

# 基于空间金字塔的视频超分辨率重建算法\*

蔡非凡<sup>1,2</sup> 万旺根<sup>1,2</sup>

(1.上海大学通信与信息工程学院 上海 200072; 2.上海大学智慧城市研究院 上海 200072)

**摘要:** 为了保证重建视觉质量的同时提高重建速率,提出了一种基于空间金字塔生成对抗网络的视频超分辨率重建算法(SPyGAN),该方法在 TccoGAN 的基础上使用更轻量级的空间金字塔网络结构 SPyNet 和更高效的上采样方法,能够快速重建图像的高频纹理细节。主要对生成对抗网络 TccoGAN 的光流预测网络、图像重建模块和损失函数部分进行改进,实验结果表明,该算法与 TccoGAN 相比,PSNR 和 SSIM 的平均值均有一定提高,此外参数量减少为 53.86%,并且重建速率提高至 239%,有效提升了模型的重建速率。

**关键词:** 视频超分辨率;深度学习;生成对抗网络;光流预测网络;SPyGAN

**中图分类号:** TP391.41 **文献标识码:** A **国家标准学科分类代码:** 520.60

## Video super-resolution reconstruction algorithm based on spatial pyramids

Cai Feifan<sup>1,2</sup> Wan Wanggen<sup>1,2</sup>

(1. School of Communication and Information Engineering, Shanghai University, Shanghai 200072, China;

2. Institute of Smart City, Shanghai University, Shanghai 200072, China)

**Abstract:** In order to guarantee the reconstruction visual quality while improving the reconstruction rate, this paper proposes a video super-resolution reconstruction algorithm (SPyGAN) based on spatial pyramid generative adversarial network, which uses a more lightweight spatial pyramid network structure SPyNet and a more efficient upsampling method based on TecoGAN, and can quickly reconstruct the high-frequency texture details of images. We mainly improve the optical flow prediction network, image reconstruction module and loss function part of the generative adversarial network TecoGAN. Experimental results show that the algorithm has improved the mean values of PSNR and SSIM compared with TecoGAN, in addition to the reduction of the parameter amount to 53.86%, and the reconstruction rate is improved to 239%, which effectively improves the reconstruction rate of the model.

**Keywords:** video super-resolution; deep learning; generative adversarial networks; optical flow prediction networks; SPyGAN

### 0 引言

视频超分辨率(VSR)<sup>[1]</sup>技术是计算机视觉研究领域内的重要研究方向之一。VSR 的主要研究内容就是从多帧低分辨率(LR)图像中恢复高分辨率(HR)图像,主要利用帧间连续信息来完成帧特征对齐。但目前现有的视频超分辨率重建算法均为达到提高重建视觉质量从而舍弃重建速率,因此本文主要研究内容聚焦于保证视觉质量和降低运算量的同时来提高重建速率。给定  $2N+1$  个连续的低分辨率帧序列,目标是根据参考低分辨率帧  $x_t^{LR}$  及其相邻帧  $x_{t-N:t+N}^{LR}$  生成指定倍率的高分辨率图像  $g_t^{SR}$ 。

基于深度学习的视频超分辨率重建算法主要包括基于时空相关性的方法<sup>[2]</sup>、基于注意力机制的方法<sup>[3]</sup>、基于运动

估计和运动补偿的方法<sup>[4]</sup>和基于生成对抗网络的方法<sup>[5]</sup>。由于 SPyGAN 主体架构为生成对抗网络,这里主要介绍基于生成对抗网络的超分辨率重建方法。

#### 1) 残差生成对抗网络-SRGAN

Ledig 等<sup>[6]</sup>首次将生成对抗网络引入到超分辨率重建任务中。且作者等人提出了感知损失<sup>[7]</sup>、内容损失<sup>[8]</sup>和对抗损失<sup>[9]</sup>来提高最终生成图像的清晰度和高频细节,但该方法不足之处在于最终生成的图像仍然存在伪影。

#### 2) 增强残差生成对抗网络-ESRGAN

基于提高最终生成图像的锐度和去除伪影的目的,Wang 等<sup>[10]</sup>改进了 SRGAN 的网络结构、对抗损失和感知损失。首次引入了去掉 BN 层的残差块,能有效去除伪影,

收稿日期:2021-12-04

\* 基金项目:安徽省自然科学基金(1908085MF178)、安徽省重点研究和开发计划(202104b11020031)、中国博士后基金(2020M681264)项目资助

