

# 深度学习实时视频超分辨率重建实验设计

彭智勇<sup>1</sup>, 黄扬铤<sup>1</sup>, 秦祖军<sup>1</sup>, 梁红珍<sup>2</sup>

(1. 桂林电子科技大学 光电工程学院, 广西 桂林 541004;

2. 桂林生命与健康职业技术学院, 广西 桂林 541001)

**摘要:**针对现有视频超分辨率重建深度学习网络结构复杂、实时性差的实验问题,设计了基于深度学习的实时视频超分辨率重建实验。提出将 Ghost-Module 结构应用于循环递归神经网络,并借助改进的残差学习结构加速网络收敛,实现了视频超分辨率的实时重建。实验结果表明:新算法的运算速度为 FRVSR 算法的 5.4 倍;在参数量与 TOFlow 算法相近的基础下,运行速度是 TOFlow 模型的 34.5 倍,是高性能算法 DUF-52L 的 28.8 倍;针对  $480 \times 270$  低分辨率视频影像可实现 24 f/s 的实时 4 倍超分辨率重建。所设计的轻量级循环卷积视频超分辨率重建网络在保证视频超分辨率重建质量的基础上极大地提高了重建速度。

**关键词:**实时超分辨率重建;循环卷积神经网络;残差网络;深度学习

中图分类号:TP 391.41;TP 18

文献标志码:A

文章编号:1006-7167(2023)09-0035-05



## Experiment Design of Real-time Video Super-resolution Reconstruction Based on Deep Learning

PENG Zhiyong<sup>1</sup>, HUANG Yangpi<sup>1</sup>, QIN Zujun<sup>1</sup>, LIANG Hongzhen<sup>2</sup>

(1. School of Optoelectronic Engineering, Guilin University of Electronic Technology, Guilin 541004, Guangxi,

China; 2. Guilin Life and Health Career Technical College, Guilin 541001, Guangxi, China)

**Abstract:** Aiming at the problem of complex structure and long processing time in traditional video super-resolution reconstruction algorithm, this paper proposes a new algorithm, in which the GhostModule structure is applied in the recurrent recursive neural network and the residual structure is used to accelerate the convergence of the network. Experimental results show that the speed of the new algorithm is 5.4 times faster than that of FRVSR algorithm; the parameters of the new algorithm are similar to those of TOFlow algorithm, but the running speed is 34.5 times; the speed of new algorithm is 28.8 times faster than that of DUF-52L. It realizes the real-time 4x super-resolution reconstruction with the speed of 24 frames per second for the low-resolution video with resolution of  $480 \times 270$ . The proposed lightweight recurrent convolution video super-resolution reconstruction network not only has good reconstruction quality, but also greatly improves the reconstruction speed.

**Key words:** real-time super-resolution reconstruction; recurrent convolution neural network (RCNN); residual network; deep learning

收稿日期:2023-01-20

**基金项目:**广西自然科学基金项目(2020GXNSFAA159091);广西高等教育本科教学改革工程重点项目(2022JGZ125, 2023JGZ126);广西研究生教育创新计划项目(JGY2022131)

**作者简介:**彭智勇(1981-),男,湖南娄底人,博士,副教授,主要从

事图像处理、计算机视觉方向的研究。

Tel.:0773-2291522;E-mail:pzy@guet.edu.cn

**通信作者:**梁红珍(1986-),女,广西博白人,学士,讲师,主要从事资料整理、电路测试。

Tel.:0773-5580615;E-mail:13024953@qq.com

## 0 引言

影像超分辨率重建是指通过软件算法从低分辨率的图像、视频中恢复出纹理清晰的高分辨率图像或视频,最早由 Harris<sup>[1]</sup> 和 Goodman<sup>[2]</sup> 提出。超分辨率重建是后期图像识别、处理、测量的基础,是图像处理领域中的基础热点问题,特别是基于深度学习的影像超分辨率重建技术由于其高性能,近年很多学者进行了深入的研究。基于深度学习的视频超分辨率重建技术可以分为基于单帧图像的帧内超分辨率重建和基于多帧视频的帧间超分辨率重建。

