

---

# Proposal

---

## 1 Problem Background

超分辨率 (super-resolution, SR) 问题希望从低分辨率多媒体数据重建出相应的高分辨率数据。本次比赛的场景为视频超分辨率问题, 恢复降质视频本身的内容, 提高视频的清晰度。

随着视频行业进入“超高清时代”, 许多硬件设备已经支持超高分辨率媒体, 却缺乏对应的高分辨率视频资源。视频超分辨率技术可以用来提升早期胶片或者用户自媒体视频的质量和清晰度, 满足用户对于高分辨率媒体资源的需求。

超分辨率技术在图像、视频以及高维数据上均得到应用。基于深度学习的 SR 方法, 主要是基于卷积神经网络 (CNN) 实现。在本次比赛中, 处理视频数据, 会存在以下挑战。

**输出分辨率限制** 应用 CNN 的 SR 方案中, 运算量会随着输出分辨率的大小而增加。

**时序特征应用** 视频数据存在时序特征, 如果同时处理多帧数据, 运动可能会造成部分区域内容的剧烈变化, 所以需要更加好的时序特征提取方式。

**上采样空洞** 因为输出为高分辨率数据, 所以必定会经过上采样过程。上采样过程中存在数据空洞, 使用插值补齐方式无法达到很好的效果。

**综合的评价指标** 现有的大部分方法, 都是尽可能提高 PSNR 指标, 然而由于 PSNR 是按像素计算的, 更高的 PSNR 分数有可能导致感知上更大的差异。本次比赛将结合 PSNR 指标和 VMAF 指标进行综合评价, 这对我们模型的选择与训练增加了挑战。

## 2 Dataset Specification

阿里巴巴优酷视频增强和超分辨率挑战赛 [8] 涉及 1000 个视频, 每个视频的时间长度为 4 ~ 6 秒。每个样本由低分辨率视频 (LR) 和高分辨率 (HR) 视频组成的视频对构成。低分辨率视频为算法的输入, 高分辨率视频为增强和超分后的真值。视频数据为无压缩的 y4m 格式。

初赛视频 250 个, 训练集、验证集、测试集占比为 150:50:50; 复赛视频 750 个, 训练集、验证集、测试集占比为 450:150:150。

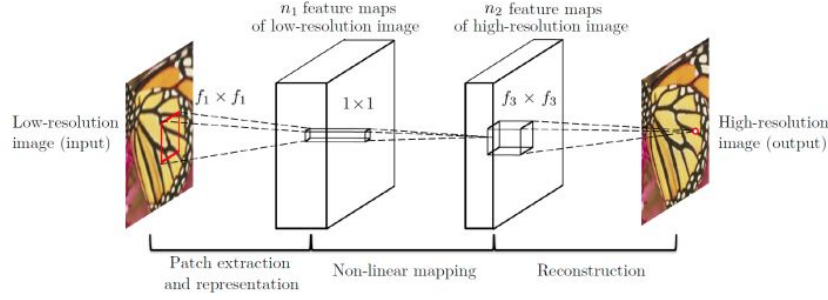


图 1:

复赛视频 750 个，训练集，验证集，测试集 (A 榜 +B 榜) 分别 450:150:150。

绝大部分的 LR 视频分辨率为 480x270，HR 视频为 1920x1080。

### 3 Evaluation Metrics

挑战赛已经给出了视频恢复的评价指标。计算方法概括如下：

1. 10% 的提交视频为完整视频结果，90% 为抽帧后的视频结果。
2. score = PSNR 指标得分  $\times 80\%$  + VMAF 指标得分  $\times 20\%$ ，其中 VMAF 结果为完整视频所有帧 VMAF 结果的均值，PSNR 是所有视频中所有帧的平均值。
3. PSNR(Peak Signal To Noise Ratio) 是一种评价图像的客观标准, 以 dB 为单位，PSNR 值越大表示图像失真越小。然而，PSNR 的评价结果无法与人眼看到的视觉品质完全一致。

$$PSNR = 10 \cdot \log_{10} \frac{MAX_I^2}{MSE} \quad (1)$$

4. VMAF(Video Multimethod Assessment Fusion) 将多种质量指标结合在一起预测主观质量。通过机器学习算法将不同指标“融合”为一个最终指标。

## 4 Survey

### 4.1 SRCNN

SRCNN[1] 是深度学习在超分辨率领域的开山之作。通过双三次插值 (见图1) 将低分辨率图像放大成目标尺寸，然后通过三层卷积网络拟合高分辨率图像。

### 4.2 ESPCN

ESPCN[5] 直接通过端到端的方式将低分辨率图像转换成高分辨率图像，而不需要经过插值等额外处理。具体方式见图2，该层也被称为 sub-pixel convolution layers。