并且首次提出残差密集网络块 RRDB,能有效提高真实度和清晰度。改进后的对抗损失和感知损失函数使得最终生成的图像细节更丰富。

### 3)时间一致性生成对抗网络-TecoGAN

由于视频帧之间存在更多的时域信息,因此 Chu 等<sup>[11]</sup>提出了一种基于循环生成网络和生成对抗网络的视频超分辨率重建方法 TecoGAN。

但 TecoGAN 主要集中在时域信息的循环利用,并未有效利用视频帧的空间信息,因此本文提出了基于空间金字塔的生成对抗网络 SPyGAN,在 TecoGAN 的基础上使用更轻量级的空间金字塔网络结构 SPyNet 和更高效的上采样方法,能够快速重建图像的高频纹理细节。以此来达到提高重建速率的目的。

## 1 网络模型架构及损失函数

本文方法主要参考方法 TecoGAN,该方法提出的循环结构能有效提高生成照片的真实度以及纹理细节,并且能

提高帧之间的时间连贯性。同时该方法使用了一种双向损失函数<sup>[12]</sup>,能有效去除对应帧之间的伪影。TecoGAN 具体网络结构如图 1 所示。

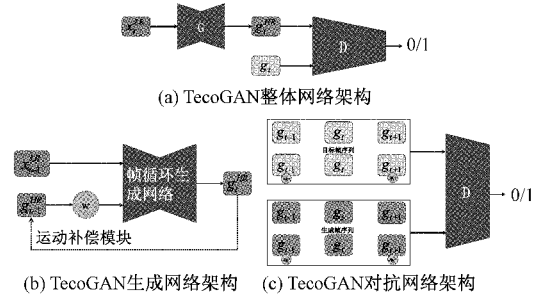


图 1 TecoGAN 结构

而本文方法所提出的生成对抗网络架构主要包含由基于空间金字塔的光流估计模块和 SE 重建模块组成的生成器和时空判别器网络。具体网络架构如下所示。

### 1.1 生成器网络架构

生成器整体架构如图 2 所示。

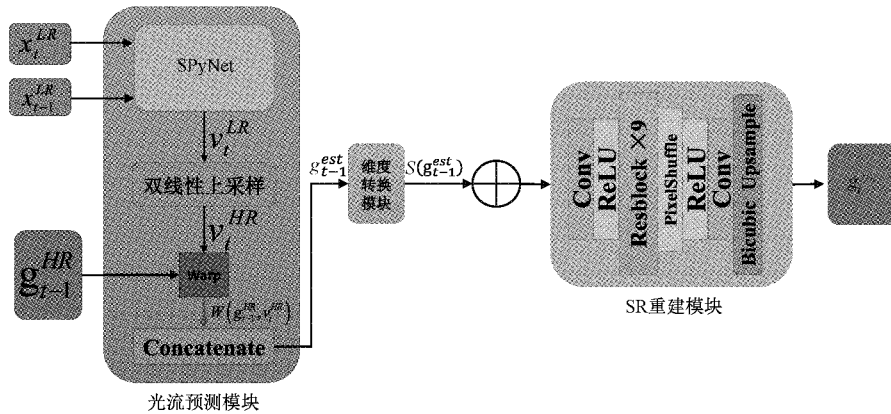


图 2 SPyGAN 结构

SPyGAN 生成器部采用的是帧循环网络架构<sup>[13]</sup>,让先前 HR 估计帧  $g_{t-1}^{HR}$  作为输入能参与到后面的帧  $g_t^{HR}$  的重建过程,这种架构有助于模型提高重建帧的纹理细节,同时能保证生成视频的时间一致性。

首先该网络将输入的参考帧  $x_t^{LR}$  及上一帧  $x_{t-1}^{LR}$  输入到 SPyNet 网络生成低分辨率光流图  $v_t^{LR}$ ,具体公式表示如下:

$$v_t^{LR} = \text{SPynet}(x_t^{LR}, x_{t-1}^{LR}) \in [-1, 1]^{H \times W \times 2} \quad (1)$$

然后利用比例因子为  $s$  的双线性插值对低分辨率光流图  $v_t^{LR}$  进行放大,从而得到高分辨率光流图  $v_t^{HR}$ ,具体公式如下:

$$v_t^{HR} = \text{Up}(v_t^{LR}) \in [-1, 1]^{sH \times sW \times 2} \quad (2)$$

利用前一帧到当前帧的光流  $v_t^{HR}$  来 Warp 之前生成的 HR 估计帧  $g_{t-1}^{HR}$ ,然后将 Warp 之后的特征图与参考帧  $x_t^{LR}$  合并 0 生成  $I_{t-1}^{est}$ ,公式如下:

$$I_{t-1}^{est} = \text{Concat}(\text{Warp}(g_{t-1}^{HR}, v_t^{HR}), x_t^{LR}) \quad (3)$$

