

双目图像超分辨研究综述

王应谦¹, 王龙光¹, 梁政宇¹, 安玮¹, 杨俊刚^{1*}

¹ 国防科技大学电子科学学院, 长沙 410073;

摘要 随着双摄像头成像技术的发展, 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图的互补信息进一步提升图像重建质量。本文对近年来双目图像超分辨领域的研究进展进行了综述: 首先介绍了双目成像的基本理论, 并对现有的双目图像超分辨算法和数据集进行归纳总结。之后在公开数据集上对多个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, 并对未来的研究方向进行了展望。

关键词 图像超分辨; 双目视觉; 综述; 性能评测; 数据集

中图分类号 TP391 **文献标志码** A

Stereo Image Super-Resolution: A Survey

Yingqian Wang¹, Longguang Wang¹, Zhengyu Liang¹, Wei An¹, Jungang Yang^{1*}

¹ *College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China;*

Abstract With the development of stereo imaging technology, stereo image super-resolution (SR) have attracted increasing attention in recent years. Different from single image SR, stereo image SR can use the complementary cross-view information to further improve the SR performance. In this paper, we present a comprehensive survey on recent advances of stereo image SR. We first introduce the fundamental theory of stereo imaging, and then review the existing stereo image SR methods and stereo image datasets. Afterwards, we comprehensively evaluate several milestone deep learning based stereo image SR methods on benchmark datasets, and investigate the impact of different training sets to the SR performance. Finally, we summarize the challenges in stereo image SR and discuss the future research directions.

Key words super-resolution; stereo vision; survey; performance evaluation; dataset;

OCIS codes 100.6640; 330.1400; 150.4232; 110.1758;

1 引言

图像超分辨率重建 (super-resolution, SR) 旨在利用一张或多张低分辨率的图像计算生成相应高分辨率的图像并恢复输入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提高图像的分辨率与视觉效果, 具有较低的图像获

取成本, 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

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)

*E-mail: yangjungang@nudt.edu.cn;

人工智能与深度学习的迅猛发展，深度卷积神经网络（deep convolutional neural networks）被广泛应用到图像超分辨任务中，并在重建精度与视觉效果上取得了突破性提升。



图1 常见的商用双目相机

Fig. 1 Examples of commercial stereo cameras.

随着双摄像头成像技术的发展，双目相机（见图 1）在近年来被成功应用于虚拟现实^[12]、增强现实^[13]、自动驾驶^[14]等领域。由于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，有助于后续任务（如深度估计、三维物体检测）达到更高的性能，许多学者尝试利用图像超分辨算法提升双目图像的分辨率^[15-16]。

实现双目图像超分辨的一个直接的方式是将单图超分辨算法分别应用于左右两幅图像。然而，对左右图分别进行单图超分辨忽略了双目图像之间的关联。实际上，双目相机从左右两个视角记录当前场景，视角间的互补信息（cross-view information）使得左右图在超分辨的过程中可以互为参考，从而为图像超分辨提供额外的先验约束。合理利用左右图的互补信息对提升双目图像超分辨的性能具有关键作用。

目前，图像超分辨领域正处于快速发展时期，每年有大量研究成果发表于计算机视觉与图像处理的顶级期刊与会议上。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综述总结^[16-18]。2019 年，Yang 等^[16]从网络结构和目标函数出发，回顾并评测了基于深度学习的单图超分辨算法，并在此基础上对若干关键问题进行总结与分析。同年，Wang 等^[18]从有监督、无监督和典型应用领域三个方面阐述了基于深度学习的单图超分辨方法。2020 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，并以有监督和不成对两个类别梳理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，缺少对于双目图像超分辨的针对性总结。随着双目相机的广泛应用，近年来领域里涌现了许多双目图像超分辨的研究，因此需要一篇相应的综述工作对这些研究进行总结与讨论。本文在上述研究的基础上，对双目图像超分辨率

重建展开综合论述。首先介绍双目相机成像的基本原理，而后按照建立左右图关联的方式对现有双目超分辨的方法进行分类回顾，并对领域内主要的双目图像数据集进行总结。之后，在领域公开数据集上对若干主流方法进行评测，并分析了不同训练集对基于深度学习算法性能的影响。最后，本文指出了当前双目图像超分辨所面临的挑战，并展望了未来的研究方向。

