Super Resolution

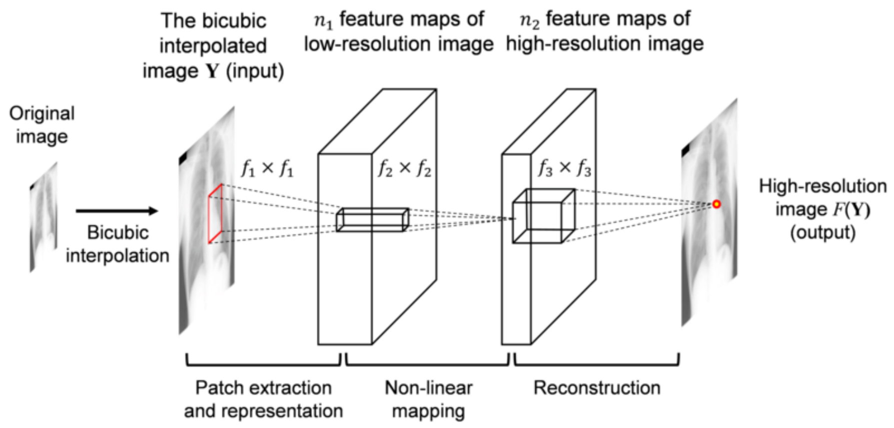
Super resolution is the process of upscaling and or improving the details within an image. Often a low resolution image is taken as an input and the same image is upscaled to a higher resolution


LR – Low Resolution Image
HR – High Resolution Image
SR – Image produced by model

DEEP SISR

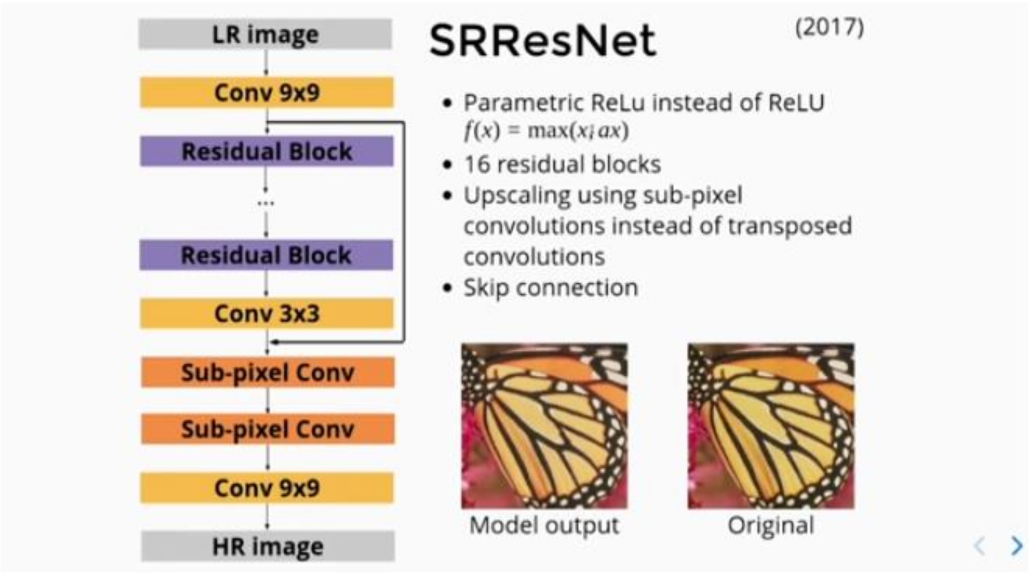
image-to-image mapping
Spatial Alignment

