Reconstruction of High Resolution Images Using Deeping Learning

Final Presentation

Zhipeng Zhu, CUHK
Tianxiang Gao, CityU HK
Mentors:
Kwai Wong, UTK
Rick Archibald, ORNL

CONTENTS

PART

01

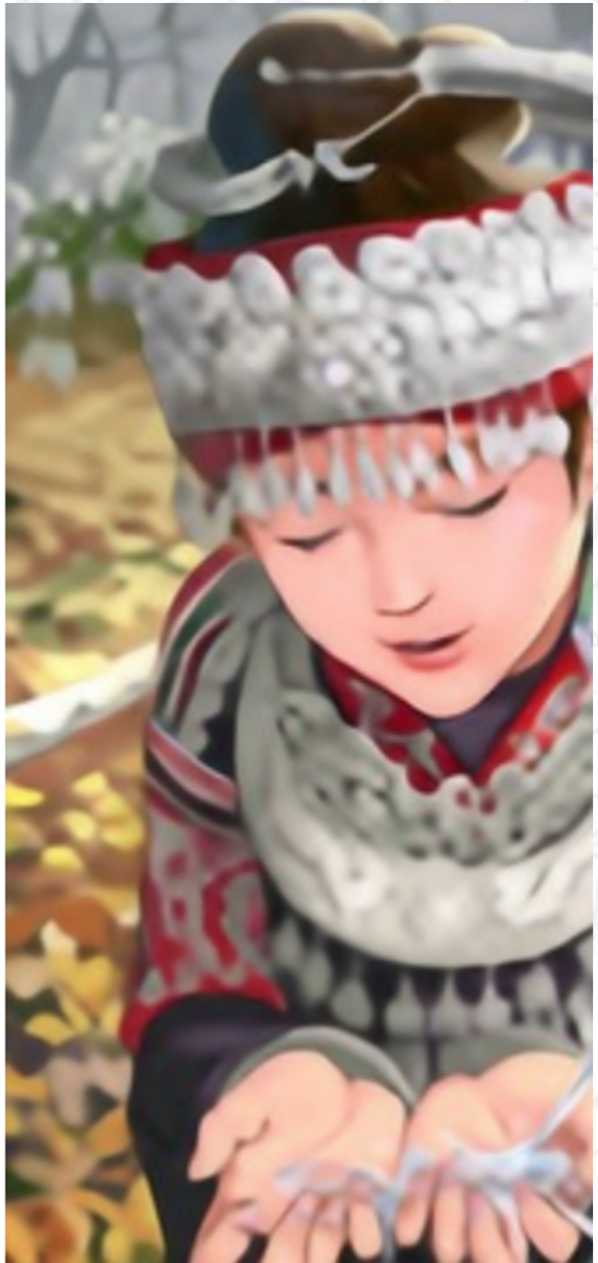
Introduction

01 Introduction

Background

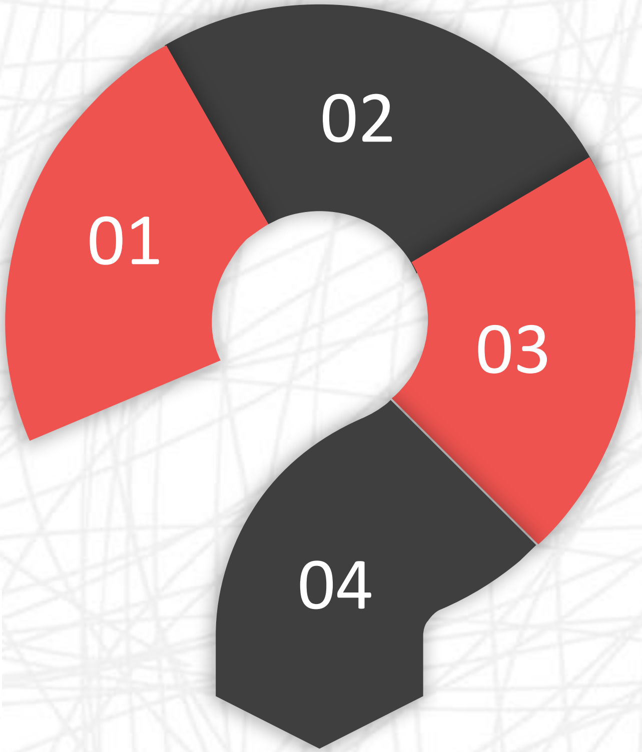
In the field of digital image processing, High-resolution images or videos are commonly needed; but in many cases, people could only obtain low-resolution images.

Image Super Resolution is a class of techniques that turn a low-resolution image into a high-resolution one for further analysis and processing.



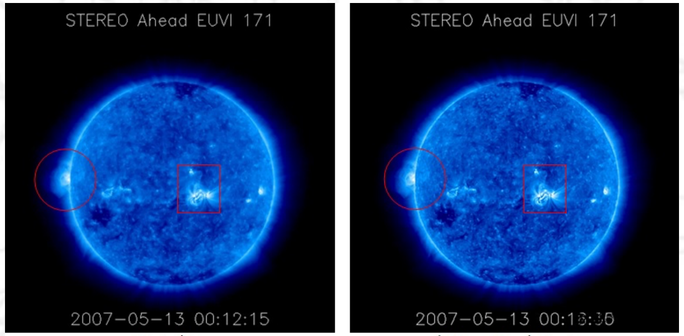
01 Applications of super-resolution

Regular video super-resolution



which applications

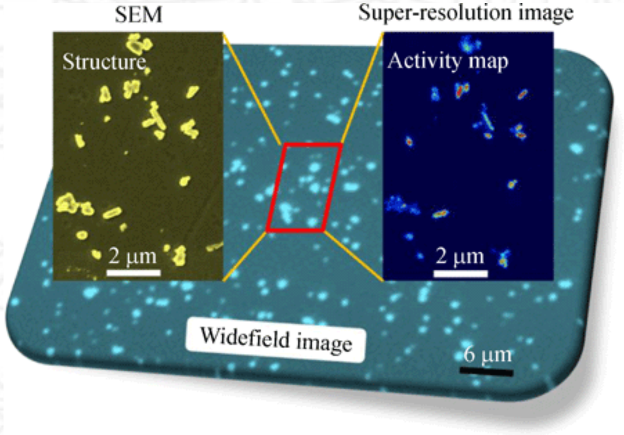
Astronomy



a. Original image

b. Super-Resolution image

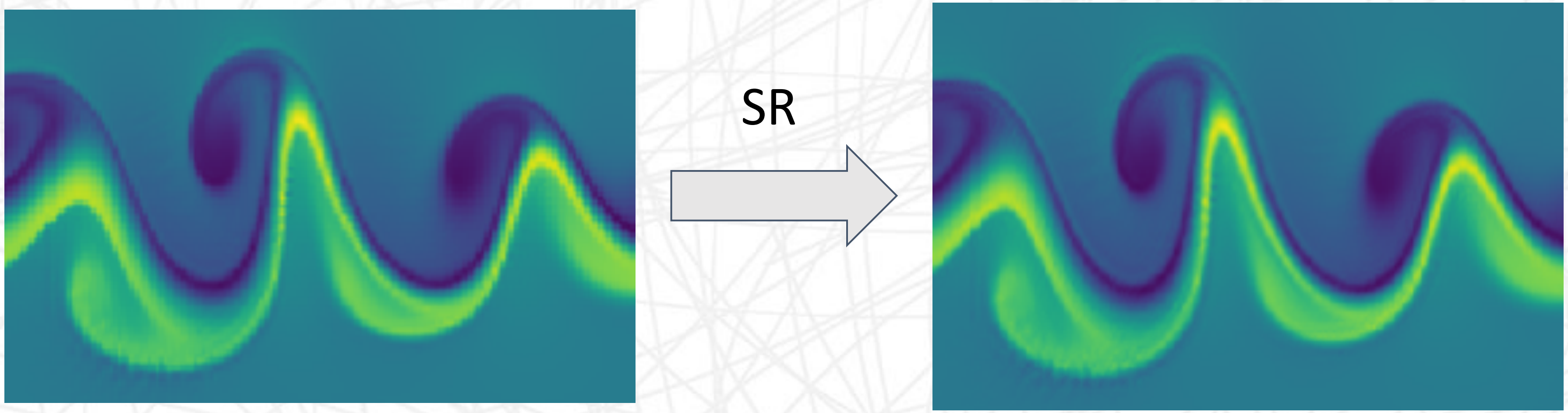
Microscopy



Surveillance

01 What do we want to super-resolute?

We mainly focus on climate data generated by super computers using shallow-water equations. They are videos that simulate basic dynamics on earth.



