

双目图像超分辨研究综述

王应谦, 王龙光, 梁政宇, 安玮, 杨俊刚*

国防科技大学电子科学学院, 湖南 长沙 410073

摘要 随着双目成像技术的发展, 双目图像超分辨在近年来得到了广泛关注。不同于单幅图像超分辨, 双目图像超分辨可以利用左右视角图像间的互补信息进一步提升图像重建质量。对双目图像超分辨领域的最新进展进行了综述。首先, 介绍了双目成像的基本理论。接着, 对现有的双目图像超分辨算法和双目图像数据集进行归纳总结。随后, 在基准数据集上对几种基于深度学习的主流算法的性能进行了评测, 并探究了不同训练集对超分辨算法性能的影响。最后, 总结了双目图像超分辨所面临的挑战, 并展望了其未来的研究方向。

关键词 成像系统; 图像超分辨; 双目视觉; 性能评测; 数据集

中图分类号 TP391

文献标志码 A

doi: 10.3788/LOP202158.1811014

Stereo Image Super-Resolution: A Survey

Wang Yingqian, Wang Longguang, Liang Zhengyu, An Wei, Yang Jungang*

College of Electronic Science and Technology, National University of Defense Technology, Changsha, Hunan 410073, China

Abstract With the development of stereo imaging technology, stereo image super-resolution has attracted increasing attention in recent years. Different from single image super-resolution, stereo image super-resolution can further improve the quality of image reconstruction by using the complementary information between left and right view images. In this paper, we present a comprehensive survey on recent advances of stereo image super-resolution field. Firstly, we introduce the fundamental theory of stereo imaging. Then, we review the existing stereo image super-resolution algorithms and stereo image datasets. Afterwards, we comprehensively evaluate several mainstream deep learning-based stereo image super-resolution algorithms on benchmark datasets, and investigate the influences of different training sets on the performances of super-resolution algorithms. Finally, we summarize challenges and prospect future research directions in stereo image super-resolution.

Key words imaging systems; image super-resolution; stereo vision; performance evaluation; datasets

OCIS codes 100.6640; 330.1400; 150.4232; 110.1758

1 引言

图像超分辨率重建可以通过一张或多张低分辨率图像, 计算得到相应的高分辨率图像, 并恢复输入的低分辨率图像中所缺失的纹理细节。图像超分辨可以在不提升传感器硬件水平的情况下, 通过计算成像的方式来提高图像的分辨率与视觉效果, 其具

有图像获取成本低的优点, 因此被广泛应用于遥感观测^[1-3]、医学诊断^[4-8]、视频监控^[9-11]等领域。近年来, 随着人工智能与深度学习的迅猛发展, 深度卷积神经网络被广泛应用于图像超分辨任务, 该网络大幅提升了图像的重建精度与视觉效果。

随着双目成像技术的发展, 双目相机在近年来被成功应用于虚拟现实^[12]、增强现实^[13]、自动

收稿日期: 2021-06-01; 修回日期: 2021-06-21; 录用日期: 2021-06-26

基金项目: 国家自然科学基金创新研究群体(61921001)、湖湘青年英才(2020RC3026)

通信作者: *yangjungang@nudt.edu.cn

驾驶^[14]等领域。图 1 为商用双目相机示例。高分辨率的双目图像包含更多的场景细节,具有更好的视觉效果,且对后续任务(如深度估计、三维物

体检测)的性能有显著增益,因此许多学者尝试利用图像超分辨算法来提升双目图像的分辨率^[15-16]。



图 1 商用双目相机示例

Fig. 1 Examples of commercial stereo cameras

实现双目图像超分辨的直接方式是将单图超分辨算法分别应用于左右视角图像,但这会忽略双目图像之间的关联。实际上,双目相机利用左右两个视角来记录当前场景,视角间的互补信息使得左右视角图像在超分辨的过程中互为参考,这为图像超分辨提供了额外的先验约束。因此,合理利用左右视角图像间的互补信息对提升双目图像超分辨的性能至关重要。

目前,图像超分辨领域正处于快速发展时期。基于现有的研究成果,领域内许多学者从不同角度对图像超分辨进行了总结^[16-18]。2019年, Yang 等^[17]从网络结构和目标函数出发,回顾了基于深度学习的单图超分辨算法的发展历程并分析了该算法的优劣势。2020年, Wang 等^[18]阐述了基于深度学习的单图超分辨算法在有监督、无监督领域中的应用。同年,唐艳秋等^[19]从模型类型、网络结构、信息传递技术等方面对图像超分辨算法进行了详细评述。2021年,曲延云等^[20]分析了当前图像超分辨所面临的挑战,并从有监督和不成对两个类别出发研究了基于深度学习的单图超分辨算法。

上述综述论文更多聚焦于单图超分辨重建,缺少对双目图像超分辨的针对性总结,故本文对近年来双目图像超分辨率重建方面的研究进行了综述。首先,介绍了双目成像的基本原理。接着,按照建立左右视角图像间关联的方式对现有双目图像超分辨算法进行分类回顾,并对双目图像超分辨率领域内主要的双目图像数据集进行总结。随后,在基准数据集上对主流算法进行了评测,并分析了不同训练集对基于深度的双目图像超分辨算法性能的影响。最后,指出了当前双目图像超分辨所面临的挑战,并展望了其未来的研究方向。

2 双目成像基本原理

双目相机可以从左右两个视角记录当前场景,其左右视角图像所包含的几何关系可以解析场景的三维结构。图 2(a)为双目相机成像模型示意图,两个光轴平行的相机分别从左右两个视角对当前场景进行拍摄,所记录的图像如图 2(b)所示。双目相机的左右视角存在偏差,故场景中的同一物体在左视角图和右视角图中的位置具有一定差异,距离相机较近的物体视差较大,距离相机较远的物体视差较小。

在双目成像中,通常采用焦距相同、指向相同(即光轴平行)且两相机连线(基线)与光轴垂直的双目相机。在上述条件下,同一物体在左右视角图像中仅存在水平方向的视差。该视差与场景深度(物体与相机的距离)的关系可以表示为

$$\gamma = \frac{Bf}{d}, \quad (1)$$

式中: γ 表示场景深度; B 表示左右两个相机之间的基线长度; f 表示相机的焦距; d 表示该物体在左视角图和右视角图中的视差。图 2(c)展示了 KITTI 2015 数据集中真实的双目图像,可以发现,场景深度不同区域具有不同的视差,图中已用圆圈圈出对应区域。

为了利用左右视角图像间的互补信息提升双目图像超分辨的重建质量,需要建立左右视角图像对应区域间的关联。在不考虑遮挡的情况下,左右视角图像的对应区域应当处于同一高度的水平线上,该水平线称为极线。双目图像超分辨算法应当结合极线约束关系建立左右视角图像间的关联。双目图像的极线约束关系如图 3 所示,即对于左视角图中

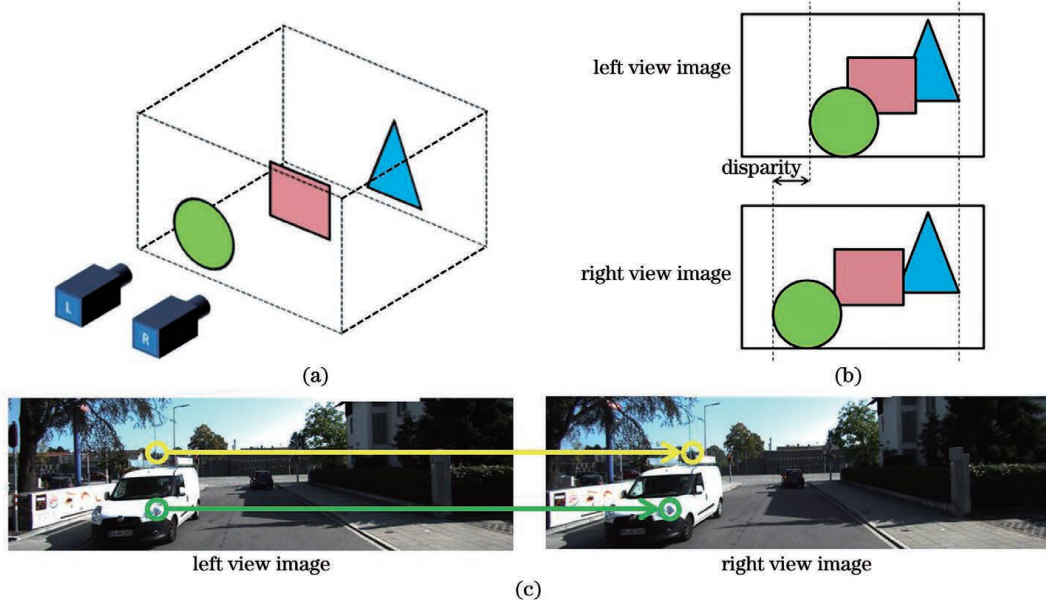


图 2 双目相机与双目图像。(a)双目相机成像模型示意图;(b)图 2(a)中的双目相机所记录的图像;(c)取自 KITTI 2015 数据集的双目相机所记录的真实场景图像^[21]