2 双目成像基本原理

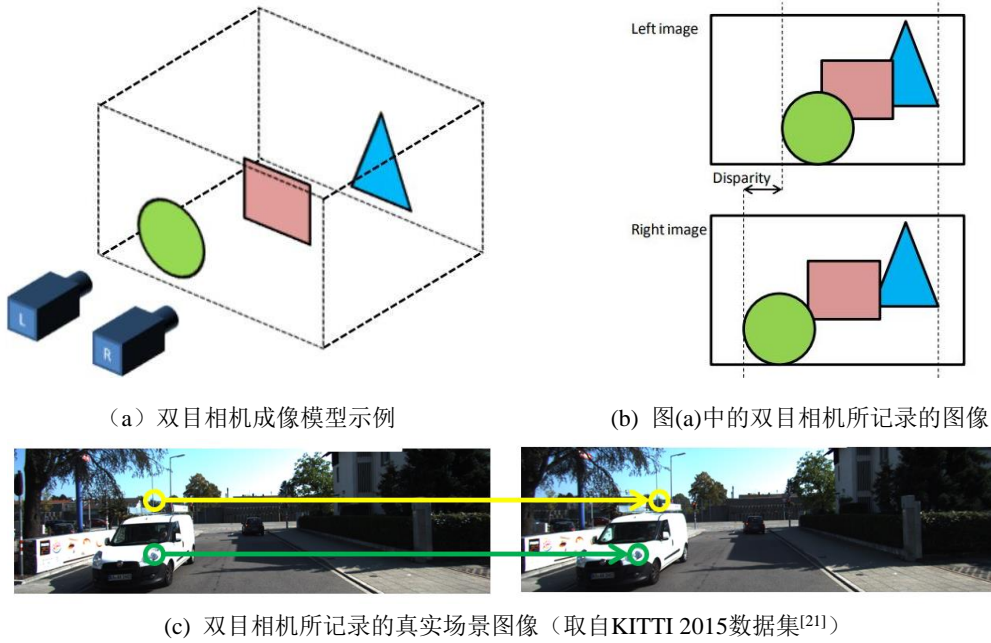


图2 双目相机与双目图像示例

Fig. 2 Illustrations of the stereo camera and stereo images

双目相机可以从左右两个视角记录当前场景，利用左右图所包含的几何关系可以解析场景的三维结构。图 2 展示了双目相机成像的一个模型示例，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。由于左右视角的偏差，场景中的同一物体在左图和右图的位置具有一定差异，距离相机较近的物体视差（disparity）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，同一物体在左右图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）的关系可以建模为：

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

式中， γ 表示物体与相机的距离（即场景深度）， B 为左右两个相机之间的基线长度， f 表示相机的焦距， d 表示该物体在左图和右图之间的视差。图 2(c)展示了 KITTI 2015 数据集上真实的双目图像，如图可见，真实场景不同深度区域具有不同的视差。

为了结合左右图的互补信息提升双目图像超分辨的重建质量，需要建立左右图对应区域的关联。由上述分析可知，在不考虑遮挡的情况下，左右图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（epipolar line），如图 3 所示。双目图像超分辨算法应当结合极线约束（epipolar constraint）建立左右图之间的关联。下一小节将按照建立左右图关联的方式对双目图像超分辨算法进行分类总结。

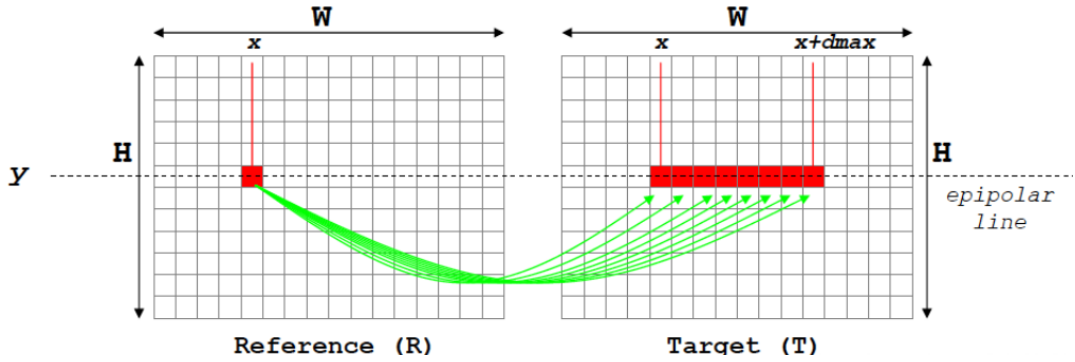


图3 双目图像的极线约束。对于左图的任意一点，其在右图中的对应点应处于极线上

Fig. 3 Epipolar constraint of stereo images. For any pixel in the left image, its corresponding pixel in the right image lies on the epipolar line.

3 方法回顾与总结

本小节梳理总结现有双目图像超分辨的主要方法¹。按照建立左右图关联方式的不同，将现有双目图像超分辨方法分为基于平移副本堆叠的方法、基于视差估计的方法以及基于视差注意力的方法。

3.1 基于平移副本堆叠的方法

2018 年，Jeon 等^[22]提出双目图像超分辨算法 StereoSR，采用平移副本堆叠的方式实现不同视差下左右图互补信息的融合。该方法将右图依次向右平移 1 至 64 像素并生成 64 幅副本图像，而后将 64 幅副本图像与左图在通道维度上级联（concatenation）并送入网络进行重建。在不考虑遮挡的情况下，当左右图的最大视差小于 64 像素时，对于左图中的任一区域，一定存在某个平移之后的副本图像在对应区域与之粗略对齐，从而可以通过卷积操作对左右图的互补信息进行融合。

StereoSR 算法是领域首个基于深度学习的双目图像超分辨方法，该算法通过对右图的

¹ 本文汇总整理了领域内现有方法的代码，相关链接：<https://github.com/YingqianWang/Stereo-Image-SR>

水平移动合理利用了左右图之间的极线约束,使得左图中大部分区域都能与右图的某个副本对应。然而,该方法需要人为设置一个最大的视差范围(论文中设为 64),对于超过该视差范围的信息则无法利用。同时,该算法未对遮挡等区域做特殊处理,限制了其重建性能。