基于单帧图像的帧内超分辨率重建,有很多学者基于深度学习提出了超分辨率重建模型,如: VDSR<sup>[3]</sup>、ESPCN<sup>[4]</sup>、ClassSR<sup>[5]</sup> 等。基于帧内图像的超分辨率重建算法由于仅考虑帧内特征信息,忽略了视频序列的帧间相关,性能较好的超分辨率重建算法计算量和内存损耗过大;结构简单算法则特征信息不能充分利用,重建效果不好。基于多帧视频的帧间超分辨率重建,有 VESPCN<sup>[6]</sup>、FRVSR<sup>[7]</sup>、RBP<sup>[8]</sup> 等,此类算法往往将已经完成超分辨率重建的帧作为后续特征提取的输入,当连续帧之间存在较大变化时,往往存在严重伪影及误差累积,并且现有深度学习超分辨率重建算法运算复杂,还不能实现实时超分辨率重建。

本文设计了基于深度学习的实时视频超分辨率重建研究型实验。针对视频特点,在利用帧内特征的同时也利用帧间特征,解决处理速度慢的问题,实现实时超分辨率重建。该实验将 GhostModule 结构应用于循环递归神经网络并结合残差结构设计一个快速的视频超分辨率网络,采用具有多级跳跃连接残差映射,以解决循环卷积神经网络中存在梯度消失的风险,从而能长时间保存高频纹理信息,在取得较好重建效果的情况下实现实时视频超分辨率重建。

## 1 实验设计

### 1.1 实验算法原理

#### 1.1.1 整体结构设计

循环卷积神经网络(RCNN)<sup>[9]</sup> 可以对时域特征进行很好的建模,广泛应用于自然语言处理的研究,并可用于提炼视频帧间时域特征。Kim 等<sup>[3]</sup> 提出的残差结构超分辨率重建网络 VDSR,该结构加深了卷积神经网络的深度,具有较好的超分辨率重建性能,VDSR 采用的全局的残差结构,尽管有效地把低层特征传递到高层,但没有充分利用中间层的特征,导致重建的图像纹理细节不够清晰。本文在传统 RCNN 的基础上,受到 VDSR 的启发,设计了结合残差结构的 RCNN,实时视频超分辨率重建网络整体结构(见图1)。

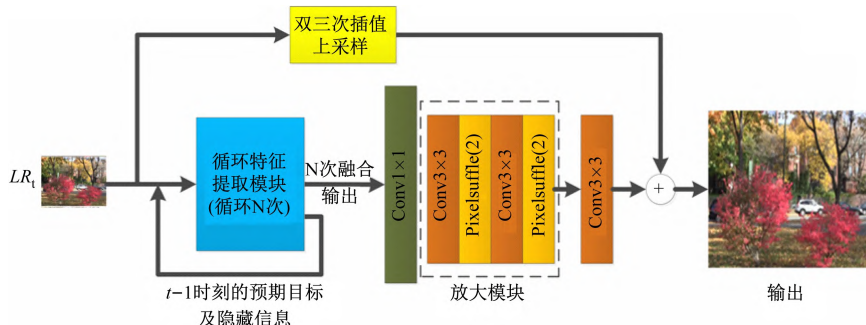


图1 实时视频超分辨率重建网络整体结构示意图

新的网络结构与 VDSR 处理方法相似,网络分为两条支路。在初始阶段,上面的支路对视频当前帧  $LR_t$  进行双三次方线性插值得到初始的高分辨率图像;下面的支路进行残差学习,对输入进行循环特征提取,并把得到的高频特征信息输入到亚像素卷积层对残差图像放大。最后通过残差图像和初始的高分辨率图像相加,得到最后的当前帧超分辨率图像  $SR_t$ 。其中,PixelShuffle(2) 是一放大 2 倍的上采样模块,在图 1 结构中以放大 4 倍为例通过 2 个 PixelShuffle(2) 共放大到 4 倍。在整体结构上,新网络与 VDSR 不同的是,VDSR 在当前图像输入的特征行提取模块之前就进行放大,因此后面的特征提取是在高分辨率图像上进行运算,这无疑增加了计算量。而本文实验针对低分辨率影像通过 RNN 网络循环进行特征提取,最后才

