VIDEO SUPER RESOLUTION BASED ON DEEP CNN WITH TWO-STAGE MOTION COMPENSATION

https://ieeexplore.ieee.org/document/8551569

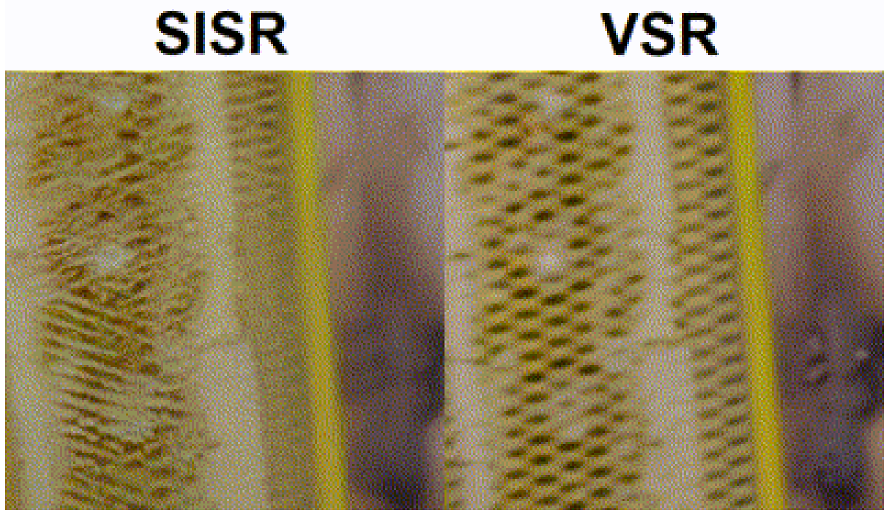
Department: AI/Software

Name: Cho, HyunMin

Background

- **Video Super-Resolution** is the process of generating high-resolution video frames from the given low-resolution ones.

  Unlike single image super-resolution (SISR), the main goal is not only to restore more fine details while saving coarse ones, but also to preserve motion consistensy.



- The most research works consider degradation process of frames.

