Video Super Resolution with Integrated 2D and 3D Convolution Neural Network

CheukHin HO, The Chinese University of Hong Kong; JiaXin LI, The Chinese University of Hong Kong

Mentor: Dr. Kwai Wong, Joint Institute of Computational Sciences

Introduction

Background

In most deep-learning based video super-resolution algorithms, a 2D CNN model is applied to each frame for image super-resolution ahead of the 3D CNN.

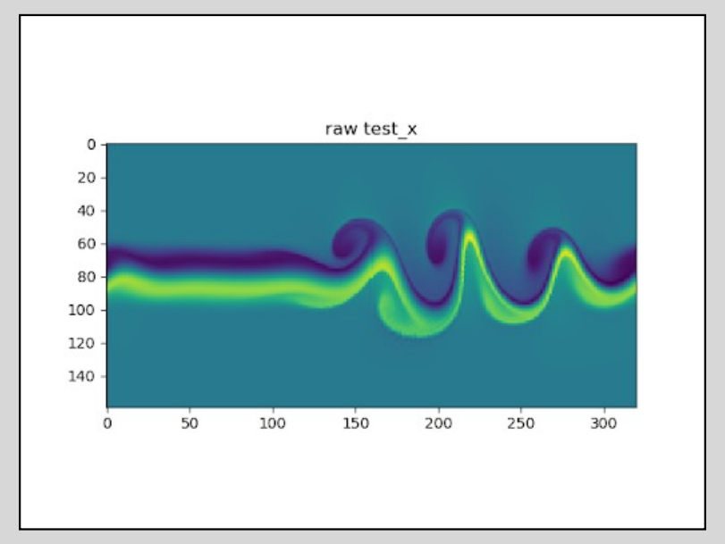
However, the temporal information in each frame, which is useful to super-resolve the middle frame, is different. As a result, using a single 2D model to pre-process all the frames in the sequence may not preserve the diverse information in each frame.

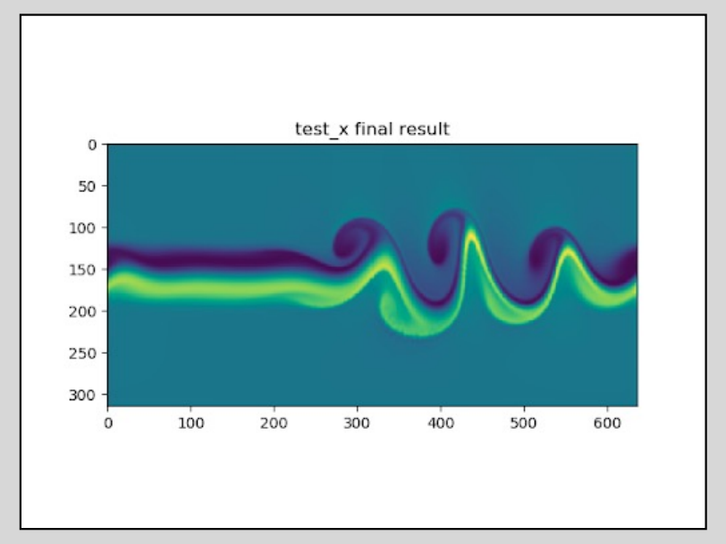
Objective

Our objective is to design a method to do video super-resolution with 2D and 3D convolution neural network, which pre-processes each frame differently.