对残差图像进行放大,能有效减少计算量并保证处理效果。算法实现的重点为循环特征提取模型的设计。

#### 1.1.2 循环特征提取模块设计

GhostModule<sup>[10]</sup> 网络基于一组原始的特征图,结合一系列线性变换,以很小的计算代价生成许多从原始特征发掘所需信息的“幻影”特征图,从而通过低的计算成本生成了更多的特征图。GhostModule 网络结构与传统结构参数量对比如下式表示:

$$\left. \begin{aligned} p_1 &= f \times f \times k \times k \\ p_2 &= (f \times f/r) \times k \times k + (f \times f/r) \times 1 \times 1 \end{aligned} \right\} \quad (1)$$

式中: $p_1$  为传统结构的参数量; $p_2$  为 GhostModule 网络的参数量;假设特征提取模块的输入和输出通道数量均为  $f$ ;  $k$  为卷积核的大小(这里取  $3 \times 3$ );  $r$  为 GhostModule 的比例参数。由式(1)2 个参数之比得

$$c = \frac{p_1}{p_2} = \frac{f \times f \times k \times k}{(f \times f / r) \times k \times k + (f \times f / r) \times 1 \times 1} \quad (2)$$

当  $k$  取值较大时,分母后半部分可忽略不计,此时  $c \approx r$ ,即相比于传统网络。GhostModule 具有更深的网络结构,更大的感受野,且参数量是普通卷积层的  $1/r$ ,从而 GhostModule 网络具有强的特征提取能力及小的算法量。

图2所示为运用 GhostModule 设计循环特征提取模块示意图,其中图2(b)为本文特征提取模块内部结构,与传统的循环块[见图2(a)]相比,设计的循环块由 GhostModule 和隐藏信息与输出分支结构两部分组成。该循环块采用具有多级跳跃连接层之间的残差映射,解决了循环卷积神经网络中存在梯度消失的缺陷,同时确保了流畅的信息流,并能够长时间的保存高频纹理信息,从而网络更容易地处理更长的视频帧序列。

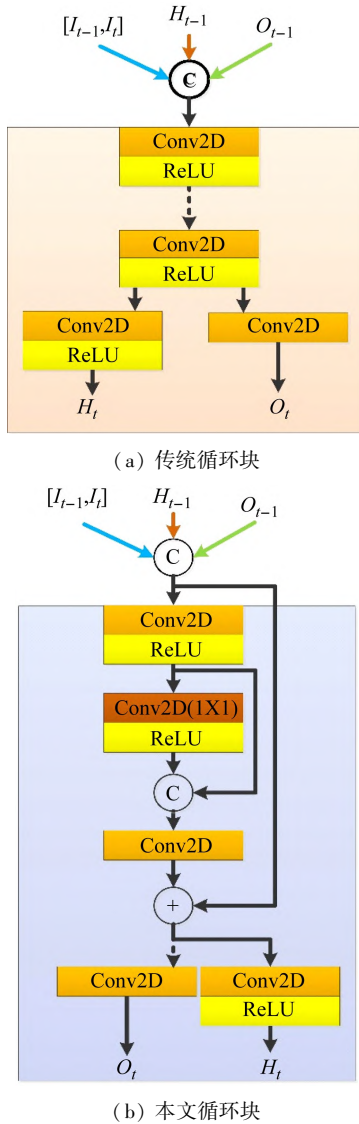


图2 循环特征提取模块示意图

由图2(b)可知,新的循环特征提取模块把当前帧  $I_t$ 、前一帧  $I_{t-1}$  ( $t$  为影像帧编号,  $t=0,1,\dots,n-1$ ), 它们之间的隐藏信息和上一个循环特征提取模块的输出作