$$\{y\} = (\{x\} * k) \downarrow_s + \{n\}$$ Reverse

Background

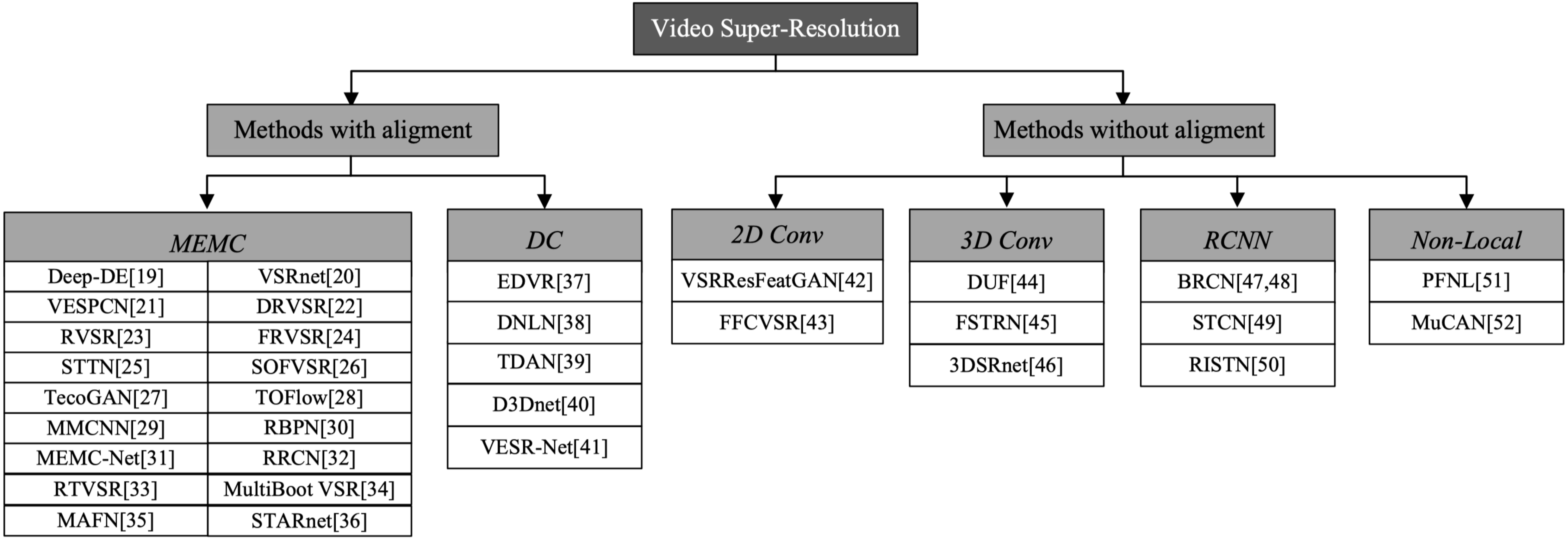


Fig. 2: A taxonomy for existing state-of-the-art video super-resolution methods. Here, $MEMC$ stands for motion estimation and compensation methods, DC is deformable convolution methods, $3D\ Conv$ is 3D convolution methods, and $RCNN$ denotes recurrent convolutional neural network based methods.

Introduction

- Video SR aims to retrieve a HR video based on the inputs from a LR video.
- In contrast to SISR, that the details can rely on external examples.

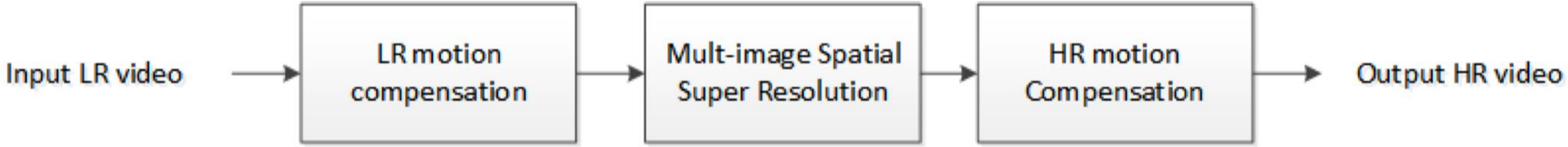
- <u>In general, a video SR system consists of two parts</u>

  > the temporal alignment module which applies **motion compensation**

  > to keep the **temporal consistency** of the output video

  > the spatial super resolution module

  > aims to retrieve the HR texture from the LR input.



Introduction

- Performance of VSR is strongly depends on how accurate the motion is estimated.

  > Dense *Optical flow / Neural Network based > **Very expensive!**

- Motion compensation is only performed at HR domain > limits the quality

- To solve this problem, VSR-TMC suggest…

  1. operate motion compensation in LR, HR domain
  2. get HR optical flow in two ways