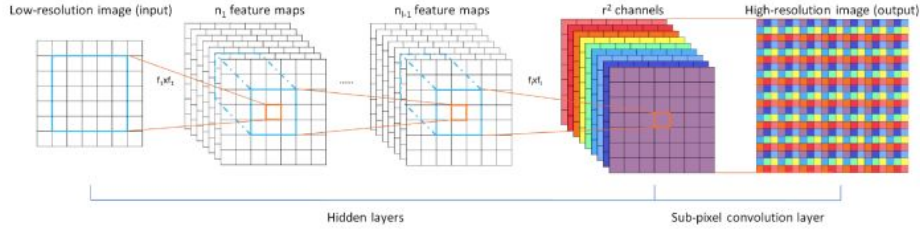


图 2:

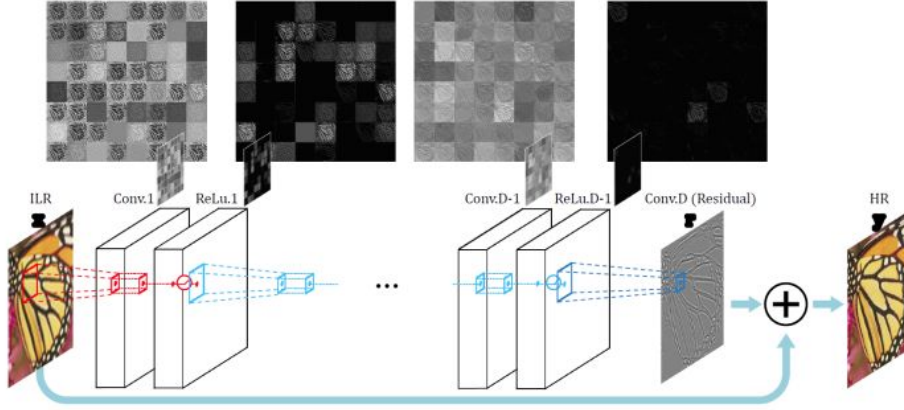


图 3:

设输入图像维度为 1，空间大小为  $H \times W$ ，需要放大的倍数为  $r$ ，那么经过若干次卷积得到一个维度为  $r^2$ ，空间大小为  $H \times W$  的特征图。将特征图中每个像素的  $r^2$  个通道排列成为  $r \times r$  的区域，最终得到维度为 1，空间大小为  $rH \times rW$  的高分辨率图像。

### 4.3 VDSR

VDSR[2] 的想法与残差网络相似。由于低分辨率图像和高分辨率图像的低频信息相似，因此通过只学习高频部分的残差（见图3）可以大大降低学习的难度，提高收敛速度。

此后 SR 领域的工作基本都基于图像残差进行学习。

### 4.4 SRResnet

SRResNet 来源于 SRGAN[3]。图4分为多个残差模块，每个残差模块内部学习图像的高频信息。在网络的最后使用 sub-pixel convolution layers 来增大特征尺寸。残差的使用使得网络的深度增加，生成的图像更加逼真。

### 4.5 EDSR

EDSR[4] 的架构非常类似 SRResNet，主要的不同在于去除了其中多余的 BN 模块（见图5）。BN 会忽略图像特征的绝对差异，而只考虑相对差异，所以可以提高分类等高层

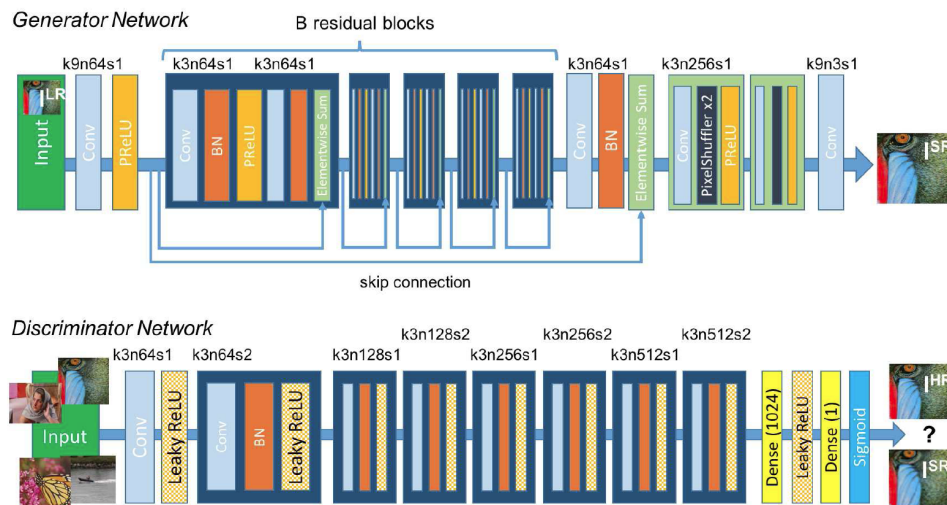


图 4:

