Reconstruction of High Resolution Images Using Deep Transfer Learning

Tianxiang Gao - City University of Hong Kong
Zhipeng Zhu - Chinese University of Hong Kong

Mentors:
Dr. R. Archibald (ORNL),
Dr. K. Wong (UTK)

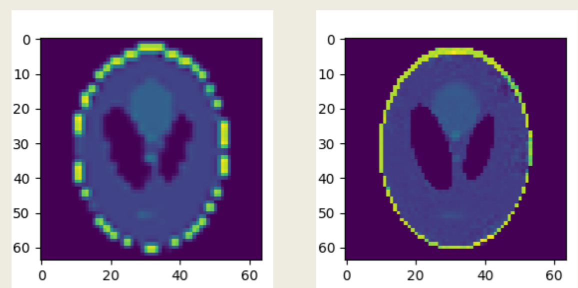
Introduction

Background:

Most current image super-resolution methods focus on single image processing with 2D CNN. Compared to 2-D CNN, 3-D CNN enables us to extract spatial and temporal correlations between consecutive frame of a video, therefore is more suitable for video image super-resolution.

Main Objective:

Train a neural network to do super-resolution on computer-generated video data, such as climate data, in the field of physics.



Data:

- Shepp-logan Phantom
  Model of a human head in the development and testing of image reconstruction algorithms.
- Image from Real Cases:
  Single images, independent to each other.
- Video from Real Cases
  Videos with main object slight changing and moving between frames, used for sequential neural network testing.
- Climate Data
  Computer-generated video based on Shallow Water Equations, which are usually used for describing the flow under a pressure surface.

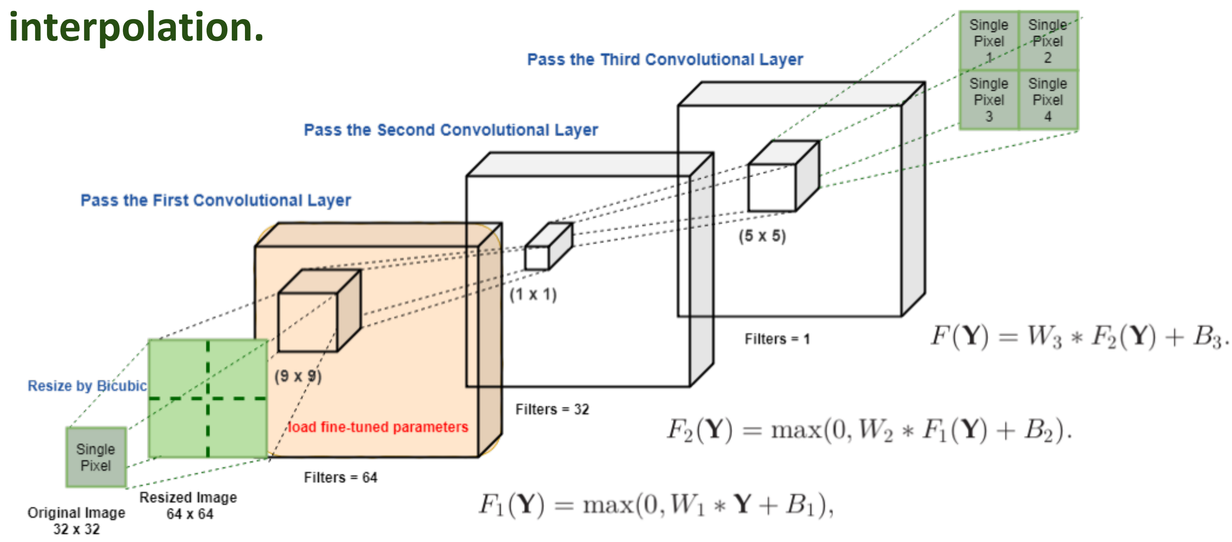
Models and Algorithms

Single Image Super-resolution (2-D CNN)

