计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.211000007

基于深度学习的视频超分辨率重构进展综述

冷佳旭[1,2]　王　佳[1]　莫梦竟成[1]　陈泰岳[1]　高新波[1]
1 重庆邮电大学图像认知重庆市重点实验室　重庆 400065
2 南京理工大学江苏省社会安全图像与视频理解重点实验室　南京 210094
(lengjx@cqupt.edu.cn)

摘　要　视频超分辨率是根据给定的低分辨率视频序列恢复其对应的高分辨率视频帧的过程。近年来,VSR 在深度学习的驱动下取得了重大突破。为了进一步促进 VSR 的发展,文中对基于深度学习的 VSR 算法进行了归类、分析和比较。首先,根据网络结构将现有方法分为两大类,即基于迭代网络的 VSR 和基于递归网络的 VSR,并对比分析了不同网络模型的优缺点。然后,全面介绍了 VSR 数据集,并在一些常用的公共数据集上对已有算法进行了总结和比较。最后,对 VSR 算法中的关键问题进行了分析,并对其应用前景进行了展望。

关键词:视频超分辨率;深度学习;卷积神经网络;帧间信息
中图法分类号　TP183

Survey on Video Super-resolution Based on Deep Learning

LENG Jia-xu[1,2],WANG Jia[1],MO Meng-jing-cheng[1],CHEN Tai-yue[1] and GAO Xin-bo[1]
1 Key Laboratory of Image Cognition,Chongqing University of Posts and Telecommunications,Chongqing 400065,China
2 Jiangsu Key Laboratory of Image and Video Understanding for Social Safety,Nanjing University of Science and Technology,Nanjing 210094,China

Abstract　Video super-resolution (VSR) aims to reconstruct a high-resolution video from its corresponding low-resolution version. Recently,VSR has made great progress driven by deep learning. In order to further promote VSR,this survey makes a comprehensive summary of VSR,and makes a taxonomy,analysis and comparison of existing algorithms. Firstly,since different frameworks are very important for VSR,we group the VSR approaches into two categories according to different frameworks:iterative- and recurrent-network based VSR approaches. The advantages and disadvantages of different networks are further compared and analyzed. Secondly,we comprehensively introduce the VSR datasets,summarize existing algorithms and further compare these algorithms on some benchmark datasets. Finally,the key challenges and the application of VSR methods are analyzed and prospected.

Keywords　Video super-resolution,Deep learning,Convolutional neural network,Inter-frame information

1　引言

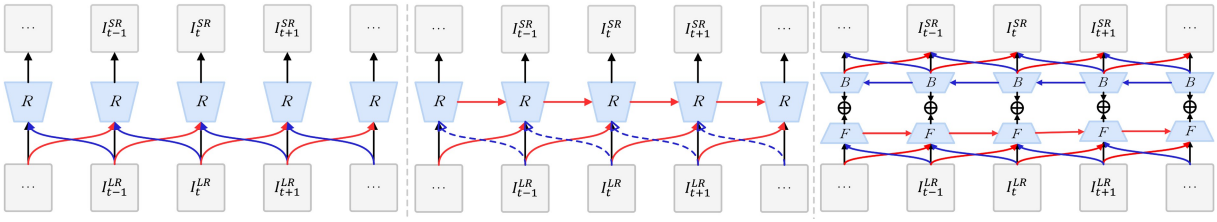
超分辨率重建(Super-Resolution,SR)旨在根据一幅或者多幅低分辨率(Low-Resolution,LR)图像恢复其对应的高分辨率(High-Resolution,HR)图像,该问题已经发展成为计算机视觉和图像处理中一个经典且具有挑战性的研究方向。SR 在医学影像、视频监控和卫星影像等领域有重要的应用前景,因此该技术吸引了越来越多研究者的关注。

SR 由 Harris[1]于 20 世纪 60 年代提出。根据需超分图像个数的不同,该领域可以分为单图像超分辨率(Single Image Super-Resolution,SISR)和多帧图像超分辨率。本文主要关注属于多帧图像超分辨率的视频超分辨率(Video Super-Resolution,VSR)。作为 SR 领域中最基本的问题,SISR 已得到较为深入的研究。对于 SISR,HR 图像的恢复是通过探索单个图像的内在特性实现的。而对于 VSR,帧内的空间信息和帧间的时间信息都可以用来增强 LR 的细节。在 SISR 的基础上,VSR 也得到了极大的发展,目前涌现出了许多优秀的 VSR 算法。现有的方法可以分为两大类:传统方法和深度学习方法。在早期的研究中,许多有实用价值的传统方法被提出,如 Bayesian 分析法[2-5]和凸集投影法[6]。由于它们只是

到稿日期:2021-09-30　返修日期:2021-11-04
基金项目:国家自然科学基金(62036007,62050175,62102057);重庆市教委科学技术研究项目(KJQN-202100627)
This work was supported by the National Natural Science Foundation of China(62036007,62050175,62102057) and Science and Technology Research Program of Chongqing Municipal Education Commission(KJQN-202100627).
通信作者:高新波(gaoxb@cqupt.edu.cn)

简单地将 LR 图像的像素点根据固定的数学表达式映射到 HR 图像对应的像素点中，因此这种显式的模型无法很好地处理视频中的各种场景。近年来，随着深度学习技术在广大领域中的成功应用，基于深度学习的 VSR 得到了研究者的广泛关注。许多深度神经网络，如卷积神经网络(Convolutional Neural Network，CNN)、生成对抗网络(Generative Adversarial Network，GAN)和循环神经网络(Recurrent Neural Network，RNN)[7-10]被应用到该领域。基于深度学习的 VSR 通常会一次输入多张 LR 视频帧到神经网络中，并将其中一帧视作目标帧，余下的视频帧视作相邻帧，其旨在通过深度神经网络进行非线性映射得到目标帧的 HR 视频帧。

尽管目前已有一些关于 SISR 算法的综述[11-13]，但是关于 VSR 算法的综述仍然较为匮乏，如 Daithankar 等[14]根据频率和空间域方法对传统 VSR 进行了简单的回顾。进一步地，Wu 等[15]则从时间、空间、时空这 3 种不同的角度出发，对传统的 VSR 进行了分类及总结。随后，Liu 等[16]根据处理相邻帧的不同方法，对基于深度学习的 VSR 算法进行了分类和较为全面的总结。近年来，有大量的基于深度学习的 VSR 算法被提出，本文对近年的 VSR 算法进行了全面的总结。不同于文献[16]将 VSR 算法按如何利用相邻帧的时间信息来进行分类介绍，本文从网络架构的角度来对已有算法进行分类和分析，如图 1 所示。相比文献[16]，本文的分类方法更能体现该领域的现状和发展趋势，能够为研究人员带来新的思考角度。



(a)迭代网络 (b)单向递归网络 (c)双向递归网络

注：其中 R 表示重构网络，F 表示前向网络，B 表示后向网络，红色、黑色和蓝色的线分别表示过去、现在和未来的信息

图 1 基于深度学习的视频超分辨率的不同网络框架(电子版为彩色)

Fig. 1 Different network frameworks for video super-resolution based on deep learning

本文首先介绍了 VSR 问题的背景，然后根据不同的网络架构全面总结和分析了现有的 VSR 算法。此外，本文还阐述了 VSR 数据集与评价指标，总结了 VSR 算法在当前面临的挑战、未来发展方向以及应用前景。

2 视频超分辨率重构背景

VSR 算法是 SISR 算法的一个扩展，旨在从给定的 LR 视频序列中恢复其对应的 HR 视频帧。这两个领域的共通之处在于都需要利用图像内的空间信息。除了图像内部信息外，VSR 算法还需要关注相邻帧的时间信息来帮助重构 HR 视频帧，比 SISR 算法更具挑战性。

本文令 I_t^{LR} 表示 LR 视频帧，I_t^{HR} 为 HR 视频帧，则退化过程可以表示为：

$$I_t^{LR}=\Phi(I_t^{HR},\delta) \tag{1}$$

其中，Φ 表示退化函数，δ 表示退化参数和各式各样的退化因子，如噪音、运动模糊和下采样因子。在现实生活中，I_t^{LR} 是很容易得到的，而退化函数 Φ 和退化参数 δ 很难确定。

