# Single Image Super Resolution

https://arxiv.org/pdf/1712.06116v2.pdf

February 15, 2021

## 1 Image Degradation Model

$$\boldsymbol{y} = (\boldsymbol{x} \otimes \boldsymbol{k}) \downarrow_s + \boldsymbol{n} \tag{1}$$

where $\boldsymbol{y}$ is the low-resolution (LR) image, $(\boldsymbol{x} \otimes \boldsymbol{k})$ is convolution between high-resolution (HR) image $\boldsymbol{x}$ and blur kernel $\boldsymbol{k}$, $\downarrow_s$ is downsampling operation with scale factor $s$, and $\boldsymbol{n}$ is additive while Gaussian noise with standard deviation $\sigma$

## 2 SISR using MAP

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2\sigma^2} \left\| (\boldsymbol{x} \otimes \boldsymbol{k}) \downarrow_s - \boldsymbol{y} \right\|^2 + \lambda \Phi(\boldsymbol{x}) \tag{2}$$

estimate HR image $\boldsymbol{x}$ using LR image $\boldsymbol{y}$, use regularisation term $\Phi(\boldsymbol{x})$ to constrain the solution since SISR has ill-posed nature.
We can write this more generally as

$$\hat{\boldsymbol{x}} = \mathcal{F}(\boldsymbol{y}, \boldsymbol{k}, \sigma, \lambda; \Theta) \tag{3}$$

By taking $\lambda$ common, we can absorb $\lambda$ into $\sigma$ to get

$$\hat{\boldsymbol{x}} = \mathcal{F}(\boldsymbol{y}, \boldsymbol{k}, \sigma; \Theta) \tag{4}$$

So, the goal of SISR is to learn $\hat{\boldsymbol{x}} = \mathcal{F}(\boldsymbol{y}, \boldsymbol{k}, \sigma; \Theta)$, rather than $\hat{\boldsymbol{x}} = \mathcal{F}(\boldsymbol{y}; \Theta)$ Here, the complexity arises that $\boldsymbol{y}$, $\boldsymbol{k}$, $\sigma$ each have different dimensions.

## 3 Dimensionality Stretching

1. Blur kernel is vectorised: $\boldsymbol{p} \times \boldsymbol{p} \longrightarrow \boldsymbol{p^2} \times \boldsymbol{1}$
2. Project vectorised blur kernel into $\boldsymbol{t} - \textit{dimensional}$ space using PCA
3. Concat noise $\sigma$ to $\boldsymbol{t} - \textit{dimensional}$ vector to get vector $\boldsymbol{v}$ of size $(t+1)$
4. Strech $\boldsymbol{v}$ into degradation maps $\mathcal{M}$ of size $W \times H \times (t+1)$, where all the elements of $i$-th map are $\boldsymbol{v}_i$

## 4 Model

1. Concat LR image and degrarion maps to get the input of size $W \times H \times (C + t + 1)$ for the network
2. Each layer in network has 3 operations: "Conv + BN + ReLU"
3. Last layer: only "Conv"
4. sub-pixel convolution layer, to convert multiple HR subimages of size $W \times H \times s^2C$ into a single image of size $sW \times sH \times C$