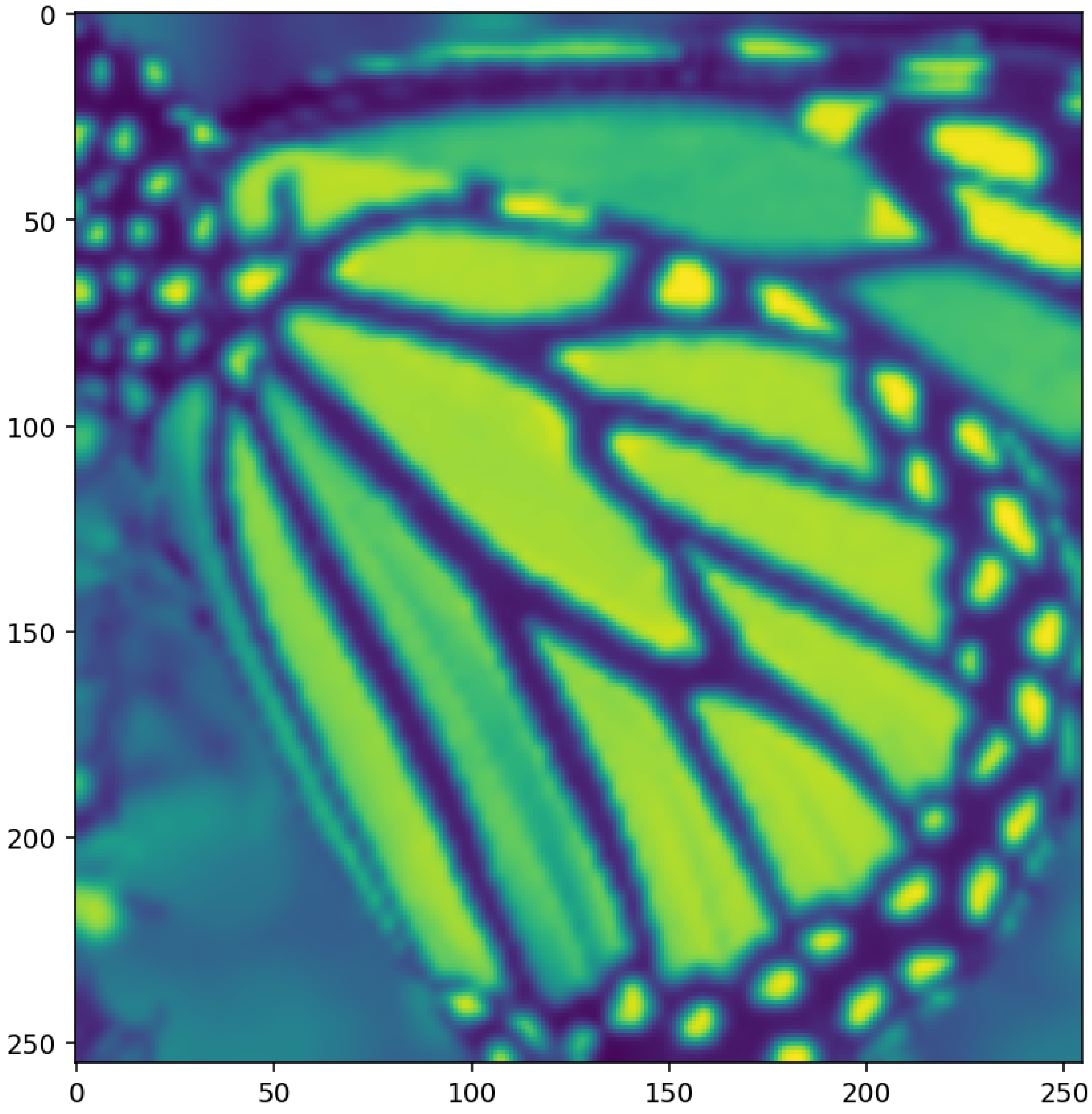
- **SRCNN**



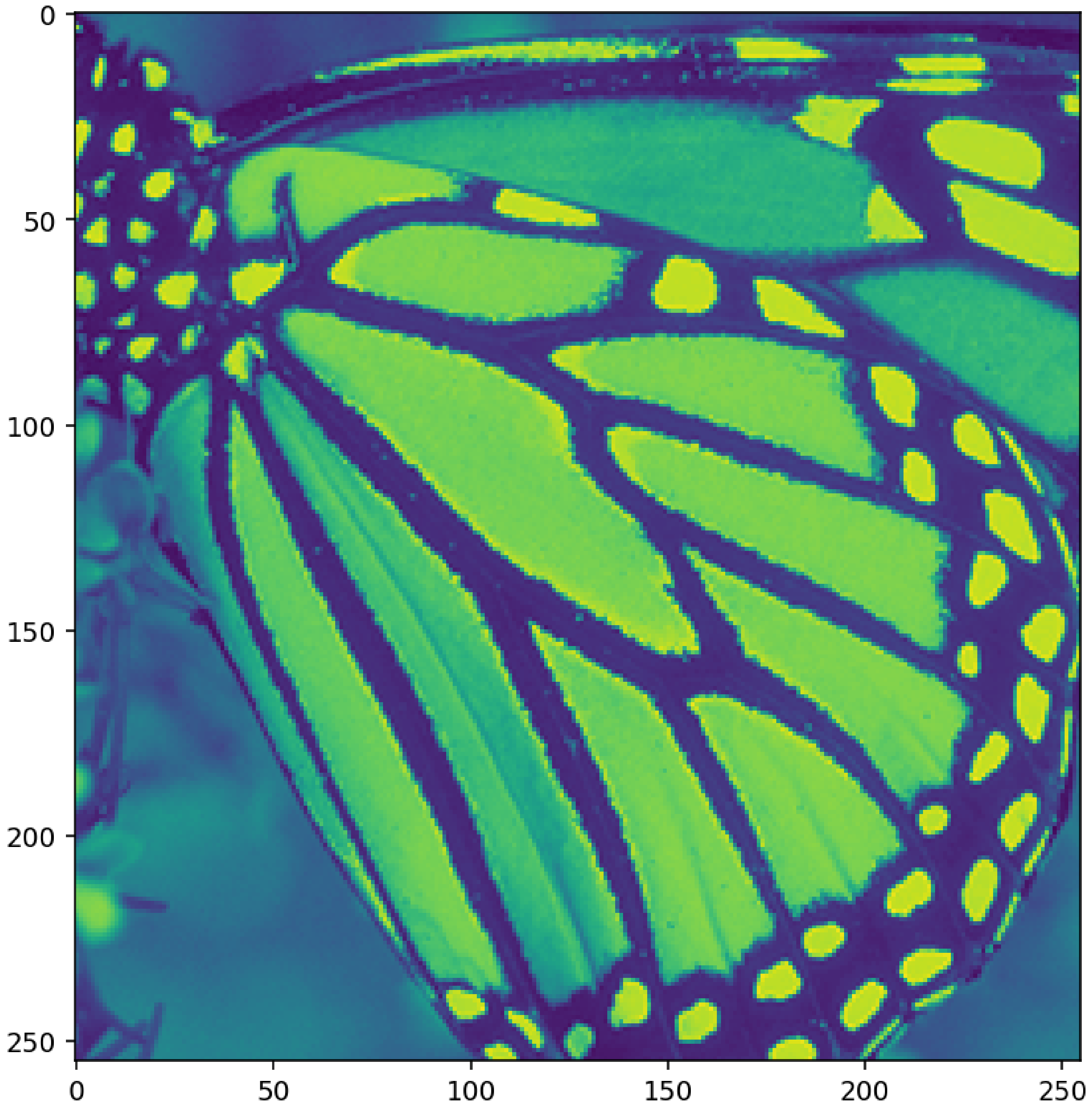
- Bicubic interpolation was replaced by Transpose convulation later
- Multiple Recursive layers