     i) upsample LR's optical flow

     ii) applying FlowNet2 to HR img≈

Overview of VSR-TMC

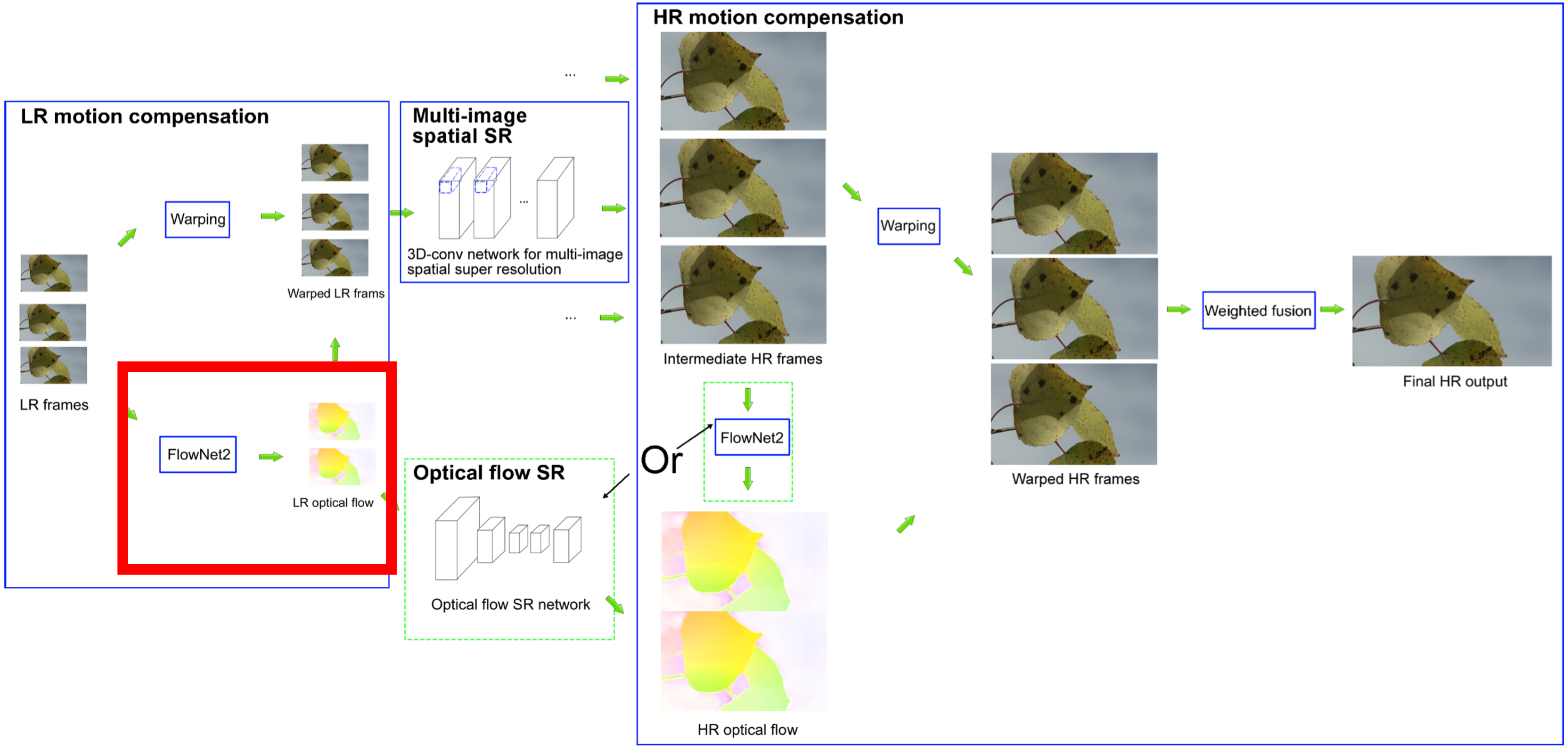


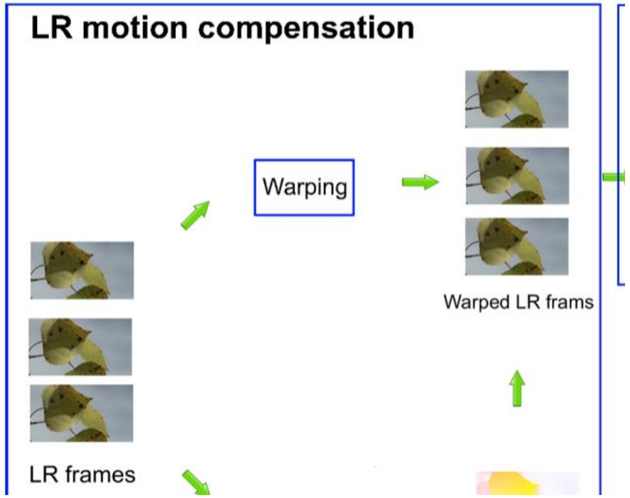
Fig. 2. Overview of the proposed VSR-TMC network. It consists of 3 parts: LR motion compensation, multi-image spatial SR, and HR motion compensation. In the HR motion compensation, the optical flow can be generated either by applying FlowNet2 on intermediate HR frames, or by a optical flow SR network, shown as the two dashed boxes.