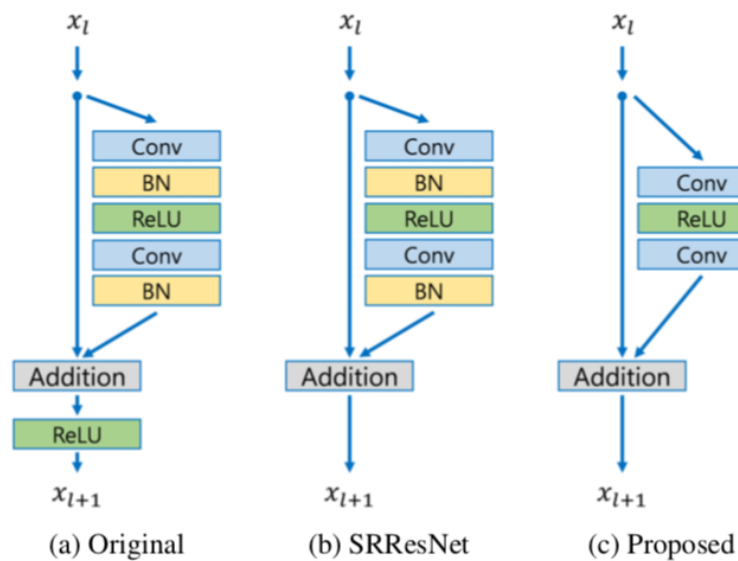


Figure 2: Comparison of residual blocks in original ResNet, SRResNet, and ours.

图 5:

任务的效果。但是在图像分辨率这种低层计算机视觉任务中，BN 反而会破坏图像原有的对比度等信息，使得训练变慢、不稳定。

去除 BN 以后也可以节省计算开销，从而增加网络深度。

5.10-5.12	Data preprocessing(finished)
5.10-5.20	Literature research
5.15-5.24	Read the code of EPSR and design our baseline model Design and train our own model Use data augmentation, learning rate strategy and different loss functions to babysit our model Evaluate our model based on PSNR and VMAF
5.25-6.18	Preliminary contest Adjust our model, do more experiments and analysis the results
6.19-6.25	Hand in the final model, code and paper

表 1: Implementation Plan

#### 4.6 EPSR

EPSR[7] 在 EPSR 的基础上, 改进了损失函数的设计。EPSR 结合了 MSE 损失, 感知损失和对抗损失, 从而达到了不同指标之间的权衡。

### 5 Preliminary Technical Solution

#### 5.1 Model

#### 5.2 Training

- 尝试不同的损失函数。例如结合 MSE、感知损失、纹理匹配损失和对抗损失, 并根据比赛的指标进行调整。
- 比赛要求不能使用额外的数据, 数据增强变得非常重要。LR 图像和 HR 图像均可用于训练, 只需将图片进行目标倍率的下采样, 即可得到训练的输入数据, 而原图像则作为标签 ([6])。也可以尝试旋转、翻转等增强操作。

### 6 Implementation Plan

实现计划见表6。

### References

- [1] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision – ECCV 2014*, pages 184–199, Cham, 2014. Springer International Publishing.

- [2] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
- [3] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew P. Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. *CoRR*, abs/1609.04802, 2016.
- [4] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. *CoRR*, abs/1707.02921, 2017.
- [5] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P. Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. *CoRR*, abs/1609.05158, 2016.
- [6] Assaf Shocher, Nadav Cohen, and Michal Irani. "zero-shot" super-resolution using deep internal learning. *CoRR*, abs/1712.06087, 2017.
- [7] Subeesh Vasu, Nimisha T.M., and A.N. Rajagopalan. Analyzing perception-distortion tradeoff using enhanced perceptual super-resolution network. In *European Conference on Computer Vision (ECCV) Workshops*, 2018.
- [8] 优酷. 阿里巴巴优酷视频增强和超分辨率挑战赛. Website, 2019. <https://tianchi.aliyun.com/competition/entrance/231711/information>.
- [9] 桂花糖. 从 srcnn 到 edsr, 总结深度学习端到端超分辨率方法发展历程. Website, 2019. <https://zhuanlan.zhihu.com/p/31664818>.