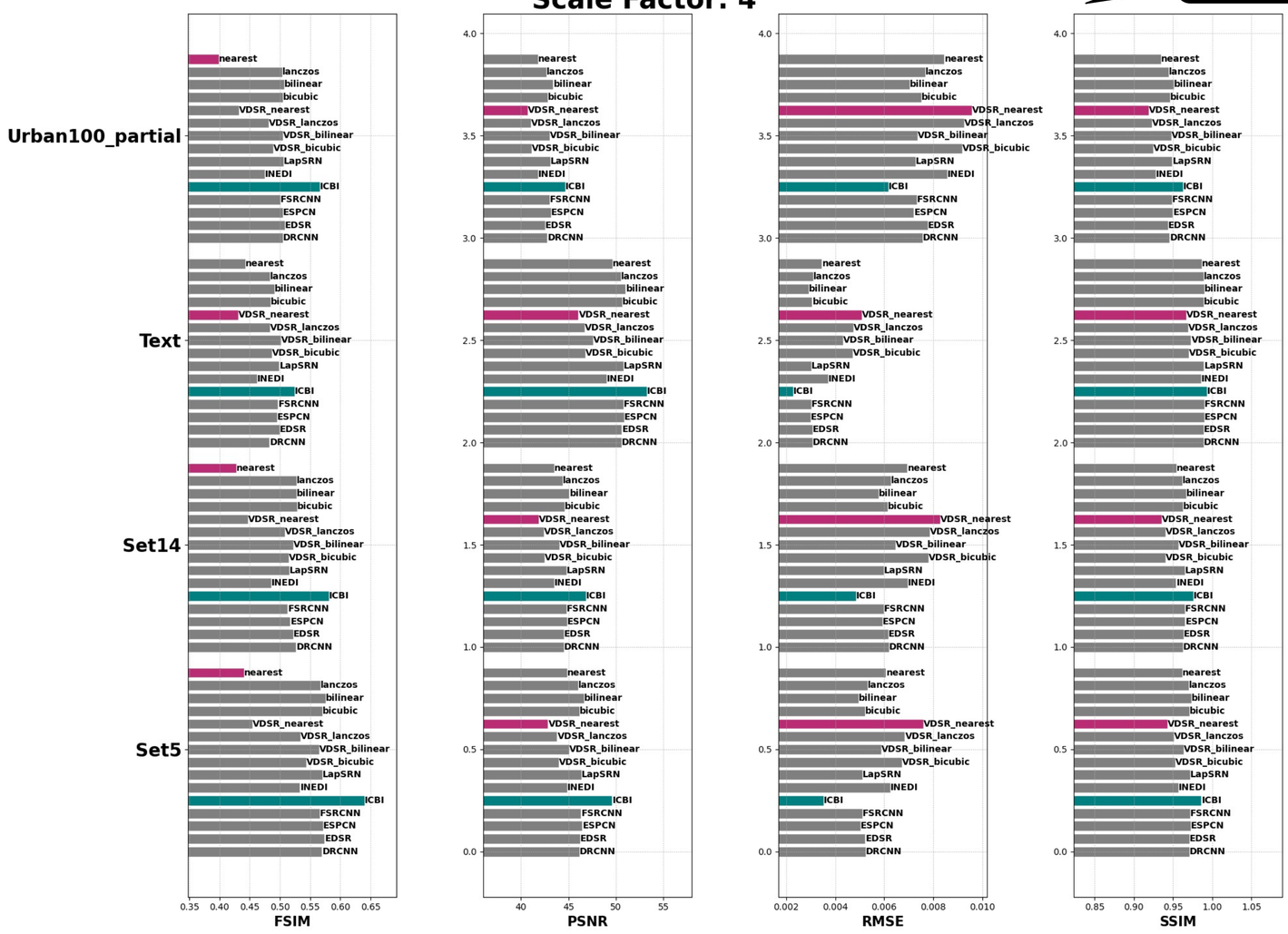
- Programming Language: Python and MATLAB.
- Graphic User Interface (GUI) is built to provide a window to select multiple up-sampling methods and parameters (eg. Scale factor) for non-programmer users.
- Besides GUI, a command-line interface was used to gather IQA results of all images and all scale factors.



6. Results

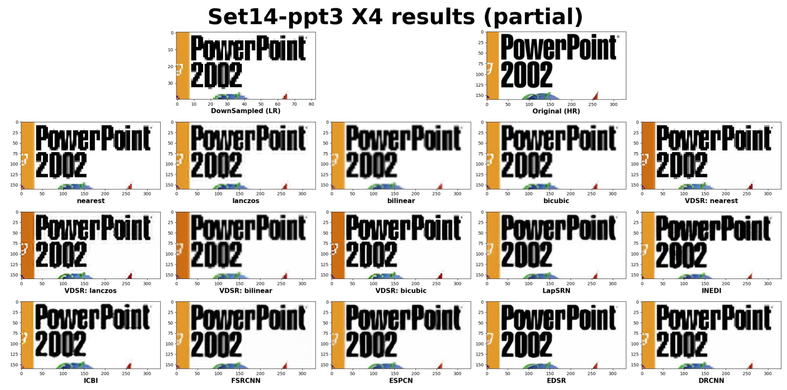
Overall Image Quality Analysis
Compared by image sets (only for scale by 4)
Scale Factor: 4

"iCBI" shows best performance for all image-sets and all IQA metrics.