LR Motion Compensation

$$I'^{L}_{t-1} = W^{L}_{t-1,t} I^{L}_{t-1}$$

$$I'^{L}_{t} = I^{L}_{t}$$

$$I'^{L}_{t+1} = \boxed{W^{L}_{t+1,t}} I^{L}_{t+1}$$



- Notations

  I: Input image

  I': Warped Input image

  L/H: Low/High resolution

  T: frame

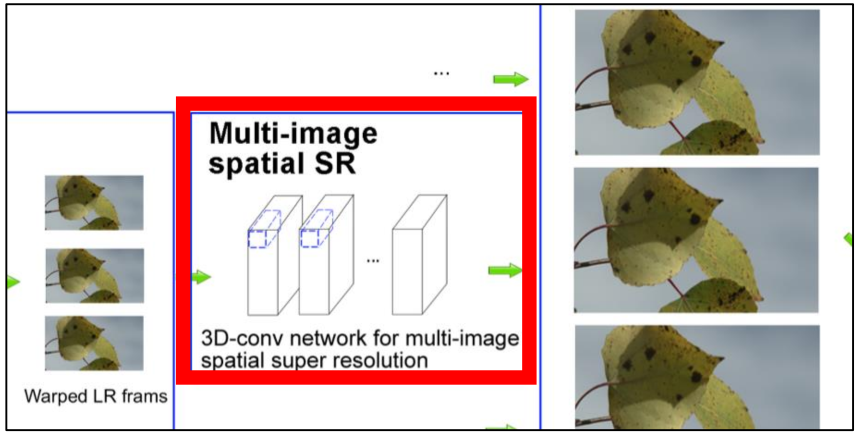
HOW TO 'WARP'?

-> USE OPTICAL FLOW

$$F^{L}_{t-1,t}, F^{L}_{t+1,t}$$

Multi-image Spatial Super Resolution

- Apply a multi-image spatial SR network on every three warped LR, to generate single intermediate HR frame.



- use a 3D convolution (3D-conv) network.
  > keep more temporal information
  > based on the deep SRCNN. (*FIG. 3*)

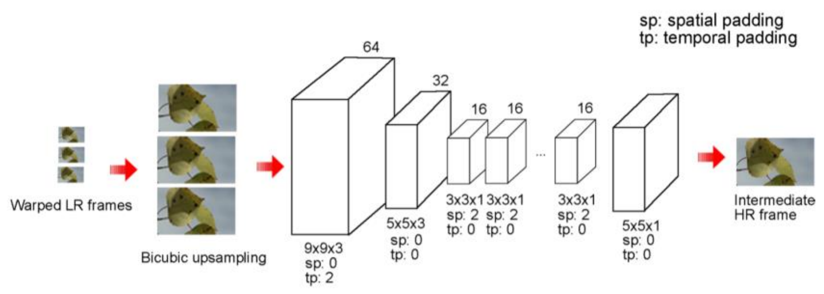


Fig. 3. Examples of multi-image spatial SR network based on 3D convolution. 'sp' is the zero padding in the spatial direction, 'tp' is the zero padding in the temporal direction.

- 2-pixels [*zero padding] in the temporal direction is applied in the first 3D-conv layer to keep the temporal depth.

HR Motion Compensation – I

- Many of the previous video SR research focus on LR motion compensation only.

  > However, using HR motion compensation can enhance the quality of the output video.

- Introduce second motion compensation stage in the HR domain.

- Warped HR image is created through HR optical flow obtained by

  1) applying FlowNet2 to Intermediate HR frame

  2) HR optical flow by SR of LR optical flow image.

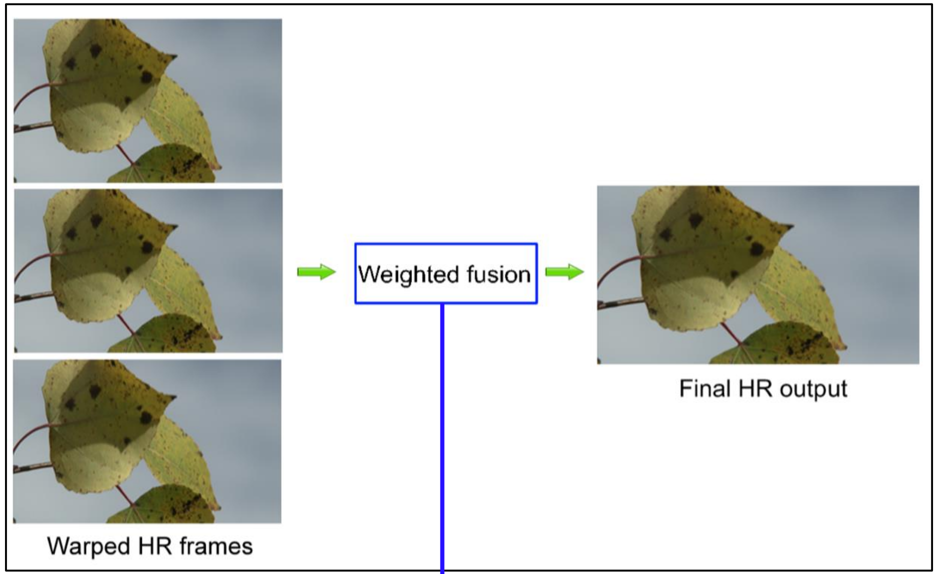
$$I'^{H}_{t-1} = W^{H}_{t-1,t} I^{H}_{t-1}$$