VSR 算法的目的是将 LR 视频帧 I_t^{LR} 重构为无限接近于其对应的 HR 视频帧 I_t^{HR}，即式(1)的逆向过程，其过程可表示为：

$$I_t^{SR}=\Phi^{-1}(I_t^{LR},\theta) \tag{2}$$

其中，I_t^{SR} 表示重建的 HR 视频帧。

由于实际生活中的退化过程未知且复杂，因此现有方法通常将式(1)的退化过程描述为：

$$I_t^{LR}=(I_t^{HR}\otimes k)\downarrow_s+n \tag{3}$$

其中，k 表示模糊核，$\otimes$ 指卷积操作，$\downarrow_s$ 指缩放因子为 s 的下采样操作，n 表示高斯噪声。

3 视频超分辨率算法

由于视频是一个动态视觉图像序列，因此 VSR 算法可以借鉴 SISR 算法。Dong 等[17]提出的 SRCNN(Super-Resolution Convolutional Neural Network)方法，首次将三层的卷积神经网络引入到 SISR 任务中。随后，学者在此基础上作了进一步的研究和扩展，如 FSRCNN[18]、VDSR[19]、ESPCN[20]和 RCAN[21]。随着卷积神经网络在 SISR 以及其他图像恢复任务中的成功应用，Kappeler 等[7]于 2016 年提出了基于卷积神经网络的 VSR 方法，首次将深度学习应用于该领域。随后，大量的 VSR[8,22-25]采用滑动窗口的方式将有限的 LR 视频序列输入到迭代网络框架中，如图 1(a)所示。，然而，迭代的网络框架只能利用窗口内的视频帧，这对长距离信息依赖是不利的且只能通过增大窗口尺寸来解决此问题。进一步地，基于递归网络框架的 VSR 算法[26-28]被提出，它按时间顺序处理所有视频帧，可以捕获长时间依赖，弥补了基于滑动窗口 VSR 只能利用窗口内视频帧的缺陷。具体地，本文将该类方法详细分为了两个子类：单向递归网络(见图 1(b))和双向递归网络(见图 1(c))。

从文献[9,22,27]可以发现，不同的网络框架对 VSR 的性能影响很大。因此，与文献[16]根据相邻帧信息的利用方式来分类的 VSR 综述不同，本文根据不同的网络框架对该领域进行了全面的归类总结，并将已有方法分成了两类：迭代网络和递归网络(见图 2)。

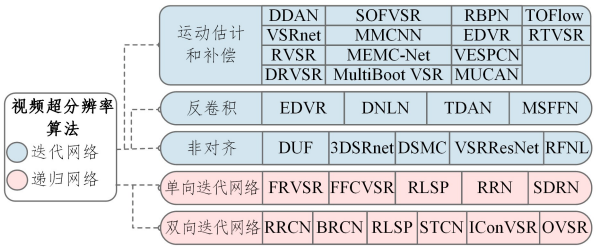


图 2 现有的视频超分辨率算法的分类

Fig. 2 Taxonomy of existing video super-resolution algorithms

3.1 迭代网络

目前，研究人员主要采用迭代网络框架的方式来处理输入的 LR 视频帧，如图 1(a)所示。基于迭代网络的 VSR 方法一般包含一个特征提取、一个对齐模块、一个融合模块和一个重构模块，如图 3 所示。具体地，以滑动窗口方式，取连续 n 个 LR 视频帧作为迭代网络的输入，其中一帧是需要重建的目标帧，余下的是相邻帧，通过不断迭代得到完整的 HR 视频序列。在一个视频序列中，迭代网络框架将整个 VSR 重建的过程视为多个相互独立的子过程。这些子过程在时间上是不相关的，这意味着它们可以并行计算从而节约大量的时间。现有基于迭代网络的 VSR 算法已经取得不错的成效，但是该类方法只有通过增加滑动窗口大小才可以考虑到长时间依赖关系，且它并未考虑到先前预测的 HR 视频帧对后续帧的辅助作用，这很可能是导致该类方法无法获得更好性能的原因。研究人员主要关注该类方法如何提高帧间信息利用率，同时这是获得高质量超分辨率结果的关键，因此，本文根据帧间信息的方式来全面介绍基于迭代网络的 VSR 算法。

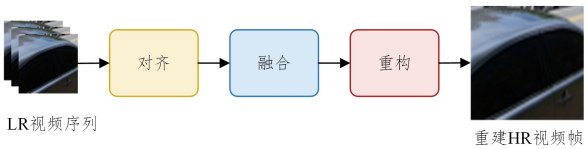


图 3 基于迭代网络的 VSR 算法一般使用的网络框架

Fig. 3 General flowchart of iterative based VSR

3.1.1 运动估计和补偿方法

在基于迭代网络的 VSR 方法中，大多数的方法都采用运动估计和补偿的方法来利用帧间信息。具体地，运动估计旨在提取帧间的运动信息，而运动补偿根据提取的运动信息来执行扭曲操作，使相邻帧与目标帧对齐。

运动估计可以分为传统方法(如 LucasKanade 算法[29]和 Druleas 算法[30])和深度学习方法(如 FlowNet[31]，FlowNet 2.0[32]和 SpyNet[33])，其中光流法是目前最流行的运动估计方法。一般地，光流法将两个视频帧(I_t^{LR} 和 I_i^{LR})当作输入，其中一个是目标帧 I_t^{LR}，另一个是相邻帧 I_i^{LR}。此方法通过这两个视频帧在时域中的相关性和变化来计算两者之间的运动信息(即光流矢量场)。上述过程可以描述为：

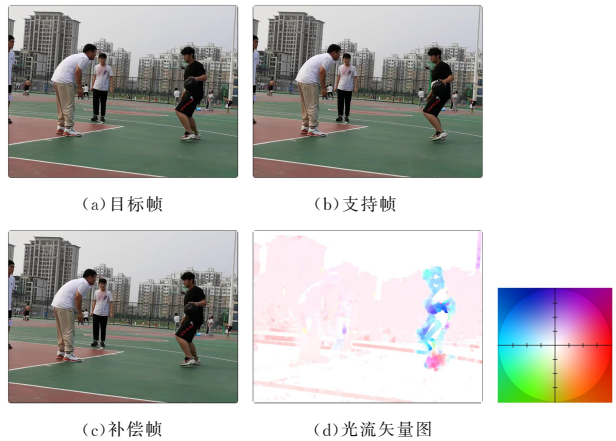
$$F_{i\to t}=(h_{i\to t},v_{i\to t})=ME(I_i^{LR},I_t^{LR};\theta_{ME}) \tag{4}$$

其中，$h_{i\to t}$ 和 $v_{i\to t}$ 分别是水平和垂直的光流矢量场 $F_{i\to t}$，ME 是用于计算光流矢量场的函数，θ_{ME} 是函数 ME 所需的参数。