Fig. 2 Stereo camera and stereo images. (a) Schematic diagram of imaging model of stereo camera; (b) images recorded by stereo camera in Fig. 2 (a); (c) real scene images recorded by stereo camera taken from KITTI 2015 dataset^[21]

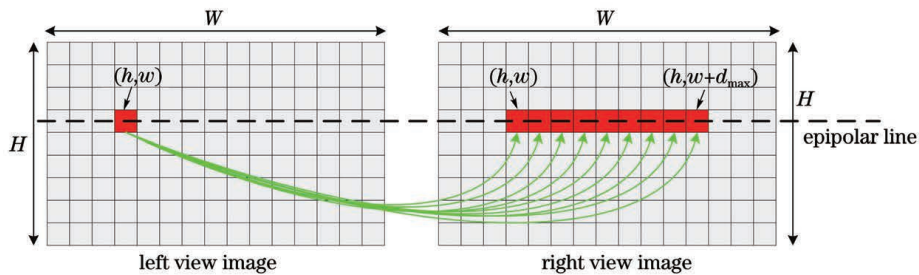


图 3 双目图像的极线约束关系

Fig. 3 Epipolar constraint relation of stereo images

的任意一点 (h, w) , 其在右视角图中的对应点应位于 (h, w) 和 $(h, w + d_{\max})$ 之间。其中, W 和 H 代表左右视角图像的宽度和高度, d_{\max} 为左右视角图像间视差的最大值。

3 算法回顾与总结

本节总结了现有主流的双目图像超分辨算法。按照建立左右视角图像间关联方式的不同, 将现有双目图像超分辨算法分为基于平移副本堆叠的算法、基于视差估计的算法以及基于视差注意力的算法。

3.1 基于平移副本堆叠的算法

2018 年, Jeon 等^[22] 提出了双目图像超分辨算法 StereoSR, 该算法采用平移副本堆叠的方式来实现不同视差下左右视角图像间互补信息的融合。首先, StereoSR 将右视角图依次向右平移 1 ~

64 pixel, 生成 64 幅副本图像。随后, 将 64 幅副本图像与左视角图在通道维度上级联并送入网络进行重建。在不考虑遮挡的情况下, 当左右视角图像间的最大视差小于 64 pixel 时, 对于左视角图中的任一区域, 一定存在某个右视角图平移之后的副本图像在对应区域与之粗略对齐, 从而该算法可以通过卷积操作来实现左右视角图像间互补信息的融合。

StereoSR 算法是双目图像超分辨领域首个基于深度学习的算法。该算法利用了左右视角图像间的极线约束关系, 使得左视角图中大部分区域都能与右视角图平移之后的某个副本对应。然而, 该算法需要人为设置一个最大的视差, 如文献^[22]中设置最大的视差为 64 pixel, 超出该视差的信息则无法利用。此外, 该算法未对遮挡等区域进行特殊处理, 限制了其重建性能。

3.2 基于视差估计的算法

2010 年, Bhavsar 等^[23]提出了一种同时对双目图像和其对应的视差图进行超分辨操作的算法, 指出双目图像超分辨与视差估计是相互耦合的两个问题。一方面, 双目图像超分辨需要建立左右视角图像间像素级的关联, 因此需要较为精确的视差估计以提供配准信息。另一方面, 精确的视差估计又依赖于细节清晰的高分辨率双目图像。基于此, Bhavsar 等将双目图像超分辨与视差估计建模至同一个能量函数中。在求解最小化能量函数的过程中, Bhavsar 等采用迭代优化的方式。首先, 估计初始场景的视差分布。接着, 利用该视差对左右视角图像进行配准, 重建出高分辨率的左右视角图像。随后, 利用重建的左右视角图像继续求解更高精度的视差图, 不断迭代直至收敛。最后便可得到高分辨率的双目图像与视差图的最优估计。

随着深度学习的发展, 研究人员将深度卷积神经网络成功应用于图像超分辨和双目视差估计算法中, 这些算法的性能均取得了突破性提升。此外, 深度学习也被广泛用于实现基于视差估计的双目图像超分辨。2020 年, Yan 等^[24]提出了双目图像重建网络 StereoIRN, 该网络的超分辨、去模糊和去噪性能在当时处于领先水平。StereoIRN 由单目特征提取子网络、视差估计子网络以及双目重建子网络三部分组成。首先, StereoIRN 利用单目特征提取子网络提取低分辨率左右视角图像的特征。接着, 采用双目立体匹配网络 StereoNet^[25]对左右视角图像的特征进行匹配, 估计出对应的视差图。最后, 按照估计的视差图将右视角图的信息与左视角图的信息进行融合, 进而实现图像的超分辨率重建。

2021 年, Dai 等^[26]利用深度卷积神经网络实现了双目图像超分辨和视差估计的联合优化, 指出双目图像超分辨和视差估计是互为促进的两个任务。高分辨率的双目图像可以提升视差估计的精度, 而高精度的视差估计又可以引导左右视角图像间建立更加准确的关联, 从而进一步提升双目图像超分辨的性能。基于此, Dai 等提出了视差估计与双目图像超分辨互相促进的反馈网络 SSRDE-FNet。首先, SSRDE-FNet 对输入的低分辨率双目图像进行超分辨重建, 生成高分辨率双目图像特征。接着, 基于高分辨率双目图像特征构建并聚合匹配代价以实现像素级视差估计。最后, 将估计出来的视差图反馈给超分辨重建网络以引导左右视角图像间的准确关联, 从而进一步提升图像超分辨的性能。SSRDE-

FNet 采用递归结构实现了左右视角图像与视差图的迭代优化, 实现了目前双目图像超分辨领域内最好的重建性能。

基于视差估计的双目图像超分辨算法首先计算左右视角图像中每一个空间位置的视差, 然后在视差图的引导下对左右视角图像进行像素级的配准。相比于平移副本堆叠的方式, 此类算法可以更加高效地利用左右视角图像间的互补信息, 实现更加优越的超分辨重建性能。但实际场景的复杂结构经常使得此类算法在处理遮挡区域以及深度跳变区域时产生较大误差, 进而此类算法无法在这些区域获得可靠的左右视角图像匹配信息, 这会导致超分辨重建结果失真。此外, 此类算法将视差估计作为一个独立的计算过程, 这会增加计算开销。文献[26]进一步利用递归网络来进行迭代优化, 虽然该算法取得了优越的重建性能, 但是难以满足实时性要求较高的场景, 如自动驾驶、机器人导航等。

3.3 基于视差注意力机制的算法

2019 年, Wang 等^[27]将双目视觉中的对极几何约束引入到注意力机制中, 提出用于计算图像中不同视差位置间相似性的视差注意力机制, 并基于计算得到的相似性测度实现了左右视角图像特征的融合。Wang 等^[28]进一步指出, 视差注意力机制能够对双目图像极线方向全局范围内的信息进行关联, 不需要预先设置最大视差, 具有更强的灵活性与鲁棒性。基于视差注意力机制的双目超分辨网络 PASSRnet 具有当时领先的超分辨重建性能。