为输入,得到循环块前部分 (GhostModule) 的输出;同时,将 GhostModule 的输出分别输入到 2 条支路:左边的支路作为本模块的预期目标输出  $O_t$  (RGB 三通道);右边支路作为本模块的隐藏 (遗留) 信息  $H_t$ 。把每次通过循环块得到的预期目标输出融合到  $n-1$  个历史预期目标中,即

$$S_t = [O_0, \dots, O_{n-1}] \quad (3)$$

式中:  $n$  为循环块执行的次数;  $S_t$  为  $n$  次循环块预期目标输出的融合后的结果,其通道数为  $n \times 3$ 。进一步把融合后的结果  $S_t$  输入到放大模块,即可得到放大后的残差图像  $H_{\text{RES}}$ 。最后把  $H_{\text{RES}}$  与初始双三次插值放大的高分辨率  $H_{\text{BICUBIC}}$  相加即可得到最终的超分辨率图像

$$\text{SR} = H_{\text{RES}} + H_{\text{BICUBIC}} \quad (4)$$

## 1.2 实验环境及设置

### 1.2.1 系统环境

实验采用的硬件环境为:CPU 型号为英特尔 i9-10900K,内存:32GB;系统:Win 7;GPU 为 NVIDIA RTX2080TI, GPU 主要用于对算法并行加速计算。实验软件环境为:ubuntu16.04、NVIDIA-3090、pytorch1.7.1、CUDA11.0 等。

### 1.2.2 测试数据及参数设置

本实验使用的训练数据集为 Vidmeo-90K<sup>[11]</sup>, 该数据集大约包含 90 000 个视频序列,每个视频序列包含 7 帧视频剪辑与各种运动和不同的场景图像,其中随机抽取 8 000 个视频序列用于训练,其余用于测试;实验的验证集为 Vid4 (包含 4 个视频序列) 和 UDM10 (包含 10 个视频序列)。由于 Vidmeo-90K 只提供高分辨率的目标图像,通过对高分辨率图像进行高斯模糊 (均方差  $\sigma=1.6$ ) 下采样到  $1/4$  像素得到  $64 \times 64$  大小的低分辨率输入块。经顺时针随机旋转  $0^\circ$ 、 $90^\circ$ 、 $180^\circ$ 、 $270^\circ$ ,左右与上下翻转得到 8 倍于原始数据集的增强数据集。本文以 Y 通道下重建高分辨率图像的 PSNR 与 SSIM 作为实验的评价指标,同时评估模型的参数量、计算复杂度和运算速度。

在训练循环卷积神经网络时,学习速率最初被设置为 0.000 1,经过每 20 个 epoch 学习率降低 0.1 倍,共训练了 60 个 epoch。本文实验选用的优化器为 Adam,优化器的相关参数设置:  $\beta_1=0.9$ ,  $\beta_2=0.999$  和权重衰减为 0.000 5,以 L1 作为损失函数进行监督训练,训练的批量数据大小为 4。

## 2 实验结果与分析

### 2.1 峰值信噪比 (PSNR) 与结构相似性 (SSIM) 性能对比

本文以通道数  $C=128$ ,循环特征提取模块数量分别取  $B=5$  和  $B=10$  两种网络结构与目前主流视频超分辨率重建网络 VESPCN, FRVSR, RBPN、RVSR-



LTD<sup>[12]</sup>, DRVSR<sup>[13]</sup>, DUF-52L<sup>[14]</sup>, PFNL<sup>[15]</sup> 在 VID4 数据集上进行了 PSNR 与 SSIM 评价指标对比测试,结果如表 1 所示。由表 1 可见,本文算法在 VID4 数据集 4