01 Objective and steps

01 Test and compare current super-resolution models.

02 Build our own model based on current models

03 Train, test, and optimize our model using climate data

04 Implement our model on Magma DNN

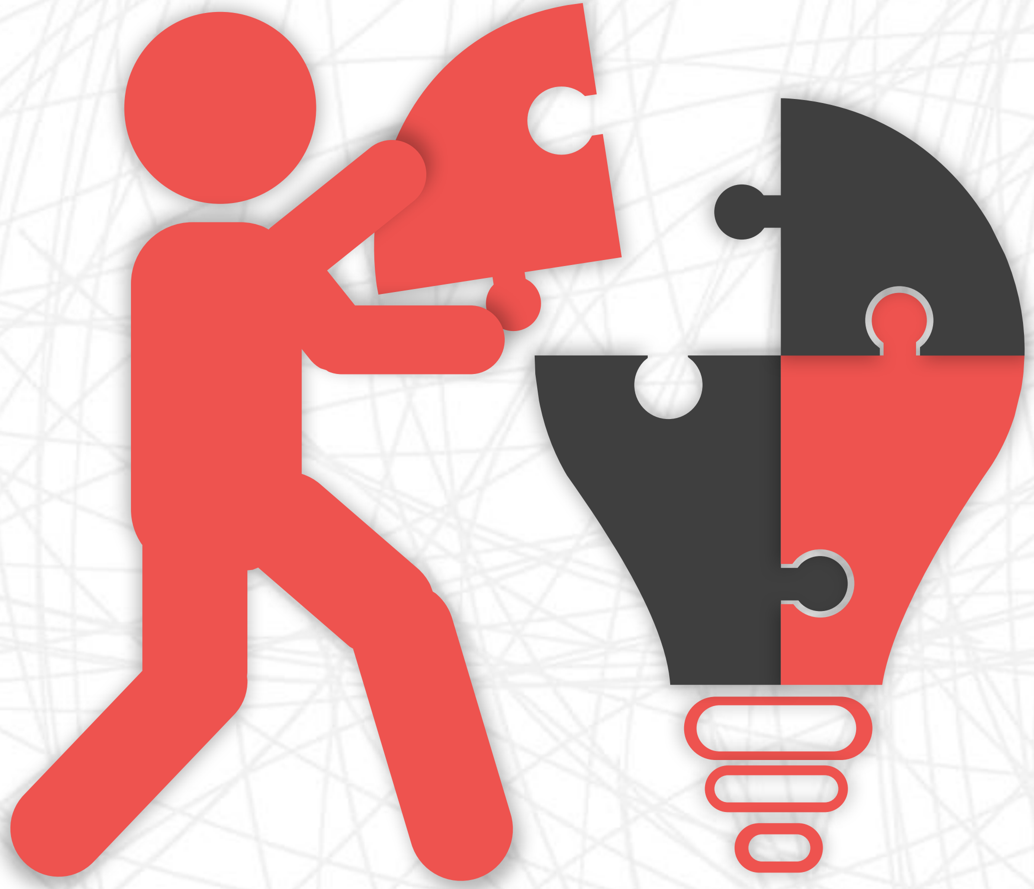
PART

02

Methods

02

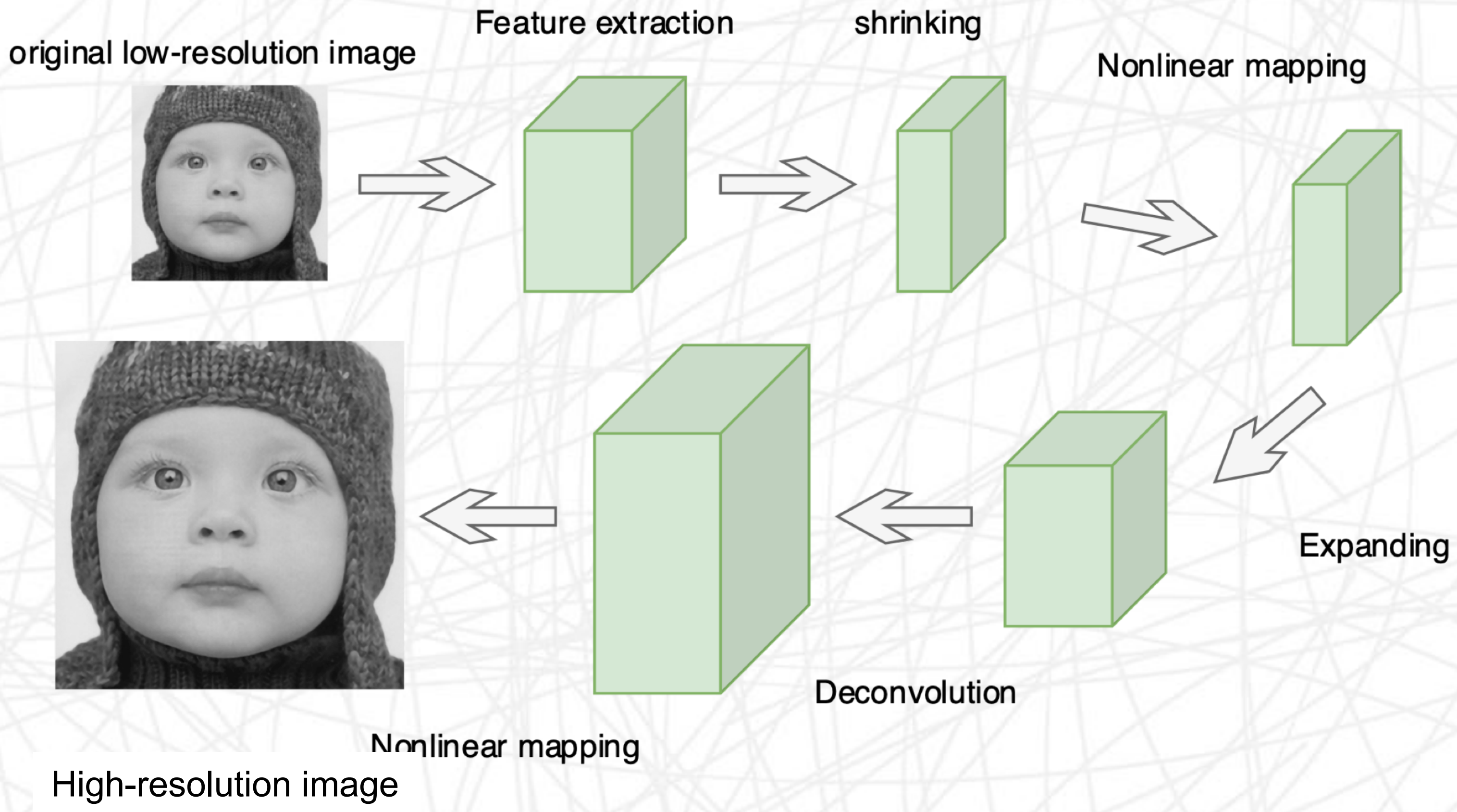
Main idea -- Do super-resolution twice



Use 2D network to super-resolute the video frame-by-frame and save the interim results as the input of 3D model.

Use 3D model to do sequential-image super-resolution, leveraging the spatial correlations between consecutive frames.

02 2D model -- structure of the network

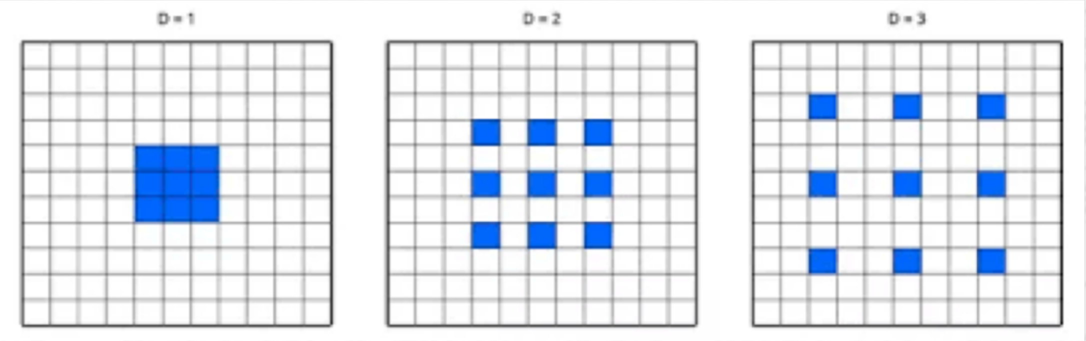


All layers use ‘ReLU’ activation function

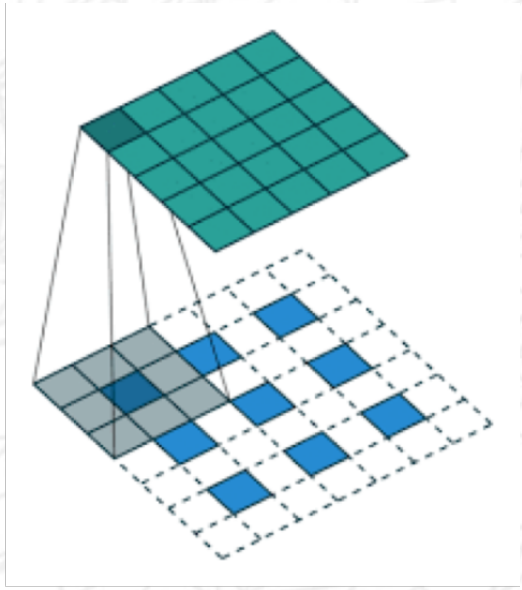
02

2D model – what is a deconvolution layer?

