Elements of SISR

Description

Single Image Super-Resolution is a problem that trying to super resolve a single image. SISR problem assumes Low-Resolution data to be a low-pass filtered(blurred), downsampled and noisy version of High-Resolution data, which means it could be treated as a reconstruction problem(LR ➡ HR) and evaluated by functions like PSNR.

Btw, due to the loss of high-frequency information, it is a highly ill-posed problem and the solution(maps between LR and HR) is non-trivial.

Key assumption

Much of the high-frequency data is redundant and thus can be accurately reconstructed from low frequency components.

Papers

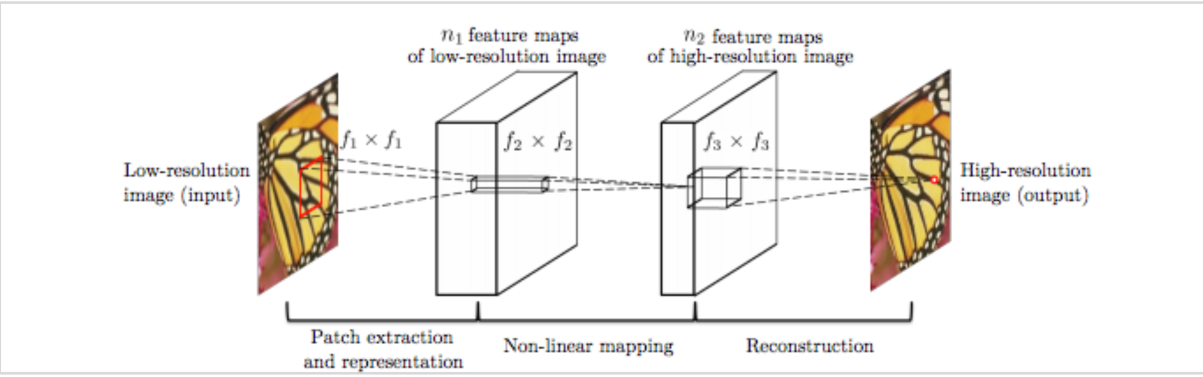
- 16 ESPCN

Target:

improve reconstruction accuracy and computational performance

Previous method:

Using a single filter(like bicubic) to upscale LR into HR space and doing the SR operation in HR space.



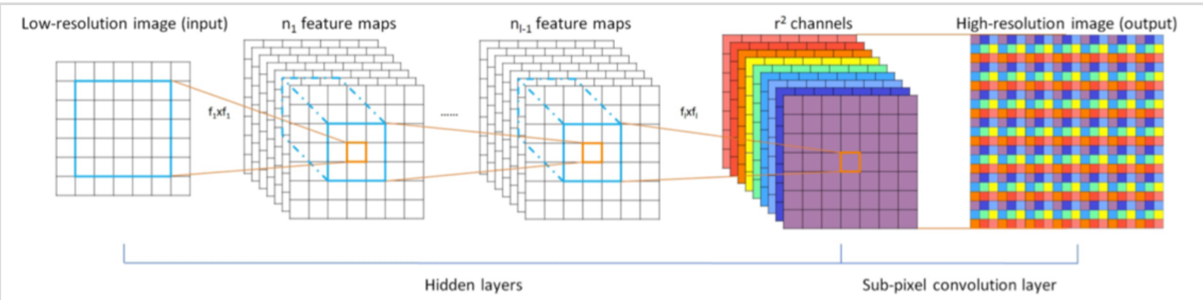
(Get pic from: https://arxiv.org/pdf/1501.00092.pdf —— SRCNN)

Achievements:

Demonstrate previous methods are sub-optimal;

A novel CNN architecture which extract feature maps in the LR space;

An efficient sub-pixel convolutional layer which learns an array of filters to upscale the final LR feature maps into the HR output.



About Recurrent

- 16 VDSR

Target & Previous methods are as above.