运动补偿根据帧间的运动信息来对视频帧进行空间转换，使相邻帧与目标帧在空间域中对齐。它可以通过双线性插值、亚像素卷积[7]和 STN[12]等方法来实现。运动补偿的过程可以描述为：

$$J=MC(I,F,\theta_{ME}) \tag{5}$$

其中，MC 表示运动补偿函数，I 是相邻帧，F 是光流矢量场，θ_{MC} 为 MC 函数的参数。运动估计(光流法)和补偿的可视化例子如图 4 所示。



(a)目标帧 (b)支持帧

(c)补偿帧 (d)光流矢量图

注：不同的颜色表示不同的运动方向，颜色的强度表示运动的范围

图 4 一个运动估计和补偿的实例

Fig. 4 Example of motion estimation and compensation

目前，许多基于运动估计和补偿的优秀 VSR 算法不断涌现。在早期的 VSR 算法中，Kappeler 等[7]首先提出了一个具有三层的卷积神经网络，用于解决 VSR 任务。该方法使用传统的运动估计算法(Druleas[30])来计算运动信息，并提出了一个滤波对称强制执行机制和自适应运动补偿机制，分别用于加速训练和减小不正确的运动信息的影响，然后将得到的补偿帧和目标帧输入到网络中生成预测帧。然而，该方法的运动补偿机制较为低效，并且对 LR 视频帧进行预处理的操作严重影响了模型的速度。为了改进运动补偿操作，Tao 等[34]受 ESPCN[20]的启发提出了亚像素运动补偿层，它通过利用光流信息和亚像素信息来达到同时进行上采样和运动补偿的目的。为了得到更精准的光流信息，Wang 等[35]首先提出了一种光流重构网络，由粗到细地推断 HR 光流。然后将 HR 光流按空间位置重新排列下采样成 LR 光流，并将其作用到相邻帧，从而得到补偿帧。最后将目标帧和补偿帧一同输入重构网络，从而得到预测的 HR 视频帧。进一步地，Bare 等[36]提出了一种运动卷积核估计网络，以它是一种全卷积编码器-解码器结构，用于估计目标帧和相邻帧之间的运动信息，并生成与当前目标帧和相邻帧相对应的一对一维卷积核，然后利用该卷积核对相邻帧进行扭曲得到补偿帧，最后将补偿帧和目标帧送入后续的超分辨率网络，以得到最终的 HR 重建视频帧。为了同时利用 LR 视频帧的时空相关性，Li 等[10]提出了一个深度双重注意网络，将运动补偿网络和重建网络级联起来，其中运动补偿网络利用金字塔框架获取相邻帧之间的多尺度光流信息，并将得到的补偿帧和目标帧一同输入到重构网络中。同时，该方法将基于通道和空间维度的双重注意机制与残差块相结合，以关注对恢复高频细节有意义的特征。

由于视频帧内空间信息对 VSR 也是至关重要的，因此

一些学者们提出了全新的生成网络。MMCNN[37]，RBPN[9]和 Multiboot VSR[38]等将递归网络引入到重构模块中，从而提高视频帧内空间信息的利用率。Wang 等[37]提出了一种多记忆卷积神经网络，通过使用一系列利用帧间空间相关性的残差块来达到特征提取和重建的目的。具体地，该方法将卷积长短期记忆神经网络嵌入到残差模块中，从而形成一个多记忆残差块(Convolutional Long Short-Term Memory，Conv-LSTM[39])来逐步提取和保留连续 LR 视频帧之间的时间相关性。与文献[37]采用长短期记忆网络不同，Haris 等[9]受反向投影算法[40]的启发提出了递归反向投影网络，使用循环编码器-解码器模块整合连续视频帧的空间和时间上下文信息。随后，Kalarot 等[38]提出了多阶段多参考引导网络。与文献[37]和文献[9]在重构网络中使用迭代的 VSR 算法不同，多阶段多参考引导网络是将重构网络进行迭代。该网络的一个阶段由输入子网络、混合主干网和空间上采样子网络组成，并反复迭代该网络来重建 HR 视频帧，再将它们重新排列为多个低分辨率图像，重新用作附加参考帧，再次送入到网络中，以达到逐步引导和提升 VSR 算法性能的目的。

除了前文总结的方法外，在迭代网络中还存在其他的基于运动估计与补偿的方法，这些方法使用特殊架构来进行运动估计与补偿。例如，Liu 等[41]提出了空间对齐模块和时间自适应模块。空间对齐模块首先通过局部化网络来估计相邻帧和目标帧之间的转换参数，然后根据所得参数通过空间转换层将相邻帧与目标帧对齐。时间自适应模块将时间域划分为多个独立的子域，每个子域有一个重构网络负责生成相应 HR 视频帧的重建，最终的输出结果就是每个子域重建 HR 视频帧的加权总和。Caballero 等[25]提出了一种用于运动估计和补偿的空间转换器网络。首先将该模块得到的补偿帧送入到一系列卷积层进行特征提取和融合，然后通过亚像素卷积层获得最终的重建 HR 视频帧。由于以往的运动估计只能得到两帧之间的运动信息，因此，Kim 等[42]提出了时空转换模块，以达到同时计算多帧的光流信息并进一步对相邻帧进行空间转换的目的。近期，为了解决传统方法面对大运动视频带来结果不连续和伪影的问题，Liu 等[43]提出了一种双子网多级通信上采样网络的方法，用于大运动的 VSR。

基于运动估计和补偿的 VSR 算法可以较好地利用相邻帧的时间信息，是该领域中最常用的方法之一。然而，它带了较大的计算量并且依赖于准确的运动估计，不准确的运动估计会影响到后续的融合、重构等操作。为了改进上述问题，研究学者尝试将可变形卷积引入到 VSR 算法，并取得了不错的效果。下节，本文将详细介绍基于可变形卷积的 VSR 算法。

3.1.2 可变形卷积方法

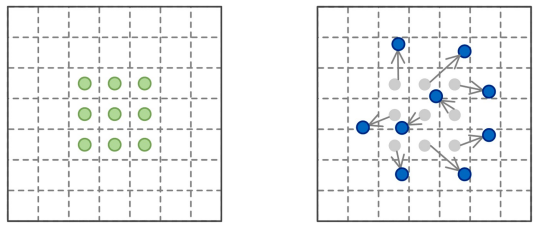
在普通卷积神经网络中，卷积核只对输入特征图的固定位置进行采样，且同一层卷积的感受野是一样的，这就造成卷积核无法适应多尺度或形状各异的对象。为了突破普通卷积神经网络的限制，Dai 等[44]于 2017 年第一次提出可变形卷积，并于 2019 年提出了改进的版本[45]。可变形卷积与普通卷积的区别如图 5 所示。具体地，基于可变形卷积的 VSR 通过将目标帧与相邻帧两个视频帧同时输入到卷积层中，以获得偏移值，然后将其通过可变形卷积作用到相邻帧，最终得到对齐的特征图。其中，可变形卷积详细操作是针对每个特征图的每个特征点 p_0 及对应的偏移值 Δp_k，具体公式如下：

$$F^a_{t+i}(p_0)=\sum_{k=1}^{k} w_k \cdot F_{t+i}(p_0+p_k+\Delta p_k)\cdot \Delta m_k \tag{6}$$

其中，$p_k \in \{(-1,-1),(-1,0),\cdots,(1,1)\}$，$F^a_{t+i}$ 为对齐后的特征图，Δp_k 和 Δm_k 是根据目标帧和参考帧串联起来的特征图来预测得到的，具体公式如下：

$$\Delta p_k = f([F_{t+i},F_t]),i\in[-N,N] \tag{7}$$

其中，f 是由几个卷积层组成的一般函数，F_{t+i} 和 F_t 分别表示第 $t+i$ 帧和第 t 帧的 LR 特征图。此外，为了方便描述，本文只考虑 Δp_k，忽略 Δm_k，上述过程如图 6 所示。



(a)普通卷积 (b)可变形卷积

注：(a)为普通卷积的采样方式；(b)为可变形卷积加上偏移量后采样点的变化

图 5 不同卷积的采样方式

Fig. 5 Different convolution sampling methods

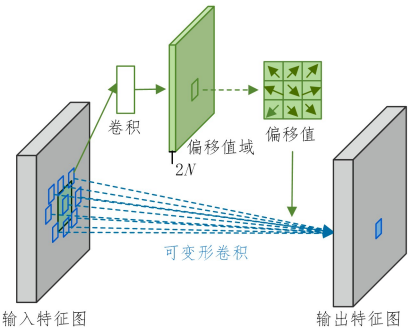


图 6 可变形卷积网络的具体实现[44]

Fig. 6 Implementation of deconvolution network[44]