Contrary to most super-resolution models, 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 upscale the images.

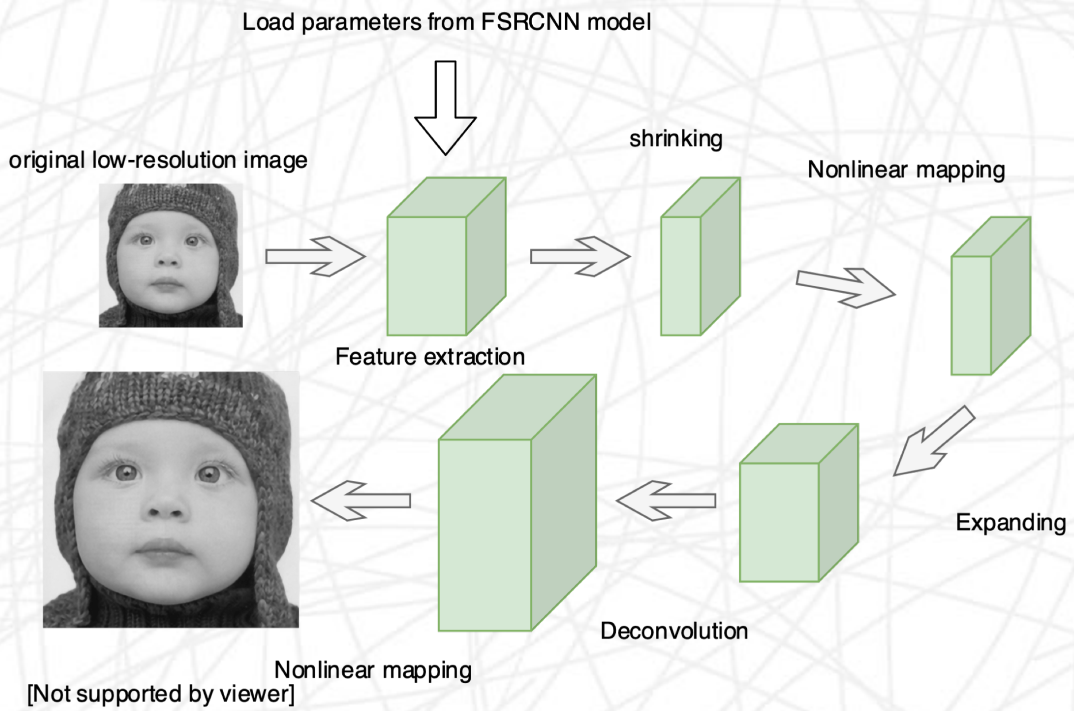


Add zeros between input entries



2D model -- transfer learning

We obtained the idea of deconvolution layer from the FSRCNN model. As indicated by the authors of FSRCNN, their first layer is for feature extraction and can be reused for other models. So we load the parameters of the first layer of FSRCNN, which is provided by the authors of FSRCNN, as the initializers of our first layer, and then, we fine tune them on our dataset.



02 3D model -- motivation

01 Currently, one second of video contains 50 or 60 frames

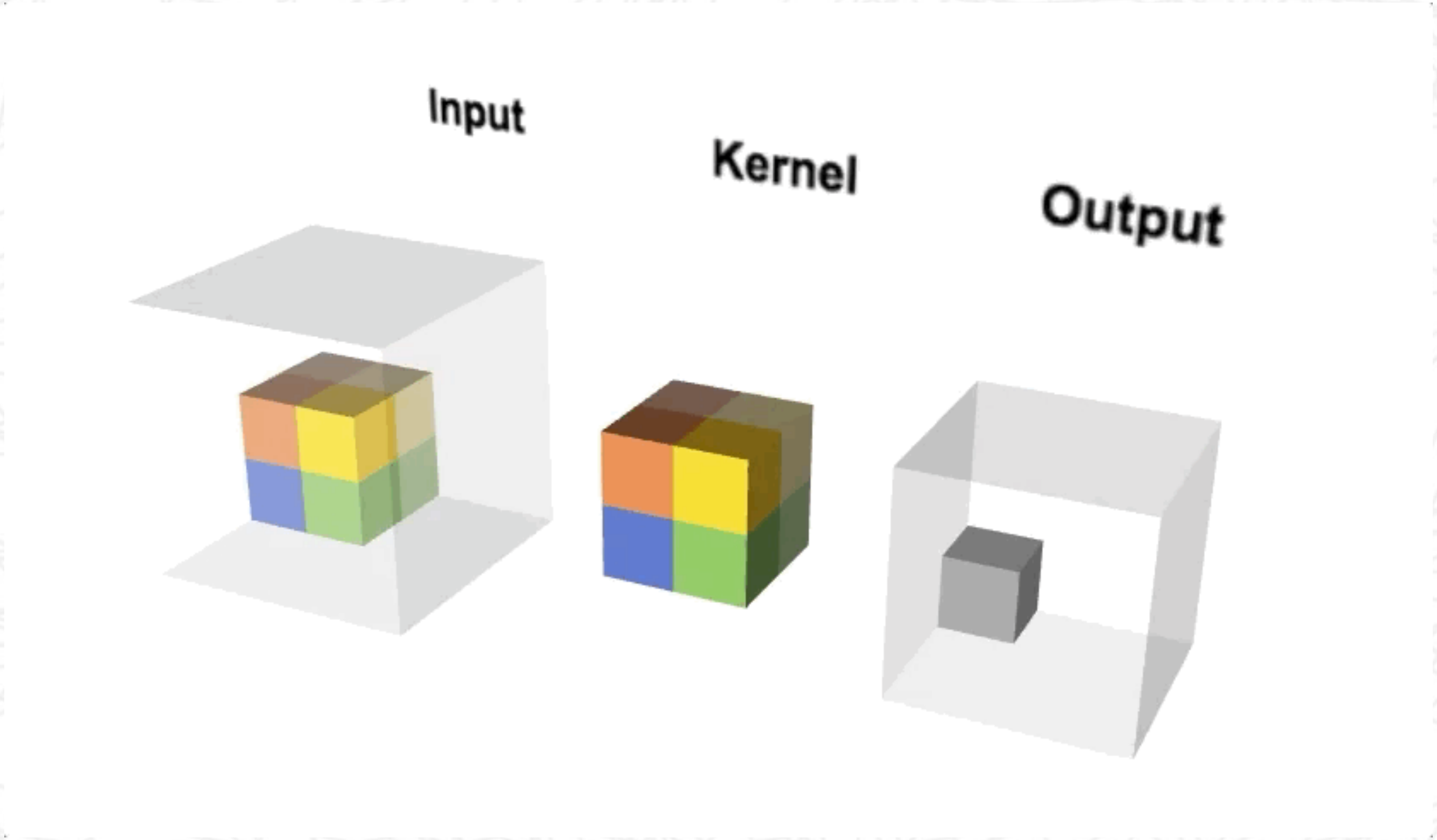
02 Consecutive frames in a video share many similarities

03 3D convolution is suitable for finding correlations between consecutive frames

04 We can leverage the correlations between frames to enhance the quality of final results

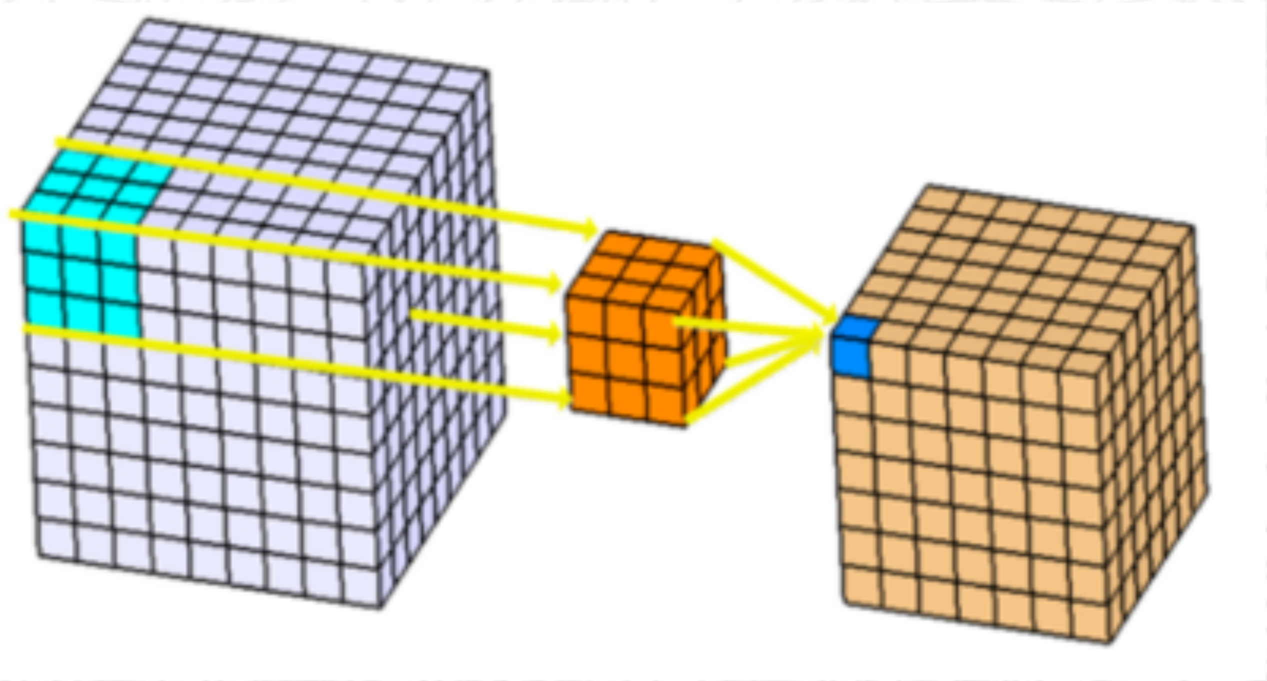
02

3D model -- what is 3D convolution?



