中图法分类号:TP391　　**文献标识码**:A　　**文章编号**:1006-8961(2024)06-1667-18

论文引用格式:Wang W J, Yang W H, Fang Y M, Huang H and Liu J Y. 2024. Visual perception and understanding in degraded scenarios. Journal of Image and Graphics, 29(06):1667-1684(汪文靖, 杨文瀚, 方玉明, 黄华, 刘家瑛. 2024. 恶劣场景下视觉感知与理解综述. 中国图象图形学报, 29(06):1667-1684)[DOI:10.11834/jig.240041]

恶劣场景下视觉感知与理解综述

汪文靖[1],杨文瀚[2],方玉明[3],黄华[4],刘家瑛[1*]

1. 北京大学王选计算机研究所,北京　100871; 2. 鹏城实验室战略与交叉前沿研究部,深圳　518055;
3. 江西财经大学信息管理学院,南昌　330032; 4. 北京师范大学人工智能学院,北京　100875

摘　要:恶劣场景下采集的图像与视频数据存在复杂的视觉降质,一方面降低视觉呈现与感知体验,另一方面也为视觉分析理解带来了很大困难。为此,系统地分析了国际国内近年恶劣场景下视觉感知与理解领域的重要研究进展,包括图像视频与降质建模、恶劣场景视觉增强、恶劣场景下视觉分析理解等技术。其中,视觉数据与降质建模部分探讨了不同降质场景下的图像视频与降质过程建模方法,涵盖噪声建模、降采样建模、光照建模和雨雾建模。传统恶劣场景视觉增强部分探讨了早期非深度学习的视觉增强算法,包括直方图均衡化、视网膜大脑皮层理论和滤波方法等。基于深度学习模型的恶劣场景视觉增强部分则以模型架构创新的角度进行梳理,探讨了卷积神经网络、Transformer模型和扩散模型等架构。不同于传统视觉增强的目标为全面提升人眼对图像视频的视觉感知效果,新一代视觉增强及分析方法考虑降质场景下机器视觉对图像视频的理解性能。恶劣场景下视觉理解技术部分探讨了恶劣场景下视觉理解数据集和基于深度学习模型的恶劣场景视觉理解,以及恶劣场景下视觉增强与理解协同计算。论文详细综述了上述研究的挑战性,梳理了国内外技术发展脉络和前沿动态。最后,根据上述分析展望了恶劣场景下视觉感知与理解的发展方向。

关键词:恶劣场景;视觉感知;视觉理解;图像视频增强;图像视频处理;深度学习

Visual perception and understanding in degraded scenarios

Wang Wenjing[1], Yang Wenhan[2], Fang Yuming[3], Huang Hua[4], Liu Jiaying[1*]

1. Wangxuan Institute of Computer Technology, Peking University, Beijing　100871, China;
2. Department of Strategic and Advanced Interdisciplinary, PengCheng Laboratory, Shenzhen　518055, China;
3. School of Information Management, Jiangxi University of Finance and Economics, Nanchang　330032, China;
4. School of Artificial Intelligence, Beijing Normal University, Beijing　100875, China

Abstract: Visual media such as images and videos are crucial means for humans to acquire, express, and convey information. The widespread application of foundational technologies like artificial intelligence and big data has facilitated the gradual integration of systems for the perception and understanding of images and videos into all aspects of production and daily life. However, the emergence of massive applications also brings challenges. Specifically, in open environments, various applications generate vast amounts of heterogeneous data, which leads to complex visual degradation in images and videos. For instance, adverse weather conditions like heavy fog can reduce visibility, which results in the loss of details. Data captured in rainy or snowy weather can exhibit deformations in objects or individuals due to raindrops, which result in

收稿日期:2024-01-16;**修回日期**:2024-02-20;**预印本日期**:2024-02-27

*通信作者:刘家瑛　liujiaying@pku.edu.cn

基金项目:国家自然科学基金项目(62332010)