Achievements:

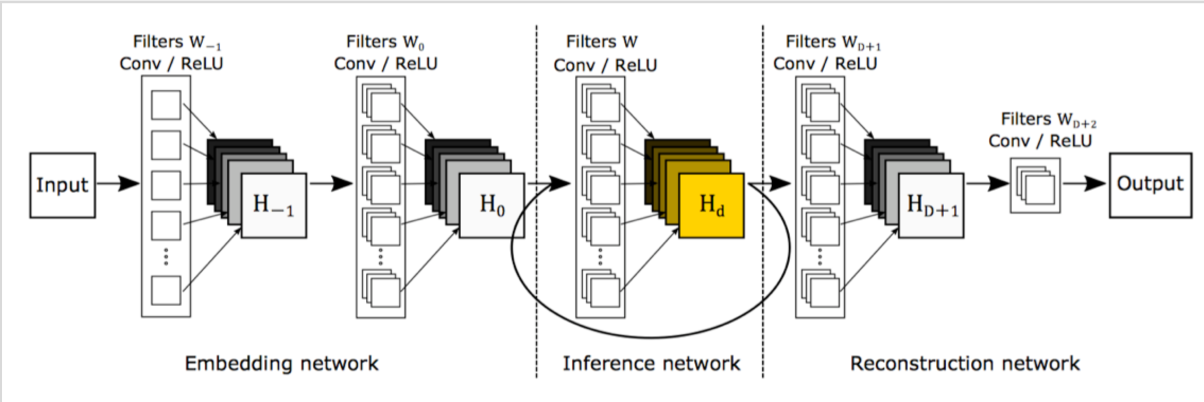
A highly accurate SR method based on **a very deep convolution network**, which has 20 weight layers(3*3 for each layer) and need to be trained with a very big initial learning rate(About 10^4 times as SRCNN proposed, in case it's too slow to converge. They do have the problem of gradient exploding, but they resolved it with residual-learning and gradient clipping).

Coping with multi-scale SR in a single network(even include fractional).

- 16 DRCN

Target & Previous methods are as above

Architecture:



Two important ways:

recursive-supervision & skip-connection

- 16 PSyCo

没看懂

- 16 7WaysToImprove example-based SISR

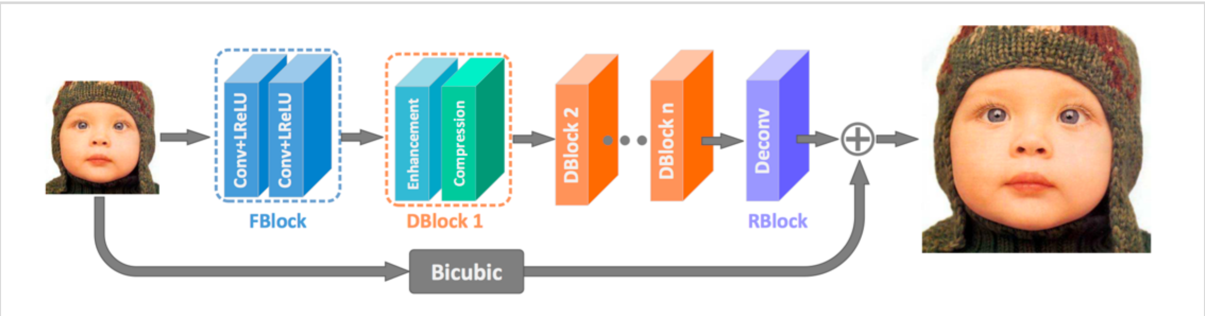
Interesting

Fast & Accurate SISR via Information Distillation Network

- A compact convolutional network directly reconstruct HR from LR
- Q: what is example-based?

Three main parts of IDN

1. FBlock: extract features directly from LR
2. (stacked)DBlocks: distill residual information
3. RBlock: aggregate feature information and reconstruct Elements of SISR images



Formulated as y = R(Fn(Bn-1)) + U(x)

- FBolck (feature extraction block)

  composed of two 3*3 Conv(LReLU)

  Expressed as B0 = f(X), represent extraction function and it is used as input to the next stage respectively.

- (stacked) DBlocks (distillation block):

  composed of multiple information distillation blocks using chained mode

  Formulated as Bk = F(Bk-1), k = 1…n, represent k-th DBlock function and input-output of the k-th DBlock.

  Each DBlock is composed of "enhancement unit + compression unit"

  enhancement unit:

  Do a process simulation in ppt, as showed in 3.2-paper

  residual network?

  compression unit

  use 1*1 convolution layer, act as dimensionality reduction or distilling relevant information for later network, formulated as

$$B_k = f_F^k\left(P^k\right) = \alpha_F^k\left(W_F^k\left(P^k\right)\right)$$

- RBlock (reconstruction block)

  Formulated as y = R(Fn(Bn-1)) + U(x), where R, U represent the RBlock and bicubic interpolation respectively.