这里采用维度转换模块将  $I_{t-1}^{est}$  转换到 LR 域内:

$$S(I_{t-1}^{est}): [0, 1]^{sH \times sW \times C} \rightarrow [0, 1]^{H \times W \times s^2 C} \quad (4)$$

最后将得到的特征图  $S(I_{t-1}^{est})$  与输入的参考帧  $x_t^{LR}$  在通道维度上合并送入重建模块得到  $g_t^{HR}$ :

$$g_t^{HR} = \text{SRNet}(S(I_{t-1}^{est}) \oplus x_t^{LR}) \quad (5)$$

### 1.2 光流预测网络 SPyGAN

Ranjan 等<sup>[1]</sup>首次提出 SPyNet 将经典的空间金字塔方法与深度学习结合来计算光流,本文方法在其基础上进行改进:1)去除 SPyNet 中的 BN 层,提高生成速率;2)使用更多的空间金字塔预测模块,使计算出的光流更为准确。具体的网络结构如图 3 所示。

### 1.3 判别器网络架构

时空判别器整体网络架构如图 4 所示。

图 4(a)为时空判别器的整体输入输出情况。时空判别器接收了两组输入:真值和生成结果。而时空判别器的整体网络架构如图 4(b)所示,本文使用了 9 个由 Conv2d、

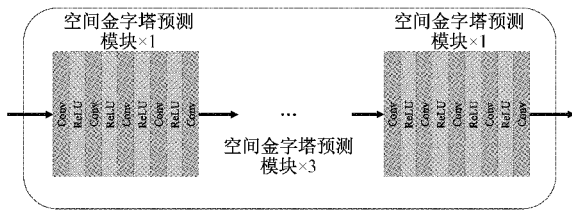
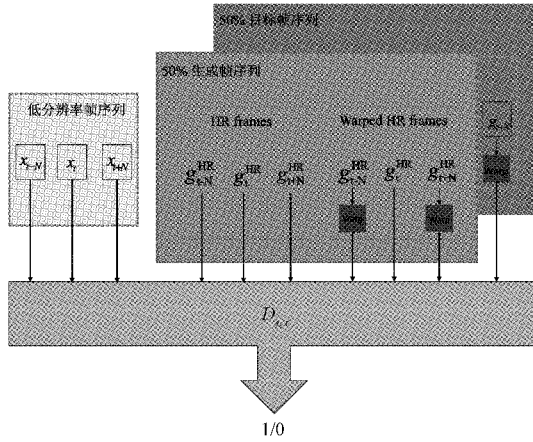
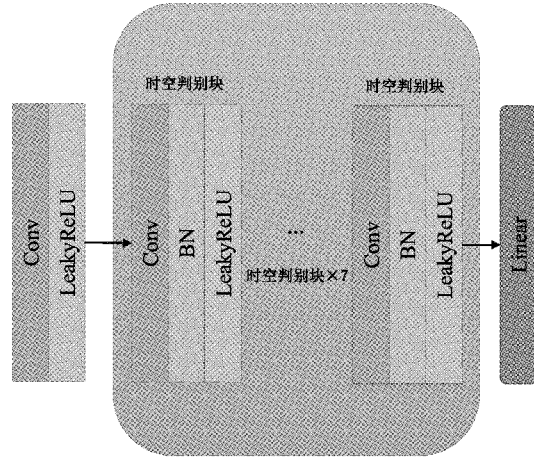