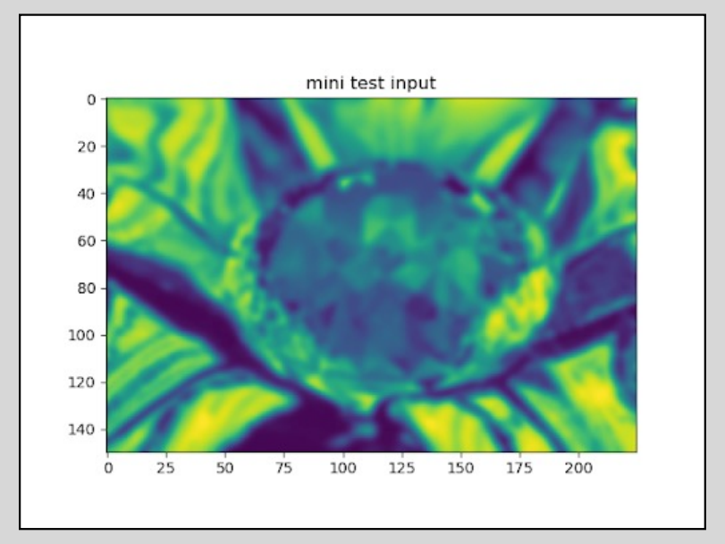
Datasets

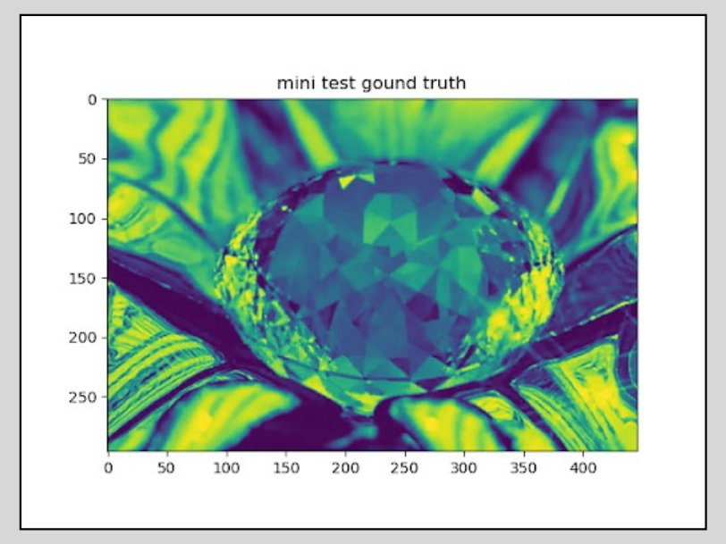
- climate data: to compare our models with previous ones





- diamond data: to adjust our model for more complex cases





Definition

- Mean squared error (MSE)

$$MSE = \frac{1}{mn}\sum_{i=0}^{m}\sum_{j=0}^{n} I(i,j) - K(i,j)$$

- Peak Signal to Noise Ratio (PSNR)

$$PSNR = 10\log_{10}(\frac{MAX^2}{MSE})$$

- Structure Similarity Index(SSIM)

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

Model One: 5-parallel

The model 5-parallel has two parts: the first part consists of 5 2D convolution layers in parallel and the second part consists of a stack of 3D convolution layers, as shown in Figure 1.