基于 Wang 等^[28]的工作, 双目图像超分辨领域中涌现出了许多基于视差注意力机制的双目图像超分辨算法。2020 年, Song 等^[29]将视差注意力机制与自注意力机制结合, 在建立左右视角图像间关联的基础上加强对单个视角图像内非局部上下文信息的利用。同时, Song 等指出双目图像的对应区域在未被遮挡时应当具有相同的纹理特征, 基于此, 其设计了双目一致性损失函数以约束超分辨重建后的左右视角图像的一致性。该算法的超分辨性能比 PASSRnet 算法更加优越。同年, Ying 等^[30]指出, 双目图像超分辨网络需要同时利用左右视角图像间的互补信息与单个视角图像内的上下文信息。与单图数据集相比, 现有双目图像数据集存在图像数量少、质量低的缺点, 因此双目图像超分辨算法难以从双目图像数据集中学到丰富的图像结构与纹理表征。但纹理丰富的单图数据集又无法直接应用于双目图像超分辨算法的训练中。为解决此问题, Ying

等提出了基于通用的双目注意力模块(SAM)的双目图像超分辨算法,其中 SAM 可用于交互左右视角图像的信息。该算法将 SAM 安插到预训练好的单图超分辨网络中,将单图超分辨网络扩展为双目图像超分辨网络并在双目图像数据集上进行微调,在结合左右视角图间互补信息的同时,实现了对单个视角图像内上下文信息的充分利用。Ying 等指出,SAM 可以应用于多个单图超分辨网络并且可以使得网络的性能得到提升。此外,SAM 与单图超分辨网络 SRResNet 结合后,可以达到优于 PASSRnet 的超分辨重建精度。

2021 年,Xu 等^[31]指出图像在不同的空间位置通常具有不同的纹理与内容,而现有的双目图像超分辨网络均采用固定的卷积核对空间中不同位置的区域进行相同的运算,因此现有网络不能适应图像的空变特性。因此,Xu 等将双边滤波的思想引入到双目图像超分辨中并提出 BSSRnet 算法。首先,利用视差注意力机制生成空变的动态双边网格。随后,利用双边滤波在高分辨率空间对左右视角图像的特征进行融合与重建。BSSRnet 算法比 StereoSR^[22]

算法、PASSRnet^[28]算法和 SRResNet+SAM^[30]算法的超分辨性能更优越。同年,Wang 等^[32]提出了 iPASSR 算法,并在交互模块、网络结构以及损失函数三个方面对 PASSRnet^[28]算法进行改进。iPASSR 算法采用孪生网络对左右视角图像同时进行超分辨操作,并利用双向视差注意力模块(BiPAM)同时交互左右视角图像的信息。此外,iPASSR 算法采用的是基于残差形式的照度一致性损失、循环一致性损失以及重建一致性损失,这较好地解决了真实场景中左右视角图像照度不一致的问题。iPASSR 算法充分地利用了双目图像之间的对称性先验约束,实现了先进的超分辨重建性能。

4 数据集

在双目视觉领域中,研究人员建立了多个双目图像数据集并将其应用于立体匹配、双目测距、光流估计、三维检测等任务中。本节简要介绍了双目视觉领域内的双目图像数据集,不同双目图像数据集的样本图像^[33]如图 4 所示。

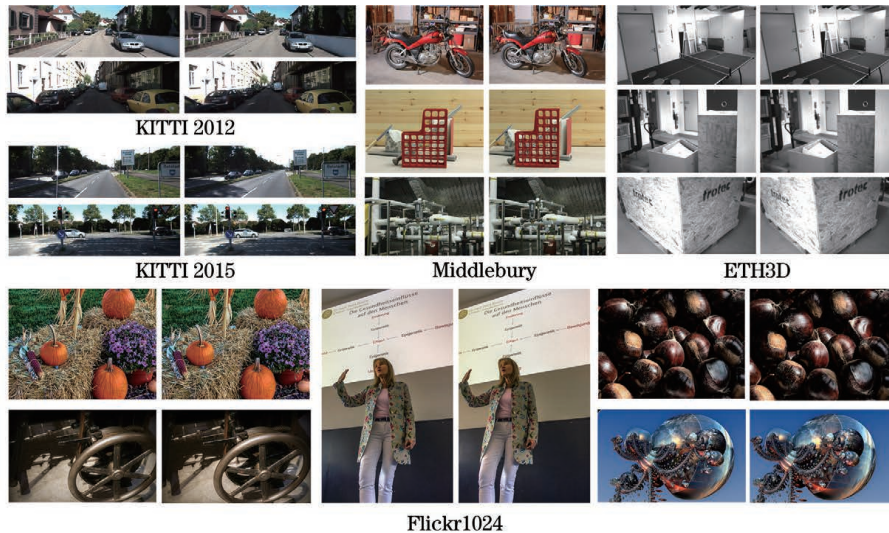


图 4 不同双目图像数据集的样本图像^[33]

Fig. 4 Sample images in different stereo image datasets^[33]

KITTI 数据集是双目视觉领域内面向自动驾驶的大规模双目图像数据集。KITTI 数据集分为 KITTI 2012^[34]和 KITTI 2015^[21]两个部分,数据集中的双目图像由车载双目相机获取。同时,该数据集利用车载激光雷达获取场景点云并将场景点云重投影至图像上,进而得到深度分布,再结合相机参数(基线、焦距等)将深度分布转化为视差真值,并将该值用于视差估计算法的训练与评测。KITTI 2012 数据集包含 194 个训练场景与 195 个测试场景,图

像的平均分辨率为 4.6×10^5 pixel。KITTI 2015 数据集包含 200 个训练场景与 200 个测试场景,图像的平均分辨率为 4.7×10^5 pixel。

Middlebury 双目数据集是双目视觉领域内用于视差估计的基准评测平台。该数据集共包含 5 个子数据集,这 5 个子数据集分别建立于 2001 年(6 个场景)^[35]、2003 年(2 个场景)^[36]、2005 年(9 个场景)^[37]、2006 年(21 个场景)^[38]与 2014 年(33 个场景)^[39]。Middlebury 数据集主要面向室内场景的

视差估计,数据集建立者利用结构光传感器将场景的深度分布转换为视差真值图,视差真值图的标注误差不超过 0.2 pixel。Middlebury 数据集中图像的平均分辨率为 6×10^6 pixel,视差分布范围为 200~800 pixel。

ETH3D 数据集^[40]是一个面向三维重建的多视角真实场景图像数据集。用于视差估计的子数据集包含 47 个灰度图场景,其中 27 个场景用于训练,20 个场景用于测试,图像的平均分辨率为 3×10^5 pixel。该数据集的建立者利用激光扫描仪将场景的点云分布转换为深度及视差真值图,并将深度和视差真值图用于视差估计算法的训练与评测。

KITTI、Middlebury 和 ETH3D 是双目深度估计领域广泛采用的三个数据集。然而,上述三个数据集主要面向视差估计与光流计算,故它们在场景数量与多样性、图像质量等方面无法满足双目图像超分辨任务的需求。KITTI 数据集主要面向自动驾驶,因此只包含驾驶员视角的道路场景,且图像噪声水平较高、质量较差。Middlebury 数据集虽然具有较高的分辨率与图像质量,但是仅包含室内场景,且图像数量较少。ETH3D 数据集的双目图像为灰度图形式,纹理不够丰富且场景数量有限。相比之下,在单图超分辨领域,研究者们建立的多个纹理丰富的大规模数据集,如 DIV2K 数据集^[41]、Flickr2K 数据集^[42],使得单图超分辨算法能够实现很高的重建性能。

为满足双目图像超分辨算法的训练以及评测需求,2019 年,Wang 等^[33]从 Flickr 网站上采集了大量双目图像,构建并公开了一个用于双目图像超分辨的大规模数据集 Flickr1024。该数据集共包含 1024 对双目图像,其中训练集 800 对、验证集 112 对、测试集 112 对,数据集图像的平均分辨率为 7.3×10^5 pixel。相比于 KITTI、Middlebury 和 ETH3D 数据集,Flickr1024 数据集中的场景更加多样,涵盖了日常生活与摄影中的各类情形。

5 算法性能评测

首先,本节对主流的双目图像超分辨算法的性能进行评测,并与先进的单图超分辨算法进行比较。接着,探究不同数据集对双目图像超分辨算法性能的影响。参与评测的算法包括 StereoSR^[22]、PASSRnet^[28]、SRResNet + SAM^[30]、BSSRnet^[31]、iPASSR^[32]、SSRDE-FNet^[26] 这 6 个双目图像超分辨算法以及 VDSR^[43]、EDSR^[44]、RDN^[45]、RCAN^[46]