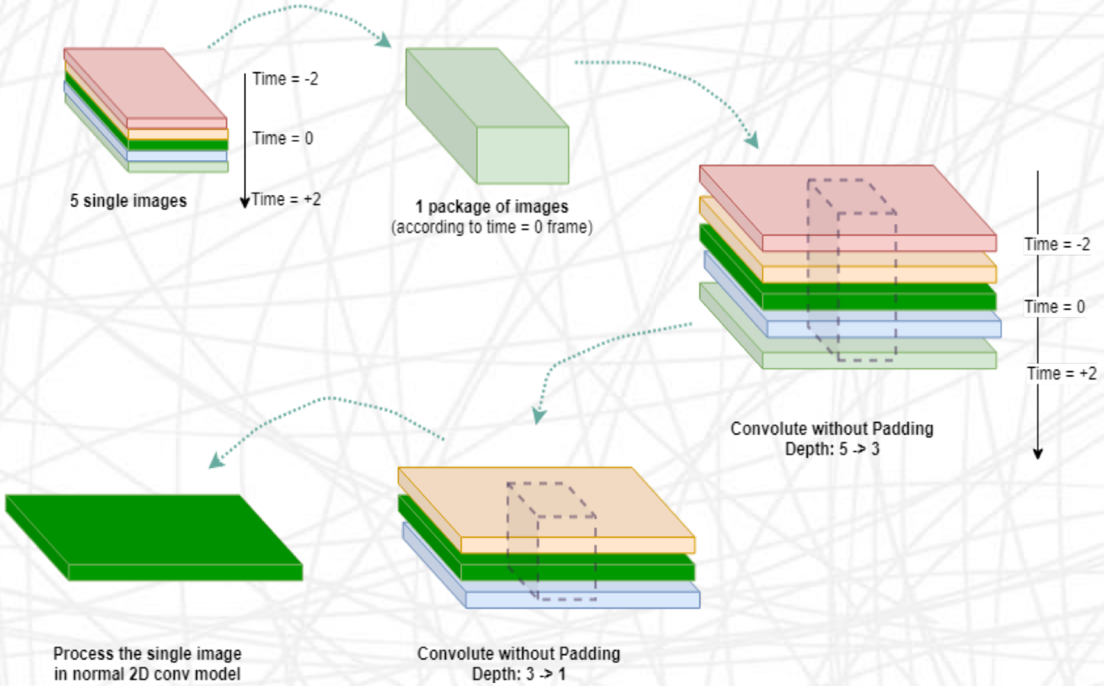
02

3D model -- 3D convolution without padding



The input size is 9*9*9, but the output size is 7*7*7. We can use this method to obtain one single image from the package of 5 consecutive images.

02 3D model -- structure of network

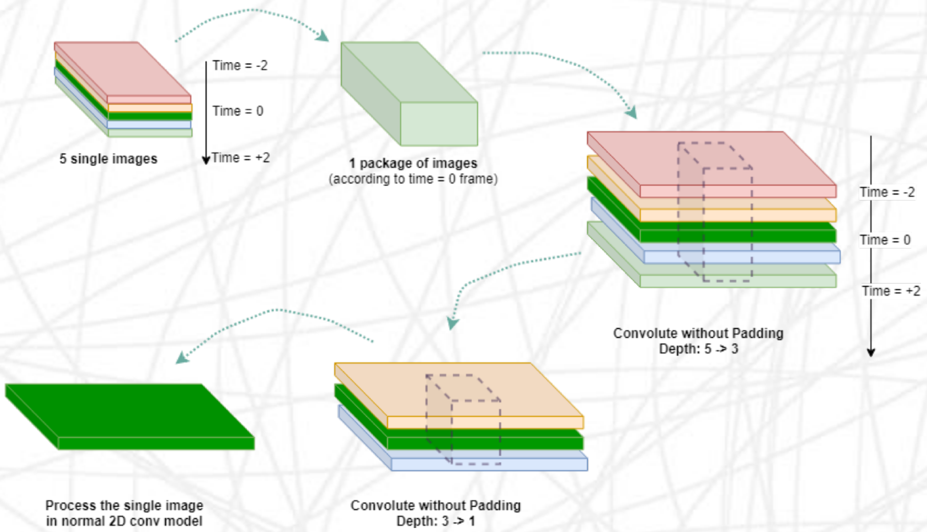


Layer (type)	Output Shape	Param #
conv3d_1 (Conv3D)	(None, 5, 318, 640, 32)	2432
conv3d_2 (Conv3D)	(None, 5, 318, 640, 8)	776
conv3d_3 (Conv3D)	(None, 5, 318, 640, 8)	1736
conv3d_4 (Conv3D)	(None, 5, 318, 640, 8)	1736
conv3d_5 (Conv3D)	(None, 5, 318, 640, 32)	800
conv3d_6 (Conv3D)	(None, 3, 316, 638, 32)	27680
conv3d_7 (Conv3D)	(None, 1, 314, 636, 32)	27680
reshape_1 (Reshape)	(None, 314, 636, 32)	0
conv2d_1 (Conv2D)	(None, 314, 636, 1)	289

All layers use ‘ReLU’ activation function

02

3D model -- structure of network



Layer (type)	Output Shape	Param #
conv3d_1 (Conv3D)	(None, 5, 318, 640, 32)	2432
conv3d_2 (Conv3D)	(None, 5, 318, 640, 8)	776
conv3d_3 (Conv3D)	(None, 5, 318, 640, 8)	1736
conv3d_4 (Conv3D)	(None, 5, 318, 640, 8)	1736
conv3d_5 (Conv3D)	(None, 5, 318, 640, 32)	800
conv3d_6 (Conv3D)	(None, 3, 316, 638, 32)	27680
conv3d_7 (Conv3D)	(None, 1, 314, 636, 32)	27680
reshape_1 (Reshape)	(None, 314, 636, 32)	0
conv2d_1 (Conv2D)	(None, 314, 636, 1)	289

Note: 3D model does not increase the resolution of input images. It is designed for improving the quality of images processed by 2D model by leveraging the correlations between consecutive frames.

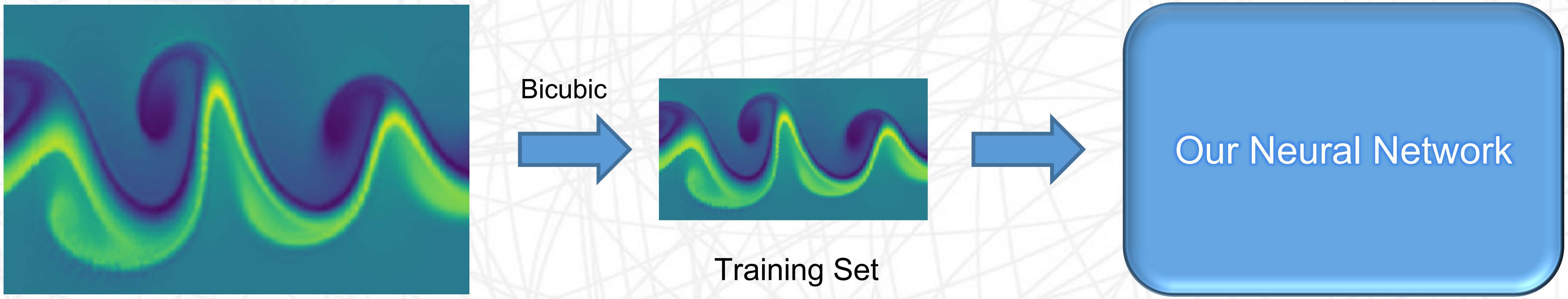
02 Preprocessing before training

Ground Truth

Our mentor provide us a video of climate simulation, which is used as ground truth.