图 3 SPyNet 结构



(a) 时空判别器整体输入输出



(b) 时空判别器网络 \$D\_{st}\$ 网络架构

图 4 时空判别器网络结构与整体输入输出

表 1 Loss 函数汇总

Loss 函数名	Loss 函数公式
$\mathcal{L}_D$	$\mathcal{L}_{feature\_matching}(real\_pred\_D, 1) + \mathcal{L}_{feature\_matching}(fake\_pred\_D, 0)$
$\mathcal{L}_G$	$\lambda_p \mathcal{L}_{pixel} + \lambda_w \mathcal{L}_{warp} + \lambda_f \mathcal{L}_{feat} + \lambda_p \mathcal{L}_{pp} + \lambda_{fm} \mathcal{L}_{feature\_matching}$
$\mathcal{L}_{pix}$	$\frac{1}{hwc} \sum_t \sqrt{(g_t^{HR} - g_t)^2 + \epsilon^2}$
$\mathcal{L}_{warp}$	$\sum \ x_t - Warp(x_{t-1}, v_t^{LR})\ _2$
$\mathcal{L}_{feat}$	$\frac{g_t^{HR} \cdot g_t}{\max(\ g_t^{HR}\ _2 \cdot \ g_t\ _2, \epsilon)}$
$\mathcal{L}_{PP}$	$\sum_{i=1}^{n-1} \ g_i^{HR} - \mathcal{L}_i^{HR}\ _2$
$\mathcal{L}_{feature\_matching}$	$-\omega_n [y_n \cdot \log \sigma(x_n) + (1 - y_n) \cdot \log(1 - \sigma(x_n))]$

所示。本文对判别器进行了改进,对所有真实预测数据和虚假预测数据采用具有对数损失的二元交叉熵损失函数<sup>[15]</sup>,这样能增强判别器的稳定性。

生成器部分采用多个损失函数结合起来提高重建视频帧的清晰度和真实程度,如表 1 的第 2 行所示。其中  $\mathcal{L}_{pix}$  和  $\mathcal{L}_{warp}$  均采用的是改进后的 Charbonnier loss,有助于丰富高频细节; $\mathcal{L}_{PP}$  采用的是 TecGAN 中提出的 Ping-Pong

BN 层和 LeakyReLU 组成的时空判别块来学习时空信息的分布,从而避免传统生成对抗网络算法遇到的时域不稳定效应。

### 1.4 损失函数

SPyGAN 对应的损失函数共有 5 部分,表中  $\lambda$  均为设定参数,具体定义如表 1 所示。

其中,判别器部分的损失函数为  $\mathcal{L}_D$ ,如表 1 的第 1 行

损失函数,能有效地移除漂移伪影以及保留高频细节。

## 2 实验结果与分析

### 2.1 数据集

训练集:本文采用和 TecGAN 相同的视频数据作为训练数据集,整个数据集包括 440 个场景,每个场景由 120 帧图像组成,所有数据均来源于真实拍摄场景,具体包括各类运动、各类天气、各类地点大范围场景,同时本文算法通过随机翻转和旋转来扩增高分辨率图像数据。同时采用 MATLAB 对上述数据进行双线性下采样处理得到原始图像 1/4 大小的低分辨率图像,作为训练过程中成对使用的低分辨率图像。

测试集:本文采用测试集为超分辨率重建中经典的三大数据集,其中 Vid4 为视频超分辨率重建任务中常用视频测试数据集,其中包括日历(calendaer)、城市(city)、树叶(foliage)和行人(walk)4 个视频序列;同时我们用 TecGA 采用的截取自电影部分的数据集 ToS3 和 Gvt72,这两部分数据集具备多个光线变换和多类运动场景,能有效地检测出不同模型在修复能力上的差异。

### 2.2 评价指标

日前重建图像质量的评价指标主要采用峰值信噪比(peak signal-to-noise ratio, PSNR)和结构相似性系数(structural similarity index, SSIM),其中 PSNR 和 SSIM 可以表示为:



$$PSNR = 20 \times \lg\left(\frac{MAX}{MSE}\right) \quad (6)$$

$$SSIM = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (7)$$

其中,  $MAX$  为图像中可能的最大像素值, 一般为 255,  $MSE$  代表的是生成图像与参考图像之间的均方误差, 如式(8)所示。

$$MSE = \frac{1}{N} \sum_{i=1}^N (y^{(i)} - \mu_y^{(i)})^2 \quad (8)$$

其中,  $\mu_x$  和  $\mu_y$  分别代表  $x$  和  $y$  的平均值,  $\sigma_x$  和  $\sigma_y$  分别代表  $x$  和  $y$  的方差,  $\sigma_{xy}$  代表两者之间的协方差,  $c_1$  和  $c_2$  是两个设定常数, 主要避免分母为 0,  $N$  代表的是样本总数。

其中两个评价指标的特性: PSNR 值越大代表生成图像的重建质量越高; SSIM 值越接近 1 代表生成图像更为接近参考图像。

### 2.3 对比实验

本文采用 VimeoTecoGAN 数据集作训练数据集, 同时通过随机翻转和旋转来增加训练数据。具体训练实验使用 Adam 方法对网络进行优化, 初始学习率为  $5 \times 10^{-4}$ , 共迭代 50 000 次, batch\_size 设置为 8。本文分别在 Vid4、ToS3 和

