Deep Learning for Super Resolution

Arsalan Syed, Aimee Montero, Fabrice Guibert

Abstract

Please include a brief abstract of this paper. Avoid using figures or equations in the abstract.

Introduction

The task of super resolution involves taking a low resolution and estimating a high resolution counterpart. Traditionally, deterministic methods such as bicubic interpolation have been used. However, as computing power and available data has increased over the years, using deep neural networks for this task has become more viable and has been shown to produce superior results to traditional methods. The issue with traditional methods is that when upscaling and image, a lot of the finer details are lost and this results in the image being blurry and appearing pixelated. However, using sophisticated techniques it is possible to reduce this effect.

One important application of super resolution is to make images more memory efficient. As you increase the size of the image in both dimensions, the number of pixels that need to be stored will grow quadratically. It would be much more efficient if one could store a smaller version of an image and have an algorithm that could upscale it when the image needs to be displayed. This would allow for faster file transfer times over networks for example.

Super resolution has many applications within digital image processing and one example would be within microscopy. Light and electron microscopes have a much higher resolving power than the human eye which is why they can show extremely small objects with great detail. However they have their physical limitations, for example due to the size of the wavelengths of light it is difficult to see anything smaller than 200nm with these microscopes[4]. Relying on super resolution would allow you to improve the details of the obtained images to extrapolate the details. Super resolution also has its usage in facial recognition, for example trying to enhance an image of a person in a crowd. It can also be used to upscale digital content like movies so that they appear much better on larger screens.

Objective

The purpose of our project is to implement a CNN that will perform the task of super resolution. The CNN will be trained on discrete wavelet transformations of images and we will observe if the transformations help the network to learn features better.

Methods

The architecture we propose is based on wavelet transforms. Normally, a single network would take as an objective, an upscaled image and would from downscaled versions, try to approximate the former; indeed, we could say it tries to find the inverse of a downscaling function - only where said downscaling can add destructive noise and artifacts as well. Instead of such a network, we explore the possibility of four subnetworks: an image can be decomposed (through wavelet transform) into four subbands of frequencies. As a consequence, it is possible to train a network per frequency subband. If every network reaches the global minimum of its function, then the overall error between the super resolution image generated and the actual objective may also be the minimal one. Furthermore, every subband might behave slightly differently; having one network per subband allows to take into account such differences – if they exist at all. To account for more information, the network training on a particular subband will use the DWT subband of the downsampled input picture to try to approximate the objective DWT subband. As the subbands can be regarded as images, the networks are CNNs. This 4 nets network will output the subbands necessary to construct the upscaled picture, through the inverse discrete wavelet transform.

Training the networks

In order to prepare the images for training upon the networks, we need to decompose them into smaller patches as the network cannot handle variable size inputs and to reduce the model's complexity. Each image is turned into a list of patches of size 64 x 64. The loss function used was mean squared error.

Hardware and libraries

The networks were trained on the ?? cluster. To implement the networks the Keras framework was used which is a python library for neural networks built upon TensorFlow. To analyze the results, built in libraries such as pandas and scikit-learn were used in order to calculate PSNR, RMSE and SSIM as well as to combine the individual measures into tables.

CNN architecture

Results

The metrics used for comparing the performance of the networks were peak signal-to-noise ratio (PSNR), root mean-square (RMSE) and the structural similarity index (SSIM).

Discussion

When comparing the results of bicubic interpolation with other papers, a dsicrepancy in the values was noticed. For example on the monarch butterfly image in Set 14, when resized with a scale of 3, the method we use (from the python OpenCV module) achieved a PSNR of ?? whereas several other papers got a PSNR of 23.21. The reason for this is because they utilise a bicubic interpolation method implemented in MatLab and the difference between the methods is the weighting strategy used, thus causing the final PSNR to be different.