Training set

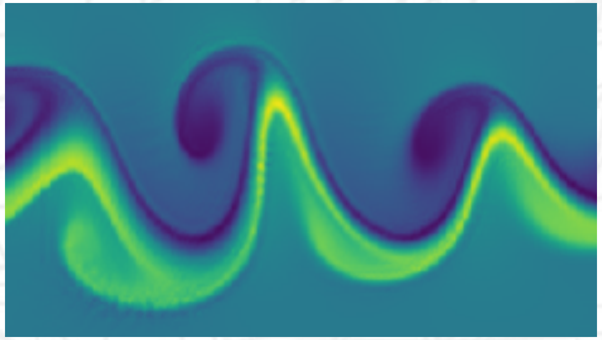
We downscale the ground truth using bicubic interpolation to obtain our training set

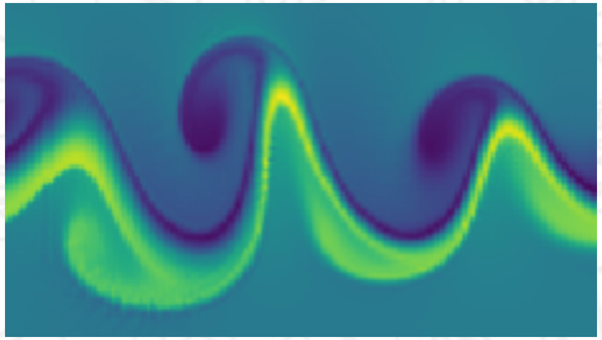


Ground Truth

02 Training – metrics we use

Final result of our network





Ground Truth

C
O
M
P
A
R
E

MSE: mean squared error

$$MSE = \frac{1}{n}\Sigma\underbrace{\left(y - \widehat{y}\right)^2}_{\text{The square of the difference between actual and predicted}}$$

PSNR: peak signal-to-noise ratio

$$\begin{aligned} PSNR &= 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right) \\ &= 20 \cdot \log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right) \\ &= 20 \cdot \log_{10}(MAX_I) - 10 \cdot \log_{10}(MSE) \end{aligned}$$

SSIM: structural similarity index

$$\mathrm{SSIM}(\mathbf{x},\mathbf{y}) = [l(\mathbf{x},\mathbf{y})]^{\alpha}[c(\mathbf{x},\mathbf{y})]^{\beta}[s(\mathbf{x},\mathbf{y})]^{\gamma},$$

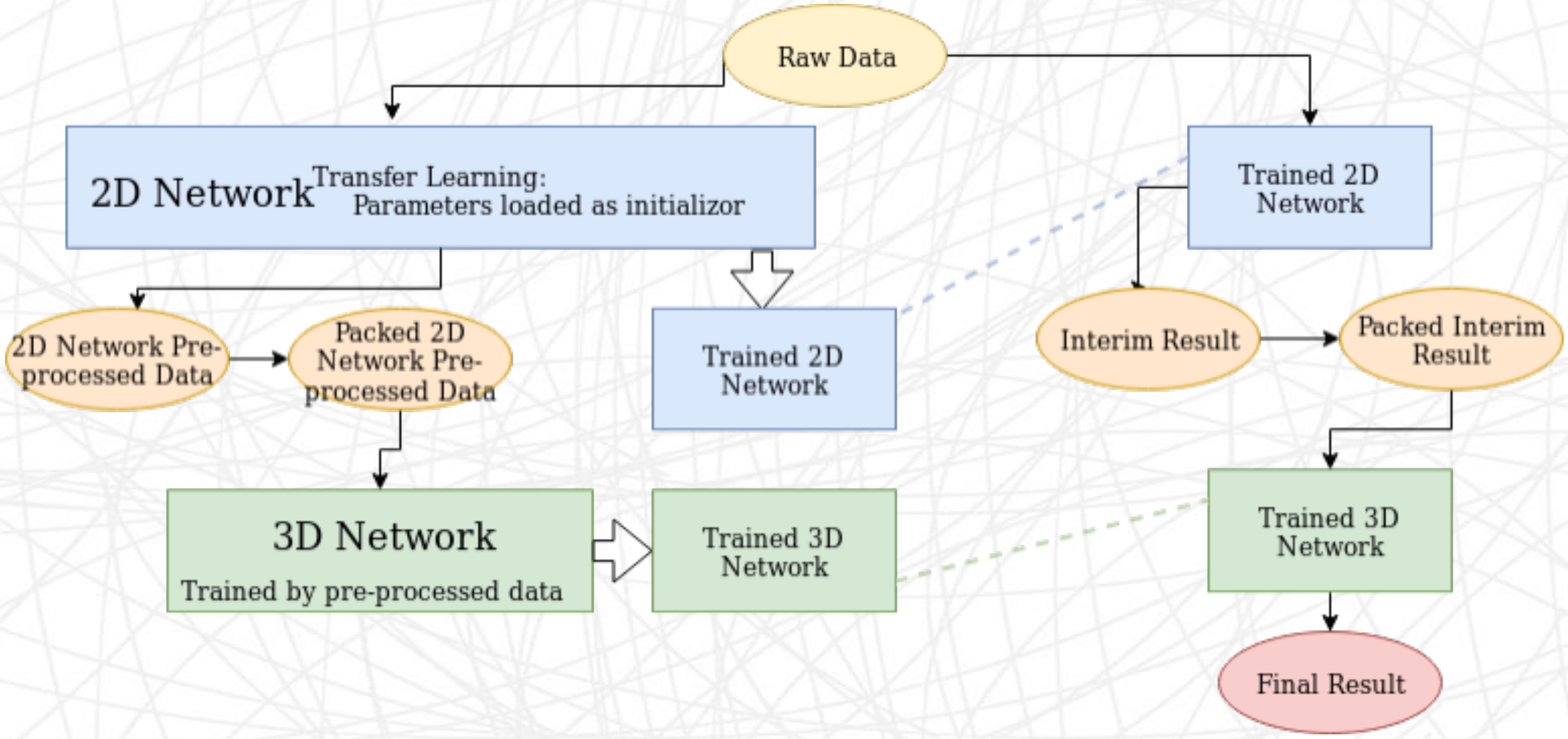
$$l(\mathbf{x},\mathbf{y}) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2\mu_y^2 + C_1},$$

$$c(\mathbf{x},\mathbf{y}) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2\sigma_y^2 + C_2},$$

$$s(\mathbf{x},\mathbf{y}) = \frac{2\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}.$$

02 Train 2D and 3D networks separately

The overall structure of this method



02 Train 2D and 3D networks separately

Easy to train and implement. Can set different parameters for two models to obtain better result.

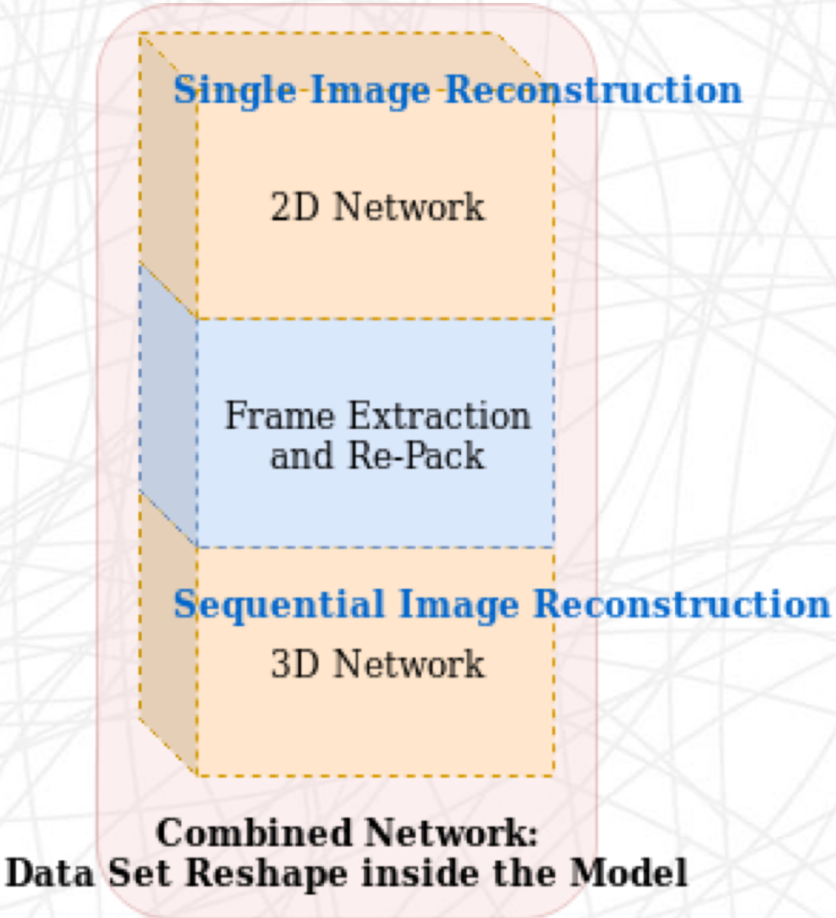
Advantages

Two separate networks mean less interaction between 2D and 3D models, which may influence the overall performance.

Disadvantages

02 Concatenate 2D and 3D models in one network

Add a Lambda Layer as the bridge of our 2D model and 3D model. This layer aims to pack the output of the 2D model to be small batches (each batch contains 5 consecutive frames), and then pass the packed frames to the 3D model.