Gvt72 三个测试集上进行了对比实验, 同时统计测试集图像的平均 PSNR 值和 SSIM 值, 具体数据如表 2 所示。

表 2 三大测试数据集上 PSNR/SSIM 结果

测试集	TecoGAN	SPyGAN
Vid4	23.99/0.693 6	24.13/0.738 3
ToS3	30.23/0.841 5	30.36/0.887 9
Gvt72	26.56/0.774 0	27.34/0.824 9

通过表 2 中数据可以发现, SPyGAN 相比 TecoGAN 在 3 个数据集上的 PSNR 值和 SSIM 值均有提高, 其中 PSNR 值分别在 Vid4 数据集上提升了 0.14 dB, 在 ToS3 数据集上提升了 0.13 dB 以及在 Gvt72 数据集上提高了 0.78 dB。

SSIM 值分别在 Vid4 数据集上提升了 0.044 7, 在 ToS3 数据集上提升了 0.046 4 以及在 Gvt72 数据集上提升了 0.050 9。

同时本文分别在 3 个测试集上完成了超分辨率重建视觉效果对比实验, 其中选取 Vid4 数据集作为视觉效果展示对比实验, 具体结果如图 5 所示。

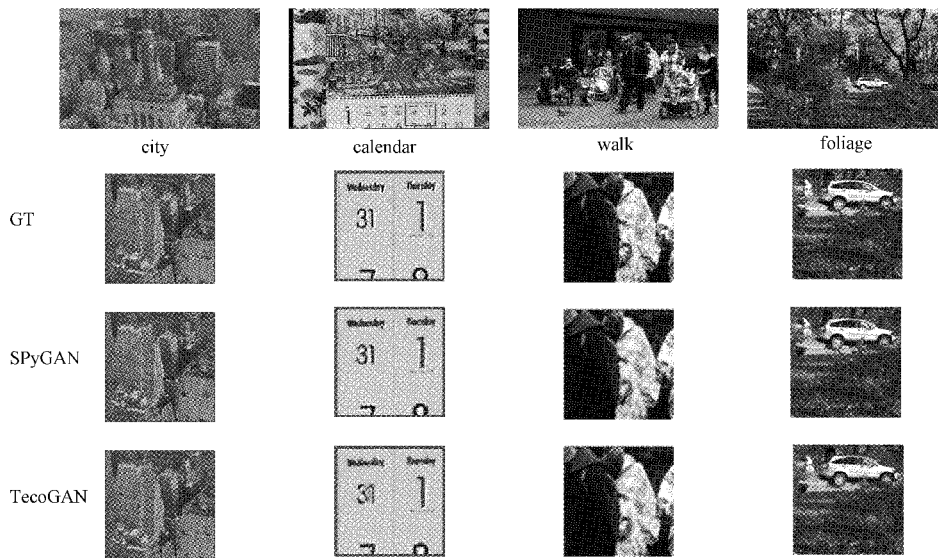


图 5 超分视觉效果对比图

图 5 中从左至右依次选取的是 Vid4 数据集中 city、calendar、walk 和 foliage 4 个视频序列中的中间第 8 位的序列帧, 分别取原始高分辨率图像 (GT)、SPyGAN 和 TecoGAN 生成图像进行对比, 从 city 组中能很明显的看出框选区域内建筑的纹理细节, SPyGAN 更为贴近原始 HR 图像, 并且图像更加清晰; 从 calendar 组中对比效果图可知, SPyGAN 生成的日历中的 31 这个数字更为真实且字母也更为锐利, 并且图像的清晰度远超 TecoGAN; 从 walk 组中可明显看出 SPyGAN 生成的人物衣服纹理细节相比 TecoGAN 生成衣服更为丰富; 从 foliage 组中可看出

SPyGAN 重建车辆的纹理结构相比 TecoGAN 重建图像中的车辆更为清晰。综上所述, SPyGAN 在视觉重建效果方面已略微高出 TecoGAN 重建视觉质量。

最后本文设计了一个重建参数量和重建速率的对比实验, 分别给 SPyGAN 和 TecoGAN 输入分辨率为  $3 \times 134 \times 320$  的低分辨率图像, 均在一块 Nvidia1080Ti 加速的情况下输出分辨率为  $3 \times 536 \times 1280$  的高清图像, 利用每秒内传输帧数 FPS 来验证本文提出的 SPyGAN 在重建视频速率方向的可靠性, 利用相同输入条件下两种算法的浮点计算量 FLOPs 来验证本文提出的 SPyGAN 在降低计算量方向的