这 4 个单图超分辨算法。其中,EDSR、RDN 和 RCAN 的单图超分辨重建精度处于超分辨领域的领先水平。此外,将低分辨率图像直接进行双三次插值(Bicubic)操作后的结果作为这些测评算法的基准结果。

5.1 实验设置与实施细节

参与评测的算法都是基于深度学习的方法,为实现算法之间的公平比较,本文在相同的数据集上对这些算法进行重训。参照文献[26, 30-32]中的设置,本文采用的训练集为 Flickr1024 数据集中的 800 幅训练图片以及 Middlebury 数据集中的 60 幅图片。对于 Middlebury 数据集中的图像,采用两倍双三次下采样生成训练所用的高分辨率图像;对于 Flickr1024 数据集中的图像,直接将原图作为训练所用的高分辨率图像。训练中的低分辨率图像采用双三次下采样获得。在训练过程中采用随机水平翻转、随机垂直翻转的方式来实现数据增强。

采用的测试集由 4 部分组成,分别为 Middlebury 数据集中的 5 幅图片、KITTI 2012 数据集中的 20 幅图片、KITTI 2015 数据集中的 20 幅图片以及 Flickr1024 测试集中的 112 幅图片。将峰值信噪比(PSNR)和结构相似度(SSIM)作为数值评测的指标。首先,分别计算每个场景左视角图和右视角图的数值指标。接着,对左右视角图像指标进行平均得到该场景的数值指标。最后,对数据集下所有场景的指标取平均得到该数据集的指标。

5.2 数值结果

不同超分辨算法在 $2 \times$ 放大倍数和 $4 \times$ 放大倍数下的数值结果如表 1 所示。其中,加粗数值表示最优数值指标,加下划线的数值表示次优数值指标。可以发现,SSRDE-FNet 在 4 个数据集上均取得了最高的重建精度,达到了当前双目图像超分辨领域最先进的重建性能。iPASSR 的重建精度略低于 SSRDE-FNet,但是具有更少的参数量。此外,单图超分辨算法 EDSR、RDN 和 RCAN 虽然没有使用左右视角图像间的互补信息,但是取得了很高的重建精度。例如,RCAN 在 KITTI 2012 和 KITTI 2015 数据集上的重建精度与 BSSRnet 相当,在 Middlebury 和 Flickr1024 数据集上的重建精度优于 SSRDE-FNet 之外的所有双目图像超分辨网络。其主要原因是:EDSR、RDN 和 RCAN 的模型参数量较大,这加强了这些网络对单幅图像内邻域信息的建模能力,进而其能够更好地学到训练集图像的纹理结构。而双目图像超分辨网络,如 iPASSR、

表 1 不同算法取得的数值结果
Table 1 Numerical results achieved by different algorithms

Algorithm	Magnification	Number of parameters / 10^6	KITTI 2012	KITTI 2015	Middlebury	Flickr1024
			PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
Bicubic	2×		28.51/0.8842	28.61/0.8973	30.60/0.8990	24.94/0.8186
VDSR ^[43]	2×	0.66	30.30/0.9089	29.78/0.9150	32.77/0.9102	25.60/0.8534
EDSR ^[44]	2×	38.6	30.96/0.9228	30.73/0.9335	<u>34.95/0.9492</u>	<u>28.66/0.9087</u>
RDN ^[45]	2×	22.0	30.94/0.9227	30.70/0.9330	34.94/0.9491	28.64/0.9084
RCAN ^[46]	2×	15.3	31.02/0.9232	30.77/0.9336	34.90/0.9486	28.63/0.9082
StereoSR ^[22]	2×	1.08	29.51/0.9073	29.33/0.9168	33.23/0.9348	25.96/0.8599
PASSRnet ^[28]	2×	1.37	30.81/0.9190	30.60/0.9300	34.23/0.9422	28.38/0.9038
BSSRnet ^[31]	2×	1.89	31.03/0.9241	30.74/0.9344	34.74/0.9475	28.53/0.9090
iPASSR ^[32]	2×	1.37	<u>31.11/0.9240</u>	<u>30.81/0.9340</u>	34.51/0.9454	<u>28.60/0.9097</u>
SSRDE-FNet ^[26]	2×	2.10	31.23/0.9254	30.90/0.9352	35.09/0.9511	28.85/0.9132
Bicubic	4×		24.58/0.7372	24.38/0.7340	26.40/0.7572	21.82/0.6293
VDSR ^[43]	4×	0.66	25.60/0.7722	25.32/0.7703	27.69/0.7941	22.46/0.6718
EDSR ^[44]	4×	38.9	26.35/0.8015	26.04/0.8039	29.23/0.8397	23.46/0.7285
RDN ^[45]	4×	22.0	26.32/0.8014	26.04/0.8043	<u>29.27/0.8404</u>	<u>23.47/0.7295</u>
RCAN ^[46]	4×	15.4	26.44/0.8029	26.22/0.8068	29.30/0.8397	<u>23.48/0.7286</u>
StereoSR ^[22]	4×	1.08	24.53/0.7555	24.21/0.7511	27.64/0.8022	21.70/0.6460
PASSRnet ^[28]	4×	1.42	26.34/0.7981	26.08/0.8002	28.72/0.8236	23.31/0.7195
SRResNet+SAM ^[30]	4×	1.73	26.44/0.8018	26.22/0.8054	28.83/0.8290	23.27/0.7233
BSSRnet ^[31]	4×	1.91	26.47/0.8049	26.17/0.8075	29.08/0.8362	23.40/0.7289
iPASSR ^[32]	4×	1.42	<u>26.56/0.8053</u>	<u>26.32/0.8084</u>	29.16/0.8367	23.44/0.7287
SSRDE-FNet ^[26]	4×	2.24	26.70/0.8082	26.43/0.8118	29.38/0.8411	23.59/0.7352

SSRDE-FNet,则是通过合理利用左右视角图像间的互补信息实现了轻量化条件下较好的图像重建效果。

5.3 视觉效果

图 5 和图 6 分别展示了不同算法在 2×放大倍数和 4×放大倍数下的视觉效果。可以发现,单图超分辨率算法仅利用了单个视角图像内的邻域信息,故在重建细节纹理时存在不同程度的模糊与失真现象。相比之下,双目图像超分辨率算法利用左右视角图像间的互补信息,可以重建出更加真实的纹理细节。通过不同算法之间的比较可得:在 2×放大倍数下,SSRDE-FNet 和 iPASSR 算法的图像重建性能优于其他算法。在 4×放大倍数下,BSSRnet 和 SSRDE-FNet 算法的图像重建性能优于其他算法。

5.4 不同数据集对双目图像超分辨率算法性能的影响

在深度学习时代,数据集对算法的性能具有重要影响。在超分辨率领域,高质量的大规模数据集可以显著提升超分辨率算法的重建精度,降低算法过拟合的风险。本节探究了不同数据集对双目图像超分辨率算法性能的影响。

将 StereoSR^[22] 与 PASSRnet^[28] 算法分别在 KITTI 2015^[21]、Middlebury^[35-39]、Flickr1024^[33] 三个数据集上进行训练。对于 KITTI 2015 以及 Flickr1024 数据集,仍沿用其原数据集中训练集的划分;对于 Middlebury 数据集,遵循文献[22, 28, 30, 32]中的设置,选取 60 幅双目图像作为训练集。在 4×放大倍数下,将算法在不同的训练集上训练 60 轮(epoch),将得到的模型在 KITTI 2015、Middlebury、Flickr1024 以及 ETH3D 数据集的测