- **Basic SRCNN Model:**

Feature Extraction -> **Non-linear Mapping** -> **Reconstruction**

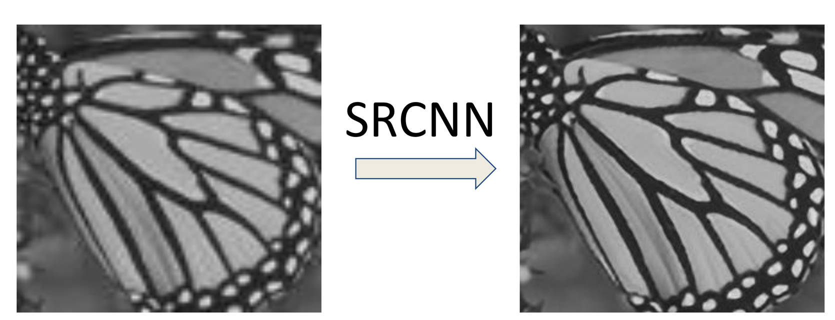
Preprocessing: Resize the low-resolution image to the size of the high-resolution image using bicubic interpolation.



- **SRCNN Model with Transfer Learing**

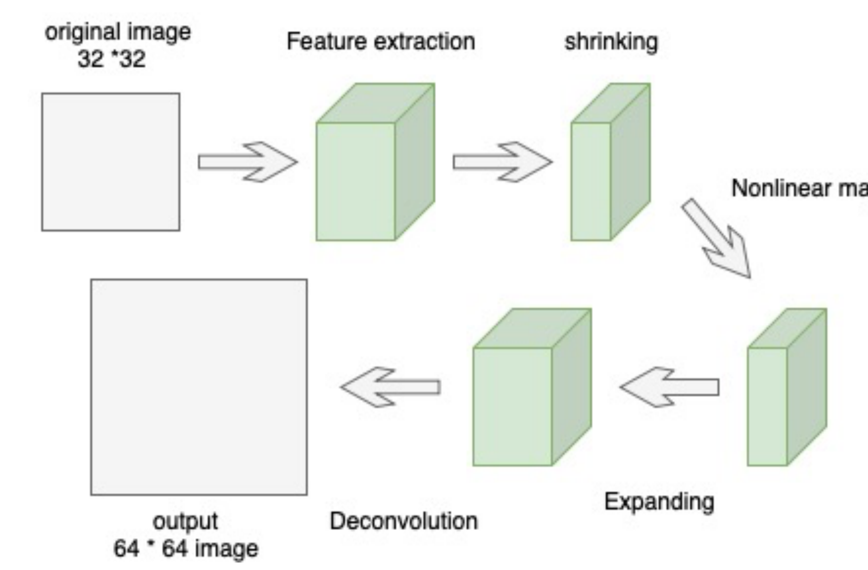
Feature Extraction -> **Non-linear Mapping** -> **Reconstruction**

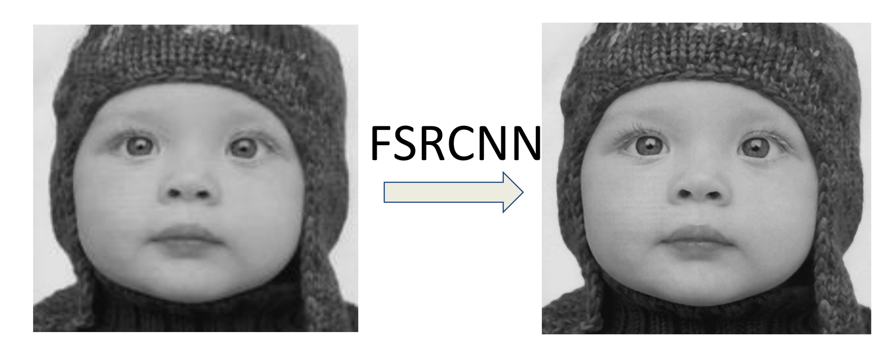
Freeze the Feature Extraction layer;
Use pre-trained parameters as initializers for the later two layers;
Fine-tune the network using our own dataset.