General Image example

Text Image example



7. Conclusion/Discussion

- The iCBI method shows the best performance in all image categories.
- However, IQA metrics used for this project are objective which does not have a high correlation with human visual recognition.
- This means the best performance on objective IQAs does not fully guarantee the readability of the text on the up-scaled images.
- The deep-learning method is state-of-the-art in SISR, but pre-trained models assessed on this project may be trained with the image set in a specific category, so the performance result showed lower than expected.
- VDSR may showed worst performance on most cases due to pre-trained model used different type of color scale which is YMCK color scale whereas other methods used RGBA scale.

8. Future Works

- In a future project, all deep-learning models will be trained with text-based images using higher computing resources, which can improve the performance of each model.
- A literature review of the project showed the SISR method for upscaling by a factor of 8 is still under development, unlike the factor of 4 and 2 where the methods show similar performances. I would like to research and study in-depth deep-learning methods and propose a better method for up-scaling, especially for scaling by a factor of 8.

9. References

[1] P. Getreuer, "Linear Methods for Image Interpolation", *Image Processing On Line*, vol. 1, pp. 252-253, 2011.
[2] A. Giachetti and N. Asuni, "Real-Time Artifact-Free Image Upscaling", *IEEE Transactions on Image Processing*, vol. 20, no. 10, pp. 2760-2768, 2011.
[3] N. Asuni and A. Giachetti, "ACCURACY IMPROVEMENTS AND ARTIFACTS REMOVAL IN EDGE BASED IMAGE INTERPOLATION", *Proceedings of the Third International Conference on Computer Vision Theory and Applications*, vol. 1, pp. 58-65, 2008.
[4] Y. LeCun, Y. Bengio and G. Hinton, "Deep learning", *Nature*, vol. 521, no. 7553, pp. 436-444, 2015.
[5]J. Kim, J. Lee and K. Lee, "Accurate Image Super-Resolution Using Very Deep Convolutional Networks", *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
[6]W. Shi, J. Caballero, F. Huszar, J. Totz, A. Aitken, R. Bishop, D. Rueckert and Z. Wang, "Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1874-1883, 2016 [Accessed: 04- Mar- 2021]
[7]C. Dong, C. Loy and X. Tang, "Accelerating the Super-Resolution Convolutional Neural Network", *European conference on computer vision*, pp. 391-407, 2016 [Online]. Available: https://arxiv.org/pdf/1608.00367.pdf. [Accessed: 04- Mar- 2021]
[8]B. Lim, S. Son, H. Kim, S. Nah and K. Lee, "Enhanced Deep Residual Networks for Single Image Super-Resolution", *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1132-1140, 2017.
[9]J. Kim, J. Lee and K. Lee, "Deeply-Recursive Convolutional Network for Image Super-Resolution", *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1637-1645, 2016 [Online]. Available: https://arxiv.org/pdf/1511.04491.pdf. [Accessed: 04- Mar- 2021]
[10]W. Lai, J. Huang, N. Ahuja and M. Yang, "Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution", *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5835-5843, 2017 [Online]. Available: https://arxiv.org/pdf/1704.03915.pdf. [Accessed: 04- Mar- 2021]
[11]Bevilacqua, M., Roumy, A., Guillemot, C., Morel, M.L.A.: Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In: BMVC. (2012)
[12] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Curves and Surfaces*. Springer, 2012, pp. 711–730.
[13] J.-B. Huang, A. Singh, and N. Ahuja. Single image super-resolution from transformed self-exemplars. In *CVPR*, 2015.
[14] Jehyun Jung, SeongHun Lee, Min Su Cho, and Jin Hyung Kim, "Touch TT: Scene Text Extractor Using Touch Screen Interface", ETRI Journal 2011
[15] SeongHun Lee, Min Su Cho, Kyomin Jung, and Jin Hyung Kim, "Scene Text Extraction with Edge Constraint and Text Collinearity Link," 20th International Conference on Pattern Recognition (ICPR), August 2010, Istanbul, Turkey.
[16]Z. Wang, A. Bovik, H. Sheikh and E. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity", *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, 2004.
[17] L. Zhang, L. Zhang, X. Mou and D. Zhang, "FSIM: A Feature Similarity Index for Image Quality Assessment", *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378-2386, 2011.
[18]R. Dosselmann and X. Yang, "Existing and emerging image quality metrics", *Canadian Conference on Electrical and Computer Engineering, 2005.*, p. 1907.