个场景的测试中,本文 5-128 和 10-128 比大多数传统算法有更高的评价指标。

表 1 在 VID4 上关于 PSNR/SSIM 性能的对比

模型	×2				×4			
	Clendar	City	Foliage	Walk	Clendar	City	Foliage	Walk
Bicubic	22.65/0.54	28.12/0.63	25.89/0.54	26.83/0.88	18.83/0.49	23.84/0.52	21.52/0.44	23.01/0.70
TOFlow	26.86/0.80	31.68/0.93	30.72/0.88	33.91/0.97	22.29/0.72	26.79/0.77	25.31/0.71	29.02/0.87
VESPCN	26.59/0.77	31.23/0.86	30.63/0.83	34.23/0.97	22.21/0.71	26.48/0.72	25.07/0.69	28.40/0.87
RVSR-LTD	26.15/0.74	31.21/0.84	30.66/0.84	33.03/0.96	21.91/0.69	26.32/0.71	25.07/0.69	28.24/0.86
DRVSR	27.22/0.82	31.75/0.91	30.86/0.87	33.88/0.97	22.88/0.75	27.06/0.76	25.58/0.73	29.11/0.88
FRVSR	27.51/0.85	32.57/0.92	31.05/0.89	34.07/0.97	23.44/0.88	27.65/0.80	25.97/0.75	29.70/0.89
DUF-52L	27.68/0.86	32.93/0.92	32.19/0.90	34.26/0.98	23.85/0.80	27.97/0.82	26.22/0.76	30.47/0.91
RBPN	28.18/0.86	32.87/0.92	32.17/0.89	34.36/0.98	24.02/0.80	27.93/0.80	26.21/0.75	30.62/0.91
PFNL	27.66/0.87	33.01/0.90	32.28/0.88	34.28/0.97	23.56/0.82	28.11/0.83	26.42/0.77	30.55/0.91
本文 5-128	28.89/0.86	32.94/0.92	32.25/0.88	34.07/0.96	23.87/0.80	27.81/0.80	26.41/0.77	29.97/0.90
本文 10-128	28.93/0.88	32.96/0.92	32.38/0.89	34.14/0.97	24.05/0.81	27.83/0.81	26.54/0.77	30.27/0.90

2.2 速度对比

与 TOFlow, FRVSR 10-128, DUF-52L, RBPN 在 UDM10 数据集上进行参数量、速度对比测试,结果如表 2 所示。由表 2 可见,本文 5-128 的方法在参数量和计算量较小的前提下,测试指标 PSNR 比 TOFlow 高 1.81 dB,比 FRVSR 高 0.98 dB;与较大的模型相比,本文 10-128 结构在 PSNR 上比 DUF-52L 低 0.01 dB,比 RBPN 低 0.19 dB。

表 2 在 UDM10 上关于速度的对比测试(×4)

模型	参数量	速度/ms	PSNR/SSIM
Bicubic			28.47/0.852 3
TOFlow <sup>[11]</sup>	1.41Mb	1 658	36.26/0.943 8
FRVSR <sup>[7]</sup>	5.05Mb	129	37.09/0.952 2
DUF-52L <sup>[14]</sup>	5.82Mb	1 382	38.48/0.960 5
RBPN <sup>[8]</sup>	12.2Mb	3 482	38.66/0.959 6
本文 5-128	1.54Mb	48	38.07/0.957 8
本文 10-128	2.67Mb	62	38.47/0.960 3