Single Image Super-resolution (2-D CNN)

- **FSRCNN model:**
  Contrary to SRCNN model, we do not resize the input before putting them into the neural network. Instead, we put a deconvolution layer as the last layer of the network to upsample the images.
- **Since the size of input is smaller, the computational cost of FCRCNN is much smaller than that of SRCNN model.**

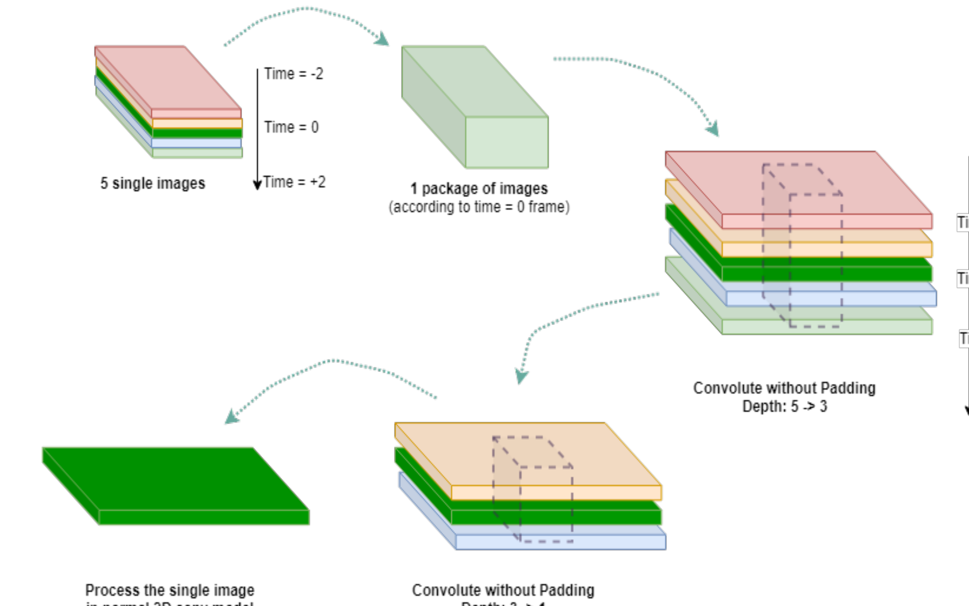




Sequential Image Reconstruction (3-D CNN)

- **Basic 3D SRnet Model:**

Time-domain Information Extraction -> **Downgrade** -> **Output**



Use information from pre-frames and post-frames to compensate the lost information for one frame.

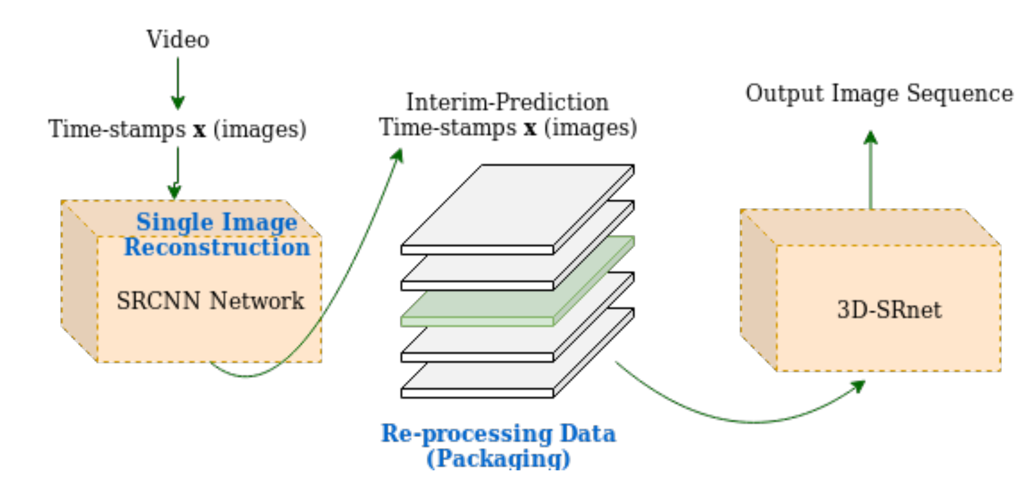
```
Layer (type)                 Output Shape              Param #
==================================================================
conv3d_1 (Conv3D)            (None, 5, 64, 64, 32)     896
conv3d_2 (Conv3D)            (None, 5, 64, 64, 32)     27680
conv3d_3 (Conv3D)            (None, 5, 64, 64, 32)     27680
conv3d_4 (Conv3D)            (None, 3, 62, 62, 32)     27680
conv3d_5 (Conv3D)            (None, 1, 60, 60, 32)     27680
reshape_1 (Reshape)          (None, 60, 60, 32)        0
conv2d_1 (Conv2D)            (None, 60, 60, 4)         1156
==================================================================
```

