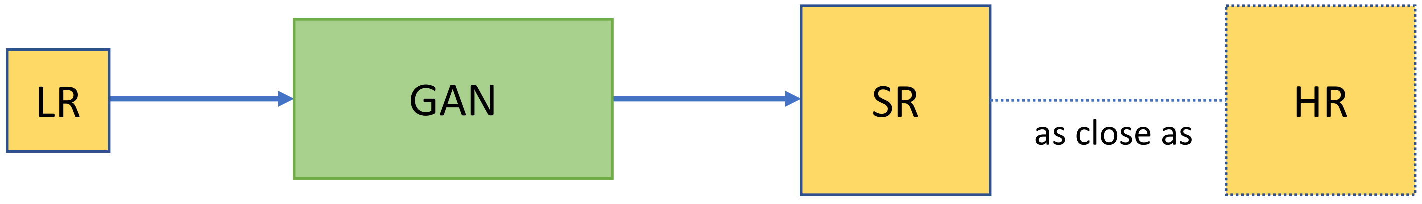
SPATIALLY ADAPTIVE LOSSES FOR VIDEO SUPER-RESOLUTION WITH GANS

Xijun Wang[a], Alice Lucas[a], Santiago Lopez-Tapia[b], Xinyi Wu[a], Rafael Molina[b], Aggelos K. Katsaggelos[a]

a) Dept. of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, USA
b) Dpto. de Ciencias de la Computación e Inteligencia Artificial, Universidad de Granada, Spain

Video Super-Resolution



$$Loss = L_{GAN} + \underbrace{\lambda_1 L_{Pixel} + \lambda_2 L_{feature}}_{\text{Spatially Adaptive}}$$

$$1 \times (flat_{SR} - flat_{HR})^2 + 100 \times (edge_{SR} - edge_{HR})^2$$

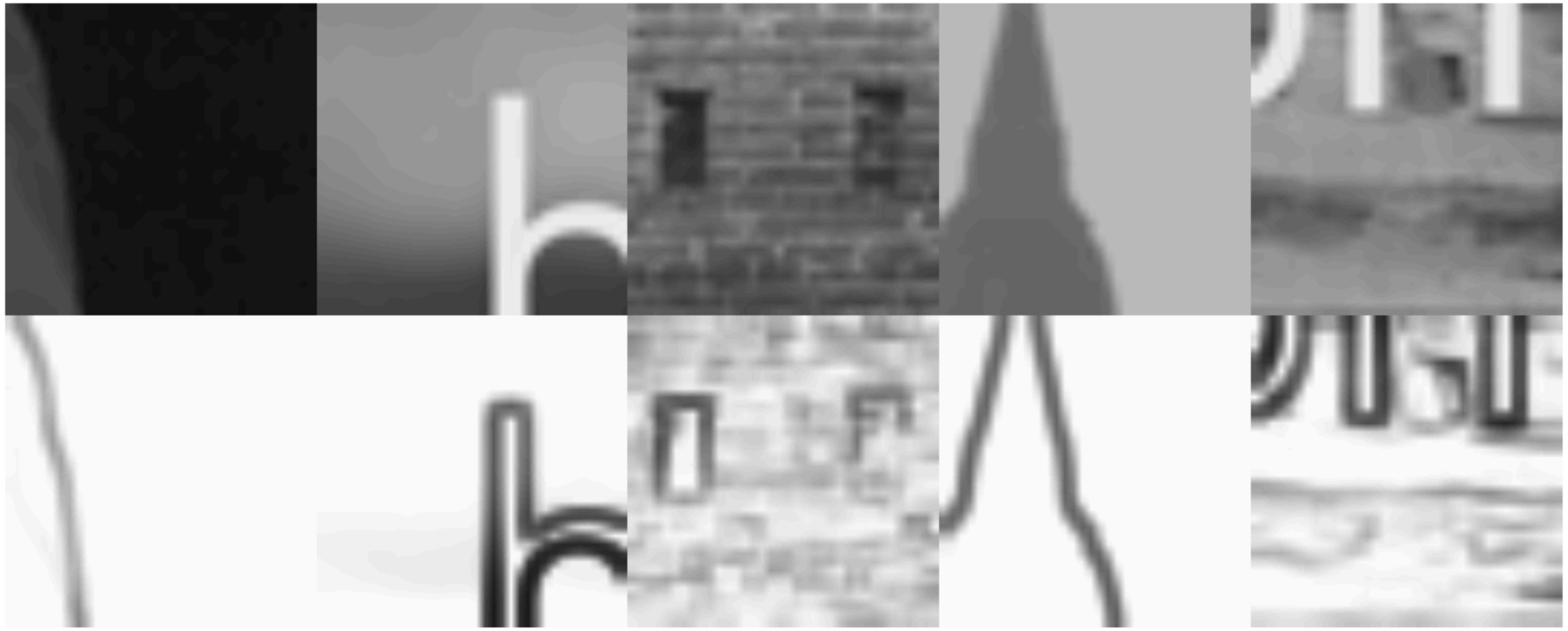
Distance Definition

- Normal Charbonnier Loss:

$$\gamma(u, v) = \sum_k \sum_i \sum_j \sqrt{(u_{k,i,j} - v_{k,i,j})^2 + \epsilon^2},$$

- Modified Charbonnier Loss:

$$\gamma_w(u, v, W(u)) = \sum_k \sum_i \sum_j w_{k,i,j}(u) \sqrt{(u_{k,i,j} - v_{k,i,j})^2 + \epsilon^2}$$



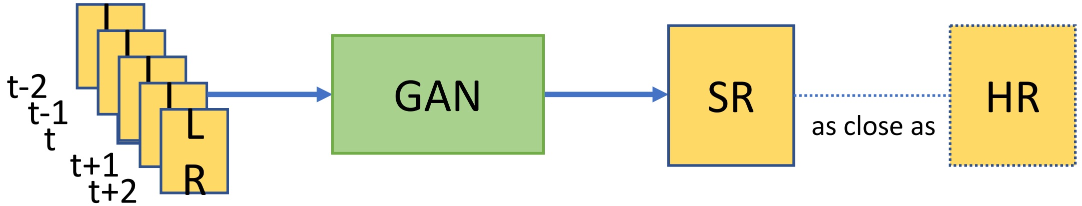
Top row: image patches; Bottom row: corresponding values of the visibility function.

- Spatially adaptive pixel-wise loss in pixel space:

$$L_{pixel} = \sum_{(x,Y)} \gamma_w \left(x, G_\theta(Y), \alpha + \beta W(x)\right),$$

- Spatially Adaptive Perceptual Loss in Feature Space:

$$L_{feature} = \sum_{(x,Y)} \gamma_w(VGG(x), VGG(G_\theta(Y)), \alpha$$
$$+ \beta VGG(W(x)))$$



$$\gamma_w(u, v, W(u)) = \sum_k \sum_i \sum_j w_{k,i,j}(u) \sqrt{(u_{k,i,j} - v_{k,i,j})^2 + \epsilon^2}$$

GAN Loss

- The adversarial min-max problem

$$\min_{\theta}\max_{\phi} L_{GAN}(\phi,\theta) = \mathbb{E}_x\left[\log D_{\phi}(x)\right] + \mathbb{E}_Y\left[\log\left(1 - D_{\phi}\left(G_{\theta}(Y)\right)\right)\right],$$

- The generator loss:

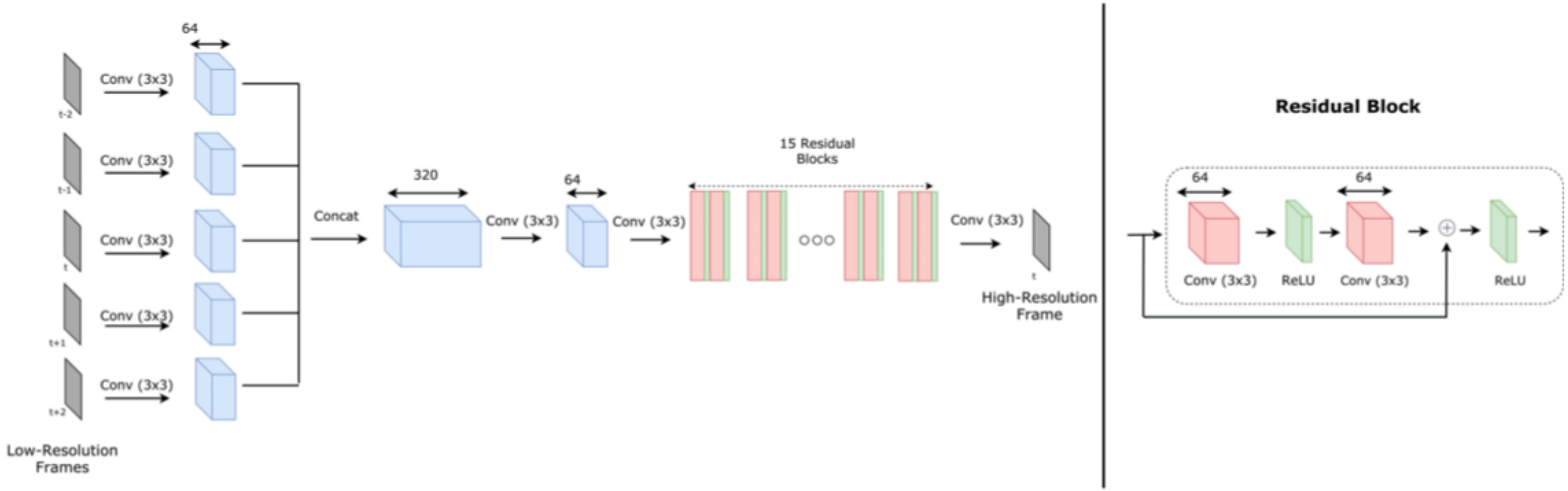
$$L_{gen} = \mathbb{E}_Y\left[-\log D_{\phi}\left(G_{\theta}(Y)\right)\right],$$