以 320 × 180 的低分辨率视频序列为输入进行 4 倍放大输入 1280 × 720 的高分辨率视频序列。与传统的快速的 FRVSR 模型相比,FRVSR 速度为 129 ms/帧,本文 5-128 速度为 24 ms/帧,故本文的重建速度为 FRVSR 的 5.4 倍。本文的模型参数为 FRVSR 的 30.5%,在参数量上与 TOFlow(1.41Mb)相近的基础上,运行速度是 TOFlow(1 658 ms/帧)模型的 34.5 倍,是高性能算法 DUF-52L 的 28.8 倍。将本文算法应用于超分辨率系统,以 1 920 × 1 080 作为高分辨率图像,取  $B=5, C=128$  的网络结构,可以达到 24 f/s 的速度实现视频超分辨率重建。

2.3 真实效果对比

本文在效果图上进行了对比测试,在 VID4 上对几种经典算法与本文算法的 2 幅图(共 4 个场景)实际视觉效果对比,如图 3 所示。

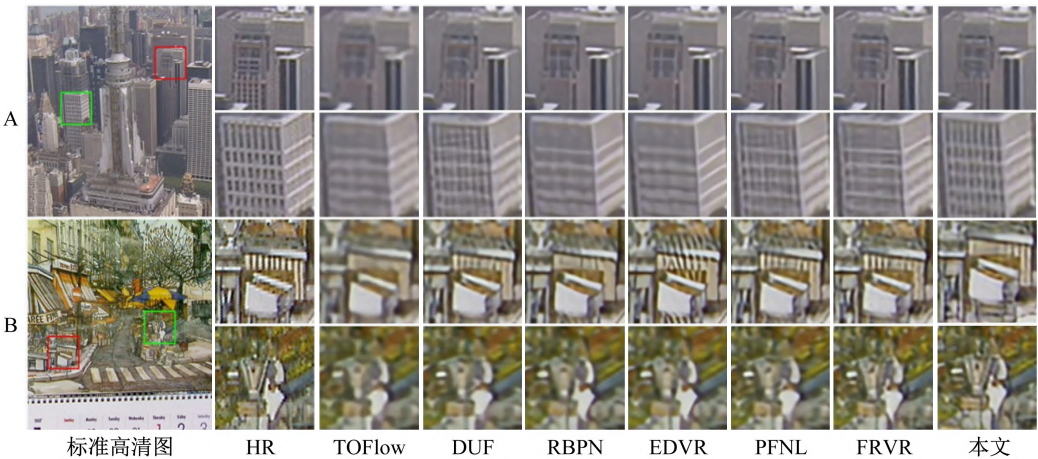


图 3 VID4 数据测试效果图(A:第 1 幅图,B:第 2 幅图)

由第1、2幅效果图的绿框场景可见,本文算法纹理明显比其他的算法更清晰;在第2幅图的红色场景区域,所有算法相比于原始的高清图像,均存在一定的纹理丢失。

### 3 结 语

本文基于RCNN,引入GhostModule代替传统的卷积层,改进了传统的残差结构,并搭建一个轻量级的循环卷积神经网络进行视频的超分辨率重建,实现视频的实时超分辨率重建。实验结果表明:本新网络的重建速度为FRVSR的5.4倍;在参数量上与TOFlow相近的基础上,运行速度是TOFlow模型的34.5倍;是高性能算法DUF-52L的28.8倍。

同时,本文基于前沿技术与工程应用角度,将科研项目与教育教学结合,积极探索和开设研究型教学实验<sup>[16-17]</sup>,开展了基于深度学习的实时视频超分辨率重建研究型实验项目设计,很好地锻炼了学生的探索精神,培养学生运用理论知识和实践技能探索解决复杂问题的能力,锻炼学生积极探索精神与工程创新能力。