Sequential Image Reconstruction after Pre-Processing using Single Image Super-Relosution Model

- **Separate 2-D and 3-D Networks**:

SRCNN Network -> Interim Results -> **3D SRnet Network**
Advantage: Can set different hyper-parameter for two models;
Easy to train and implement
Disadvantage: Missing backpropogation between two networks

Do super-resolution twice:
Use SRCNN network to super-resolute the video frame-by-frame and save the interim results as the input of 3-D model.
Use 3D SRnet network to do sequential-image reconstruction, leveraging the spatial correlations between consecutive frames.

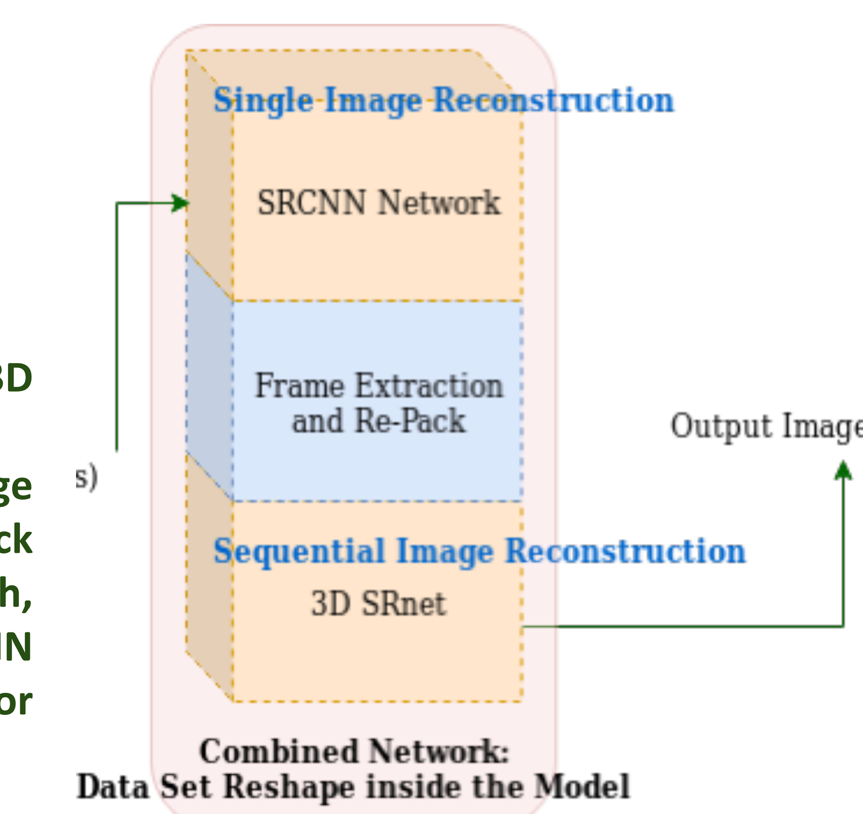


Current Problem:
Interim results are packed for 3D processing. After putting the processed data into the second network, performance is facing a sudden drop. The combination of two neural network does not make an '1+1>2' effect as we expected.