前文提到基于运动估计和补偿的 VSR 方法严重依赖于精准的运动估计，效果也较差。为了缓解这个问题，Tian 等[24]提出了时域可变形对齐网络(Tempral Deformable Alignment Network，TDAN)，通过动态的估计特征空间上的偏移量来自适应地给相邻帧与目标帧对齐，以避免运动估计带来的误差。随后，Wang 等[46]在 TDAN 的基础上进一步提出了基于增强可变形卷积网络的 VSR 方法，同时设计了金字塔可变形卷积模块(PCD)和时空注意力融合模块(TSA)。具体地，PCD 受 TDAN 的启发，使用可变性卷积在特征空间上将每个相邻帧与目标帧对齐。与 TDAN 不同，PCD 使用金字塔结构首先对低尺度特征进行粗略对齐，然后将偏移值和对齐的特征传播到高尺度，以进行精确的运动补偿。为了有效融合多个对齐特征信息，TSA 通过计算相邻帧和目标帧特征之间的相关系数来引入时间注意机制，然后通过相关系数对每个相邻特征进行加权，最后将所有加权特征进行卷积融合。此后，Wang 等[47]提出了可变形非局部网络，并设计了对齐模块和非局部注意力模块。其中，对齐模块使用可变形卷积的分层特征融合模块来生成卷积参数，且级联多个可变形卷积，

使帧间对齐更加准确。其次,非局部注意力模块通过不同扩张率的 8 个 3×3 卷积从相邻帧中选择有用的特征。最近,Song 等[48]提出了端到端多阶段特征融合网络,它在前馈神经网络结构的不同阶段融合了目标帧和对齐的相邻帧的特征图。此外,该方法还设计了时间对齐分支,通过使用多尺度可变形卷积的方法来达到降低对齐误差的目的。

与基于运动估计和补偿的方法相比,上述方法通过可变形卷积计算偏移量,提升了网络空间变换的能力,但计算代价也随之增大。

3.1.3 非对齐方法

在迭代网络中,还有一类方法采取了较为隐式的方法来利用时间信息。例如,Jo 等[23]受动态滤波网络的启发,提出了动态上采样滤波器的方法,该方法使用提取时空信息的三维卷积滤波器,以避免使用运动估计和补偿。Lucas 等[49]将生成对抗网络引入到视频超分领域,并取得了不错的结果。特别地,该方法不对目标帧和相邻帧进行对齐操作,而是先上采样再进行一系列的卷积操作和特征融合操作,然后将结果送入到重构网络得到 HR 预测帧,最后鉴别器对该预测帧进行评价,帮助重构模块产生理想的结果。Yi 等[50]使用非局部残差块提取时空特征,同时也提出了渐进融合的网络框架,并在性能和速度上都取得了不错的效果。Isobe 等[51]通过将 LR 视频序列按时间间隔分为若干组的方式,有效地利用了时间信息,并将其输入到具有 2 维残差块和 3 维残差块的混合模块中进行组间融合,以达到充分利用帧间的互补信息来恢复目标帧丢失细节的目的。由于相邻视频帧存在大量相似的图案,因此 Li 等[52]设计了时间多对应聚合策略,以跨帧的方式利用相似的图案,并采用跨尺度的方法和非局部对应聚合策略来关注不同尺度视频帧的自相似性。

与对齐方法相比,非对齐方法避免了计算量增加和对齐误差的问题。然而,非对齐的方法未有效地利用相邻帧的时间信息,从而导致性能难以达到应用需求,因此如何高效地利用相邻帧的时间信息是需要继续探讨的。本文将在 3.2 节详细阐述基于递归网络的 VSR 算法。

3.2 递归网络

由于 RNN 可以用于处理序列数据,因此该网络框架能非常契合地应用到自然语言、视频、音频等领域。鉴于此,通过递归网络来解决 VSR 任务的算法[27-28,56]纷纷被提出。相比迭代网络以滑动窗口的方式处理视频序列,递归网络按顺序对视频帧进行处理。该网络可以很好地弥补迭代网络只考虑窗口内视频帧的缺陷,并且可以利用预测的上一帧信息,具体网络框架如图 1(b)、图 1(c)所示。在早期的递归网络中,模型不仅太大,而且没有体现出该框架的优势。近几年,一些方法针对此问题进行了改进,并在性能和速度上都取得了不错的效果。

3.2.1 单向递归网络

单向递归网络将 LR 输入帧按顺序从第一帧传播到最后一帧(见图 1(b))。Sajjadi 等[53]首次提出了一种基于递归网络的 VSR 算法。该方法不直接将光流信息作用到相邻帧,而是采用双线性插值将 LR 光流信息放大到与 HR 视频帧相同的大小,并将放大后的光流信息作用到上一次的预测帧,然后将 HR 帧按空间位置重新排列下采样成 LR 帧,并与 LR 帧一同输入到生成网络中,从而获得最终的结果。此后,Yin 等[54]提出了一个以一种新颖的方式利用帧间信息的基于上下文特征的 VSR 网络,它将 LR 未对齐视频序列和上一帧的输出作为网络的输入,以达到恢复高频细节信息和保持时间一致性的目的。Zhu 等[55]受可逆模块[56]的启发,设计了残差可逆模块,该模块可以有效地利用视频帧的空间信息,并提出了一个具有残差密集模块的长短期记忆网络,以提取时空信息和稀疏融合策略用于自适应地选择特征。最近,为了高效地利用前一帧的信息对目标帧进行超分,Isobe 等[57]提出了新颖的递归 VSR 方法,将 LR 视频序列分为结构和纹理两个部分并分别送入递归单元。此外,该方法还提出了隐藏状态自适应模块,使当前帧可以从隐藏状态中选择有用的信息辅助重构目标帧,以提升对场景切换和误差累积的鲁棒性。

尽管上述方法可以弥补迭代网络只能利用滑动窗口中的视频帧的缺陷,但是在单向递归网络中不同帧接收到的信息是不平衡的。例如,第一帧除了自身之外,不考虑来自视频序列的信息,而最后一帧却接收来自整个视频序列的所有信息。因此,单向递归网络会导致较早的帧的预测结果较差。接下来,下文将详细阐述基于双向递归网络的 VSR 算法,它不仅考虑了前面的视频帧同时也考虑了后面的视频帧。

3.2.2 双向递归网络

双向递归网络不仅可以考虑前面时间序列的视频帧信息,还可以利用未来视频序列的视频帧信息。双向递归网络由两个子网络构成:前向网络和后向网络(见图 1(c))。具体地,前向网络建模正序时间序列视频帧,反向网络加墨反向时间序列视频帧。在早期的研究中,Huang 等[58]提出了双向递归卷积网络,其中前向和后向网络采用一致的架构。最终的重建 HR 视频帧是两个子网络结果的融合。为了更好地利用时空信息,Guo 等[59]在 LSTM[39]的基础上将其扩展为双向、多尺度和多层次的卷积方式,并在性能上取得了不错的结果。Li 等[60]提出了一个基于非同步双向递归卷积网络,该方法通过非常深的双向递归卷积层来建模时空非线性映射,并取得了较好的效果。然而,它在生成不同 HR 视频帧时无法共享信息,导致效率低下。尽管上述方法均能有效提取时域上下文信息,但在速度上却表现不佳,在实时性要求高的任务上实用性不强。最近,基于双向递归网络的超分辨率算法在提高性能的同时也减少了额外的计算量。具体地,Fuoli 等[61]提出了一种递归潜在空间传播算法,将高维的潜在信息以隐式的方式在帧间传播时间信息,使其在速度上得到了提升。Chan 等[28]重新分析了视频超分网络的 4 大模块(网络结构、对齐、聚合和上采样)的作用以及优缺点,并提出了可能成为该领域基础的网络框架。Yi 等[27]提出了全知网络,该方法通过前向网络和后向网络,不仅利用了前面的输出,同时也利用了当前和未来的输出,因此在速度和性能上都取得了突破。

基于双向递归网络的 VSR 的研究较少,因此该方法还存在着一定的潜力待挖掘。该网络架构不仅可以解决迭代网络只能利用滑动窗口视频帧的问题,同时也弥补了单向递归网络无法考虑未来视频帧的缺陷。鉴于此,如何更好地将双向递归网络应用于 VSR 重建领域值得深入研究。