SRResNets





LR- Interpolated



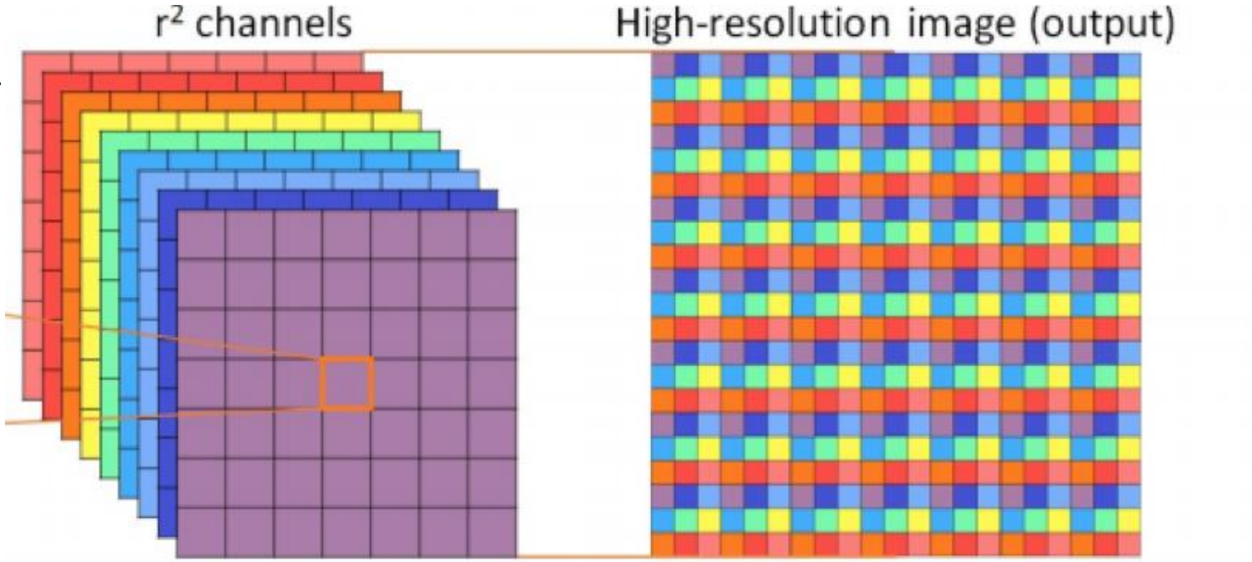
SRCNN

- **Sub-pixel convolution for upsampling (pixel shuffle)**

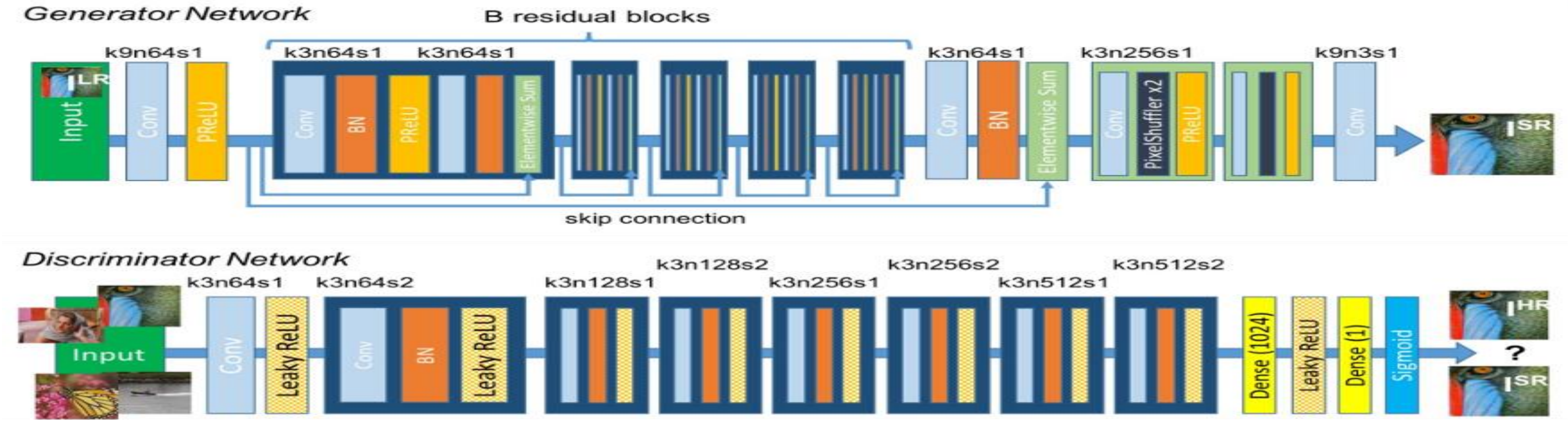
1) R squared channels are produced by convolutions on LR is then pixel reshuffeled to produce xR SR.