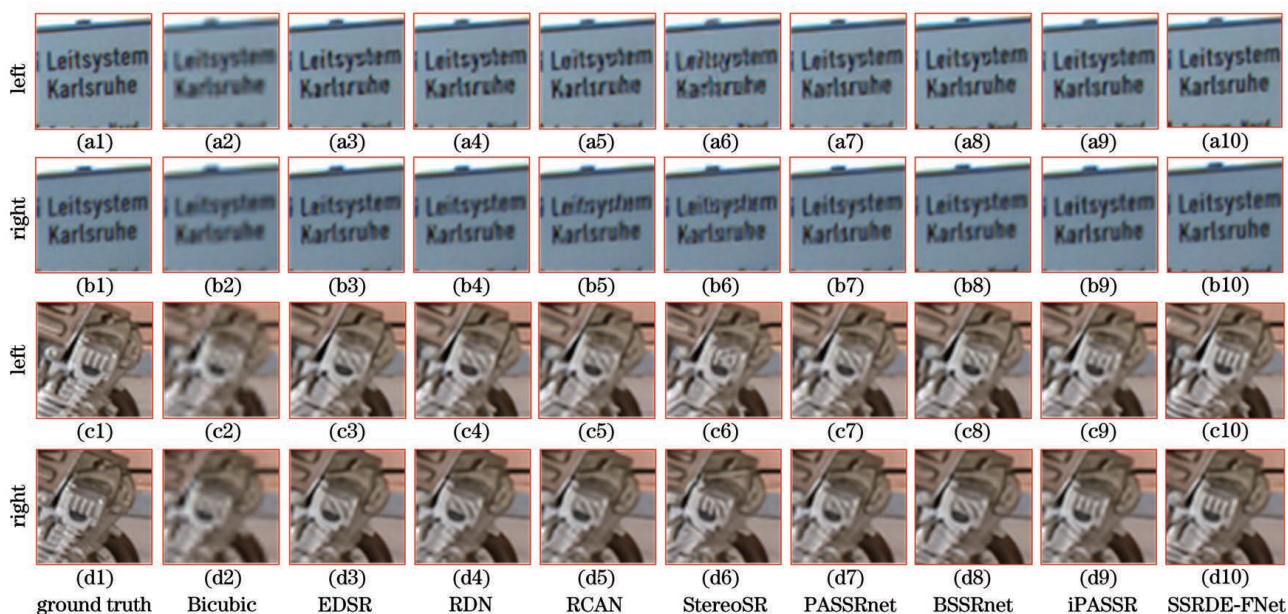


图 5 不同算法在 2×放大倍数下的视觉效果比较

Fig. 5 Comparison of visual effects of different algorithms at 2× magnification

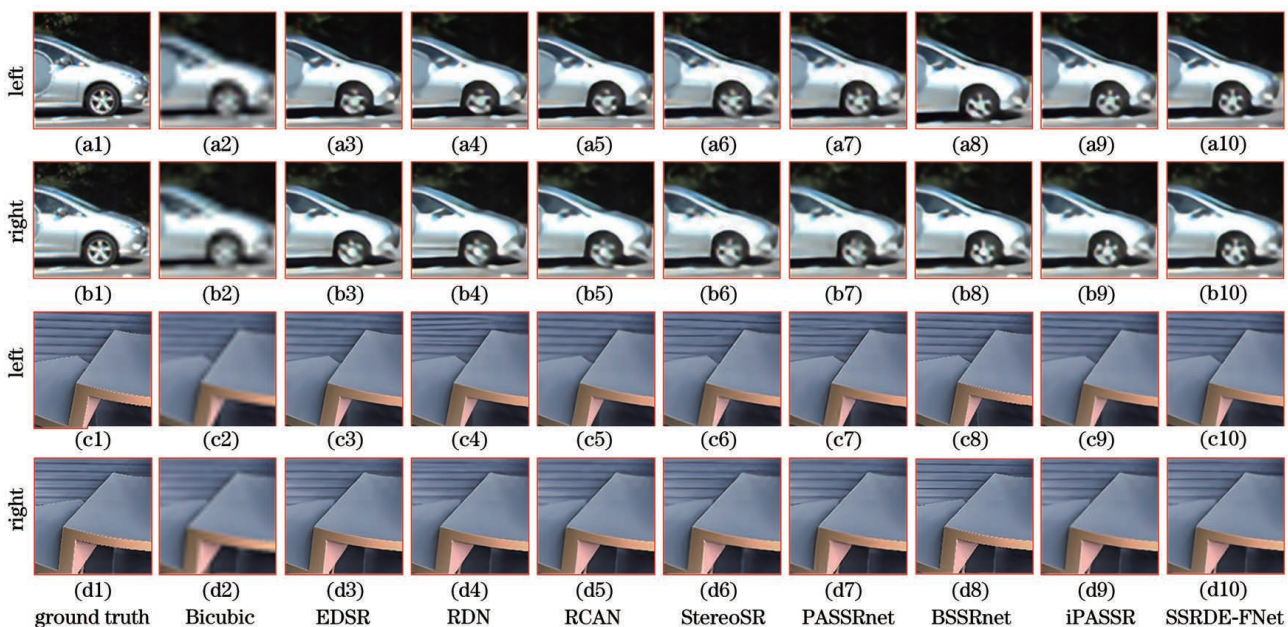


图 6 不同算法在 4×放大倍数下的视觉效果比较

Fig. 6 Comparison of visual effects of different algorithms at 4× magnification

试集上进行测试。表 2 和表 3 分别展示了在 4×放大倍数下, StereoSR 和 PASSRnet 算法在不同数据集上训练 60 轮所取得的数值结果, 其中加粗数值表示最优数值指标。可以发现, 在 Flickr1024 数据集上训练得到的算法模型在各个测试集上均取得了最高的重建精度, 这表明 Flickr1024 数据集中高质量的图像与多样化的场景可以提升双目图像超分辨率算法的重建精度。

接着, 研究了在 4×放大倍数下 PASSRnet 算

法的 PSNR 值与训练轮数间的关系, 如图 7 所示。可以发现, PASSRnet 算法在 Flickr1024 数据集上训练任意轮数后的 PSNR 值均优于 KITTI 2015 和 Middlebury 数据集。当采用 KITTI 2015 数据集的训练集来训练 PASSRnet 算法时, 算法在各测试集上的重建精度随训练轮数的增加呈下降趋势, 这说明 PASSRnet 算法在 KITTI 2015 数据集上训练时发生了过拟合。发生过拟合的原因为 KITTI 2015 数据集是面向自动驾驶的数据集, 该数

表 2 在 4×放大倍数下, StereoSR 算法在不同训练集上训练 60 轮所取得的数值结果

Table 2 Numerical results achieved by StereoSR algorithm trained on different training sets for 60 epochs at 4× magnification

Training set	Testing set	Testing set	Testing set	Testing set
	in KITTI 2015	in Middlebury	in Flickr1024	in ETH3D
	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
Training set in KITTI 2015	24.28/0.741	26.27/0.749	21.77/0.617	29.63/0.831
Training set in Middlebury	23.64/0.743	26.62/0.773	21.64/0.646	28.66/0.843
Training set in Flickr1024	25.08/0.779	27.85/0.807	22.64/0.692	30.55/0.860

表 3 在 4×放大倍数下, PASSRnet 算法在不同训练集上训练 60 轮所取得的数值结果

Table 3 Numerical results achieved by PASSRnet algorithm trained on different training sets for 60 epochs at 4× magnification

Training set	Testing set	Testing set	Testing set	Testing set
	in KITTI 2015	in Middlebury	in Flickr1024	in ETH3D
	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
Training set in KITTI 2015	23.13/0.703	25.42/0.762	21.31/0.600	26.95/0.789
Training set in Middlebury	25.18/0.774	28.08/0.853	22.54/0.676	31.39/0.864
Training set in Flickr1024	25.62/0.791	28.69/0.873	23.25/0.718	31.94/0.877