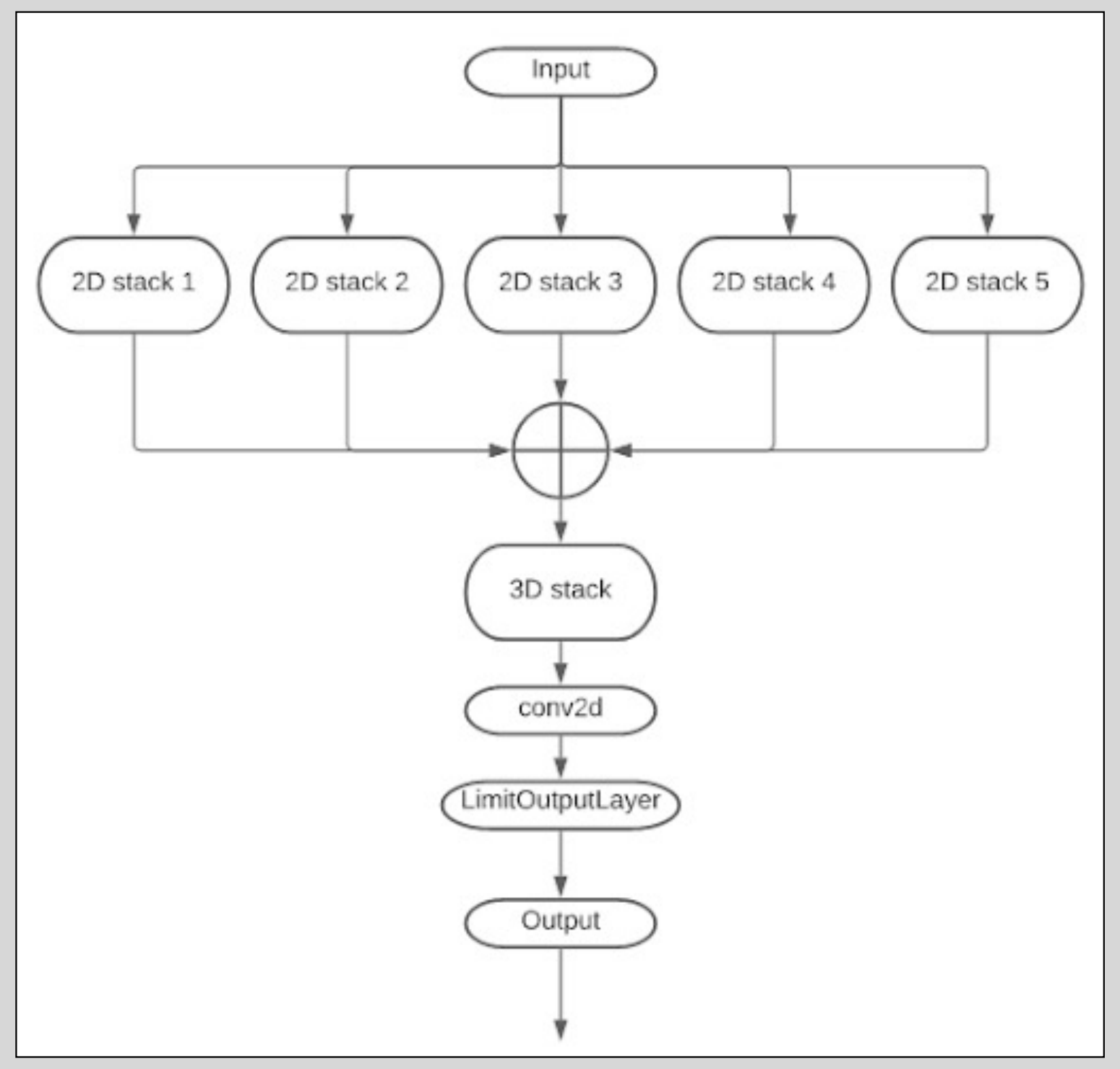


Figure 1: Structure of model 5-parallel

For every input stack of 5 frames, it splits up and each frame goes through different 2D stacks for independent 2D feature extraction of each frame (Figure 2). In particular, there is a small recursive block in the middle of each 2D stack.

The 5 outputs are then concatenated together and fed into the 3D part (Figure 3). The 3D part gathers the information obtained from 2D parts (of each frame) to perform super-resolution. Similarly, there are also two small 3D recursive blocks in the middle of the 3D stack. Throughout 2D and 3D parts, ReLU layers are used to add non-linearity to the model after each convolution layer.