$$I'^{H}_{t} = I^{H}_{t}$$

$$I'^{H}_{t+1} = W^{H}_{t+1,t} I^{H}_{t+1}$$

HR Motion Compensation – II

- After the warped image is created, a final HR output image is created by applying <u>weights</u>.



$(weight\colon\ w_t)$

: weight matrix having the same size as the warped HR frame

$$w_k = \{\boxed{w_{k,i,j}}, 1 <= i <= C_k, 1 <= j <= D_k, k = \{t-1, t, t+1\}\}$$

$$I_t^{*H} = \frac{w_{t-1} \cdot I'^{H}_{t-1} + w_t \cdot I'^{H}_{t} + w_{t+1} \cdot I'^{H}_{t+1}}{w_{t-1} + w_t + w_{t+1}}$$

$$w_{k,i,j} = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{\boxed{-z^2_{k,i,j}}}{2\sigma^2}}$$

magnitude of the HR optical flow at pixel (i, j) in frame k

Optical Flow Super Resolution - Overview

- The only question left is how to generate the HR optical flow.

    As we saw earlier, we can apply FlowNet2 again on the intermediate HR frames.

    >> Too much computational cost

- Therefore, VSR-TMC suggest…

    > Generate the HR optical flow by applying an optical flow SR on the LR optical flow.

    [Advantages]

    1) the cost is much smaller compared to using FlowNet2 on the RGB images.

    2) optical flow SR could be parallel executed with multi-image spatial SR, which further reduces the computational cost.

Differences from image SR

- In image SR…

  1) the LR intensity is the same as the HR intensity, while the LR optical flow is smaller compared to the HR optical flow.

  2) the intensity in image SR ranges from [0,255], while such boundary does not exist in optical flow

  −> directly applying interpolation such as will not work well in optical flow SR

Solution

−> train a neural network, constructs a mapping from the LR optical flow to the HR optical flow

Implementation

- Use 5-layer Convolution layer. (9-5-3-3-5, 2D kernels / 64-32-16-16-1 filters).
- 2-pixels zero padding is applied in the two 3×3 layers

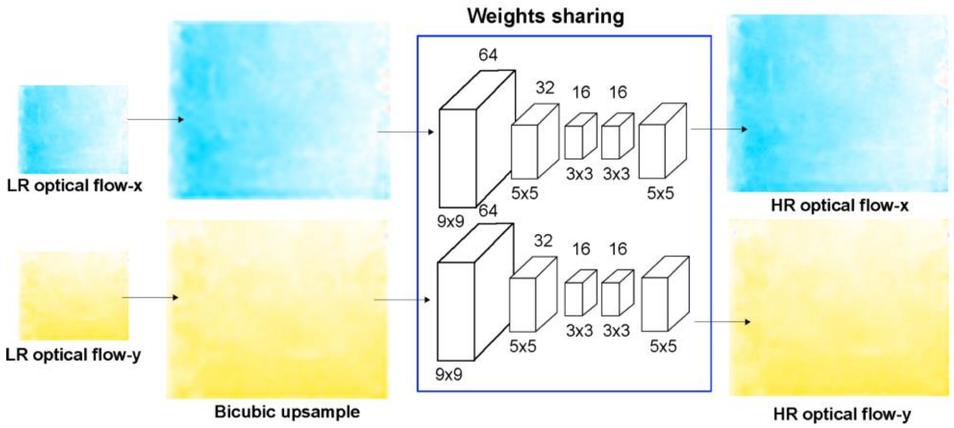


Fig. 4. 5-layer SRCNN for optical flow SR.