3.2 基于视差估计的方法

2009 年, Bhavsar 等^[23]提出了一种同时超分辨双目图像与对应视差图的方法。Bhavsar 等指出双目图像超分辨与视差估计是相互耦合的两个问题:双目图像超分辨需要建立左右图之间的像素级关联,从而需要估计较为精确的视差提供配准信息;而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,并通过基于图割(graph cut)的方法求解最小化能量函数。在求解的过程中, Bhavsar 等采用迭代优化的方式,首先估计初始场景的视差分布,而后利用该视差对左右图进行配准,重建出高分辨率的左右图,之后再利用重建图像继续求解更高精度的视差图,不断迭代直至收敛,最终得到高分辨率双目图像与视差图的最优估计。

随着深度学习的发展,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任务,并取得了突破性的性能提升。领域内多个方法也采用深度学习的范式来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,在超分辨、去模糊和去噪三个任务上取得了当时领先的效果。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共三部分组成,该方法首先采用单目特征提取子网络提取低分辨率左右图特征,而后采用双目立体匹配算法 StereoNet^[25]对左右图特征进行匹配,估计出对应的视差图(disparity map)。之后,该方法将右图按照估计的视差图投影(warping)至左边视角,与左图的信息进行融合,并进行超分辨率重建。

2021 年, Dai 等^[26]利用深度卷积神经网络实现双目图像超分辨和视差估计的联合优化。该工作指出,双目图像超分辨和双目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,而高精度的视差又可以引导左右图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,他们提出了双目视差估计与双目图像超分辨互相促进的反馈网络 SSRDE-FNet。该网络首先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,而后基于高分辨率双目图像特征构建并聚合匹配代价(cost volume),实现像素级视差估计。之后,该方法将估计出来的视差图反馈给超分辨重建网络,引导左右图的准确关联,进一步提升图像超分辨的性能。该方法采用递归(recurrent)结构实现左右图与视差图的迭代优化,达到了目前领域内最好的超分辨重建性能。

基于视差估计的双目图像超分辨方法首先计算左右图中每一个空间位置的视差,而后在视差图的引导下对左右图进行像素级的配准。相比于平移副本堆叠的方式,该类方法可以更加高效地利用左右图的互补信息,从而取得更为优越的超分辨重建性能。然而,双目图像的视差估计同样是一个具有挑战性的任务,实际场景的复杂结构经常使得算法在处理遮挡区域以及深度跳变区域时产生较大误差,进而无法在这些区域可靠地匹配左右图的信息,导致超分辨重建结果存在失真(artifact)。此外,此类方法将视差估计作为一个独立的计算过程,增加了计算开销。文献[26]进一步采用了递归网络进行迭代优化,虽然取得了优越的重建性能,但是使得算法难以满足诸多实时性要求较高的场景(例如自动驾驶、机器人导航)。

3.3 基于视差注意力机制的方法

2019年,Wang等^[28]将双目视觉中的对极几何约束引入到注意力机制中,提出视差注意力机制(Parallax Attention)用于计算图像中不同视差位置间的相似性,并基于计算得到的相似性测度实现对左右图特征的融合。Wang等^[27]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搜索范围,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取得了当时领先的超分辨重建性能。

基于Wang等^[28]的工作,领域在近两年来涌现出了许多基于视差注意力机制的双目图像超分辨方法。2020年,Song等^[29]将视差注意力机制与自注意力机制(Self Attention)相结合,在建立左右图关联的基础上加强对单个视角内非局部上下文信息(non-local context information)的利用。同时作者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。基于该先验,作者设计了双目一致性损失函数以约束超分辨重建后的左右图的一致性。该算法实现了较PASSRnet算法更加优越的超分辨性能。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间的互补信息与单个视角内的上下文信息。然而,相比于单图数据集而言,现有双目图像数据集中图像的数量较少、质量较低,使得双目图像超分辨算法难以学到丰富的图像结构与纹理表征,而纹理丰富的单图数据集无法直接用于训练双目图像超分辨算法。为解决此问题,Ying等提出了一个通用的双目注意力模块(Stereo Attention Module, SAM)用于交互左右图的信息。该方法将SAM安插至预训练好的单图超分辨网络中,将其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(finetune),实现在结合左右图互补信息的同时保持对单图信息的充分利用。论文验证了SAM可以应用于多个单图超分辨网络并实现一致性的性能提升,且与单图超分辨网络SRResNet结合后的网络可以

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, 不能适应图像的空变特性。基于此, Xu 等将双边滤波 (bilateral filtering) 的思想引入双目图像超分辨, 首先利用视差注意力机制生成空变的动态双边网格 (content-aware bilateral grid), 而后利用双边滤波在高分辨率空间对左右图的特征进行融合与重建。BSSRnet 取得了相比于 StereoSR^[22]、PASSRnet^[28]和 SRRes+SAM^[30]更为优越的超分辨性能。同年, Wang 等^[32]出了 iPASSR 网络, 并在交互模块、网络结构以及损失函数三个方面对基于视差注意力机制的双目图像超分辨进行改进。iPASSR 采用孪生网络 (Siamese Network) 对左右图同时进行超分辨, 并提出双向视差注意力模块 (BiPAM) 同时交互左右图信息。此外, 该方法提出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, 较好地解决了真实场景中左右图照度不一致的问题, 更加充分地利用了双目图像之间的对称性先验约束。iPASSR 实现了当前领域先进的超分辨重建性能。