02 Concatenate 2D and 3D models in one network

Training 2D and 3D models as a whole may make them adapt to each other. Hopefully, we can obtain better result than training them separately

Advantages

Using this method, we have to set same hyperparameters for two models.
Training a deeper network is more difficult than training two small ones

Disadvantages

PART

03

Performance

3.1

2D Model

Epoch = 150

PSNR: Peak Signal-to-noise Ratio (dB)
SSIM: Structural Similarity Index
MSE: Mean Squared Error

Totally 120 images in data set	Train Set (80% of all train pictures)			Validation Set (20% of all train pictures)		
	MSE	SSIM	PSNR	MSE	SSIM	PSNR
Epoch = 0	0.1787	0.0206	7.6772	0.0853	0.0334	10.6893
Epoch = 150	6.4832e-5	0.9871	41.8839	7.0370e-5	0.9859	41.5278

3.2 ## 3D Model

Data set input into the 3D model could be trained or not trained.

RAW DATA

Climate data upscaled by interpolation method only, not pre-trained by 2D FSRCNN network.

Raw Data

⇩

3D Network

⇩

Output

PROCESSED DATA

Climate data upscaled by 2D FSRCNN network. This pre-processed set passed to 3D SRnet for training and resolution improvement.

Raw Data ⇨ 2D Network

⇩

Interim Result

⇩

3D Network

⇩

Output

3.2.1 3D Model trained with Climate Data

Climate data not pre-processed; Epoch = 100

PSNR: Peak Signal-to-noise Ratio (dB)
SSIM: Structural Similarity Index
MSE: Mean Squared Error

Raw data quality: PSNR = 13.8458 dB; SSIM = 0.5304; MSE = 0.0413

Totally 120 images in data set	Train Set (80% of all train pictures)			Validation Set (20% of all train pictures)		
	MSE	SSIM	PSNR	MSE	SSIM	PSNR
Epoch = 0	0.0342	0.6628	14.7373	0.0280	0.6572	15.5334
Epoch = 100	0.0226	0.7377	16.4686	0.0231	0.7262	16.3682

After Processed: PSNR = 16.4487 dB; SSIM = 0.6995; MSE = 0.0227

3.2.2

3D Model trained with Pre-processed Climate Data

Climate data **pre-processed** to **PSNR = 42.09dB**

PSNR: Peak Signal-to-noise Ratio (dB)
SSIM: Structural Similarity Index
MSE: Mean Squared Error

Totally 120 images in data set	Train Set (80% of all train pictures)			Validation Set (20% of all train pictures)		
	MSE	SSIM	PSNR	MSE	SSIM	PSNR
Epoch = 0	0.0750	0.7401	15.3667	0.0019	0.9306	27.2128
Epoch = 80	4.5721e-05	0.9907	43.4010	4.9283e-05	0.9896	43.0746

Test Set:

Pre-processed data quality: PSNR = 42.0952 dB; SSIM = 0.9877; MSE = 6.1786e-05

After Processed: PSNR = 43.3459 dB; SSIM = 0.9911; MSE = 4.6327e-05

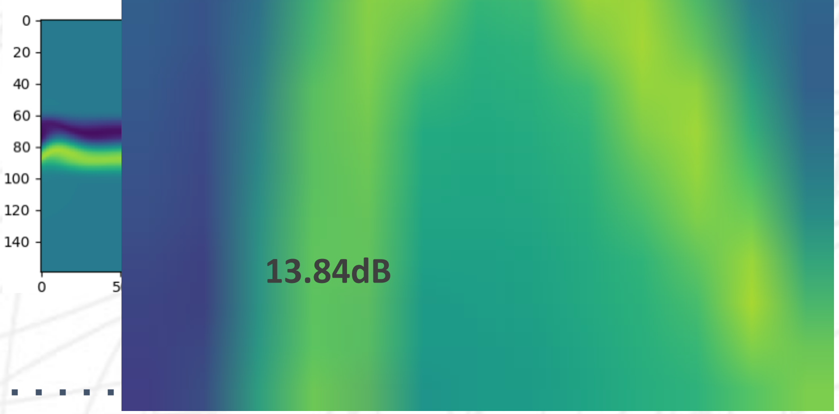
3.3 Combine 2D Network and 3D Network

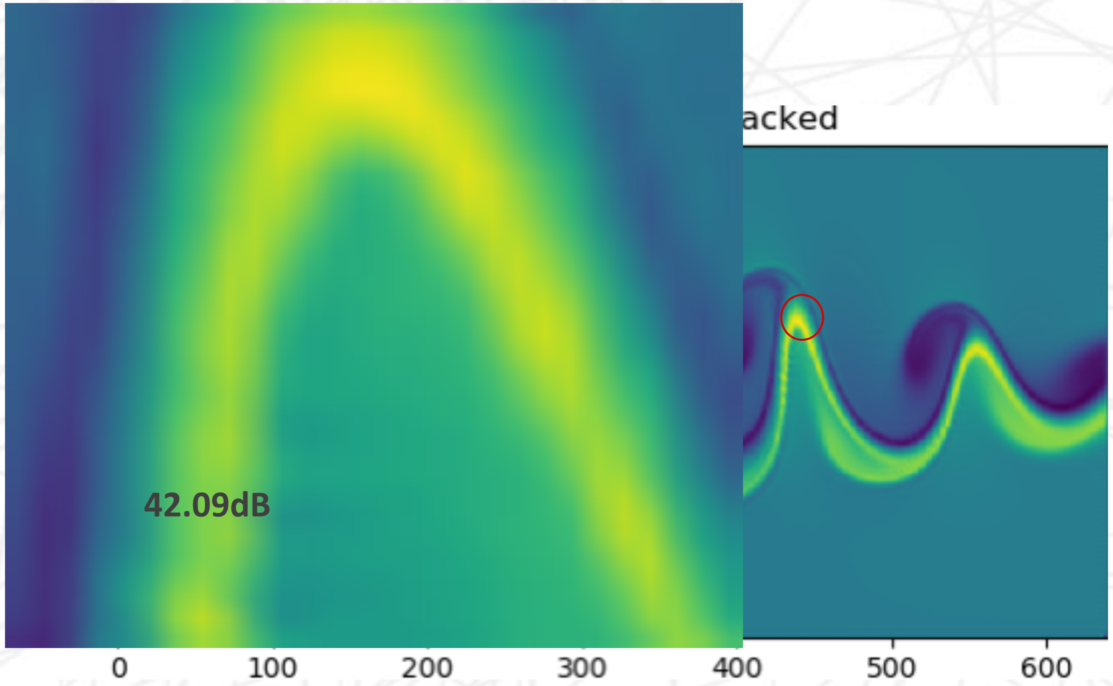
PSNR: Peak Signal-to-noise Ratio (dB)
SSIM: Structural Similarity Index
MSE: Mean Squared Error

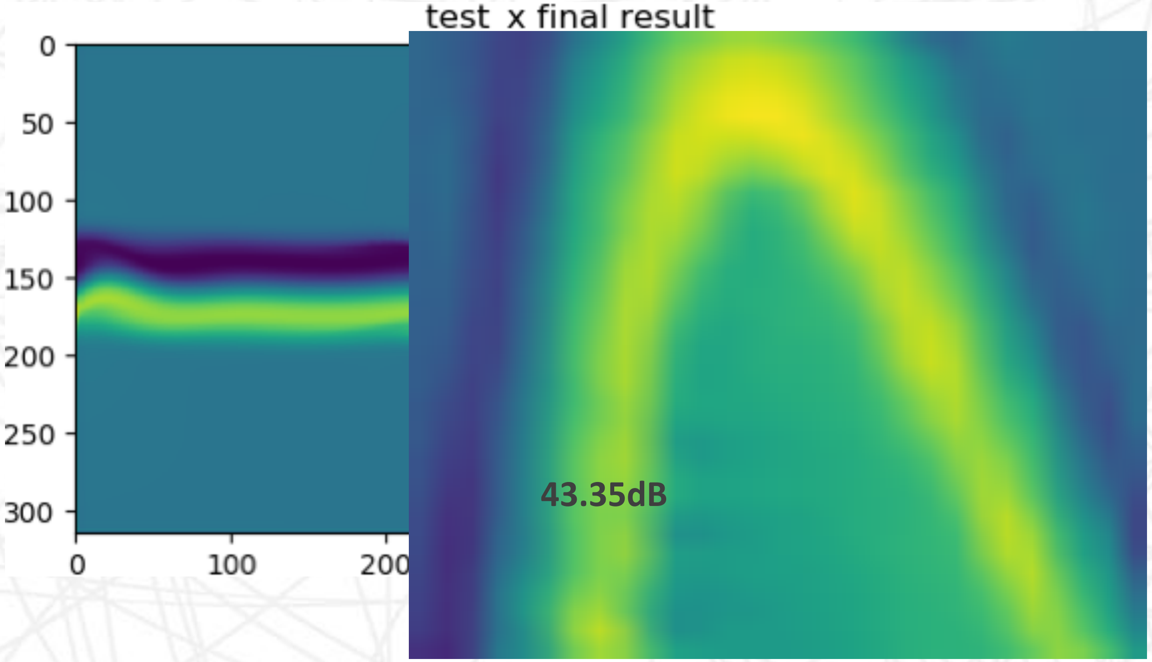
Totally 120 images in data set		MSE	SSIM	PSNR
3D network trained with **raw data**	2D network	7.593760673e-05	0.9847759507246662	41.201901955249504
	After data re-packing	8.124760623839217e-05	0.9846107211562674	40.9071674094057
	3D SRnet network	0.0096011151711	0.803450561713647	20.17703292426752
3D network trained with 2D network pre-processed result	2D network	5.807663600363179e-05	0.987824159428385	42.36632787676039
	After data re-packing	6.178588381302932e-05	0.9877120117274287	42.09522054489414
	3D SRnet network	4.632699437652877e-05	0.9910728454576089	43.34586764073154