- optical flow has two channels

  : x direction & y direction

  > train SRCNNs on each channel

- Loss function: sum of MSE on optical flow x-channel & y-channel

$$loss_{flowSR} = \frac{1}{2}\sum_i (||F_{i,x}^H - G_{i,x}^H||^2 + ||F_{i,y}^H - G_{i,y}^H||^2).$$

Experiment in paper

- Testbed
  - Training Dataset: 96 videos from CDVL, LIVE Video Quality Assessment Dataset
  - Test Dataset: 4 videos from VideoSet4 (Vid4)
  - GPU: Titan X GPU

Type	Label	LR motion compensation	Spatial SR	HR motion compensation	PSNR/SSIM	Time (second/frame)
SISR	C1	-	SRCNN-19	-	27.12/0.8344	0.047
SISR	C2	-	SRCNN-19	FlowNet2	27.31/0.8398	0.218
VSR	C3	FlowNet2	3D-SRCNN-19	-	27.33/0.8414	0.068
VSR-TMC	C4	FlowNet2	3D-SRCNN-19	Bicubic	27.07/0.8308	0.068
VSR-TMC	C5	FlowNet2	3D-SRCNN-19	Flow SR	27.41/0.8451	0.070
VSR-TMC	C6	FlowNet2	3D-SRCNN-19	FlowNet2	27.57/0.8476	0.224
VSR-TMC	C7	FlowNet2	3D-SRCNN-19	Oracle	**27.63/0.8483**	0.224

Experiment in paper

- Testbed

  - Training Dataset: 96 videos from CDVL, LIVE Video Quality Assessment Dataset

  - Test Dataset: 4 videos from VideoSet4 (Vid4)

  - GPU: Titan X GPU

Type	Label	LR motion compensation	Spatial SR	HR motion compensation	PSNR/SSIM	Time (second/frame)
SISR	C1	-	SRCNN-19	-	27.12/0.8344	0.047
SISR	C2	-	SRCNN-19	FlowNet2	27.31/0.8398	0.218
VSR	C3	FlowNet2	3D-SRCNN-19	-	27.33/0.8414	0.068
VSR-TMC	C4	FlowNet2	3D-SRCNN-19	Bicubic	27.07/0.8308	0.068
VSR-TMC	C5	FlowNet2	3D-SRCNN-19	Flow SR	27.41/0.8451	0.070
VSR-TMC	C6	FlowNet2	3D-SRCNN-19	FlowNet2	27.57/0.8476	0.224
VSR-TMC	C7	FlowNet2	3D-SRCNN-19	Oracle	**27.63/0.8483**	0.224

Experiment in paper

- Testbed
  - Training Dataset: 96 videos from CDVL, LIVE Video Quality Assessment Dataset
  - Test Dataset: 4 videos from VideoSet4 (Vid4)
  - GPU: Titan X GPU

Type	Label	LR motion compensation	Spatial SR	HR motion compensation	PSNR/SSIM	Time (second/frame)
SISR	C1	-	SRCNN-19	-	27.12/0.8344	0.047
SISR	C2	-	SRCNN-19	FlowNet2	27.31/0.8398	0.218
VSR	C3	FlowNet2	3D-SRCNN-19	-	27.33/0.8414	0.068
VSR-TMC	C4	FlowNet2	3D-SRCNN-19	Bicubic	27.07/0.8308	0.068
VSR-TMC	C5	FlowNet2	3D-SRCNN-19	Flow SR	27.41/0.8451	0.070
VSR-TMC	C6	FlowNet2	3D-SRCNN-19	FlowNet2	27.57/0.8476	0.224
VSR-TMC	C7	FlowNet2	3D-SRCNN-19	Oracle	**27.63/0.8483**	0.224

Thank you!

Annex information

STARnet

A Space-Time-Aware Multi-Resolution Video Enhancement (CVPR2020)

https://github.com/alterzero/STARnet