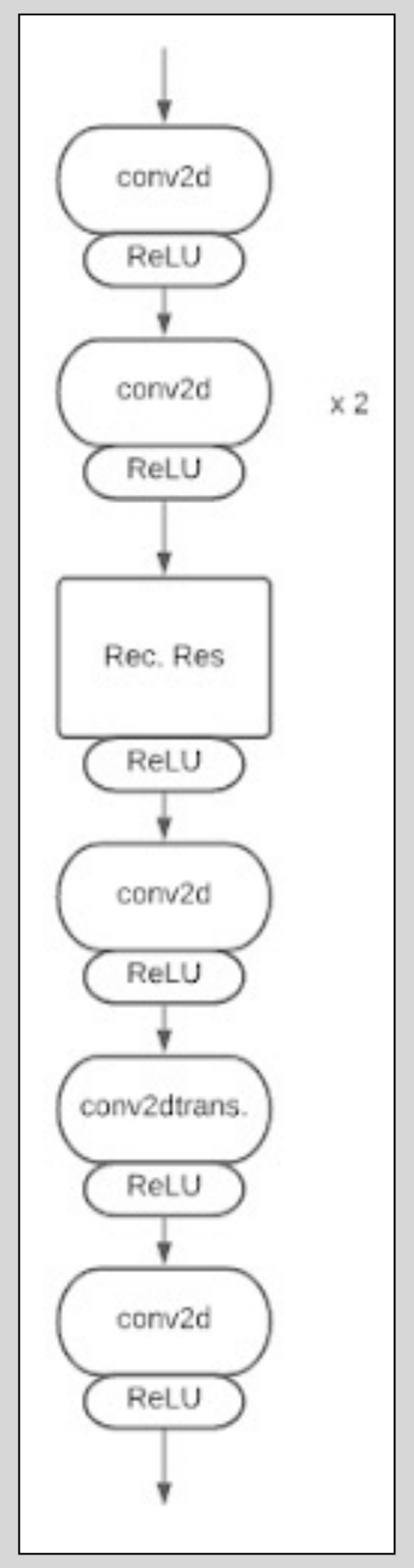


Figure 2

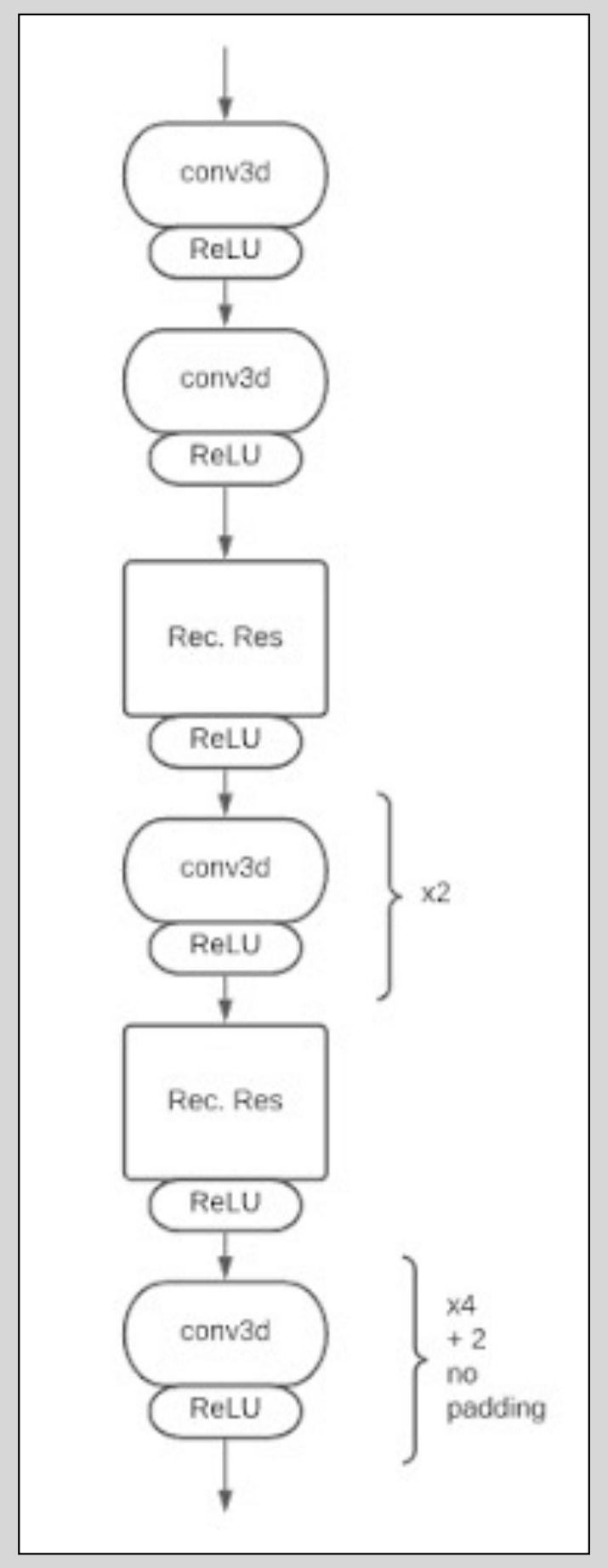


Figure 3

Through training and back propagation while updating the weight of the network, we expect the 3D part to interact with the 2D part and help each other to learn to super resolve the image.

Model Two: Delta

Similar to 5-parallel, the Delta model is also composed of 2D convolution layers and 3D convolution stack.

The biggest difference between 5-parallel and Delta lies in the design of the 2D convolution part. Instead of using 5 parallel 2D convolution networks, the Delta model achieves the target of processing each frame differently in another manner to reduce the number of parameters.(Figure 4)