2) This method is faster than other SR methods

3) ConvT causes image artifacts (distortions)



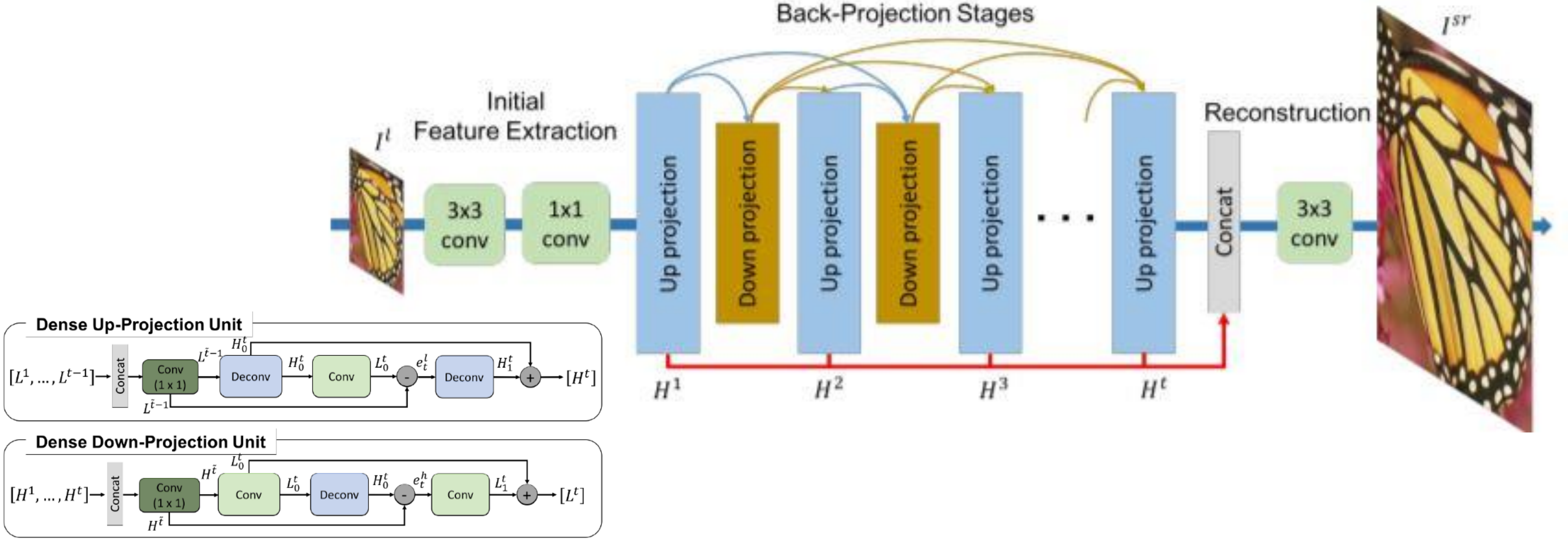
- **SRGAN**



• DBPN

1) Deep Back-Projection Networks (DBPN) exploit iterative up- and down-sampling layers

2) We use error feedbacks from the up- and down-scaling steps to guide the network to achieve a better result