Supported by: National Natural Science Foundation of China (62332010)

structured noise. Low-light conditions can cause severe loss of details and structured information in images. Visual degradation not only diminishes the visual presentation and perceptual experience of images and videos but also significantly affects the usability and effectiveness of existing visual analysis and understanding systems. In today's era of intelligence and information technology, with explosive growth in visual media data, especially in challenging scenarios, visual perception and understanding technologies hold significant scientific significance and practical value. Traditional visual enhancement techniques can be divided into two methods: spatial domain-based and frequency domain-based. Spatial domain methods directly process 2D spatial data, including grayscale transformation, histogram transformation, and spatial domain filtering. Frequency domain methods transform data into the frequency domain through models, like Fourier transform, for processing and then restore it to the spatial domain. The development of computer vision technology has facilitated the emergence of more well-designed and robust visual enhancement algorithms, such as dehazing algorithms based on dark channel priors. Since 2010s, the rapid advancement in artificial intelligence technology has enabled the development of many visual enhancement methods based on deep learning models. These methods not only can reconstruct damaged visual information but also can further improve the visual presentation, which comprehensively enhances the visual perceptual experience of images and videos captured in challenging scenarios. As computer vision technology becomes more widespread, intelligent visual analysis and understanding are penetrating various aspects of society, such as face recognition and autonomous driving. However, visual enhancement in traditional digital image processing frameworks mainly focuses on improving visual effects, which ignores the impact on high-level analysis tasks. This oversight severely reduces the usability and effectiveness of existing visual understanding systems. In recent years, several visual understanding datasets for challenging scenarios have been established, which leads to the development of numerous visual analysis and understanding algorithms for these scenarios. Domain transfer methods from ordinary scenes to challenging scenes are gaining attention in further reducing reliance on datasets. Coordinating and optimizing the relationship between visual perception and visual presentation, which are two different task objectives, are also important research problems in the field of visual computing. To address the development needs of the visual computing field in challenging scenarios, this study extensively reviews the challenges of the aforementioned research, outlines the developmental trends, and explores the cutting-edge dynamics. Specifically, this study reviews the technologies related to visual degradation modeling, visual enhancement, and visual analysis and understanding in challenging scenarios. In the section on visual data and degradation modeling, various methods for modeling image and video degradation processes in different degradation scenarios are discussed. These methods include noise modeling, downsampling modeling, illumination modeling, and rain and fog modeling. For noise modeling, Poissonian-Gaussian noise modeling is the most commonly used. For downsampling modeling, classical methods include bicubic interpolation and blurring kernel. Noise including JPEG compression is also considered. A recent comprehensive model jointly uses blurring, downsampling, and noise. For illumination modeling, the Retinex theory is one of the most widely used. It decomposes images into illumination and reflectance. For rain and fog modeling, images are generally decomposed into rain and background layers. In the traditional visual enhancement section, numerous visual enhancement algorithms have been developed to address the degradation of image and video information in adverse scenarios. Early algorithms often employed simple strategies, such as super-resolution methods primarily based on interpolation techniques. However, these methods are constrained by linear models and struggle to restore high-frequency details. Researchers have proposed more sophisticated algorithms to address the complex degradation issues in adverse scenarios. These algorithms include techniques such as histogram equalization, Retinex theory, and filtering methods. Deep neural networks have shown remarkable performance in various fields such as image classification, object detection, and facial recognition. Simultaneously, in low-level computer vision tasks such as super-resolution, style transfer, color conversion, and texture transfer, they also demonstrate excellent performance. With the continuous evolution of deep neural network frameworks, researchers have proposed diverse visual enhancement methods. The section on visual enhancement based on deep learning models takes an innovative approach to model architecture. It discusses architectures like convolutional neural networks, Transformer models, and diffusion models. Unlike traditional visual enhancement, which aims to comprehensively improve human visual perception of images and videos, the new generation of visual enhancement and analysis methods considers the interpretive performance of machine vision in degraded scenarios. The section on visual understanding

technology in challenging scenarios discusses visual understanding and its corresponding datasets in challenging scenarios based on deep learning models. It also explores the collaborative computation of visual enhancement and understanding in challenging scenarios. Finally, based on the analysis, it provides prospects for the future development of visual perception and understanding in adverse scenarios. When facing complex degradation scenarios, real-world images may be simultaneously influenced by various factors such as heavy rain and fog, dynamic changes in lighting, low-light environments, and image corruption. This condition requires models to handle unknown and diverse image features. The current challenge lies in the fact that most existing models are designed for specific degradation scenarios. This complexity introduces a significant amount of prior knowledge and causes difficulty in adapting to other degradation scenarios. The construction of existing visual understanding models in adverse scenarios relies on downstream task information, including target domain data distribution, degradation priors, and pre-trained models for downstream tasks. This reliance causes difficulty in achieving robustness for arbitrary tasks and analysis models. Moreover, most methods are limited to a specific machine analysis downstream task and cannot generalize to new downstream task scenarios. Finally, in recent years, large models have achieved significant accomplishments in various fields. Currently, many studies have demonstrated unprecedented potential for large models in tasks like enhancing reconstruction and other low-level computational visual tasks. However, the high complexity of large models also presents challenges, including substantial computational resource requirements, long training times, and difficulties in model optimization. At the same time, the generalization capability of models in adverse scenarios is a pressing challenge that requires more comprehensive data construction strategies and more effective model optimization methods. How to improve the performance and reliability of large visual models in visual perception and understanding tasks in adverse scenarios is a key problem that remains unsolved.