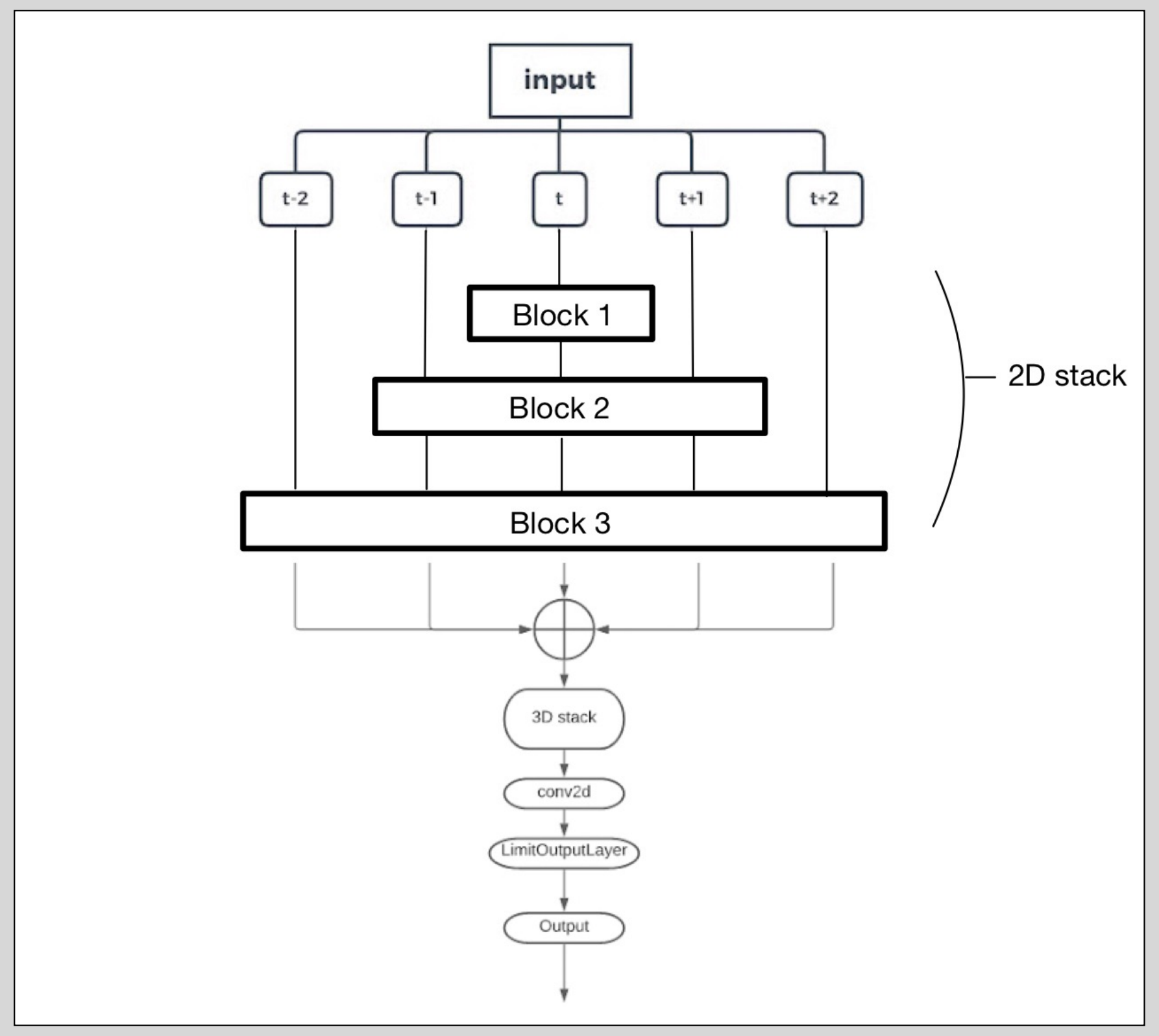


Figure 4: Structure of model Delta

With the idea of symmetry, the Delta model lets each frame pass through different numbers of 2D layers in the 2D convolution network. In this way, the 2D stack extracts the information of each frame respectively without introducing extra parameters.

In addition, Delta also adopts recursive learning to add complexity and depth.

Transfering from Tensorflow to Pytorch

The previous work was implemented in Tensorflow. There are few reasons for transferring from Tensorflow to Pytorch.

- Many source code available are written in pytorch. If our code is also written in pytorch, it will be easier for us to combine or test our work with others' models.
- Some features are not directly supported in Tensorflow directly, whereas Pytorch has no such problem. It allows development of structure of complex networks easily.
- Parallel computation is emphasized in Pytorch and it has good compatibility with cuda packages.
- Memory-wise, Pytorch has better memory allocation than Tensorflow, which facilitates testing and training on local computers before moving on to XSEDE.

Performance and Analysis

- On climate data

	MSE	SSIM	PSNR
Raw	0.03296	0.6517	14.8202
Previous	4.9283e-05	0.9896	43.0746
Delta	1.9801e-05	0.9943	46.0013
5-parallel	2.8866e-05	0.9953	45.6618

- On diamond data

	MSE	SSIM	PSNR
Raw	9193.6647	0.0392	8.5000
Delta	125.1786	0.9155	27.1675
5-parallel	127.5626	0.9184	27.0820

Indeed, the number of parameters in 5-parallel is far more than that in Delta, which costs much more memory .