4 数据集

在双目视觉领域, 研究者建立了多个双目图像数据集用于立体匹配、双目测距、光流估计、三维检测等任务。本节简要介绍领域内的双目图像数据集。

KITTI 数据集是领域内首个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^[33]和 KITTI 2015^[21]两个部分, 分别于 2012 年和 2015 年建立。数据集中的双目图像由车载双目相机获取。同时, 该数据集通过车载激光雷达获取场景点云并重投影至图像上得到深度分布 (depth map), 最终结合相机参数 (基线、焦距等) 转化为视差真值 (groundtruth disparity), 用于对视差估计算法进行训练与评测。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 46 万像素。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 47 万像素。



图4 领域内双目图像数据集场景示例^[34]

Fig. 4 Sample images of different stereo image datasets^[34]

Middlebury 双目数据集提供了领域内首个双目视差估计的基准评测平台²。该数据集共包含 5 个子数据集，分别于 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]建立。Middlebury 数据集主要面向室内场景双目视差估计，其中场景的深度分布通过结构光传感器获取并转换为视差真值图，其标注误差不超过 0.2 像素。Middlebury 数据集图像的平均分辨率为 600 万像素，视差分布范围在 200 至 800 像素。

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，其中用于双目视差估计的子数据集包含 47 个灰度图场景（27 个场景用于训练，20 个场景用于测试），平均分辨率为 30 万像素。该数据集的建立者通过 Faro Focus X 330 激光扫描仪获取场景的点云分布，并转换为深度及视差真值图，用于对视差估计算法进行训练与评测。

KITTI、Middlebury 和 ETH3D 是双目深度估计领域广泛采用的三个数据集，并被鲁棒视觉挑战赛³用作基准评测数据集。然而，由于上述三个数据集主要面向视差估计与光流计算，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，研究者们建立了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达到了很高的重建性能。

为满足领域内双目图像超分辨算法的训练以及评测需求，2019 年，Wang 等^[34]从 Flickr

² 相关链接: <https://vision.middlebury.edu/stereo/eval3/>

³ 相关链接: <http://www.robustvision.net/>

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 73 万像素。相比于 KITTI、Middlebury 和 ETH3D 数据集，Flickr1024 数据集的场景更加多样，涵盖了日常生活与摄影中的各类情形（如动物、植物、建筑、室内、微距等）⁴。

5 算法性能评测

在本小节中，首先对领域若干主流的双目图像超分辨算法的性能进行评测，并与领域内先进的单图超分辨算法进行比较，而后探究了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]、SRRes+SAM^[30]、BSSRnet^[31]、iPASSR^[32]、SSRDE-FNet^[26]共 6 个双目图像超分辨算法以及 VDSR^[43]、EDSR^[44]、RDN^[45]、RCAN^[46]共 4 个单图超分辨算法。值得一提的是，EDSR、RDN 和 RCAN 可以取得领域先进的单图超分辨重建精度。同时，双三次插值的结果作为基准结果一并给出。

5.1 实验设置与实施细节

由于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]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 2 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，直接采用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通过双三次下采样获得。在训练过程中采用随机水平翻转、随机垂直翻转进行数据增强。

本文采用的测试集由四部分组成，分别为 Middlebury 数据集的 5 幅图片、KITTI 2012 数据集的 20 幅图片、KITTI 2015 数据集的 20 幅图片以及 Flickr024 测试集的 112 幅图片。采用峰值信噪比 PSNR 和结构相似度 SSIM 作为数值评测指标。首先分别计算每个场景左图和右图的数值指标，而后对左右图指标进行平均得到该场景的数值指标，最后对数据集下的所有场景的指标取平均得到该数据集的指标。

⁴ Flickr1024 数据集的访问链接为: <https://yingqianwang.github.io/Flickr1024>

5.2 数值结果

不同超分辨率算法取得的数值结果如表 1 所示。由表可见，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。值得一提的是，单图超分辨率算法 EDSR、RDN 和 RCAN 虽然没有使用左右图的互补信息，却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率网络。其主要原因是，EDSR、RDN 和 RCAN 采用较大的模型体量（较高的参数量）加强了对于单幅图像内邻域信息的建模能力，能够更好地学到训练集图像的纹理结构先验。相比之下，双目图像超分辨率网络（例如 iPASSR、SSRDE-FNet）通过合理利用左右图的互补信息实现了轻量化条件下较好的重建性能。

表1 不同算法取得的数值结果比较。#Params表示模型参数量。最优结果通过**加粗**表示，次优结果通过下划线表示。

Table 1 Results achieved by different methods. #Params represents the model size. The best results are shown in **bold** and the second best results are underlined.

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

5.3 视觉效果

图 5 和图 6 分别展示了不同算法 $2\times$ 和 $4\times$ 超分辨下的视觉效果。由图可见，由于单图超分辨算法仅利用单个视角的邻域信息，在重建细节纹理时存在不同程度的模糊与失真。相比之下，双目图像超分辨算法可以利用左右图之间的互补信息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，SSRDE-FNet 和 iPASSR 算法在 2 倍超分辨下相比于其他算法有较为显著的优越性。在 4 倍超分辨下，BSSRnet 和 SSRDE-FNet 取得了较为优越的视觉效果。