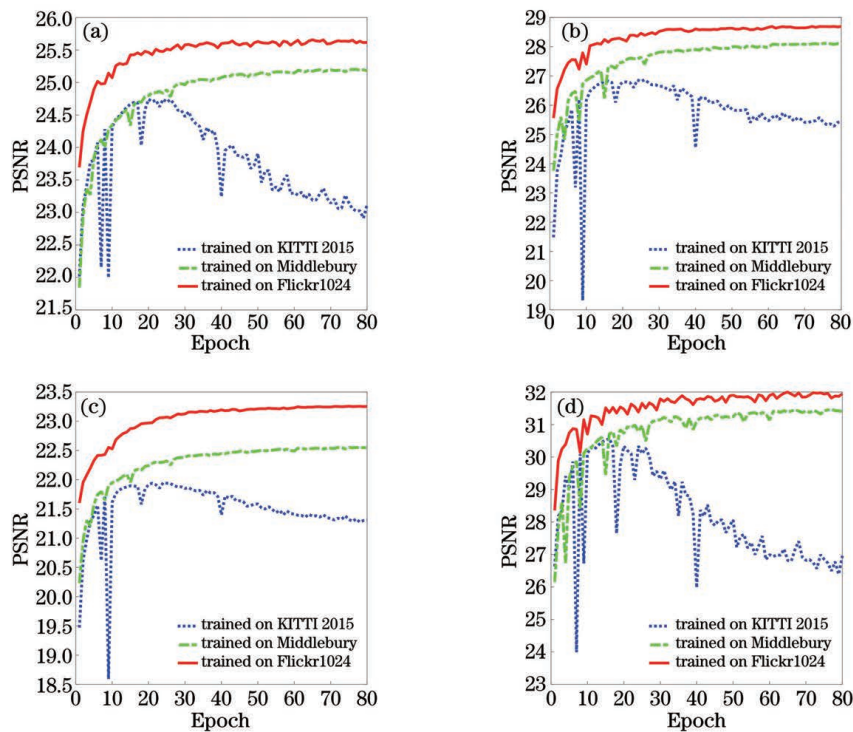


图 7 在 4×放大倍数下, PASSRnet 算法在不同测试集中取得的 PSNR 值。

(a) KITTI 2015 的测试集; (b) Middlebury 的测试集; (c) Flickr1024 的测试集; (d) ETH3D 的测试集

Fig. 7 PSNR values achieved by PASSRnet algorithm in different testing sets at 4× magnification . (a) Testing set in KITTI 2015; (b) testing set in Middlebury; (c) testing set in Flickr1024; (d) testing set in ETH3D

数据集均采用单一类型的双目相机记录驾驶员视角的道路场景图, 算法难以从中学到通用的纹理表征。相比之下, Flickr1024 数据集上训练的模型未发生

过拟合现象, 其 PSNR 值随着训练轮数的增加单调上升。表 2、表 3 和图 7 表明 Flickr1024 数据集不但可以帮助算法取得更加优异的性能, 还能有效防

止过拟合现象的发生。

6 挑战与展望

虽然双目图像超分辨在近年来被广泛研究,其重建精度和视觉效果在不断提升,但是该领域仍然存在许多挑战与亟待解决的问题。本节总结了当前双目图像超分辨领域存在的挑战,并对未来的研究方向进行了展望。

6.1 双目图像超分辨面临的挑战

6.1.1 充分利用左右视角图像间的互补信息与单个视角图像内的上下文信息提升超分辨重建精度

表 1 中的性能评测结果表明,当前大部分双目图像超分辨算法在利用左右视角图像间的互补信息的前提下,其性能方面仍然逊于先进的单图超分辨算法,如 RCAN^[46]。这说明双目图像超分辨算法在学习单幅图像纹理结构表征方面的能力与单图超分辨算法间有明显差距。因此,实现双目图像超分辨算法在利用左右视角图像间互补信息的同时充分利用单个视角图像内的上下文先验并进一步提升重建精度具有一定的挑战性。

6.1.2 左右视角间的信息交互与计算开销之间的权衡

当前双目图像超分辨算法主要采用基于视差估计以及基于视差注意力的左右视角图像信息交互机制。在基于视差估计的方法中,需要设计独立的模块来计算左右视角图像的匹配代价,并通过多个卷积层对匹配代价进行聚合,这增加了网络的参数量与计算开销。基于视差注意力机制的方法通过构建视差注意力图来关联左右视角图像的特征。然而,在网络中加入多个视差注意力模块,会占用大量显存,这使得该类算法难以部署至各类移动终端。因此,在实现左右视角图像间信息充分交互的同时,降低算法的计算开销具有一定的挑战性。

6.1.3 无监督方式下可靠的左右视角图像关联

因为含有真值视差标注的双目图像难以获取,所以现有双目图像超分辨网络大多采用无监督的方式建立左右视角图像间的关联,即利用左右视角图像间对应区域的照度一致性以及双目图像的循环对称性来设计网络训练所需的损失函数。然而,现有的无监督视差估计方法在性能方面仍然逊于有监督的方法。因此,在没有真值视差标注的双目图像数据集上训练网络,并使网络可靠地关联左右视角图像具有一定的挑战性。

6.1.4 高质量双目图像数据集

与单图超分辨数据集相比,现有的双目图像超分辨数据集在图像数量、图像质量、场景多样性方面均处于劣势,故经双目图像超分辨数据集训练的超分辨算法的性能受到了限制。当前大部分双目相机的成像质量仍低于单反相机,故高质量双目图像数据集的获取具有一定的挑战性。

6.2 未来的研究方向

6.2.1 新的双目图像信息交互机制

现有的双目图像超分辨算法普遍采用基于视差估计或基于视差注意力的方式交互双目信息。当前,Transformer^[47]与可变形卷积^[48-49]等先进技术被成功应用于单图超分辨^[50]与多图^[51-52]/视频^[53-57]超分辨中并取得了较好的性能。因此,在双目图像超分辨领域中,未来的研究可以考虑结合双目图像的极线约束条件来探索新的双目图像信息交互机制,进而实现左右视角图像对应位置的关联与信息的融合。

6.2.2 不同双目图像间的域差异

域差异是视差估计中一个极具挑战性的问题。由于不同双目图像数据集所包含的场景类型与风格具有显著差异,因此现有视差估计算法通常只能在少数数据集上取得优越性能。虽然近年来双目图像超分辨领域中有多个工作^[58-62]着重解决视差估计的域差异问题并取得了一定的进展,但是视差估计算法在不同类型数据上的鲁棒性仍然有待提升。当前,双目图像超分辨领域内还没有针对域差异方面的研究。未来的研究工作可以聚焦于探究超分辨算法在不同类型双目图像上的泛化性能、解决双目图像的域差异问题。此外,利用双目图像超分辨来解决下游任务(如视差估计)的域差异问题也值得进一步探索。

6.2.3 面向未知降质的盲超分辨

现有的双目图像超分辨方法采用双三次下采样的方式生成低分辨率图像,并用高低分辨率双目图像对来训练网络。文献^[63-65]指出,当测试阶段与训练阶段的降质过程不一致时,超分辨方法的性能会显著下降。由于真实图像的降质过程多样,因此现有的双目图像超分辨算法无法直接对真实图像进行处理。在单图超分辨领域中,学者们提出盲超分辨算法^[66-71]以解决未知降质下的图像超分辨问题。相比于采用双三次下采样方式训练得到的超分辨算法,盲超分辨算法显著提升了真实图像上的视觉效果。当前,对面向真实场景的双目图像超分辨有

较大的需求。因此,将现有单图盲超分辨算法与双目极线约束条件结合,解决未知降质下的双目图像盲超分辨问题,具有重要的研究意义与光明的发展前景。

6.2.4 面向下游任务的双目图像超分辨

当前双目图像超分辨的工作大多追求的是更高的重建精度与更好的视觉效果,而忽略了其对下游任务性能的潜在增益。虽然 Wang 等^[32]初步验证了超分辨对视差估计的增益,但是并未面向视差估计任务进行端到端训练。相比于单图超分辨被广泛应用于目标检测^[72-73]、分割^[74-76]、人脸识别^[77-78]等方向,当前双目图像超分辨领域中仍缺乏面向下游任务的双目图像超分辨研究工作。由于高分辨率的双目图像对视差估计^[79-81]、双目图像分割^[82]、三维目标检测^[83-85]等下游任务的性能具有显著增益,因此面向下游任务的双目图像超分辨可以作为下一步的研究方向。

6.2.5 基于新型双目成像系统的图像超分辨

当前双目图像超分辨算法均假定两相机参数一致且按平行光轴方式进行装配。在智能手机和机器人领域中,也常常采用更加灵活的方式对双目相机进行装配部署,如采用非平行光轴进行装配、使用不同分辨率的两个子相机以及采用“黑白+彩色”的成像模式。未来的研究可以结合实际的应用需求,发展出适合于新型双目成像系统的图像超分辨算法。

7 结束语

对双目图像超分辨研究进行了综述。首先,介绍了双目成像的基本原理。接着,回顾了双目图像超分辨领域的算法与数据集。随后,在基准数据集上对主流超分辨算法的性能进行了评测,并探究了不同训练集对超分辨算法性能的影响。最后,指出了双目图像超分辨领域中存在的挑战,并展望了未来的研究方向。