- The discriminator loss:

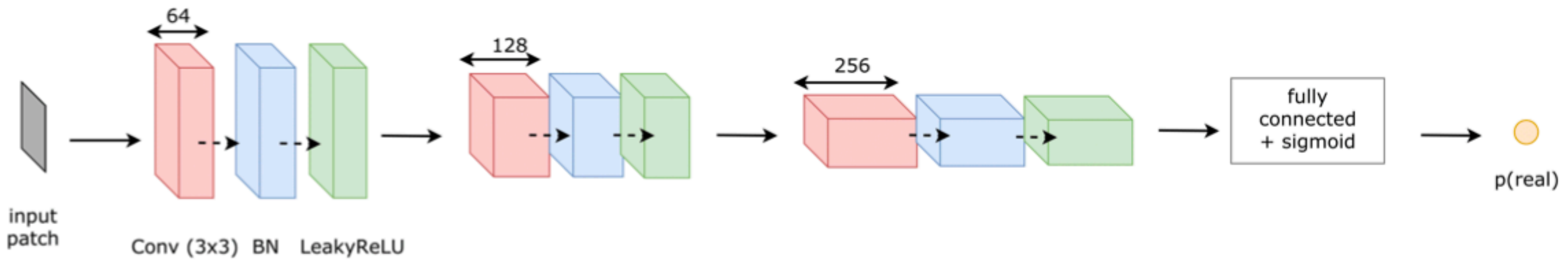
$$L_{dis} = \mathbb{E}_x\left[-\log D_{\phi}(x)\right] + \mathbb{E}_Y\left[-\log\left(1 - D_{\phi}\left(G_{\theta}(Y)\right)\right)\right]$$

I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde- Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in Advances in neural information processing systems, pp. 2672–2680, 2014.

Generator Architecture



Discriminator Architecture



The Combined Loss

$$
\begin{aligned}
L_{final} = \alpha_1 \left[\mathbb{E}_Y \left[-\log D_\phi \left(G_\theta\left(Y\right)\right)\right]\right] \\
+ \sum_{(x,\ Y)} \gamma_w \left(x,\ G_\theta\left(Y\right), \alpha_2 + \beta_2 W(x)\right) \\
+ \sum_{(x,\ Y)} \gamma_w \left(VGG(x),\ VGG\left(G_\theta\left(Y\right)\right),\ \alpha_3 + \beta_3 VGG(W(x))\right),
\end{aligned} \tag{13}
$$

Evaluation Results

- Quantitative Result

	VSRResFeatGAN PSNR/SSIM/PercepDist	Spatially Adaptive VSRGAN PSNR/SSIM/PercepDist
2	30.90/0.9241/0.0283	**31.64/0.9327/0.0257**
3	26.53/0.8148/0.0668	**26.80/0.8256/0.0641**
4	24.50/0.7023/0.1043	**24.72/0.7233/0.1010**

Comparison with state-of-the-art for VidSet4 dataset for scale factors 2,3, and 4.

For PSNR/SSIM metrics, bigger is better

For the PercepDis metric, smaller is better

- Qualitative Result

Ground Truth	LR Input
VSRResFeatGAN	Spatially Adaptive VSRGAN

- Qualitative Result

Ground Truth	LR Input
VSRResFeatGAN	Spatially Adaptive VSRGAN

Q&A