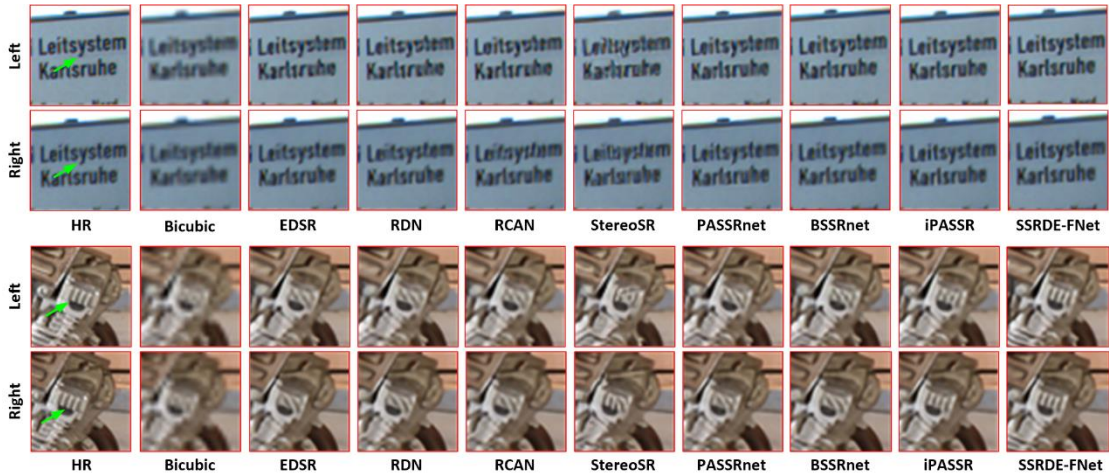


图5 不同算法在 $2\times$ 超分辨下的视觉效果比较

Fig. 5 Visual comparison of different methods for $2\times$ SR



图6 不同算法在 $4\times$ 超分辨下的视觉效果比较

Fig. 6 Visual comparison of different methods for $4\times$ SR

5. 4 不同数据集对双目图像超分辨算法性能影响

在深度学习时代，数据集对算法的性能具有重要影响。在超分辨领域，高质量的大规模数据集可以显著提升超分辨算法的重建精度，降低算法过拟合的风险。本小节探究了不同数

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^[34]三个数据集上进行训练。对于 KITTI 2015 以及 Flickr1024 数据集，本实验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。在第一个实验中，将上采样系数设为 4 并训练 60 轮（epoch），将得到的模型在 KITTI 2015、Middlebury、Flickr1024 以及 ETH3D 数据集的测试集上进行测试。表 2 和表 3 分别展示了 StereoSR 和 PASSRnet 算法的数值结果。由表可见，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。

表2 StereoSR算法^[22]在不同训练集上训练60轮所取得的PSNR与SSIM指标(4倍上采样)

| Table 2 PSNR/SSIM values achieved by StereoSR ^[22] trained on different datasets for 60 epochs under $4\times$ SR. | | | | |
|---|--------------------|--------------------|--------------------|--------------------|
| Dataset | KITTI 2015 (Test) | Middlebury (Test) | Flickr1024 (Test) | ETH3D (Test) |
| KITTI 2015 (Train) | 24.28/0.741 | 26.27/0.749 | 21.77/0.617 | 29.63/0.831 |
| Middlebury (Train) | 23.64/0.743 | 26.62/0.773 | 21.64/0.646 | 28.66/0.843 |
| Flickr1024 (Train) | 25.08/0.779 | 27.85/0.807 | 22.64/0.692 | 30.55/0.860 |

表3 PASSRnet算法^[28]在不同训练集上训练60轮所取得的PSNR与SSIM指标(4倍上采样)

| Table 3 PSNR/SSIM values achieved by PASSRnet ^[28] trained on different datasets for 60 epochs under $4\times$ SR. | | | | |
|---|--------------------|--------------------|--------------------|--------------------|
| Dataset | KITTI 2015 (Test) | Middlebury (Test) | Flickr1024 (Test) | ETH3D (Test) |
| KITTI 2015 (Train) | 23.13/0.703 | 25.42/0.762 | 21.31/0.600 | 26.95/0.789 |
| Middlebury (Train) | 25.18/0.774 | 28.08/0.853 | 22.54/0.676 | 31.39/0.864 |
| Flickr1024 (Train) | 25.62/0.791 | 28.69/0.873 | 23.25/0.718 | 31.94/0.877 |

进一步，我们将 PASSRnet 算法训练至不同的阶段，并探究其重建精度（PSNR）随训练轮数的变化趋势。由图 7 可见，采用 Flickr1024 数据集将 PASSRnet 算法训练任意轮数时的重建精度均优于采用 KITTI 2015 和 Middlebury 数据集训练相同轮数时的精度。注意到，当采用 KITTI 2015 数据集训练 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显示出下降的趋势。这说明 PASSRnet 算法在 KITTI 2015 数据集上训练时发生了过拟合。其原因为 KITTI 2015 数据集是面向自动驾驶的数据集，该数据集采用单一类型的双目相机记录驾驶员视角的道路场景。因此单独使用 KITTI 2015 数据集训练超分辨率算法使得算法难以学到通用的纹理表征，从而无法取得较好的性能。相比之下，在 Flickr1024 数据集上训练的模型未发生过拟合现象，其重建精度随着训练轮数的增加单调上升。以上实验表明 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。