参 考 文 献

- [1] Dong X Y, Sun X, Jia X P, et al. Remote sensing image super-resolution using novel dense-sampling networks[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(2): 1618-1633.
- [2] Dong X Y, Wang L G, Sun X, et al. Remote sensing image super-resolution using second-order multi-scale networks[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(4): 3473-3485.
- [3] He S T, Zou H X, Wang Y Q, et al. ShipSRDet: an end-to-end remote sensing ship detector using super-resolved feature representation [EB/OL]. (2021-03-17) [2021-05-20]. <https://arxiv.org/abs/2103.09699>.
- [4] Huang Y W, Shao L, Frangi A F. Simultaneous super-resolution and cross-modality synthesis of 3D medical images using weakly-supervised joint convolutional sparse coding[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 5787-5796.
- [5] Oktay O, Ferrante E, Kamnitsas K, et al. Anatomically constrained neural networks (ACNNs): application to cardiac image enhancement and segmentation [J]. IEEE Transactions on Medical Imaging, 2017, 37(2): 384-395.
- [6] Hu F, Lin Y, Hou M D, et al. Super-resolution reconstruction of cytoskeleton image based on deep learning[J]. Acta Optica Sinica, 2020, 40(24): 2410001.
胡芬, 林洋, 侯梦迪, 等. 基于深度学习的细胞骨架图像超分辨重建[J]. 光学学报, 2020, 40(24): 2410001.
- [7] Xiao K, Tian L J, Wang Z Y. Fast super-resolution fluorescence microscopy imaging with low signal-to-noise ratio based on deep learning [J]. Chinese Journal of Lasers, 2020, 47(10): 1007002.
肖康, 田立君, 王中阳. 基于深度学习的低信噪比下的快速超分辨荧光显微成像[J]. 中国激光, 2020, 47(10): 1007002.
- [8] Bar-Zion A, Tremblay-Darveau C, Solomon O, et al. Fastvascular ultrasound imaging with enhanced spatial resolution and background rejection[J]. IEEE Transactions on Medical Imaging, 2017, 36(1): 169-180.
- [9] Seibel H, Goldenstein S, Rocha A. Eyes on the target: super-resolution and license-plate recognition in low-quality surveillance videos[J]. IEEE Access, 2017, 5: 20020-20035.
- [10] Pang Y W, Cao J L, Wang J, et al. JCS-net: joint classification and super-resolution network for small-scale pedestrian detection in surveillance images[J]. IEEE Transactions on Information Forensics and Security, 2019, 14(12): 3322-3331.
- [11] Chen L, Pan J S, Li Q. Robust face image super-resolution via joint learning of subdivided contextual model[J]. IEEE Transactions on Image Processing, 2019, 28(12): 5897-5909.
- [12] Tan T D, Guo Z M. Research of hand positioning and gesture recognition based on binocular vision [C]//

- 2011 IEEE International Symposium on VR Innovation, March 19-20, 2011, Singapore. New York: IEEE Press, 2011: 311-315.
- [13] Oskiper T, Sizintsev M, Branzoi V, et al. Augmented reality binoculars[J]. IEEE Transactions on Visualization and Computer Graphics, 2015, 21(5): 611-623.
- [14] Jia B X, Chen J, Zhang K X. Drivable road reconstruction for intelligent vehicles based on two-view geometry[J]. IEEE Transactions on Industrial Electronics, 2017, 64(5): 3696-3706.
- [15] Gong W, Qin L, Ren G F, et al. Binocular stereo matching algorithm based on multidimensional feature fusion[J]. Laser & Optoelectronics Progress, 2020, 57(16): 161501.
龚伟, 秦岭, 任高峰, 等. 基于多维特征融合的双目立体匹配算法研究[J]. 激光与光电子学进展, 2020, 57(16): 161501.
- [16] Cheng M Y, Gai S Y, Da F P. A stereo-matching neural network based on attention mechanism [J]. Acta Optica Sinica, 2020, 40(14): 1415001.
程鸣洋, 盖绍彦, 达飞鹏. 基于注意力机制的立体匹配网络研究[J]. 光学学报, 2020, 40(14): 1415001.
- [17] Yang W M, Zhang X C, Tian Y P, et al. Deep learning for single image super-resolution: a brief review[J]. IEEE Transactions on Multimedia, 2019, 21(12): 3106-3121.
- [18] Wang Z H, Chen J, Hoi S C H. Deep learning for image super-resolution: a survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 2166: 32217470.
- [19] Tang Y Q, Pan H, Zhu Y P, et al. A survey of image super-resolution reconstruction [J]. Acta Electronica Sinica, 2020, 48(7): 1407-1420.
唐艳秋, 潘泓, 朱亚平, 等. 图像超分辨率重建研究综述[J]. 电子学报, 2020, 48(7): 1407-1420.
- [20] Qu Y Y, Chen R, Li C H, et al. Survey of single image super-resolution reconstruction with deep learning[J]. Journal of Xiamen University (Natural Science), 2021, 60(3): 555-570, 641-642.
曲延云, 陈蓉, 李翠华, 等. 深度学习单帧图像超分辨率重建研究综述[J]. 厦门大学学报(自然科学版), 2021, 60(3): 555-570, 641-642.
- [21] Menze M, Geiger A. Objectscene flow for autonomous vehicles[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 3061-3070.
- [22] Jeon D S, Baek S H, Choi I, et al. Enhancing the spatial resolution of stereo images using a parallax prior[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 1721-1730.
- [23] Bhavsar A V, Rajagopalan A N. Resolution enhancement in multi-image stereo [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(9): 1721-1728.
- [24] Yan B, Ma C X, Bare B, et al. Disparity-aware domain adaptation in stereo image restoration[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 13176-13184.
- [25] Khamis S, Fanello S, Rhemann C, et al. StereoNet: guided hierarchical refinement for real-time edge-aware depth prediction[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11219: 596-613.
- [26] Dai Q Y, Li J C, Yi Q S, et al. Feedback network for mutually boosted stereo image super-resolution and disparity estimation [EB/OL]. (2021-06-02) [2021-06-05]. <https://arxiv.org/abs/2106.00985>.
- [27] Wang L G, Wang Y Q, Liang Z F, et al. Learning parallax attention for stereo image super-resolution [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 12242-12251.
- [28] Wang L G, Guo Y L, Wang Y Q, et al. Parallax attention for unsupervised stereo correspondence learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 6899: 32976095.
- [29] Song W, Choi S, Jeong S, et al. Stereoscopic image super-resolution with stereo consistent feature [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12031-12038.
- [30] Ying X Y, Wang Y Q, Wang L G, et al. A stereo attention module for stereo image super-resolution [J]. IEEE Signal Processing Letters, 2020, 27: 496-500.
- [31] Xu Q Y, Wang L G, Wang Y Q, et al. Deep bilateral learning for stereo image super-resolution[J]. IEEE Signal Processing Letters, 2021, 28: 613-617.
- [32] Wang Y Q, Ying X Y, Wang L G, et al. Symmetric parallax attention for stereo image super-resolution [EB/OL]. (2021-04-20) [2021-05-20]. <https://arxiv.org/abs/2011.03802v2>.
- [33] Wang Y Q, Wang L G, Yang J G, et al. Flickr1024: a large-scale dataset for stereo image super-resolution [C]//2019 IEEE/CVF International Conference on

