

个人工作

MINGRUIBO

- 阅读 VESPCN 并分享。
 - Related Work: ESPCN.
 - VESPCN = Video + Efficient Sub-Pixel Convolution neural Network
 - sub-pixel convolution: 一种上采样方法。

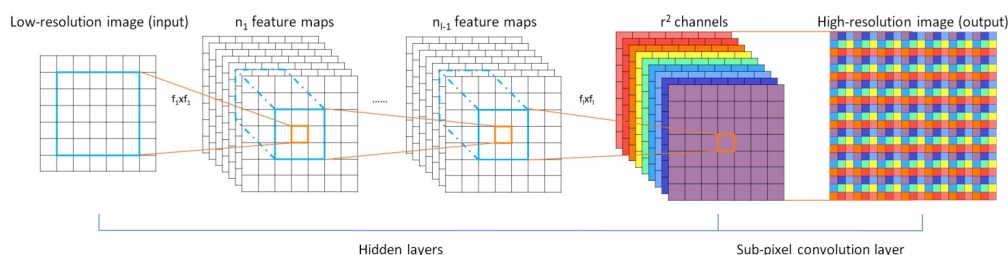


Figure 1. The proposed efficient sub-pixel convolutional neural network (ESPCN), with two convolution layers for feature maps extraction, and a sub-pixel convolution layer that aggregates the feature maps from LR space and builds the SR image in a single step.

- 常见的上采样方法：双三次插值 (bicubic interpolation)、反卷积 (deconvolution)、亚像素卷积 (sub-pixel convolution)。
- 先用两个CNN层进行特征提取，到后面彩色部分就是sub-pixel conv的操作了。首先，如果我想对原图放大3倍，那么我需要生成出 $3^2=9$ 个same size的特征图。将九个same size的特征图拼成一个 3×3 的大图，这就是sub-pixel convolution的操作了。

这是一种抽样的反思想，如果把一张放大三倍的大图，每个 3×3 小方格内取一个点，就可以得到9张原图大小的图像。于是，如果我们可以通过CNN来获得9张符合分布的低分辨率图像，那么就可以组成一张高分辨率的大图。

- spatial-temporal networks.
- Slow Fusion beats Early Fusion.

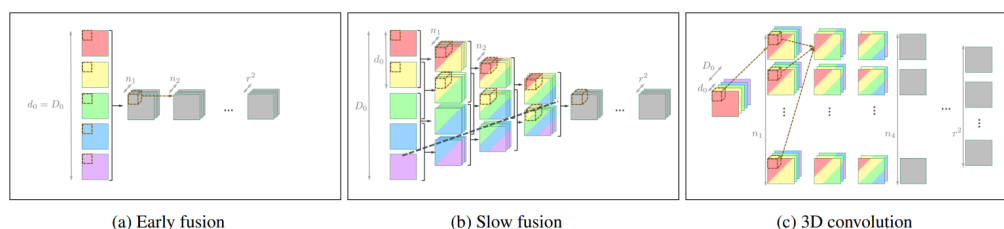


Figure 2: Spatio-temporal models. Input frames are colour coded to illustrate their contribution to different feature maps, and brackets represent convolution after concatenation. In early fusion (a), the temporal depth of the network's input filters matches the number of input frames collapsing all temporal information in the first layer. In slow fusion (b), the first layers merge frames in groups smaller than the input number of frames. If weights in each layer are forced to share their values, operations needed for features above the dashed line can be reused for each new frame. This case is equivalent to using 3D convolutions (c), where the temporal information is merged with convolutions in space and time.

- 使用 spatial transformer 进行动作补偿。
- 阅读 iSeeBetter 并分享。
 - <https://github.com/amanchadha/iSeeBetter>
 - Related Work: SRGAN, RBPN.
 - Innovations
 - 以RBPN作为Generator，继承SRGAN的Discriminator (RBPN提供时序信息，SRGAN提供真实性)。

- 4-fold loss function: MSE/content, perceptual(from [SRGAN](#)), adversarial, TV(from [RBPN](#)).
 - MSE/content: loss in pixel space
 - perceptual: 感知层面，而非像素级。
 - adversarial: GANs
 - TV: from [TV-loss](#)

Table 5 Ablation analysis for iSeeBetter using the “City” clip from Vid4

iSeeBetter Config	PSNR
RBPN baseline with L1 loss	27.73
RBPN baseline with MSE loss	27.77
RBPN generator + SRGAN discriminator with adversarial loss	28.08
RBPN generator + SRGAN discriminator with adversarial + MSE loss	28.12
RBPN generator + SRGAN discriminator with adversarial + MSE + perceptual loss	28.27
RBPN generator + SRGAN discriminator with adversarial + MSE + perceptual + TV loss	28.34

- RNN based optical flow and GAN mechanism.
- 展望
 - 基于前景 (foreground) 、后景 (background) 的帧分类
 - 基于帧相似度 (frame-similarity) 的注意力分类分配机制 (更远距离但是与参考帧属于同一个scene的注意力权重分配可能更大) 。