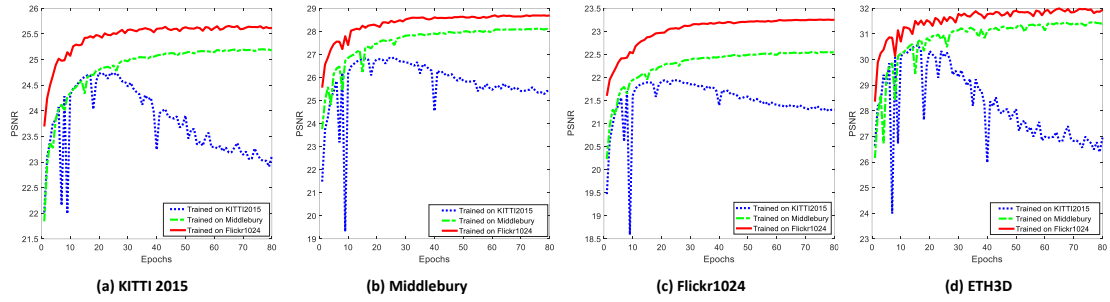


图7 PASSRnet算法^[28]训练不同轮数所取得的PSNR指标(4倍上采样)。性能分别在(a) KITTI 2015、(b) Middlebury、(c) Flickr1024和(d) ETH3D数据集上进行评估。

Fig. 7 PSNR results achieved by PASSRnet^[28] with different numbers of training epochs for $4\times$ SR. Note that, the performance is evaluated on the test sets of (a) KITTI 2015, (b) Middlebury, (c) Flickr1024, and (d) ETH3D datasets, respectively.

6 挑战与展望

尽管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也得到持续提升，该领域仍然有许多现存的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展望。

6.1 双目图像超分辨面临的挑战

(1) 充分利用左右图互补信息与单图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了左右图互补信息的前提下在性能方面仍然逊于先进的单图超分辨网络（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网络有明显差距。因此，如何使双目图像超分辨网络在利用左右图互补信息的同时能够充分利用单图上下文先验并进一步提升重建精度具有一定的挑战。

(2) 左右图信息交互与计算开销之间的权衡

当前领域内的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图信息交互机制。基于视差估计的方法需要设计独立的模块计算左右图的匹配代价并通过多个卷积层对匹配代价进行聚合，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图特征。然而，当在网络中加入多个视差注意力模块时会产生较大的显存占用，使得算法难以部署至各类移动终端。综上，如何在实现左右图信息

充分交互的同时降低算法的计算开销具有一定的挑战。

(3) 无监督范式下可靠的左右图关联

由于含有真值视差标注的双目图像难以获取,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图的关联,即利用左右图对应区域的照度一致性以及双目图像的循环对称性,通过设计相应的损失函数对网络进行训练。然而,现有无监督双目视差估计方法在性能方面仍然逊于有监督的方法。因此,如何在没有真值视差标注的双目图像数据集上训练网络,并让网络可靠地关联左右图的信息具有挑战性。

(4) 高质量双目图像数据集

相比于单图超分辨数据集(例如 DIV2K^[41]、Flickr2K^[42]),现有的双目图像超分辨数据集在图像数量、图像质量、场景多样性方面处于劣势,因而限制了在其上训练的超分辨算法的性能。然而,由于领域当前大部分双目相机的成像质量仍低于单反相机,因此获取大量高质量且场景多样的双目图像数据供双目图像超分辨算法训练具有挑战性。

6.2 未来的研究方向

(1) 新的双目图像信息交互机制

领域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。受启发于相关领域的最新工作,未来的研究可以考虑结合双目图像的极线约束探索并设计新的双目图像信息交互机制,实现左右图对应位置的关联与信息的融合。

(2) 不同双目图像间的域差异 (domain gap)

域差异是双目视差估计中的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(见第4节),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,却难以泛化至更多类型的双目图像。虽然近年来领域中有多个工作^[58-62]着重解决双目视差估计的域代沟问题并取得了一定的进展,双目视差估计算法在不同类型数据上的鲁棒性仍然有较大的提升空间。当前,领域内还没有针对双目图像超分辨域代沟方面的研究。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，能否通过双目图像超分辨率来解决下游任务（如视差估计）的域差异问题也值得进一步探索。

（3）面向未知降质的盲超分辨（blind super-resolution）

现有的双目图像超分辨率方法采用双三次下采样的方式生成低分辨率图像，并用生成的高低分辨率双目图像对网络进行训练。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，超分辨率方法的性能会显著下降。由于真实图像的降质过程多样，现有的双目图像超分辨率算法无法直接应用于真实图像。在单图超分辨率领域，学者们提出盲超分辨率算法^[66-71]用于解决未知降质下的图像超分辨率问题。相比于双三次下采样下训练得到的超分辨率算法，盲超分辨率算法显著提升了在真实图像上的视觉效果。当前，面向真实场景的双目图像超分辨率具有较大的需求。因此，如何将现有单图盲超分辨率算法与双目极线约束相结合，解决未知降质下的双目图像盲超分辨率问题，具有较大的研究意义与光明的发展前景。

（4）面向下游任务（downstream task）的双目图像超分辨

当前双目图像超分辨率的工作大多追求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的潜在增益。虽然 Wang 等^[32]初步验证了超分辨率对双目视差估计的增益，却并未面向视差估计任务进行端到端训练。相比于单图超分辨率被广泛应用于目标检测^[72-73]、分割^[74-76]、人脸识别^[77-78]等应用，当前领域仍缺乏面向下游任务的双目图像超分辨率工作。鉴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]等诸多下游任务具有显著增益，面向下游任务的双目图像超分辨率可以作为下一步研究的方向。

（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率算法均假定两相机参数一致且按平行光轴方式进行装配。在智能手机以及机器人领域，双目相机也常常采用更加灵活的方式进行装配部署，例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。未来的工作可以贴合实际的应用需求，探究适合于新型双目成像系统的图像超分辨率算法。