However, 5-parallel can be implemented parallel, which means the computational time cost of 5-parallel should be similar to that of Delta using similar layer depth.

Future work

- Optical flow

For each pixel, an optical flow with respect to n-th and n+1-th frames record has two values, x and y, which denotes the change in x and y respectively from the n-th to the n+1-th frame. (Figure 5)

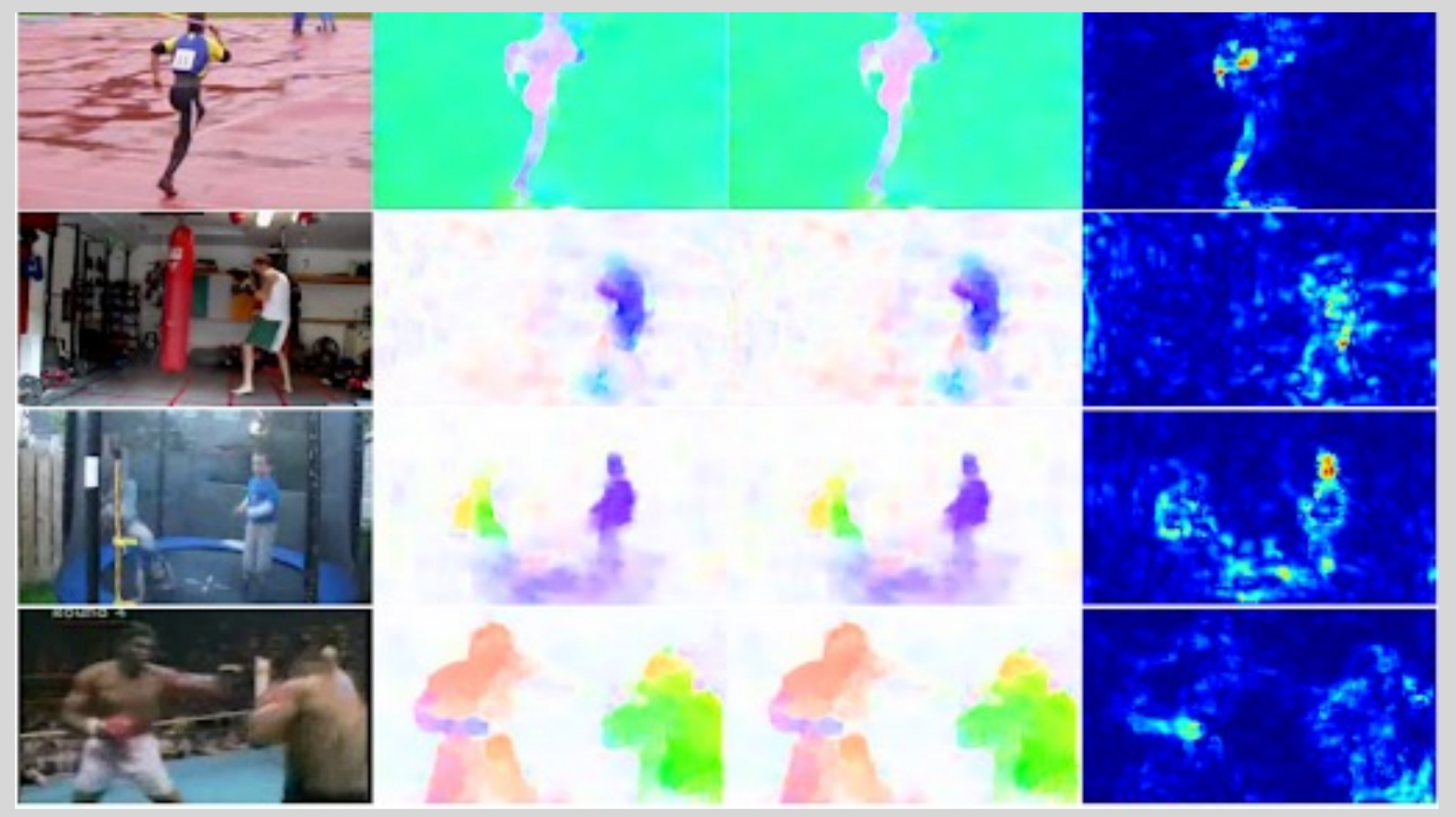
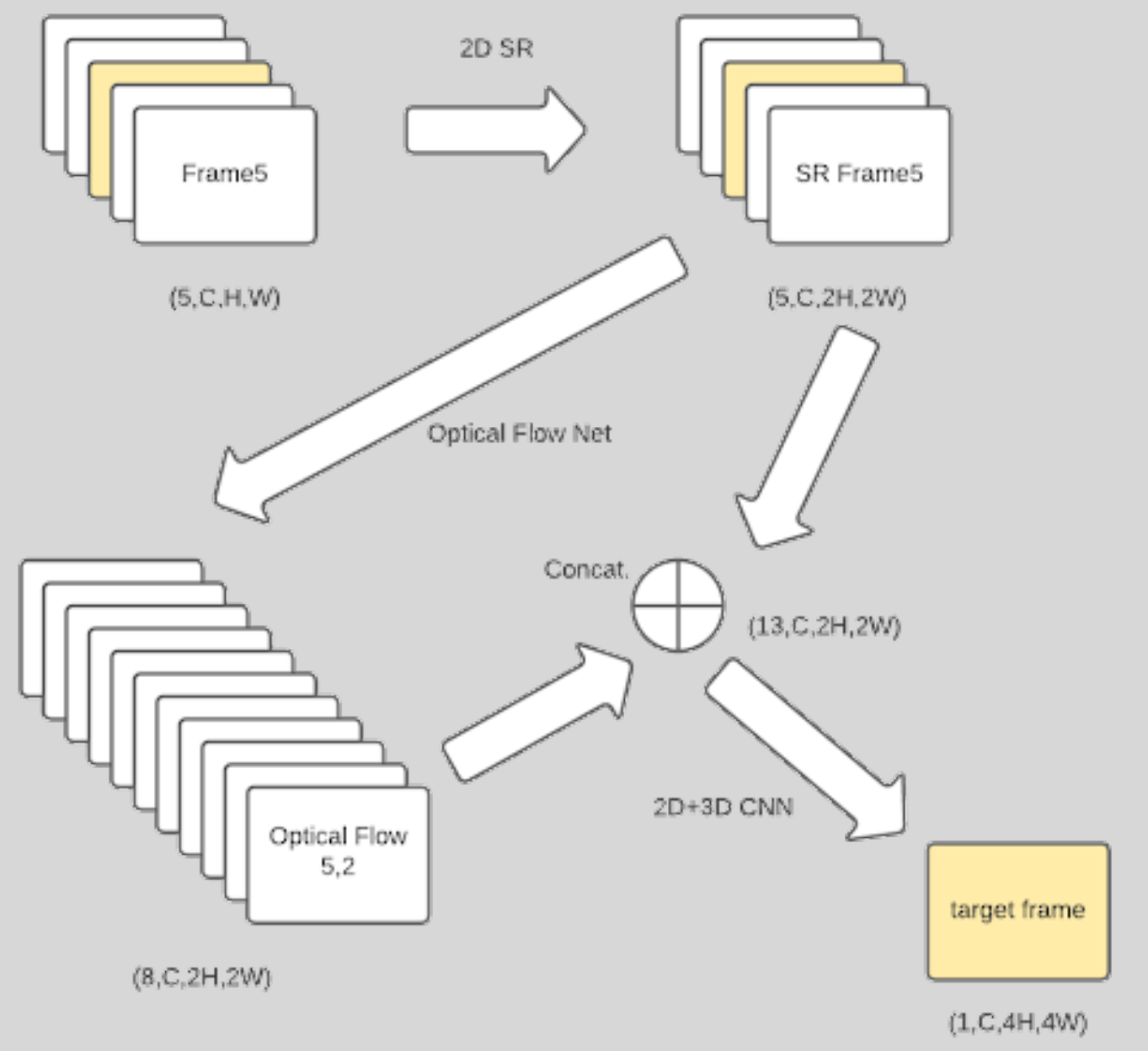


Figure 5: : Example of Optical Flow

- Proposed future model


Frame5
(5,C,H,W)
2D SR
SR Frame5
(5,C,2H,2W)
Optical Flow Net
Concat.
(13,C,2H,2W)
Optical Flow 5,2
(8,C,2H,2W)
2D+3D CNN
target frame
(1,C,4H,4W)