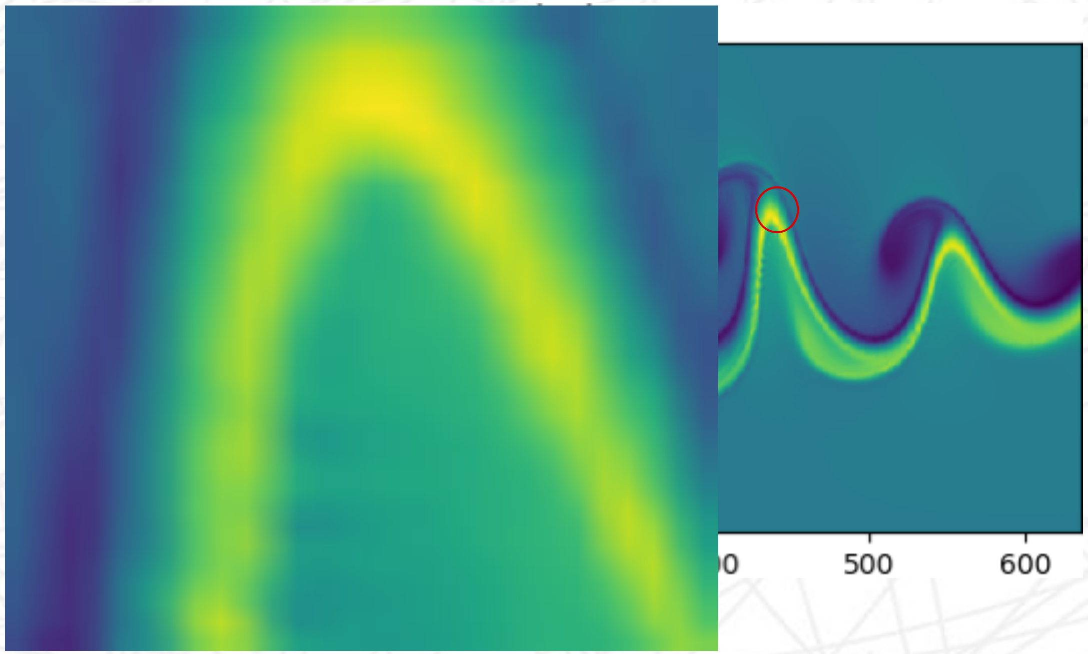
3.3 Combine

or









PART

04

Future Work

4.1 Dataset Selection

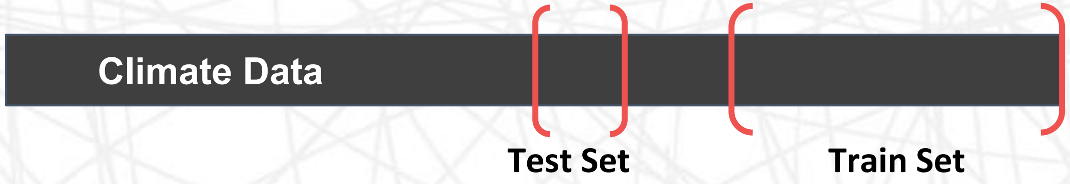
When picking train set and test set, should we use the same video?

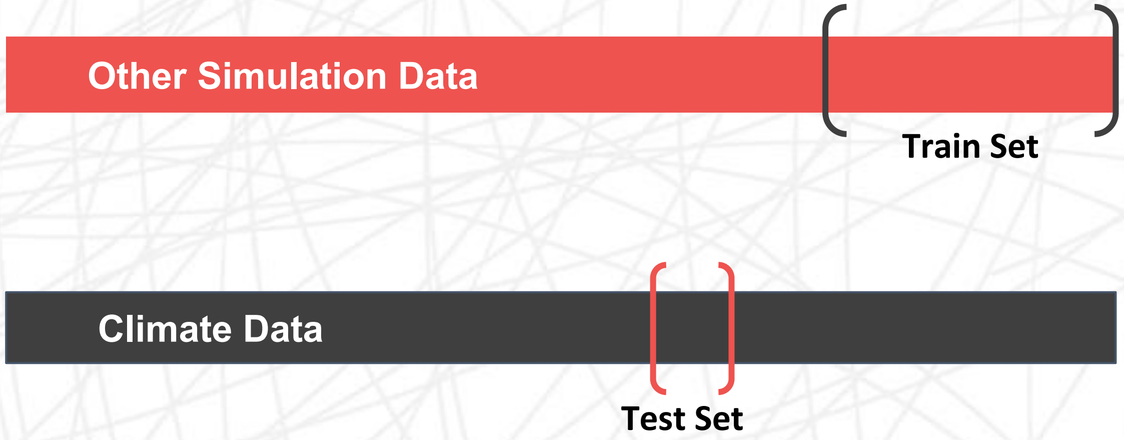
Specific

Train set and test set will have many common scenes, better for the performance; If train set and test set are too much similar, it may cause overfitting problem.

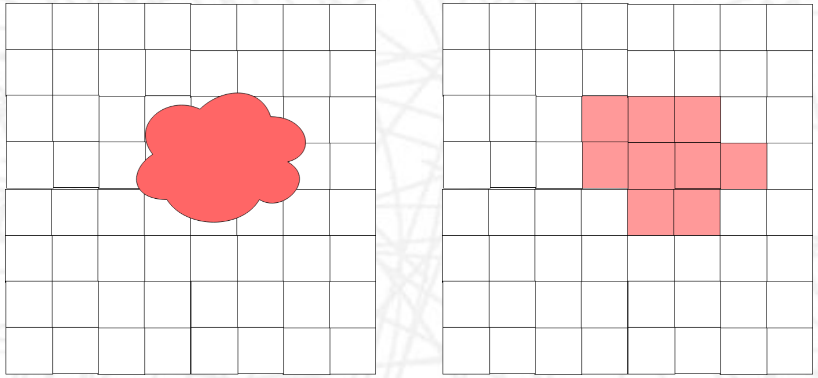
General

By generalizing the dataset, we can avoid overfitting problem. But this may compromise the performance of neural network on test set.



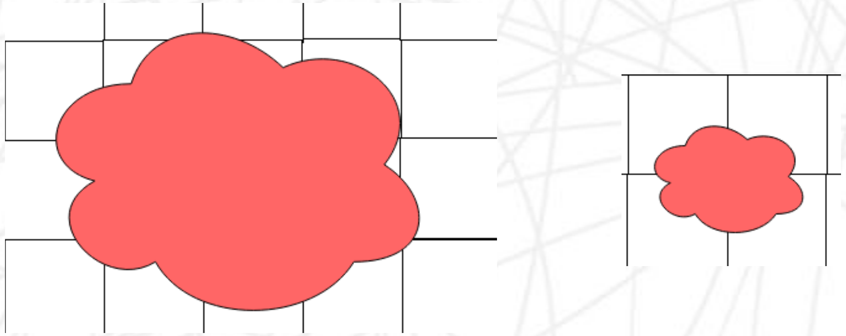


4.2 Compression and Contraction



Compression:

Including lossless compression and lossy compression, is to minimize the storage of redundant information inside images.



Contraction:

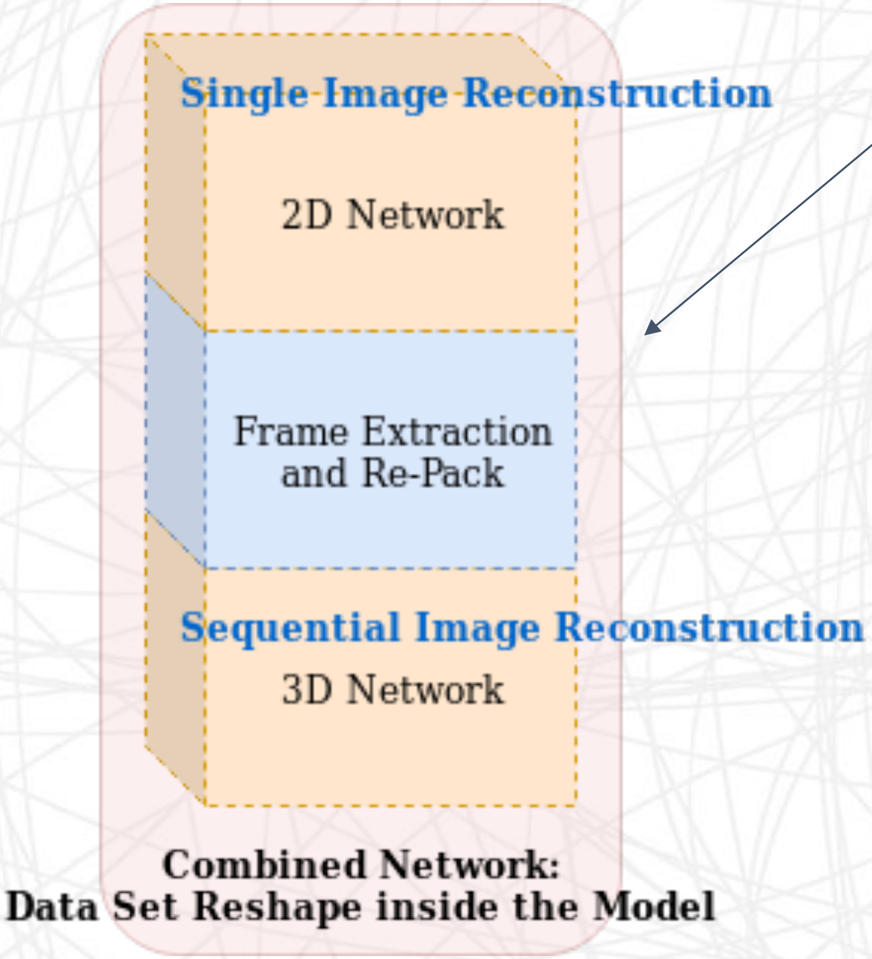
Using the natural of information lost in zoom-out process to throw information away.