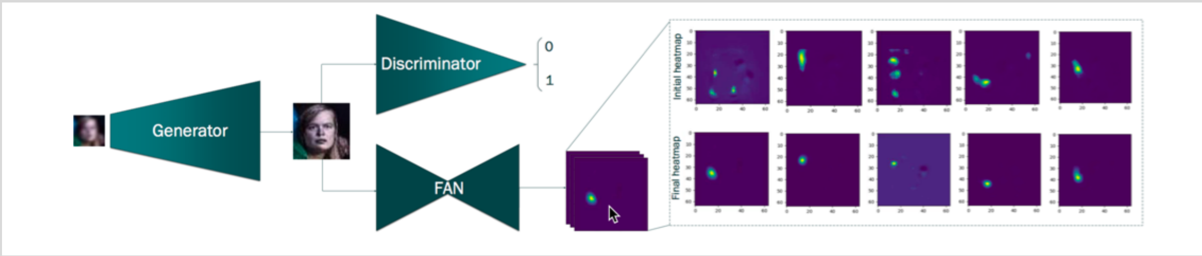
Loss function

Training with MAE, fine-tune by MSE loss.(Empirically)

Q：什么是features的自适应性adaptive?

Super-FAN: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with GANs

- Architecture



Super-Resolution network: super-resolve LRs

Discriminator: distinguish between HR & SR

FAN: localizing facial landmarks on SR

(p.s. discriminator is not used at test time.

Super-Resolution network

Layer distribution:

A residual based network, use resolution as inputs. They used a more balanced distribution “12-3-2” when “r = 4”, as single block at higher resolutions is insufficient for generate sharp details…

Building block architecture:

BatchNorm + ReLU + Deconv

Long skip connection:

No answer…

- Losses

Pixel loss

pixel-wise MSE loss

$$l_{pixel} = \frac{1}{r^2WH}\sum_{x=1}^{rW}\sum_{y=1}^{rH}(I^{HR}_{x,y} - G_{\theta_G}(I^{LR})_{x,y})^2,$$

Perceptual loss

$$l_{feature/i} = \frac{1}{W_iH_i}\sum_{x=1}^{W_i}\sum_{y=1}^{H_i}(\phi_i(I^{HR})_{x,y} - \phi_i(G_{\theta_G}(I^{LR}))_{x,y})^2, \quad (2)$$

where φ i denotes the feature map obtained after the last convolutional layer of the i−th block. Means SR & HR must also close in feature space.

Heatmap loss

$$l_{heatmap} = \frac{1}{N} \sum_{n=1}^{N} \sum_{ij} (\widetilde{M}_{i,j}^{n} - \widehat{M}_{i,j}^{n})^2,$$

does not require having access to ground truth landmark annotations just access to a pre-trained FAN.

Adversarial loss

$$l_{WGAN} = \mathop{\mathbb{E}}_{\hat{I} \sim \mathbb{P}_g} [D(\hat{I})] - \mathop{\mathbb{E}}_{I \sim \mathbb{P}_r} [D(I^{HR})] + \lambda \mathop{\mathbb{E}}_{\hat{I} \sim \mathbb{P}_{\hat{I}}} [(\|\nabla_{\hat{I}} D(\hat{I})\|_2 - 1)^2],$$

Q?

Overall

$$l^{SR} = \alpha l_{pixel} + \beta l_{feature} + \gamma l_{heatmap} + \zeta l_{WGAN}, \quad \text{(5)}$$

weighted, and trained with the consideration of all of them.

- 一些CNN based methods

  Dong SRCNN

  jointly optimized the feature extraction, non-linear mapping & image reconstruction stages in an end-to-end manner.

  deconvolution to accelerate SRCNN in combination with smaller filter sizes and more convolution layers.

  Shi ESPCN

  extract feature maps in the LR space and replace the bicubic unsampling operation with an efficient sub-pixel convolution.

  Kim

  has recursive convolution with skip connection to avoid introducing additional parameters when the depth is increasing.

  Mao

  tackle the general image restoration problem with encoder-decoder networks and symmetric skip connections

Lai LapSRN

takes the original LR images as input and progressively reconstructs the sub-band residuals of HR images.

Tai MemNet

A deep recursive residual network, effectively build a very deep network structure for SR, which weighs the model parameters against the accuracy. 根据精度权衡模型参数

Sajjadi

A novel combination of automated automated texture synthesis with a perceptual loss focusing on creating realistic textures at a high magnification of ratio of 4.建立细致逼真的纹理