7 结束语

本文对双目图像超分辨率研究进行了综述，介绍了双目相机成像的基本原理，回顾了双目

图像超分辨领域的算法与数据集，并在公开数据集上对主流超分辨算法的性能进行了评测。最后，本文指出了领域中存在的挑战，并展望了未来的研究方向。

参考文献

- [1] Dong X, Sun X, Jia X, *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, Wang L, 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, Zou H, Wang Y, *et al.* ShipSRDet: An End-to-End Remote Sensing Ship Detector Using Super-Resolved Feature Representation[J]. arXiv preprint arXiv:2103.09699, 2021.
- [4] Huang Y, Shao L, Frangi A F. Simultaneous Super-Resolution and Cross-Modality Synthesis of 3D Medical Images Using Weakly-Supervised Joint Convolutional Sparse Coding[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 6070-6079.
- [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, *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, Wang Z. 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.* Fast Vascular Ultrasound Imaging with Enhanced Spatial Resolution and Background Rejection[J]. IEEE Transactions on Medical Imaging, 2016, 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, Cao J, 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, 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] Guo Z. Research of Hand Positioning and Gesture Recognition Based on Binocular Vision[C]//2011 IEEE International Symposium on VR Innovation. IEEE, 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, Chen J, Zhang K. Drivable Road Reconstruction for Intelligent Vehicles Based on Two-View Geometry[J]. IEEE Transactions on Industrial Electronics, 2016, 64(5): 3696-3706.
- [15] Gong W, Qin L, Ren G, *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] Chen M, Gai S, Da F. 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, 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.
- [18] 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.
- [19] Tang Y, Pan H, Zhu Y, *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, Chen R, Li C, *et al.* Survey of Single Image Super-Resolution Reconstruction with Deep Learning[J]. Journal of Xiamen University(Natural Science), 2021, 60(03): 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(03): 555-570+641-642.
- [21] Menze M, Geiger A. Object Scene Flow for Autonomous Vehicles[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 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]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 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, Bare B, *et al.* Disparity-Aware Domain Adaptation in Stereo Image Restoration[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2020: 13179-13187.
- [25] Khamis S, Fanello S, Rhemann C, *et al.* Stereonet: Guided Hierarchical Refinement for Real-time Edge-Aware Depth Prediction[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 573-590.
- [26] Dai Q, Li J, Yi Q, *et al.* Feedback Network for Mutually Boosted Stereo Image Super-Resolution and Disparity Estimation, arXiv preprint arXiv:2106.00985, 2021.
- [27] Wang L, Guo Y, Wang Y, *et al.* Parallax Attention for Unsupervised Stereo Correspondence Learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020.
- [28] Wang L, Wang Y, Liang Z, *et al.* Learning Parallax Attention for Stereo Image Super-Resolution[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 12250-12259.
- [29] Song W, Choi S, Jeong S, *et al.* Stereoscopic Image Super-Resolution with Stereo Consistent Feature[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020, 34(07):

12031-12038.

- [30] Ying X, Wang Y, Wang L, *et al.* A stereo Attention Module for Stereo Image Super-Resolution[J]. IEEE Signal Processing Letters, 2020, 27: 496-500.
- [31] Xu Q, Wang L, Wang Y, *et al.* Deep Bilateral Learning for Stereo Image Super-Resolution[J]. IEEE Signal Processing Letters, 2021, 28: 613-617.
- [32] Wang Y, Ying X, Wang L, *et al.* Symmetric Parallax Attention for Stereo Image Super-Resolution[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2021.
- [33] 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. IEEE, 2012: 3354-3361.
- [34] Wang Y, Wang L, Yang J, *et al.* Flickr1024: A Large-Scale Dataset for Stereo Image Super-Resolution[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. 2019: 0-0.
- [35] Scharstein D, Szeliski R. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms[J]. International Journal of Computer Vision, 2002, 47(1): 7-42.
- [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. IEEE, 2003, 1: I-I.
- [37] Scharstein D, Pal C. Learning Conditional Random Fields for Stereo[C]//2007 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 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. IEEE, 2007: 1-8.
- [39] Scharstein D, Hirschmüller H, Kitajima Y, *et al.* High-Resolution Stereo Datasets with Subpixel-Accurate Ground Truth[C]//German Conference on Pattern Recognition. Springer, Cham, 2014: 31-42.
- [40] Schops T, Schonberger J L, Galliani S, *et al.* A Multi-View Stereo Benchmark with High-Resolution Images and Multi-Camera Videos[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 3260-3269.
- [41] Agustsson E, Timofte R. Ntire 2017 Challenge on Single Image Super-Resolution: Dataset and Study[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2017: 126-135.
- [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. 2017: 114-125.
- [43] 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.
- [44] Lim B, Son S, Kim H, *et al.* Enhanced Deep Residual Networks for Single Image Super-Resolution[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2017: 136-144.
- [45] Zhang Y, Tian Y, Kong Y, *et al.* Residual Dense Network for Image Super-Resolution[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 2472-2481.