4 数据集和性能评估

4.1 数据集介绍

在 VSR 领域中，研究人员利用各种各样的数据集对神经网络进行训练(见表 1)，这些训练集在数量、视频个数及视频内容上都有一定的区别。这些数据集的 LR 图像通常都是使用人为设置的退化模型仿真得到，如双三次插值下采样和高斯模糊下采样，分别如式(8)、式(9)所示：

$$I_t^{LR} = I_t^{HR} \downarrow_s^{bic} \tag{8}$$

$$I_t^{LR} = (I_t^{HR} \otimes k) \downarrow_s \tag{9}$$

其中，s 表示下采样因子，k 表示高斯核，$\otimes$ 表示卷积操作。此外，由于不同的退化方法对模型的性能会产生一定的影响，因此，本文在对算法的性能进行比较时，展示了算法的退化方式。本文对测试集的相关信息进行了详细介绍(见表 2)。接下来本文将分为训练集与测试集并按时间顺序介绍现有的 VSR 数据集。

表 1 视频超分辨率算法常使用的训练集

Table 1 Training set often used in video super-resolution algorithms

数据集	年份	下载地址	视频个数	分辨率	颜色空间
V-train1	—	https://media.xiph.org/video/derf/	25	—	YUV
Venice	2014	https://www.harmonicinc.com/free-4k-demo-footage/	1	3840×2160	RGB
Myanmar	2014	https://www.harmonicinc.com/insights/blog/4k-in-context/	1	3840×2160	RGB
CDVL	2016	http://www.cdvl.org/	—	1920×1080	RGB
Vimeo-90K	2019	http://toflow.csail.mit.edu/	91701	448× 256	RGB
MM522	2019	https://drive.google.com/open?id=1xPMYiA0JwtUe9GKiUa4m31XvDPnX7Juu	522	318×180	RGB
REDS	2019	https://seungjunnah.github.io/Datasets/reds.html	270	1920×1080	RGB

注：'—'表示未知的信息

表 2 视频超分辨率算法常使用的测试集

Table 2 Testing set often used in video super-resolution algorithms

数据集	年份	下载地址	视频个数	分辨率	颜色空间
UVG	2020	http://ultravideo.cs.tut.fi/	16	3840× 2160	YUV
REDS	2019	https://seungjunnah.github.io/Datasets/reds.html	270	1920×1080	RGB
Vimeo-90K	2019	http://toflow.csail.mit.edu/	91701	448×256	RGB
YUV21	2014	http://www.codersvoice.com/a/webbase/video/08/152014/130.html	21	352×288	YUV
UDM10	2020	—	—	—	RGB
SPMC	2019	https://tinyurl.com/y426-dcn9	30	480×270	RGB
Vid4	2011	https://drive.google.com/drive/folders/10-gUO6zBe-OpWEamrWKCtSkkUFukB-9W5m	4	720×480	RGB
V-test	—	https://media.xiph.org/video/derf/	5	—	YUV

注：'—'表示未知的信息

4.1.1 训练集

(1)Myanmar 数据集[62]。Myanmar 由 Liao 等[62]提出，该数据集将收集的 26 个 1080p 高清视频剪辑成 160 个视频序列，从而构建一个 VSR 数据集。这些视频涵盖了各种场景和对象，并在不同的层次上具有非常复杂的运动和不同程度的遮挡。

(2)CDVL 数据集。CDVL 是发布在因特网上的一个数字视频库。研究人员可以根据网站的剪辑描述符(如分辨率、扫描格式、帧速率、色度采样结构等)从数据库中选择数据集来进行训练。

(3)Vimeo-90K 数据集[8]。随着神经网络的兴起，VSR 重构技术取得了突破性的进展。然而，深度神经网络的学习不满足于以前数量小且内容有限的训练集，因此 Xue 等[8]从 Vimeo(一个高清视频播客网站)提取了一部分视频，并制作了 Vimeo-90K 数据集。该数据集包含 64612 个视频，其中每个视频包含 7 帧且内容各不相同。

(4)MM522 数据集[37]。MM522 数据集由 Wang 等提出。该数据集从 542 个高清纪录片中提取各种场景，包含了森林、雪地、沙漠、城市等，其中 20 个视频用作验证集。与上述 Vimeo-90K 数据集中每个视频只包含 7 个视频帧不同，MM522 的每个视频序列有 32 个视频帧。此外，由于递归网络学习的是长时间视频序列信息，因此该数据集更符合递归网络的训练。

(5)REDS 数据集[63]。REDS 在 NTIRE19 比赛中提出了包含 300 个视频(每个视频有 100 个视频帧)，其中 240 个用于训练集，30 个视频用于验证集，30 个视频用于测试集。相比其他数据集，REDS 中的视频包含更多的复杂运动。

需要注意的是，这些数据集会选取一小部分用作测试集并成为了该领域的公共基准，如 Vimeo-90K 和 REDS。此外，Vimeo-90K，REDS 和 MM522 是近几年最常用的训练集。其中，MM522 常用于递归网络的训练，并取得了较好的结果。

4.1.2 测试集

(1)Vid4 数据集[64]。Vid4 由 Liu 等提出，该数据集包含 4 个视频(每个视频包含 30～50 个视频帧)，分别为城市(city)、日历(calendar)、桥(foliage)和行走(walk)。

(2)SPMC 数据集[34]。SPMC 由 Tao 等[34]在 ICCV2017(International Conference on Computer Vision)上提出。该数据集一共包含了 32 个视频(每个视频包含 31 个视频帧)，并且有远景、近景和不同视频背景。

(3)UDM10 数据集[50]。UMD10 数据集是由 Yi 等在 ICCV2019 上提出的。该数据集包含 10 个视频类别：拱门人、拱墙、听众席、乐队、咖啡、相机、鼓掌、湖、摄影和多流。其中，每个视频包含 31 个视频帧。

需要注意的是，还有两类数据集是从 Vimeo-90K 和 REDS 数据集中各提取一部分用作测试集(见 4.1.1 节)。

4.2 性能评估

用于评价 VSR 重构结果的指标主要有两个：峰值信噪比

PSNR（Peak Signal-to-Noise Ratio）和结构相似性 SSIM（Structure Similarity）。PSNR 需要先计算均值方差 MSE，具体定义如下：

$$MSE=\frac{1}{N}\sum_{i=1}^{N}(x_i-\hat{x}_i)^2 \tag{10}$$

$$PSNR=10\log_{10}\left(\frac{x_{\max}^2}{MSE}\right) \tag{11}$$

其中，x 为真实的目标图像，$\hat{x}$ 为预测出的结果，$x_{\max}$ 为目标图像中颜色的最大像素值，如 RGB 中的最大像素值为 255。需要注意的是，MSE 损失函数的平均性质很容易导致得到的结果因缺乏空间细节而过于平滑。针对这个问题，本文提出了感知损失函数，但是限于感知损失的复杂性，VSR 领域仍采用 PSNR 作为主要评价指标。

SSIM 对两张图片进行了 3 个方面的比较，即亮度、对比度和结构，其定义如下：

$$SSIM(x,\hat{x})=\frac{(2\mu_x,\mu_{\hat{x}}+c_1)(\rho_{x\hat{x}}+c_2)}{(\mu_x{}^2+\mu_{\hat{x}}{}^2+c_1)(\rho_x{}^2+\rho_{\hat{x}}{}^2+c_2)} \tag{12}$$

其中，μ_x 和 $\mu_{\hat{x}}$ 分别为 x 和 $\hat{x}$ 的均值；ρ_x 和 $\rho_{\hat{x}}$ 为 x 和 $\hat{x}$ 的方差；$\rho_{x\hat{x}}$ 为 x 和 $\hat{x}$ 的协方差；c_1 和 c_2 为常数。

表 3 列出了代表性的 VSR 算法在 PSNR 和 SSIM 指标上的性能。由于在近几年的超分辨率领域中，大多数算法都只考虑缩放因子为 4 的情况，因此本文也重点考虑这种情况。此外，不同的训练集和退化方式会引起算法在公开测试集上产生不同性能，表 3 详细列出了其相关信息。

表 3 各种方法在缩放因子为 4 的数据集上的比较结果

Table 3 Comparison results of all methods on the dataset with a scaling factor of 4