- Computer Vision Workshop (ICCVW), October 27-28, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 3852-3857.
- [34] Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? The KITTI vision benchmark suite[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition, June 16-21, 2012, Providence, RI, USA. New York: IEEE Press, 2012: 3354-3361.
- [35] Scharstein D, Szeliski R, Zabih R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms[C]//Proceedings IEEE Workshop on Stereo and Multi-Baseline Vision (SMBV 2001), December 9-10, 2001, Kauai, HI, USA. New York: IEEE Press, 2001: 131-140.
- [36] Scharstein D, Szeliski R. High-accuracy stereo depth maps using structured light[C]//2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings, June 18-20, 2003, Madison, WI, USA. New York: IEEE Press, 2003: 1-8.
- [37] Scharstein D, Pal C. Learning conditional random fields for stereo[C]//2007 IEEE Conference on Computer Vision and Pattern Recognition, June 17-22, 2007, Minneapolis, MN, USA. New York: IEEE Press, 2007: 1-8.
- [38] Hirschmuller H, Scharstein D. Evaluation of cost functions for stereo matching[C]//2007 IEEE Conference on Computer Vision and Pattern Recognition, June 17-22, 2007, Minneapolis, MN, USA. New York: IEEE Press, 2007: 1-8.
- [39] Scharstein D, Hirschmüller H, Kitajima Y, et al. High-resolution stereo datasets with subpixel-accurate ground truth[M]//Jiang X Y, Hornegger J, Koch R. Pattern recognition. Lecture notes in computer science. Cham: Springer, 2014, 8753: 31-42.
- [40] Schöps T, Schönberger J L, Galliani S, et al. A multi-view stereo benchmark with high-resolution images and multi-camera videos[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 2538-2547.
- [41] Agustsson E, Timofte R. NTIRE 2017 challenge on single image super-resolution: Dataset and study [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 1122-1131.
- [42] Timofte R, Agustsson E, van Gool L, et al. NTIRE 2017 challenge on single image super-resolution: methods and results [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, July 21-26, Honolulu, HI, USA. New York: IEEE Press, 2017: 114-125.
- [43] Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 1646-1654.
- [44] Lim B, Son S, Kim H, et al. Enhanced deep residual networks for single image super-resolution [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 1132-1140.
- [45] Zhang Y L, Tian Y P, Kong Y, et al. Residual dense network for image super-resolution [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 2472-2481.
- [46] Zhang Y L, Li K P, Li K, et al. Image super-resolution using very deep residual channel attention networks [M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 294-310.
- [47] Liu Z, Lin Y, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows [EB/OL]. (2021-03-25) [2021-05-20]. <https://arxiv.org/abs/2103.14030>.
- [48] Dai J F, Qi H Z, Xiong Y W, et al. Deformable convolutional networks [C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 764-773.
- [49] Zhu X Z, Hu H, Lin S, et al. Deformable ConvNets v2: more deformable, better results [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 9300-9308.
- [50] Chen H T, Wang Y H, Guo T Y, et al. Pre-trained image processing transformer [EB/OL]. (2020-12-01) [2021-05-20]. <https://arxiv.org/abs/2012.00364>.
- [51] Wang Y Q, Yang J G, Wang L G, et al. Light field image super-resolution using deformable convolution [J]. IEEE Transactions on Image Processing, 2020, 30: 1057-1071.
- [52] Yang F Z, Yang H, Fu J L, et al. Learning texture

- transformer network for image super-resolution[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 5790-5799.
- [53] Tian Y P, Zhang Y L, Fu Y, et al. TDAN: temporally-deformable alignment network for video super-resolution[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 3357-3366.
- [54] Ying X Y, Wang L G, Wang Y Q, et al. Deformable 3D convolution for video super-resolution[J]. IEEE Signal Processing Letters, 2020, 27: 1500-1504.
- [55] Chan K C K, Wang X, Yu K, et al. Understanding deformable alignment in video super-resolution[EB/OL]. (2020-09-15)[2021-05-20]. <https://arxiv.org/abs/2009.07265>.
- [56] Chan K C K, Wang X T, Yu K, et al. BasicVSR: the search for essential components in video super-resolution and beyond[EB/OL]. (2020-12-03)[2021-05-21]. <https://arxiv.org/abs/2012.02181>.
- [57] Kim T H, Sajjadi M S M, Hirsch M, et al. Spatio-temporal transformer network for video restoration[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11207: 111-127.
- [58] Pang J H, Sun W X, Yang C X, et al. Zoom and learn: generalizing deep stereo matching to novel domains[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 2070-2079.
- [59] Guo X Y, Li H S, Yi S, et al. Learning monocular depth by distilling cross-domain stereo networks[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11215: 506-523.
- [60] Tonioni A, Tosi F, Poggi M, et al. Real-time self-adaptive deep stereo[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 195-204.
- [61] Liu R, Yang C X, Sun W X, et al. StereoGAN: bridging synthetic-to-real domain gap by joint optimization of domain translation and stereo matching[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 12754-12763.
- [62] Song X, Yang G R, Zhu X G, et al. AdaStereo: a simple and efficient approach for adaptive stereo matching[EB/OL]. (2020-08-09)[2021-05-20]. <https://arxiv.org/abs/2004.04627>.
- [63] Efrat N, Glasner D, Apartsin A, et al. Accurate blur models vs. image priors in single image super-resolution[C]//2013 IEEE International Conference on Computer Vision, December 1-8, 2013, Sydney, NSW, Australia. New York: IEEE Press, 2013: 2832-2839.
- [64] Yang C Y, Ma C, Yang M H. Single-imagesuper-resolution: a benchmark[M]//Fleet D, Pajdla T, Schiele B, et al. Computer vision-ECCV 2014. Lecture notes in computer science. Cham: Springer, 2014, 8692: 372-386.
- [65] Gu J J, Lu H N, Zuo W M, et al. Blind super-resolution with iterative kernel correction[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 1604-1613.
- [66] Wang L G, Wang Y Q, Dong X Y, et al. Unsupervised degradation representation learning for blind super-resolution[EB/OL]. (2021-04-01)[2021-05-20]. <https://arxiv.org/abs/2104.00416>.
- [67] Gu J J, Lu H N, Zuo W M, et al. Blind super-resolution with iterative Kernel correction[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 1604-1613.
- [68] Luo Z X, Huang Y, Li S, et al. Unfolding the alternating optimization for blind super resolution[EB/OL]. (2020-10-06)[2021-05-20]. <https://arxiv.org/abs/2010.02631>.
- [69] Michaeli T, Irani M. Nonparametric blind super-resolution[C]//2013 IEEE International Conference on Computer Vision, December 1-8, 2013, Sydney, NSW, Australia. New York: IEEE Press, 2013: 945-952.
- [70] Shocher A, Cohen N, Irani M. Zero-shot super-resolution using deep internal learning[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 3118-3126.
- [71] Yuan Y, Liu S Y, Zhang J W, et al. Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks[C]//2018 IEEE/CVF Conference

- on Computer Vision and Pattern Recognition Workshops (CVPRW), June 18-22, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 814-823.
- [72] Li Z, Zhang Y. Hyperspectral anomaly detection via image super-resolution processing and spatial correlation[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(3): 2307-2320.
- [73] Ji H, Gao Z, Mei T C, et al. Vehicle detection in remote sensing images leveraging on simultaneous super-resolution[J]. IEEE Geoscience and Remote Sensing Letters, 2020, 17(4): 676-680.
- [74] Wang L, Li D, Zhu Y S, et al. Dual super-resolution learning for semantic segmentation[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 3773-3782.
- [75] Zhang Z P, Zhang K P. FarSee-net: real-time semantic segmentation by efficient multi-scale context aggregation and feature space super-resolution[C]//2020 IEEE International Conference on Robotics and Automation (ICRA), May 31-August 31, 2020, Paris, France. New York: IEEE Press, 2020: 8411-8417.
- [76] Lei S, Shi Z W, Wu X, et al. Simultaneous super-resolution and segmentation for remote sensing images[C]//IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium, July 28-August 2, 2019, Yokohama, Japan. New York: IEEE Press, 2019: 3121-3124.
- [77] Chen Y, Tai Y, Liu X M, et al. FSRNet: end-to-end learning face super-resolution with facial priors[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 2492-2501.
- [78] Chen C F, Gong D H, Wang H, et al. Learning spatial attention for face super-resolution[J]. IEEE Transactions on Image Processing, 2021, 30: 1219-1231.
- [79] Xu B, Xu Y H, Yang X L, et al. Bilateral grid learning for stereo matching network [EB/OL]. (2021-01-01)[2021-05-20]. <https://arxiv.org/abs/2101.01601>.
- [80] Liang Z F, Guo Y L, Feng Y L, et al. Stereo matching using multi-level cost volume and multi-scale feature constancy[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(1): 300-315.
- [81] Liang Z F, Feng Y L, Guo Y L, et al. Learning for disparity estimation through feature constancy[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 2811-2820.
- [82] Li X J, Huang H, Zhao H L, et al. Learning a convolutional neural network for propagation-based stereo image segmentation[J]. The Visual Computer, 2020, 36(1): 39-52.
- [83] Chen Y L, Liu S, Shen X Y, et al. DSGN: deep stereo geometry network for 3D object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 12533-12542.
- [84] Qin Z Y, Wang J L, Lu Y. Triangulation learning network: from monocular to stereo 3D object detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 7607-7615.
- [85] Li P L, Chen X Z, Shen S J. Stereo R-CNN based 3D object detection for autonomous driving [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 7636-7644.