### 参考文献(References):

- [1] Harris J L. Diffraction and resolving power[J]. Journal of the Optical Society of America, 1964, 54(7): 931-936.
- [2] Goodman J W. Introduction to Fourier optics: McGraw-Hill[J]. Optical Engineering, 1996, 35(3): 1513-1513.
- [3] Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016:1646-1654.
- [4] Shi W Z, Caballero. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016:1874-1883.
- [5] Kong X T, Zhao H Y, Qiao Y, et al. ClassSR: A general framework to accelerate super-resolution networks by data characteristic[C]//Proceedings of 2021 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Online, IEEE, 2021:12011-12020.
- [6] Caballero J, Ledig C, Aitken A, et al. Real-time video super-resolution with spatio-temporal networks and motion compensation[C]//Proceedings of 2017 IEEE Computer Vision and Pattern Recognition (CVPR). Hawaii HI, USA: IEEE, 2017:2848-2857.
- [7] Sajjadi M S M, Vemulapalli R, Brown M. Frame-recurrent video super-resolution[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, UT, USA: IEEE, 2018:6626-6634.
- [8] Haris M, Shakhnarovich G, Ukita N. Recurrent back-projection network for video super-resolution[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, 2019: 3892-3901.
- [9] Li D, Liu Y, Wang Z. Video super-resolution using non-simultaneous fully recurrent convolutional network[J]. IEEE Transactions on Image Processing, 2019, 28(3): 1342-1355.
- [10] Han K, Wang Y, Tian Q, et al. Ghostnet: more features from cheap operations[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020: 1577-1586.
- [11] Chen B A, Wu J J, Wei D L, et al. Video enhancement with task-oriented flow[J]. International Journal of Computer Vision. 2019, 127(8): 1106-1125.
- [12] Su H, Wu Y, Zhou J. Adaptive incremental video super-resolution with temporal consistency[C]//Proceedings of 2011 18th IEEE International Conference on Image Processing. Brussels, Belgium: IEEE, 2011:1149-1152.
- [13] Tao X, Gao H, Liao R, et al. Detail-revealing deep video super-resolution[C]//Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017: 4482-4490.
- [14] Jo Y, Oh W S, Kang J, et al. Deep video super-resolution network using dynamic upsampling filters without explicit motion compensation[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 3224-3232.
- [15] Yi P, Wang Z, Jiang K, et al. Progressive fusion video super-resolution network via exploiting non-local spatio-temporal correlations[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South): IEEE, 2019: 3106-3115.
- [16] 高 禄,曹佳宁,卞 通,等. 双残差深度学习鬼成像研究[J]. 实验室研究与探索, 2021, 40(12): 13-17.
- [17] 涂淑琴,黄 健,林跃庭,等. 基于改进 Faster R-CNN 的百香果自动检测[J]. 实验室研究与探索, 2021, 40(11): 32-37.
- [7] 罗贤欣,刘光斌,王 忠. 干涉仪测向技术研究[J]. 舰船电子工程, 2018, 32(8): 74-76.
- [8] 李东虎. 干涉仪测向系统中的基线配置技术[J]. 无线电工程, 2014, 44(4): 17-19.
- [9] 高 婕,牛 刚,杜太行,等. 超短波多径信号相关干涉仪测向算法[J]. 兵工学报, 2020, 41(11): 2252-2259.
- [10] 耿 赞,谷振宇. 基于FPGA的短波相关干涉仪测向[J]. 舰船电子对抗, 2014, 37(2): 57-60.
- [11] 杨明洋,向长波,房鹏飞. 基于相关峰的理论样本干涉仪测向算法[J]. 电子测量与仪器学报, 2021, 35(10): 33-40.
- [12] 胡天翔,陈 科,赵志强. 一种基于System Generator的CORDIC算法实现[J]. 科技创新与应用, 2020(19): 23-27.
- [13] 李春娟,李 沙. 基于自适应旋转角度的CORDIC算法的设计与仿真[J]. 数字技术与应用, 2021, 39(4): 113-115.
- [14] 熊正大,张宏伟. 基于CORDIC算法的干涉仪测向的FPGA实现方法[J]. 电子技术与软件工程, 2020(3): 113-115.
- [15] 王 强,应 浩. 反正切函数的CORDIC算法及其FPGA实现[J]. 兵工自动化, 2020, 39(6): 45-48.

(上接第34页)