方法	训练集	测试集	双三次退化		高斯核退化	
			PSNR	SSIM	PSNR	SSIM
MultiBoot[38]	REDS	REDS-Test	31	0.8822	—	—
SOFVSR[35]	CDVL	DAVIS-10	34.32	0.925	—	—
		Vid4	26.01	0.771	—	—
MEMC-Net[22]	Vimeo-90K	Vimeo-90K-T	33.47	0.947	—	—
		Vid4	24.37	0.838	—	—
FRVSR[53]	vimeo.com	Vid4	—	—	26.69	0.822
RBPN[9]	Vimeo-90K	Vid4	27.16	0.819	—	—
		SPMCS	30.1	0.874	—	—
DRVSR[34]	*	Vid4	25.52	0.76	—	—
		SPMCS	29.69	0.84	—	—
FFCVSR[54]	Venice+Myanmar	Vid4	26.97	0.83	—	—
VSRResNet[49]	Myanmar	Vid4	25.51	0.753	—	—
EDVR[46]	Vimeo-90K	Vid4	27.35	0.8264	—	—
		Vimeo-90K-T	37.61	0.9489	—	—
	REDS	REDS4	31.09	0.88	28.88	0.8361
TDAN[24]	Vimeo-90K	Vid4	26.24	0.78	26.58	0.801
DNLN[47]	Vimeo-90K	Vid4	—	—	27.31	0.8257
		SPMCS	—	—	30.36	0.8794
BRCN[27]	V-trian	Vid4	—	—	24.43	0.6334
		V-test	—	—	28.2	0.7739
DUF[23]	Internet	Vid4	—	—	26.81	0.8145
PFNL[50]	*	Vid4	—	—	27.4	0.8384
OVSR[27]	MM522	Vid4	—	—	28.41	0.8724
		UDM10			40.14	0.9713
IConVSR[28]	REDS	REDS	31.67	0.8948	—	—
	Vimeo-90K	Vid4	27.39	0.8251	27.39	0.8279
		Vimeo-90K-T	37.47	0.9476	37.47	0.9476
		REDS	31.67	0.8948	—	—
		UDM10	—	—	40.03	0.9694
MuCAN[52]	REDS	REDS	30.88/0.8750		—	—
MSFFN[48]	Vimeo-90K	Vid4	27.23	0.8218	—	—
		SPMC-11	30.13	0.8769		
		Vimeo-90K-T	37.33	0.9467		
RSDN[56]	Vimeo-90K	Vid4	—	—	27.79	0.8474
		Vimeo-90K-T			37.23	0.9471
		UDM10			39.35	0.9653
DSMC[43]	REDS	REDS	25.73	0.8428	—	—
		Vid4	24.63	0.8403		
DDAN[10]	MM522	Vid4	26.48	0.7892	—	—
		Myanmar	34.46	0.9144		
		YUV21	29.18	0.799		
OFRnet[35]	CDVL	Vid4	27.34	0.8327	—	—
	Vimeo-90K	Vid4	26.86	0.814	—	—

注：‘*’表示未知数据的来源，‘—’表示未知数据集上测试该方法

5 挑战

尽管基于深度学习的 VSR 算法取得了不错的进展，但目前仍存在亟待解决的潜在问题。本文总结了以下 7 种挑战，未来可以从这 7 种角度出发进行研究。

5.1 更加轻量级的视频超分模型

虽然基于深度学习的 VSR 算法已经取得不错的效果，但是其模型含有很大的参数，需要大量的计算和存储资源，并且训练时间很长。随着移动设备的流行，人们更希望算法可以部署在移动设备中。因此，如何设计一个高性能、轻量级、更实用的超分辨率算法成为一个挑战。

5.2 更加可解释的模型

深度神经网络通常被看作一个黑匣子，无法像数学公式一样得到非常合理的解释。当模型的性能好或者不好时，就无法得知模型学到的真实信息。例如，在 VSR 领域中无法知道一个神经网络是如何将 LR 视频序列重建为 HR 视频帧的，也没有一个理论来支持。可解释性不仅是 VSR 的挑战，也是整个深度学习领域的挑战，可解释性的深入研究，将会推动基于深度学习的相关领域（如 VSR）得到进一步发展。

5.3 更大缩放因子的视频超分

现有的超分辨率方法的研究中大多关注缩放因子为 4 的情况，很少探究更具挑战性的尺度缩放因子。随着高清视频和超高清视频的普及，更大尺度缩放因子的研究得到了更多的关注。显然，随着缩放因子越来越大，预测和恢复一段视频中的未知信息将会变得越来越具有挑战性，并且会导致算法性能退化并削弱模型中的鲁棒性。因此，为了得到更大尺度的超分结果，如何设计一个适用于大尺度的 VSR 算法是一个很重要的问题。

5.4 更加合理和正确的退化方式

现有的算法通常使用双三次插值和高斯核模糊对 HR 视频进行下采样，得到 LR 视频序列，并将其输入到网络中。虽然这两种方法在理论上表现良好，但是由于在现实生活中的退化过程是多种多样的，如散焦、运动模糊、传感器噪声和压缩噪声等，因此在构建 LR 视频时应根据实际情况进行理论建模，以缩小研究与实践的差距。

5.5 无监督视频超分方法

由于大多数的 VSR 方法都是监督学习，因此需要大量配对的 LR 视频帧和 HR 视频帧来训练网络。但是配对的视频帧在现实生活中很少，虽然可以人工合成与 HR 视频帧配对的 LR 视频帧，但无法很好地模拟现实生活中的情况，限制了超分辨率的性能。鉴于此，无监督的 VSR 方法是很有价值的研究方向。

5.6 更加有效地利用帧间时间信息

如何利用视频帧间的时间信息对于 VSR 来说非常重要，且是否能有效利用也会直接影响性能。虽然在这个领域中已经提出了许多关于如何利用视频帧的方法，但是仍然存在许多缺点，如迭代网络只能利用滑动窗口内的视频帧。因此，如何有效地利用帧间信息亟待学者的进一步研究。

5.7 更加合理的视频质量评价标准

VSR 的评价指标主要包括 PSNR 和 SSIM。但它们的平均性会导致产生的结果过于平滑，从而在视觉上不符合人类视觉感知，因此，设计一个更为合理的 VSR 评价指标是非常重要的。

6 应用前景

在许多领域中，VSR 算法需要面临不同类型的视频退化问题，如光学退化和摄像机传感器有限的尺寸、劣质的光纤和天气条件、检测器和摄像机的运动等。SR 算法可以恢复由于各种情况导致退化的原始视频，因此在许多领域都有广泛的应用前景。本文将 VSR 应用前景分为 4 类并进行简要介绍。

6.1 视频成像

近些年的显示器能够渲染高动态范围图像（High-Dynamic Range，HDR）以及高达 8k 超高清（Ultra High Definition，UHD）的 HR 视频。高动态范围图像超高清成像已经发展到了很高的水平。然而，做出原始高动态范围超高清视频的价格非常昂贵，而传统 LR 标准动态范围（Standard Dynamic Range，SDR）视频在生活中仍然被大量使用。因此，当下迫切需要利用适当的转换技术将传统 LR 标准动态范围视频变换到高动态范围超高清视频中。

最直接的转换方式是将传统的 LR 视频通过视频超分辨技术得到 HR 视频。目前，文献[65-69]都对此进行了研究。具体地，文献[65]提出了基于生成性对抗网络的架构，并采用融合超分辨率技术和逆色调映射来提升图像质量。文献[68]首次提出了一个可以实时进行视频转化的卷积神经网络，在他们专用的基于卷积神经网络的超分硬件中，将深度可分离卷积与残差相结合对 LR 输入帧进行逐帧处理。