- **Two-In-One Network**:

SRCNN Layers
+
Lambda Layer
(data repacking)
+
3D SRnet Layers

Regard SRCNN network and 3D SRnet as two layers;
Add a Lambda Layer as the bridge of SRCNN and 3D SRnet. Pack consecutive frames as a small batch, resize the output tensor from SRCNN layer, and pass the modified tensor to 3D SRnet.



Challenge:
Inside the lambda layer, we need to use input batch size as index of looping. But the number is stored to be dynamic, means we cannot get the exact number to run the code.

Performance and Analysis

	Basic SRCNN (Image from Real Cases)	SRCNN with Transfer Learning (Image from Real Cases)	Basic 3D SRnet (Video from Real Cases)
PSNR	Raw Data: 33.2460 After Processing: 33.9450	Raw Data: 33.2460 After Processing: 34.5681	Raw Data: 22.1305 After Processing: 33.1454
SSIM	Raw Data: 0.91219 After Processing: 0.92870	Raw Data: 0.91219 After Processing: 0.93395	Raw Data: 0.73280 After Processing: 0.95365
MAE	Raw Data: 0.01673 After Processing: 0.01605	Raw Data: 0.01673 After Processing: 0.01488	Raw Data: 10.58593 After Processing: 5.05191
MSE	Raw Data: 0.00093 After Processing: 0.00078	Raw Data: 0.00093 After Processing: 0.00068	Raw Data: 396.006513 After Processing: 107.77480

	SRCNN network and 3D SRnet (Climate Data)
PSNR	Raw Data: 11.9369 Interim Result: 24.2009 After Reloading: 5.4658 After Processing: 16.9059
SSIM	Raw Data: 0.2889 Interim Result: 0.8801 After Reloading: 0.0752 After Processing: 0.3452
MSE	Raw Data: 5060.8718 Interim Result: 247.2685 After Reloading:36429.17 After Processing:1325.938

- Compression methods (e.g, JPEG) and Resizing impose different loss on pictures;
- Performance of the neural network also depends on the quality of raw data;
- Hyper-parameter tuning can improve performance of neural network, but the improvement is not significant;
- Deepen the network (adding more layers) may not improve performance for particular cases;
- Sudden increase of interim result reloading may be caused by data re-packing.

PSNR: Peak Singnal-to-noise Ratio
SSIM: Structural Similarity Index
MAE: Mean Absolute Error
MSE: Mean Square Error

Future Work

Due to limit of computational capacity and lack of appropriate data, we have not done enough training on our neural network. We plan to collect more data and use them to fine tune our neural network.

Also, since we have many choices for the 2-D model and 3-D model, we believe that the structure of our neural network can also be improved.

With more data, we can change the structure of our neural network, and compare their overall performance on a large dataset to get an optimized neural network.

Reference and Acknowledgement

- Dong C, Loy CC, He K, Tang X. Image super-resolution using deep convolutional networks. IEEE transactions on pattern analysis and machine intelligence. 2015 Jun 1;38(2):295-307.
- Dong C, Loy CC, Tang X. Accelerating the super-resolution convolutional neural network. InEuropean conference on computer vision 2016 Oct 8 (pp. 391-407). Springer, Cham.
- Kim SY, Lim J, Na T, Kim M. 3DSRnet: Video Super-resolution using 3D Convolutional Neural Networks. arXiv preprint arXiv:1812.09079. 2018 Dec 21.

This project was sponsored by the National Science Foundation through Research Experience for Undergraduates (REU) award, with additional support from the Joint Institute of Computational Sciences at University of Tennessee Knoxville. This project used allocations from the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation. In addition, the computing work was also performed on technical workstations donated by the BP High Performance Computing Team.