有效性,具体实验结果如表 3 所示。

表 3 重建参数量、计算量和速率对比实验

	TecoGAN	SPyGAN
Parameters	2.589 M	2.587 M
FLOPs	141.86 G	76.408 G
Speed (FPS)	24.747	59.353

由表 3 数据可知,在总体参数量相同的情况下,本文提出的空间金字塔光流预计网络 SPyGAN 产生的浮点计算量仅为 TecoGAN 的 53.86%,但重建速率高达其 239.83%倍,证明了在相同 GPU 环境下,本文提出的 SPyGAN 能有效提高重建速率。

### 3 结 论

为达到保证重建视觉质量的同时提高重建速率的目的,本文提出了基于空间金字塔生成对抗网络的超分辨率重建网络 SPyGAN,引入基于空间金字塔的轻量级光流预测模块来提高重建视觉质量;其次改进 Loss 函数用来移除漂移伪影以及保留高频细节;最后设计了帧循环生成网络架构和时空判别器网络架构来学习时空信息的分布,使最终生成图像的质量更为真实,以此来实现高速度高质量的视频超分辨率重建。通过数据集验证结果表明,本文算法能在保证重建质量的前提下有效提高重建速度。如何在降低实时重建速度的条件下进一步提升视觉质量是下一步研究的课题之一。

### 参考文献

- [1] 帅源. 基于深度学习的图像/视频超分辨率算法研究[D]. 上海:上海大学,2019.
- [2] 李金航,肖亮,韦志辉. 基于时空相关性的视频超分辨率重建算法[J]. 计算机与数字工程, 2020, 48(4): 946-951.
- [3] 卞鹏程,郑忠龙,李明禄,等. 基于注意力融合网络的视频超分辨率重建[J]. 计算机应用, 2021, 41(4): 1012-1019.
- [4] 程松盛,潘金山. 基于深度学习特征匹配的视频超分辨率方法[J]. 计算机科学, 2021, 48(7): 184-189.
- [5] 程焕新,刘文翰. 基于生成对抗网络的图像超分辨率研究[J]. 电子测量技术, 2020, 43(14): 132-135.
- [6] LEDIG C, THEIS L, HUSZAR F, et al. Photo-

realistic single image super-resolution using a generative adversarial network [J]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 105-114.

- [7] 杨娟,李文静,汪荣贵,等. 融合感知损失的生成式对抗超分辨率算法[J]. 中国图象图形学报, 2019, 24(8): 1270-1282.
- [8] 李伟,张旭东. 基于卷积神经网络的深度图像超分辨率重建方法[J]. 电子测量与仪器学报, 2017, 31(12): 1918-1928.
- [9] 姜玉宁,李劲华,赵俊莉. 基于生成式对抗网络的图像超分辨率重建算法[J]. 计算机工程, 2021, 47(3): 249-255.
- [10] WANG X T, YU K, WU SH X, et al. ESRGAN: Enhanced super-resolution generative adversarial networks[J]. ECCV, 2018: 63-79.
- [11] CHU M Y, XIE Y, MAYER J, et al. Learning temporal coherence via self-supervision for GAN-based video generation[J]. ACM Transactions on Graphics (TOG), 2020, 39(4): 75:1-75:13.
- [12] 王延年,刘航宇,刘宏涛,等. 基于小波深度残差网络的图像超分辨率重建[J]. 国外电子测量技术, 2021, 40(9): 160-164.
- [13] 刘佳,安鹤男,李蔚,等. 基于帧循环网络的视频超分辨率技术[J]. 电子技术应用, 2020, 46(9): 43-46.
- [14] RANJAN A, BLACK J. Optical flow estimation using a spatial pyramid network[J]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 2720-2729.
- [15] LAI W S, HUANG J B, AHUJA N, et al. Fast and accurate image super-resolution with deep laplacian pyramid networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(11): 2599-2613.

### 作者简介

蔡非凡,硕士研究生,主要研究方向为图像/视频超分辨率重建算法等。

E-mail: caifeifan@shu.edu.cn

万旺根,教授,博士生导师,主要研究方向为计算机图形学、信号处理和数据挖掘等。

E-mail: wanwg@staff.shu.edu.cn