Noisy	GT	BM3D-SR [54]	BM3D-SRNI [55]	Ours
$\sigma = 20$	PSNR/SSIM	25.05/0.5868	25.31/0.6206	**27.03/0.7330**

DRLN

DEEP VSR

Temporal alignment

- **Temporal Concatenation**

- SISR for all Frames

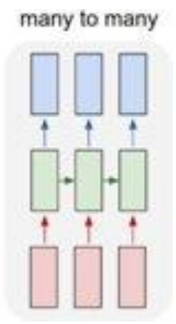
- Frames are not temporal smooth (temporal incoherence)

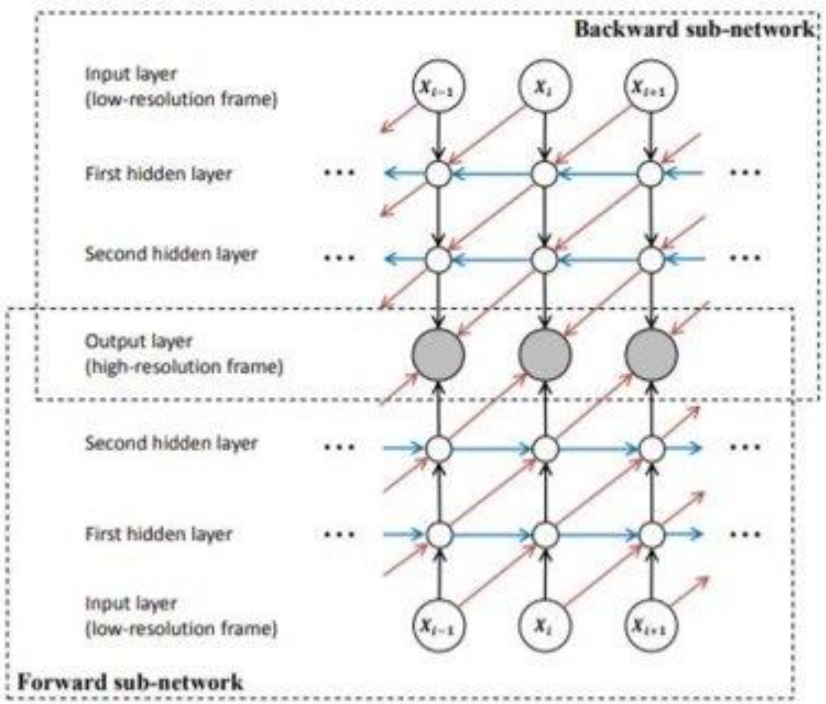
- It doesn't include details/motion from surrounding frames

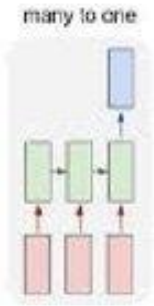
- **MISR**

- MISR utilizes the missing details available from the neighboring frames LR(t−n), ..., LR(t), ..., LR(t+n) and fuses them for super-resolving LR

- **RNNs**







A sequence of LR frames is mapped to a single target HR frame in a window fashion