Overall Document Guidelines: Head

Clear document of any fonts other than Times, Arial, and Symbol. The paper should be formatted using the tags provided in

Values for each measure upon Set14

Image	Model 1	Model 2	Model 3
Baboon	A B C	y	z
Woman	x	y	z
Bridge	x	y	z
Comic	x	y	z
Girl	x	y	z
Flowers	x	y	z
Worker	x	y	z
Lena	x	y	z
Man	x	y	z
Monarch	x	y	z
Peppers	x	y	z
Powerpoint	x	y	z
Zebra	x	y	z
Mean	X Y Z	y	z

the template, i.e., title, author, section, subsection, subsubsection, eq./fig., references, etc. in a 2 column format on US letter size paper (8.5×11 inches, or 21.6×27.9 cm).

The left and right margins are set automatically to .75 inch (1.90 cm), and the top and bottom margins to 1.0 inch (2.54 cm). The document is in a 2-column format with column widths set at 3.38 inch (8.57 cm) and the gutter -the space between columns- at .25 inch (0.635 cm).

Papers should be a maximum of 4-6 pages; longer papers will be returned for revision. Please do not place folios or page numbers in your paper. That information is inserted when we assemble the book.

Graphics and Equations

Graphics and equations should fit within one column (3.38 inches wide), but full width (7 inch) figures are also acceptable. Equations, figures and figure captions each have their own style tags. Equations are numbered using parentheses flushed right as shown below.

$$\text{IS\&T} + \text{members} \times \text{Confs.} = \text{Success} \quad (1)$$

Helpful Hints and Style Tags: Subhead

For a complete listing of the style tags for use in this template refer to Table 1. These are the style tags for conference proceedings; if you use the wrong template/style tags your paper will be sent back to you to be reformatted. All of these forms and templates related to the publication of conference papers are available at

www.imaging.org/conferences/guidelines.cfm.

Select the specific conference and download the Authors Kit. The template may vary from one conference to another.

The template is set up for MS Word and LaTeX. Please check the paper carefully to confirm that the styles have been applied correctly, then print it out and double check to ensure that the paper appears as intended.

Submitting Your Paper

The submission of your paper has to be performed through the IS&T submission website. Authors receive a login for this site by e-mail. Papers can be submitted in Postscript, Word, or WordPerfect format, and will be converted to PDF by the submission server. Please carefully review the generated PDF and verify that all the text, equations, figures and tables are displayed correctly before approving its submission.

Margins in LaTeX

Because of the differences between dvips conversion utilities, the margins of your generated PDF document might vary. Please print your document, and verify its margins. If they are incorrect, please adapt the sizes of the margins in the file *ist.sty*. Typically the top margin should be decreased.

Document Specs: Table head

Paper Size	US Letter
Left/right margin	.75 inch (1.90 cm)
Top/bottom margin	1 inch (2.54 cm)
Columns	2 at 3.38 inch (8.57 cm) wide. Spacing between columns: 0.25 inch (0.635 cm)

Figure 1. *IS&T logo.*

Please contact IS&T with any questions or requests for assistance in helping prepare the paper. We look forward to having your paper presented at the conference and published in the conference proceedings.

Figure 2. *IS&T logo.*

Reference Preparation

Use the standard LaTeX *cite* command for references in the text. You can then use the standard bibliography command to generate the list of references. Add the command *small* before the bibliography to give it the right font size. Reference [1] style should be used for books, Reference [2] style should be used for Journals, and Reference [3] style should be used for Proceedings.

References

[1] John Doe, Recent Progress in Digital Halftoning II, IS&T, Springfield, VA, 1999, pg. 173.

[2] John Doe, Digital Imaging, J. Imaging. Sci. and Technol., 42, 112 (1998).
[3] John Doe, An Inexpensive Micro-Goniophotometry You Can Build, Proc. PICS, pg. 179. (1998).

Author Biography

Please submit a brief biographical sketch of no more than 75 words. Include relevant professional and educational information as shown in the example below.

Jane Doe received her BS in physics from the University of Nevada (1977) and her PhD in applied physics from Columbia University (1983). Since then she has worked in the Research and Technology Division at Xerox in Webster, NY. Her work has focused on the development of toner adhesion and transport issues. She is on the Board of IS&T and a member of APS and SPIE.