The basic models we referred to are doing **super-resolution**, which generally enhances resolution by **enlarging the total amount of pixels**. These method should be more suitable for cases that lost information because of contraction, but not for all cases of general compressions.

Besides super-resolution, there are some other methods to enhance image resolution. For example, compression artifacts removal can sense distorted parts and reconstruct based on these entries.

4.3 Concatenated Two-in-One Combined Model

Break through the peak values of two models by designing a new one.



Challenge: need to use iterations inside the lambda layer, and the index for iteration is the batch size of each input tensor. But we cannot get the exact number of batch size since the number is dynamically stored.

Move the data packing process to the beginning of neural network training.

The structure of 2D network need to be modified respectively;

2D network needs to process a lot of duplicated frames.

Data Pre-processing

Array programming allows the application of operations to an entire set of values at once.

Avoid the usage of for-loop;

Keras and tensorflow are developed steadily. Hopefully, MagmaDNN could help.

Vectorise Operation

4.4

"3D + 2D" Network Model

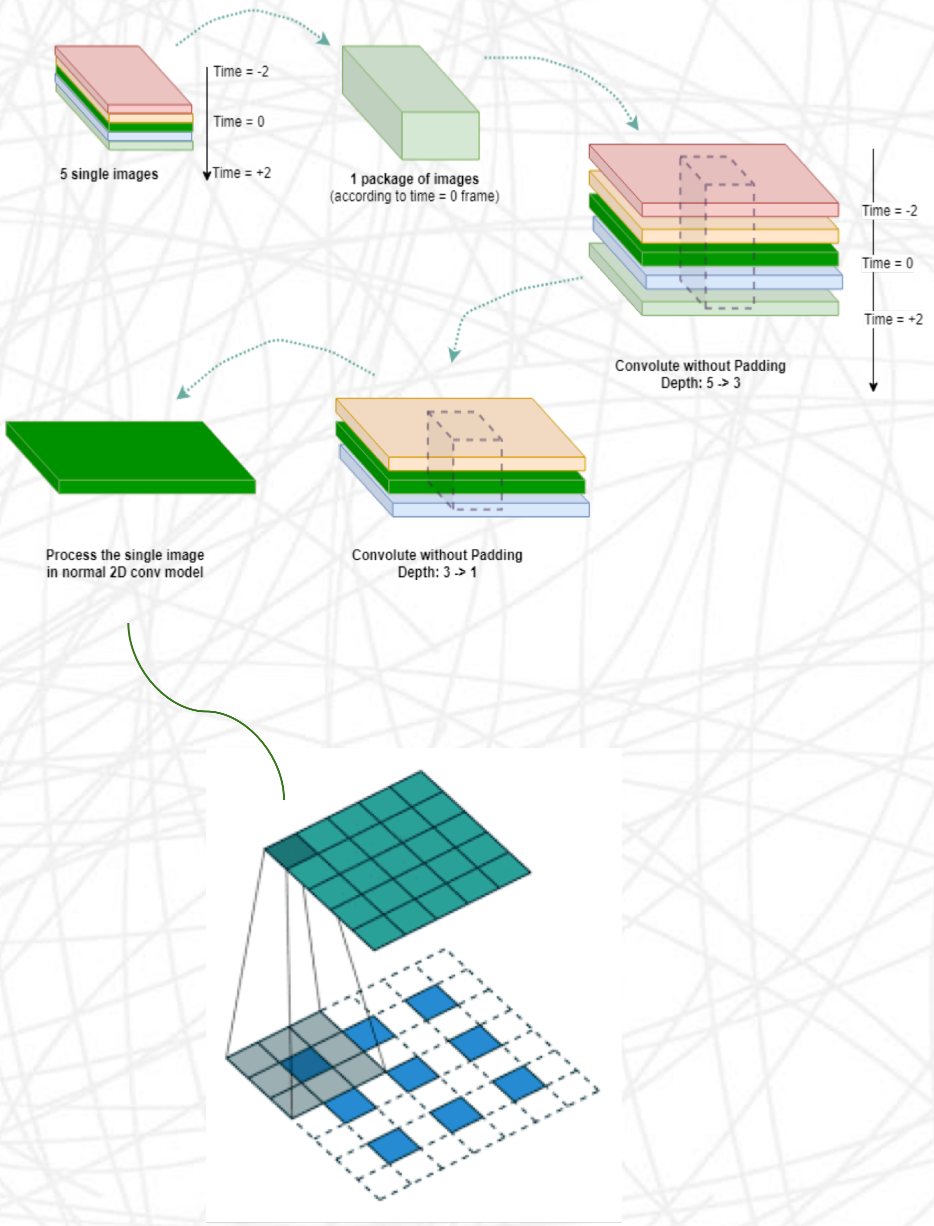
Extracting information from consecutive frames first, reconstruct from independent images later.

"Cask Effect":

For each model and dataset, there should be a peak value that the result could be improved to by neural network processing. The inter-compensation idea of combination does not mean the peak values are also combined.

3D network has more complicated layers and difficult to learn. The peak value should be lower than 2D network.

Break through the limitation of 3D network by processing 2D network after 3D network.



PART

05

Conclusion

01

Dataset Gathering and Processing

Collecting independent images, videos and simulation data as training data; Downscaling these data using bicubic interpolation.

02

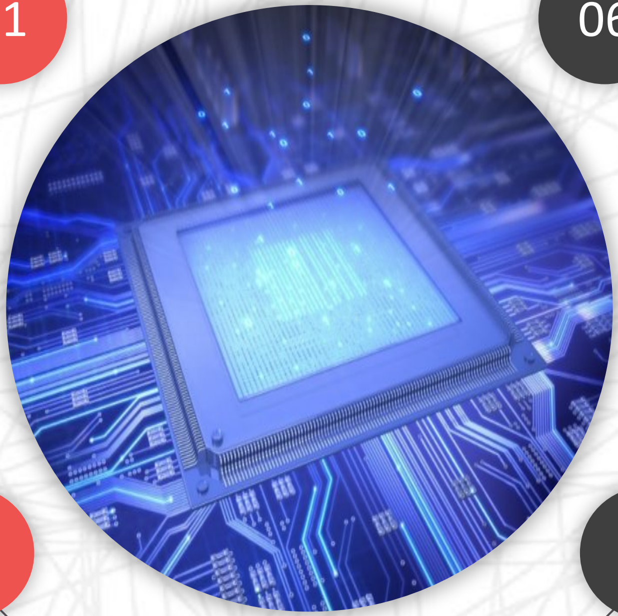
2D Network Model Construction

Using 2D network with the concept of transfer learning to scale up small images to larger ones, comparing the result with ground truth.

03

3D Network Model Construction

Packing consecutive images into packages; 3D network processes images in time domain.



04

2D and 3D Network Combination

Combining the effect of independent reconstruction and sequential image reconstructions.

05

Hyper-parameter Tuning

Tuning hyperparameters to run out more satisfying results.

06

Analysis and Discussion

According to the performance and current progress, set possible goals for future work.

Reference

- Huang T, Yang J. Image super-resolution: Historical overview and future challenges. In Super-resolution imaging 2010 Sep 28 (pp. 19-52). CRC Press.
- Yue L, Shen H, Li J, Yuan Q, Zhang H, Zhang L. Image super-resolution: The techniques, applications, and future. Signal Processing. 2016 Nov 1;128:389-408.
- 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.
- 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.
- Dong C, Loy CC, Tang X. Accelerating the super-resolution convolutional neural network. In European conference on computer vision 2016 Oct 8 (pp. 391-407). Springer, Cham.
- Dumoulin V, Visin F. A guide to convolution arithmetic for deep learning. arXiv preprint arXiv:1603.07285. 2016 Mar 23.

Q&A

MANY THANKS !