- **Optical flow methods**

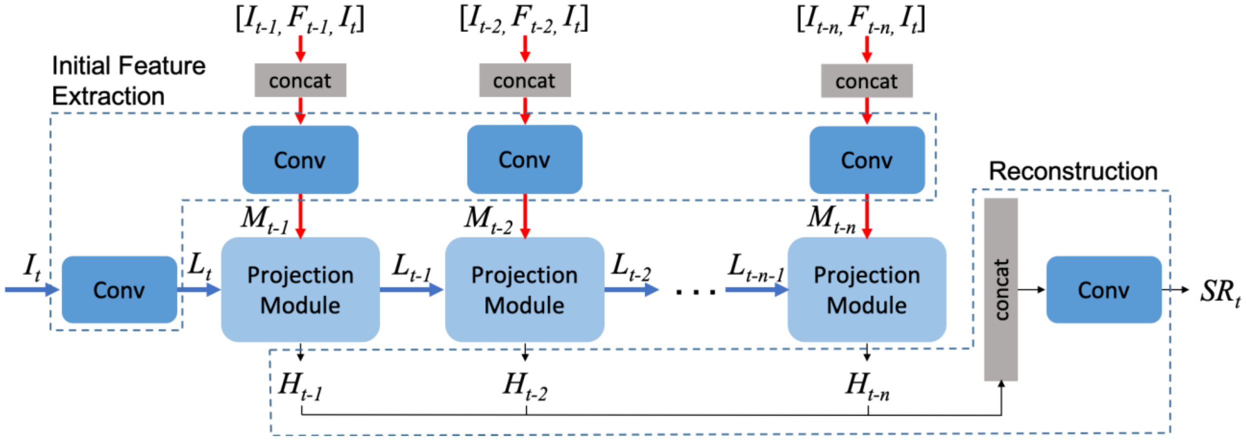
- Separate network to calculate optical flow between neighboring frames LR(t-n)..LR(t+n)
- Optical flow methods allow estimation of the trajectories of moving objects, thereby assisting in VSR.
- Video frames are Warped using the optical flow method LR(t-k), LR(t), F(t-k)