6.2 视频监控

在视频监控中，由于天气、环境、成像设备等因素的影响，使获得的图像分辨率低、模糊，因此，研究人员将基于深度学习的 VSR 方法应用到了视频监控[70-73]中。

在识别监控中的人体时，通常将人捕获为 LR 图像，有时不仅需要从人脸上检测和识别，还需要从人体外观上检测和识别。因此，文献[72]提出了基于样本的超分辨率方法，将 VSR 应用于 LR 人体识别。此外，车牌识别是视频监控中使用最广泛的应用之一，文献[70-71]提出了基于 VSR 的车牌识别方法，文献[73]提出了基于对抗生成网络的超分模块，并使用实时 YOLO（You Only Look Once）检测方法来检测车牌，并获得了不错的效果。

6.3 医学

医学视频在医学领域扮演着越来越重要的角色，是医学诊断的重要形式。然而，由于宽带和硬件设备的限制，一些医学视频的分辨率较低，而具有超分辨率的医学视频可以帮助医生更清楚地了解情况，因此，对其的研究具有重大的现实意义。文献[74]提出了基于非对称反投影网络和结合光流算法以及多帧融合策略的先进医学 VSR 方法。此后，与上述方法关注于视频帧的所有区域不同，文献[75]提出了新的医学视频压缩系统，该系统使用 VSR 算法作为预处理步骤，并使用增强层来选择性地提高感兴趣区域的质量。

6.4 其他应用

基于深度学习的 VSR 算法也应用到了其他视觉任务上。

Xiao 等[76]结合图像超分辨率和 VSR 算法 EDVR，用于卫星遥感视频。Garcia 等[77]提出了一个混合分辨率点云表示和一个超分辨率框架，从中衍生出了一些处理工具，如压缩、噪音和错误隐藏。此外，随着移动设备的普及，VSR 在现代智能手机和相机中也得到了实际应用。为了解决视频中运动模糊而导致配准错误的问题，Matsushita 等[78]提出了从视频序列中恢复 HR 视频帧的方法，并成功将其应用于摄像机中。

结束语 本文对基于深度学习的 VSR 进行了详尽的回顾。首先，从迭代网络和递归网络两种网络框架详细阐述了视频超分算法的研究进展。然后，为了明晰基于深度学习的 VSR 算法中尚未解决的潜在问题，本文从轻量级、可解释性、缩放因子等 7 个方面总结了其面临的挑战。最后，详细阐述了 VSR 在视频成像、视频监控、医学和其他应用这 4 个方面的应用前景。

参 考 文 献

[1] HARRIS J L. Diffraction and resolving power[J]. JOSA，1964，54(7)：931-936.

[2] CAPEL D，ZISSERMAN A. Super-resolution enhancement of text image sequences[C]// Proceedings 15th International Conference on Pattern Recognition. 2000：600-605.

[3] SCHULTZ R R，STEVENSON R L. Extraction of high-resolution frames from video sequences[J]. IEEE Transactions on Image Processing，1996，5(6)：996-1011.

[4] BORMAN S，STEVENSON R L. Simultaneous multi-frame MAP super-resolution video enhancement using spatio-temporal priors[C]// Proceedings 1999 International Conference on Image Processing (Cat. 99CH36348). 1999：469-473.

[5] GUNTURK B K，ALTUNBASAK Y，MERSEREAU R. Bayesian resolution-enhancement framework for transform-coded video[C]// Proceedings 2001 International Conference on Image Processing (Cat. No. 01CH37205). 2001：41-44.

[6] PATTI A J，SEZAN M I，TEKALP A M. Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time[J]. IEEE Transactions on Image Processing，1997，6(8)：1064-1076.

[7] KAPPELER A，YOO S，DAI Q，et al. Video super-resolution with convolutional neural networks[J]. IEEE Transactions on Computational Imaging，2016，2(2)：109-122.

[8] XUE T，CHEN B，WU J，et al. Video enhancement with task-oriented flow[J]. International Journal of Computer Vision，2019，127(8)：1106-1125.

[9] HARIS M，SHAKHNAROVICH G，UKITA N. Recurrent back-projection network for video super-resolution[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019：3897-3906.

[10] LI F，BAI H，ZHAO Y. Learning a deep dual attention network for video super-resolution[J]. IEEE Transactions on Image Processing，2020，29：4474-4488.

[11] WANG Z，CHEN J，HOI S C H. Deep learning for image super-resolution：A survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2020：3365-3387.

[12] SINGH A，SINGH J. Survey on single image based super-resolution—implementation challenges and solutions[J]. Multimedia Tools and Applications，2020，79(3)：1641-1672.

[13] YANG W，ZHANG X，TIAN Y，et al. Deep learning for single image super-resolution：A brief review[J]. IEEE Transactions on Multimedia，2019，21(12)：3106-3121.

[14] DAITHANKAR M V，RUIKAR S D. Video Super Resolution：A Review[C]// ICDSMLA. 2020：488-495.

[15] WU Y，FAN G H. Survey of Super - Resolution Reconstruction Techniques for Video Sequences[J]. Computer Engineering & Software，2017，38(4)：154-160.

[16] LIU H，RUAN Z，ZHAO P，et al. Video super resolution based on deep learning：A comprehensive survey[J]. arXiv：2007.12928，2020.

[17] DONG C，LOY C C，HE K，et al. Image super-resolution using deep convolutional networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2015，38(2)：295-307.

[18] DONG C，LOY C C，TANG X. Accelerating the super-resolution convolutional neural network[C]// European Conference on Computer Vision. 2016：391-407.

[19] KIM J，LEE J K，LEE K M. Accurate image super-resolution using very deep convolutional networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016：1646-1654.

[20] SHI W，CABALLERO J，HUSZAR F，et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016：1874-1883.

[21] ZHANG Y，LI K，LI K，et al. Image super-resolution using very deep residual channel attention networks[C]// Proceedings of the European Conference on Computer Vision (ECCV). 2018：286-301.

[22] BAO W，LAI W S，ZHANG X，et al. Memc-net：Motion estimation and motion compensation driven neural network for video interpolation and enhancement[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2019：933-948.

[23] JO Y，OH S W，KANG J，et al. Deep video super-resolution network using dynamic upsampling filters without explicit motion compensation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018：3224-3232.

[24] TIAN Y，ZHANG Y，FU Y，et al. Tdan：Temporally-deformable alignment network for video super-resolution[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020：3360-3369.

[25] CABALLERO J，LEDIG C，AITKEN A，et al. Real-time video super-resolution with spatio-temporal networks and motion compensation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017：4778-4787.

[26] GUO J，CHAO H. Building an end-to-end spatial-temporal convolutional network for video super-resolution[C]// Thirty-First AAAI Conference on Artificial Intelligence. 2017：4053-4060.

[27] YI P，WANG Z，JIANG K，et al. Omniscient Video Super-Resolution[J]. arXiv：2103.15683，2021.

[28] CHAN K C K, WANG X, YU K, et al. BasicVSR: The search for essential components in video super-resolution and beyond [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 4947-4956.

[29] LUCAS B. An Iterative Image Registration Technique with an Application to Stereo Vision (DARPA)[J]. Proc. IJCAI, 1981, 81(3): 674-679.

[30] DRULEA M, NEDEVSCHI S. Total variation regularization of local-global optical flow[C]// 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC). 2011: 318-323.

[31] DOSOVITSKIY A, FISCHER P, ILG E, et al. Flownet: Learning optical flow with convolutional networks[C]// Proceedings of the IEEE International Conference on Computer Vision. 2015: 2758-2766.

[32] ILG E, MAYER N, SAIKIA T, et al. Flownet 2.0: Evolution of optical flow estimation with deep networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 2462-2470.

[33] RANJAN A, BLACK M J. Optical flow estimation using a spatial pyramid network[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 4161-4170.

[34] TAO X, GAO H, LIAO R, et al. Detail-revealing deep video super-resolution[C]// Proceedings of the IEEE International Conference on Computer Vision. 2017: 4472-4480.

[35] WANG L, GUO Y, LIN Z, et al. Learning for video super-resolution through HR optical flow estimation[C]// Asian Conference on Computer Vision. 2018: 514-529.

[36] BARE B, YAN B, MA C, et al. Real-time video super-resolution via motion convolution kernel estimation[J]. Neurocomputing, 2019, 367: 236-245.