- [46] 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.
- [47] Liu Z, Lin Y, Cao Y, *et al.* Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows[J]. arXiv preprint arXiv:2103.14030, 2021.
- [48] Dai J, Qi H, Xiong Y, *et al.* Deformable Convolutional Networks[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 764-773.
- [49] 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.
- [50] Chen H, Wang Y, Guo T, *et al.* Pre-Trained Image Processing Transformer[J]. arXiv preprint arXiv:2012.00364, 2020.
- [51] Wang Y, Yang J, Wang L, *et al.* Light Field Image Super-Resolution Using Deformable Convolution[J]. IEEE Transactions on Image Processing, 2020, 30: 1057-1071.
- [52] Yang F, Yang H, Fu J, *et al.* Learning Texture Transformer Network for Image Super-Resolution[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 5791-5800.
- [53] 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.
- [54] Ying X, Wang L, Wang Y, *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[J]. arXiv preprint arXiv:2009.07265, 2020, 4: 3.
- [56] Chan K C K, Wang X, Yu K, *et al.* BasicVSR: The Search for Essential Components in Video Super-Resolution and Beyond[J]. arXiv preprint arXiv:2012.02181, 2020.
- [57] 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.
- [58] Pang J, Sun W, Yang C, *et al.* Zoom and Learn: Generalizing Deep Stereo Matching to Novel Domains[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 2070-2079.
- [59] Guo X, Li H, Yi S, *et al.* Learning Monocular Depth by Distilling Cross-Domain Stereo Networks[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 484-500.
- [60] Tonioni A, Tosi F, Poggi M, *et al.* Real-time Self-Adaptive Deep Stereo[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 195-204.
- [61] Liu R, Yang C, Sun W, *et al.* Stereogan: Bridging Synthetic-to-Real Domain Gap by Joint Optimization of Domain Translation and Stereo Matching[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 12757-12766.
- [62] Song X, Yang G, Zhu X, *et al.* AdaStereo: A Simple and Efficient Approach for Adaptive Stereo Matching[J]. arXiv preprint arXiv:2004.04627, 2020.
- [63] Efzrat N, Glasner D, Apartsin A, *et al.* Accurate Blur Models vs. Image Priors in Single Image Super-Resolution[C]//Proceedings of the IEEE International Conference on Computer Vision.

2013: 2832-2839.

- [64] Yang C Y, Ma C, Yang M H. Single-image Super-Resolution: A Benchmark[C]//European Conference on Computer Vision. Springer, Cham, 2014: 372-386.
- [65] Gu J, Lu H, Zuo W, *et al.* Blind Super-Resolution with Iterative Kernel Correction[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 1604-1613.
- [66] Wang L, Wang Y, Dong X, *et al.* Unsupervised Degradation Representation Learning for Blind Super-Resolution[J]. arXiv preprint arXiv:2104.00416, 2021.
- [67] Gu J, Lu H, Zuo W, *et al.* Blind Super-Resolution with Iterative Kernel Correction[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 1604-1613.
- [68] Luo Z, Huang Y, Li S, *et al.* Unfolding the Alternating Optimization for Blind Super Resolution[J]. arXiv preprint arXiv:2010.02631, 2020.
- [69] Michaeli T, Irani M. Nonparametric Blind Super-Resolution[C]//Proceedings of the IEEE International Conference on Computer Vision. 2013: 945-952.
- [70] Shocher A, Cohen N, Irani M. “Zero-shot” Super-Resolution Using Deep Internal Learning[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 3118-3126.
- [71] Yuan Y, Liu S, Zhang J, *et al.* Unsupervised Image Super-Resolution Using Cycle-in-Cycle Generative Adversarial Networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2018: 701-710.
- [72] Li Z, Zhang Y. Hyperspectral Anomaly Detection via Image Super-Resolution Processing and Spatial Correlation[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020.
- [73] Ji H, Gao Z, Mei T, *et al.* Vehicle Detection in Remote Sensing Images Leveraging on Simultaneous Super-Resolution[J]. IEEE Geoscience and Remote Sensing Letters, 2019, 17(4): 676-680.
- [74] Wang L, Li D, Zhu Y, *et al.* Dual Super-Resolution Learning for Semantic Segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 3774-3783.
- [75] Zhang Z, Zhang K. 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). IEEE, 2020: 8411-8417.
- [76] Lei S, Shi Z, Wu X, *et al.* Simultaneous Super-Resolution and Segmentation for Remote Sensing Images[C]//IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2019: 3121-3124.
- [77] Chen Y, Tai Y, Liu X, *et al.* Fsrnet: End-to-End Learning Face Super-Resolution with Facial Priors[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 2492-2501.
- [78] Chen C, Gong D, Wang H, *et al.* Learning Spatial Attention for Face Super-Resolution[J]. IEEE Transactions on Image Processing, 2020, 30: 1219-1231.
- [79] Xu B, Xu Y, Yang X, *et al.* Bilateral Grid Learning for Stereo Matching Network[J]. arXiv preprint arXiv:2101.01601, 2021.
- [80] Liang Z, Guo Y, Feng Y, *et al.* Stereo Matching Using Multi-Level Cost Volume and Multi-Scale Feature Constancy[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,

2019.

- [81] Liang Z, Feng Y, Guo Y, *et al.* Learning for Disparity Estimation through Feature Constancy[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 2811-2820.
- [82] Li X, Huang H, Zhao H, *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, Liu S, Shen X, *et al.* Dsgn: Deep Stereo Geometry Network for 3d Object Detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 12536-12545.
- [84] Qin Z, Wang J, Lu Y. Triangulation Learning Network: from Monocular to Stereo 3d Object Detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 7615-7623.
- [85] Li P, Chen X, Shen S. Stereo R-CNN Based 3d Object Detection for Autonomous Driving[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 7644-7652.