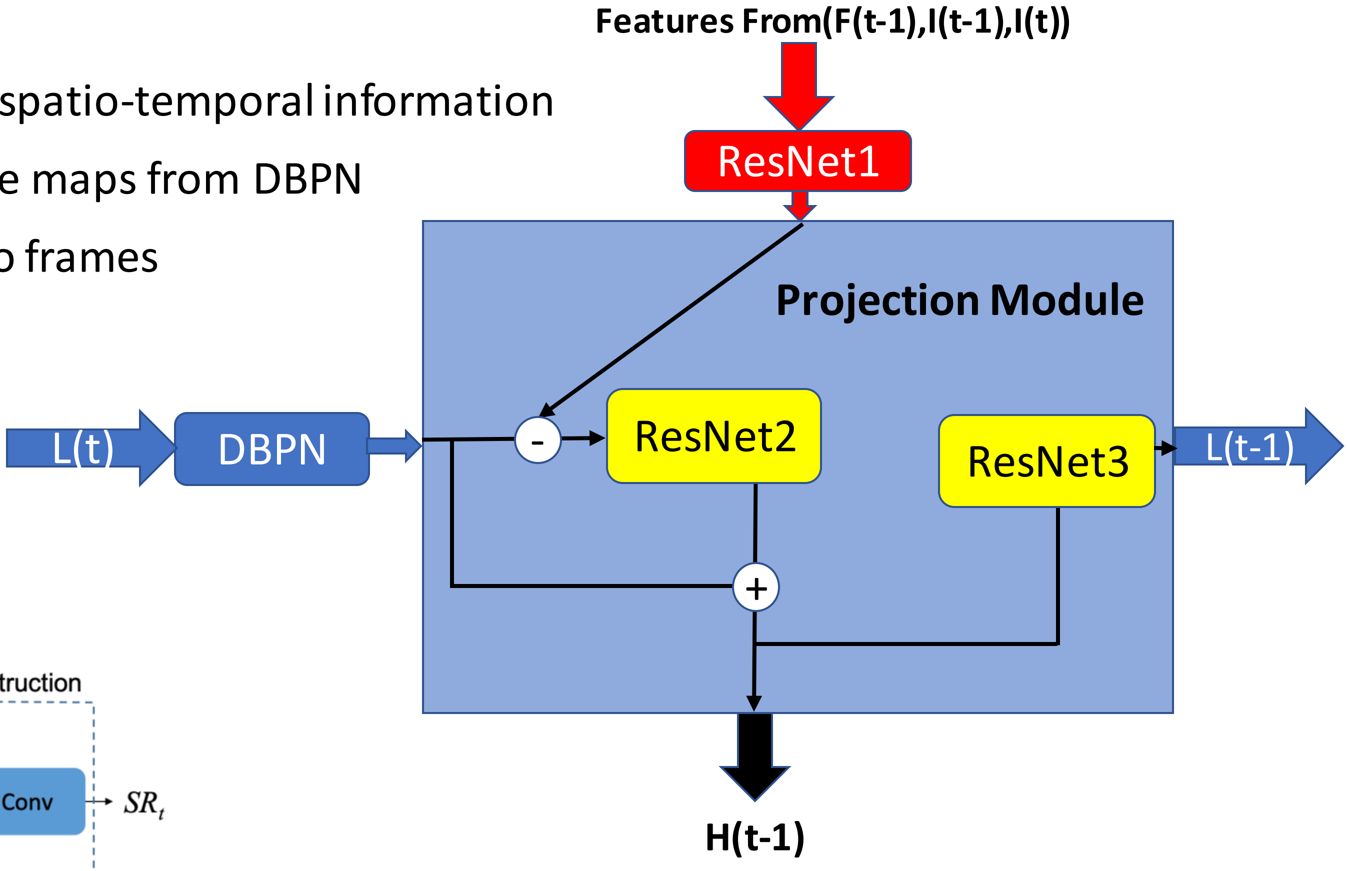
• RBPN

- An RNN-based optical flow method that preserves spatio-temporal information
- RBPN uses the idea of iteratively refining HR feature maps from DBPN
- But extracts missing details using neighboring video frames

I- LR Frames   H – Used to construct SR

L- Refined Frame M-Residual Features





MSE Loss is not enough for SR!

- While optimizing MSE during training improves PSNR and SSIM, these metrics may not capture fine details in the image leading to misrepresentation of perceptual quality
- The ability of MSE to capture intricate texture details based on pixel-wise frame differences is very limited
- Minimizing MSE encourages finding pixel-wise averages of plausible solutions that are typically overly-smooth

Perceptual Loss:

- Focuses on perceptual similarity instead of similarity in pixel space.
- Perceptual loss relies on features extracted from the activation layers of the pre-trained VGG-19 network, instead of low-level pixel-wise error measures
- It is the Mean squared distance b/w features extracted from HR and SR

4 Fold Loss

$$Loss_{G_{\theta_G}}(SR_t) = \begin{array}{l} \alpha \times MSE(SR_t, HR_t) \\ + \beta \times PerceptualLoss(SR_t, HR_t) \\ + \gamma \times AdversarialLoss(SR_t) \\ + \delta \times TVLoss(SR_t, HR_t) \end{array} \quad (5)$$

where, α, β, γ, δ are weights set as 1, 6×10^{-3}, 10^{-3} and 2×10^{-8} respectively [14].

- **AdversarialLoss**

Limit model “fantasy”, thus improving the “naturality”

AdversarialLoss(t) = −log(D(G(LR(t))) (instead of log(1 − D (G(LR(t)))

- **TV Loss**

It is defined as the sum of the absolute differences between neighboring pixels in the horizontal and vertical directions.

To De-noise the output SR

$$TV\,Loss_t = \frac{1}{WH} \sum_{i=0}^{W} \sum_{j=0}^{H} \sqrt{ \begin{array}{l} (G_{\theta_G}(LR_t)_{i,j+1,k} - G_{\theta_G}(LR_t)_{i,j,k})^2 + \\ (G_{\theta_G}(LR_t)_{i+1,j,k} - G_{\theta_G}(LR_t)_{i,j,k})^2 \end{array} } \quad (4)$$

- **MSE Loss**
- **Perceptual Loss**

Discriminator Loss = 1 − D (HR(t)) + D (SR(t))

Results:

Dataset	Clip Name	VSR-DUF [23]	**iSeeBetter**	Ground Truth
Vid4	Calendar			
SPMCS	Pagoda			
Vimeo90K	Motion