[37] WANG Z, YI P, JIANG K, et al. Multi-memory convolutional neural network for video super-resolution[J]. IEEE Transactions on Image Processing, 2018, 28(5): 2530-2544.

[38] KALAROT R, PORIKLI F. Multiboot vsr: Multi-stage multi-reference bootstrapping for video super-resolution[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2019: 2060-2069.

[39] SHI X J, CHEN Z, WANG H, et al. Convolutional LSTM network: A machine learning approach for precipitation nowcasting [C]// Advances in Neural Information Processing Systems. 2015: 802-810.

[40] IRANI M, PELEG S. Improving resolution by image registration [J]. CVGIP: Graphical Models and Image Processing, 1991, 53(3): 231-239.

[41] LIU D, WANG Z, FAN Y, et al. Robust video super-resolution with learned temporal dynamics[C]// Proceedings of the IEEE International Conference on Computer Vision. 2017: 2507-2515.

[42] KIM T H, SAJJADI M S M, HIRSCH M, et al. Spatio-temporal transformer network for video restoration[C]// Proceedings of the European Conference on Computer Vision (ECCV). 2018: 106-122.

[43] LIU H, ZHAO P, RUAN Z, et al. Large motion video super-resolution with dual subnet and multi-stage communicated upsampling[J]. arXiv: 2103.11744, 2021.

[44] DAI J, QI H, XIONG Y, et al. Deformable convolutional networks[C]// Proceedings of the IEEE International Conference on Computer Vision. 2017: 764-773.

[45] ZHU X, HU H, LIN S, et al. Deformable convnets v2: More deformable, better results[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 9308-9316.

[46] WANG X, CHAN K C K, YU K, et al. Edvr: Video restoration with enhanced deformable convolutional networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2019: 1954-1963.

[47] WANG H, SU D, LIU C, et al. Deformable non-local network for video super-resolution[J]. IEEE Access, 2019, 7: 177734-177744.

[48] SONG H, XU W, LIU D, et al. Multi-Stage Feature Fusion Network for Video Super-Resolution[J]. IEEE Transactions on Image Processing, 2021, 30: 2923-2934.

[49] LUCAS A, LOPEZ-TAPIA S, MOLINA R, et al. Generative adversarial networks and perceptual losses for video super-resolution[J]. IEEE Transactions on Image Processing, 2019, 28(7): 3312-3327.

[50] YI P, WANG Z, JIANG K, et al. Progressive fusion video super-resolution network via exploiting non-local spatio-temporal correlations[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 3106-3115.

[51] ISOBE T, LI S, JIA X, et al. Video super-resolution with temporal group attention[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 8008-8017.

[52] LI W, TAO X, GUO T, et al. Mucan: Multi-correspondence aggregation network for video super-resolution[C]// European Conference on Computer Vision. 2020: 335-351.

[53] SAJJADI M S M, VEMULAPALLI R, BROWN M. Frame-recurrent video super-resolution[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 6626-6634.

[54] YIN B, LIN C, TAN W. Frame and feature-context video super-resolution[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2019, 33(1): 5597-5604.

[55] ZHU X, LI Z, ZHANG X Y, et al. Residual invertible spatio-temporal network for video super-resolution[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2019: 5981-5988.

[56] JACOBSEN J H, SMEULDERS A, OYALLON E. i-revnet: Deep invertible networks[J]. arXiv: 1802.07088, 2018.

[57] ISOBE T, JIA X, GU S, et al. Video super-resolution with recurrent structure-detail network[C]// European Conference on Computer Vision. 2020: 645-660.

[58] HUANG Y, WANG W, WANG L. Bidirectional recurrent convolutional networks for multi-frame super-resolution[J]. Advances in Neural Information Processing Systems, 2015, 28: 235-243.

[59] GUO J, CHAO H. Building an end-to-end spatial-temporal con-

volutional network for video super-resolution[C]// Thirty-First AAAI Conference on Artificial Intelligence. 2017:4053-4060.

[60] LI D,LIU Y,WANG Z. Video super-resolution using non-simultaneous fully recurrent convolutional network[J]. IEEE Transactions on Image Processing,2018,28(3):1342-1355.

[61] FUOLI D,GU S,TIMOFTE R. Efficient video super-resolution through recurrent latent space propagation[C]// 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). IEEE,2019:3476-3485.

[62] LIAO R,TAO X,LI R,et al. Video super-resolution via deep draft-ensemble learning[C]// Proceedings of the IEEE International Conference on Computer Vision. 2015:531-539.

[63] NAH S,BAIK S,HONG S,et al. Ntire 2019 challenge on video deblurring and super-resolution: Dataset and study[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2019:1996-2005.

[64] LIU C,SUN D. On Bayesian Adaptive Video Super Resolution[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2014,36(2):346-360.

[65] ZENG H,ZHANG X,YU Z,et al. SR-ITM-GAN:Learning 4K UHD HDR With a Generative Adversarial Network[J]. IEEE Access,2020,8:182815-182827.

[66] HE Z,HUANG H,JIANG M,et al. FPGA-based real-time super-resolution system for ultra high definition videos[C]// 2018 IEEE 26th Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM). IEEE,2018:181-188.

[67] LIU Z,CUI C. A New Low Bit-Rate Coding Scheme for Ultra High Definition Video Based on Super-Resolution Reconstruction[C]// 2018 IEEE International Conference on Computer and Communication Engineering Technology (CCET). IEEE,2018:325-329.

[68] KIM Y,CHOI J S,KIM M. A real-time convolutional neural network for super-resolution on FPGA with applications to 4K UHD 60 fps video services[J]. IEEE Transactions on Circuits and Systems for Video Technology,2018,29(8):2521-2534.

[69] YANG Y,BI P,LIU Y. License plate image super-resolution based on convolutional neural network[C]// 2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC). IEEE,2018:723-727.

[70] GHONEIM M,REHAN M,OTHMAN H. Using super resolution to enhance license plates recognition accuracy[C]// 2017 12th International Conference on Computer Engineering and Systems (ICCES). 2017:515-518.

[71] MEHREGAN K,AHMADYFARD A,KHOSRAVI H. Super-resolution of license-plates using frames of low-resolution video[C]// 2019 5th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS). 2019:1-6.

[72] NISHIBORI K,TAKAHASHI T,DEGUCHI D,et al. Exemplar-based human body super-resolution for surveillance camera systems[C]// 2014 International Conference on Computer Vision Theory and Applications (VISAPP). IEEE,2014:115-121.

[73] LEE Y,YUN J W,HONG Y,et al. Accurate license plate recognition and super-resolution using a generative adversarial networks on traffic surveillance video[C]// 2018 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia). IEEE,2018:1-4.

[74] REN S,LI J,GUO K,et al. Medical video super-resolution based on asymmetric back-projection network with multilevel error feedback[J]. IEEE Access,2021,9:17909-17920.

[75] BONANNO D,DEBONO C J. A Medical Video Coding Scheme with Preserved Diagnostic Quality[C]// 2019 IEEE Global Communications Conference (GLOBECOM). IEEE,2019:1-6.

[76] XIAO A,WANG Z,WANG L,et al. Super-resolution for "Jilin-1" satellite video imagery via a convolutional network[J]. Sensors,2018,18(4):1194.

[77] GARCIA D C,FONSECA T A,QUEIROZ R L D. Example-based super-resolution for point-cloud video[C]// 2018 25th IEEE International Conference on Image Processing (ICIP). IEEE,2018:2959-2963.

[78] MATSUSHITA Y,KAWASAKI H,ONO S,et al. Simultaneous deblur and super-resolution technique for video sequence captured by hand-held video camera[C]// 2014 IEEE International Conference on Image Processing (ICIP). IEEE, 2014: 4562-4566.

LENG Jia-xu, born in 1989, Ph.D. His main research interests include object detection, face super-resolution, person re-identification and video anomaly detection.

GAO Xin-bo, born in 1972, Ph.D, professor, Ph.D supervisor. His main research interests include artificial intelligence, machine learning, computer vision and pattern recognition.

(责任编辑:李亚辉)