Key words: adverse scenarios; visual perception; visual understanding; image and video enhancement; image and video processing; deep learning

0 引　言

图像视频等视觉媒体是人类获取信息、表达信息和传递信息的重要手段。随着人工智能、大数据等基础技术在各个领域的广泛应用，图像视频感知和理解系统正在逐渐融入到生产和生活中的方方面面。但是新兴的海量应用也带来了挑战：在开放环境下，各种应用产生海量异质数据，为图像与视频带来复杂的视觉降质。例如，大雾天气会导致图像视频能见度下降，造成细节丢失；在雨雪天气下拍摄的数据，雨点的遮挡会造成人物或物体形变，或形成结构化噪声；暗光条件会造成图像细节和结构化信息丢失严重。视觉降质不仅降低图像视频的视觉呈现与感知体验，还会严重影响现有视觉分析与理解系统的可用性和有效性。在当今智能化与信息化的时代，图像视频等视觉媒体数据呈现爆炸式增长的背景下，面向恶劣场景的视觉感知与理解技术具有十分重要的科学意义和应用价值。

20世纪50年代，当人们开始利用计算机记录图像信息时，数字图像处理便已诞生。1964年，美国喷气推进实验室（jet propulsion laboratory）对徘徊者7号（ranger 7）发送回的月球图像使用计算机修正多种图像降质因素（O’Handley 和 Green，1972），包括几何校正、灰度变换和去噪等，被视为计算机视觉领域最早的视觉增强应用。1960年—1980年，视觉增强技术进一步应用于医学图像、地球遥感和天文等领域。进入21世纪，图像、视频等视觉媒体在数字信息中占比日益增长，视觉增强处理技术已用于自然科学、人类生活的方方面面。

在早期，视觉增强定义为根据特定目标要求，凸显感兴趣的信息，同时减弱或去除不必要的信息，以增强有用信息的可见性。传统的视觉增强技术可分为基于空间域和基于频率域的两种方法。空间域方法直接处理二维空间数据，包括灰度变换、直方图变换和空间域滤波等；而频率域方法则先将数据通过傅里叶变换等模型转换到频率域进行处理，再将其还原到空间域。随着计算机视觉技术的发展，出现了更多设计精良、效果稳健的视觉增强算法，如He等人（2009）提出的基于暗通道先验的去雾算法。随着人工智能技术的迅速进步，许多基于深度学习的视觉增强方法应运而生，它们不仅可以重建受损的视觉信息，还能进一步提升视觉呈现效果，全面提升在恶劣场景下捕获的图像视频的视觉感知体验。另

一方面，随着计算机视觉技术的普及，智能视觉分析与理解正深入到社会的方方面面，例如人脸识别、自动驾驶等。然而，在传统的数字图像处理框架中，视觉增强通常以提升视觉效果为主要目标，忽视了对高层次分析任务的影响，这严重降低了现有视觉理解系统的可用性和有效性。近年来，建立了一系列针对恶劣场景的视觉理解数据集，促使大量应用于这些场景下的视觉分析理解算法的诞生。为了进一步减少对数据集的依赖，从普通场景到恶劣场景的域迁移方法越来越受到关注。如何协调和优化视觉感知与视觉呈现这两种不同任务目标之间的关系，也是视觉计算领域重要的研究问题。

面向恶劣场景下视觉计算领域的发展需求，本文从视觉感知、视觉理解及其协同计算的关键技术着手进行研究，归纳其发展现状。这将对推动相关技术进步，进而确保视觉计算领域的持续发展起到积极的作用。

1 国际研究现状

1.1 恶劣场景下视觉增强技术

视觉增强的目标是恢复因降质而造成的信息丢失，并全面提升人眼对图像视频的视觉感知效果。对图像视频等视觉数据和各类降质过程进行建模是设计视觉增强算法的基础。早期的视觉增强算法采用人工设计的规则，随着人工智能技术的发展，近年来数据驱动的深度学习算法成为主流。由此出发，本文将从视觉数据与降质建模、传统算法以及深度学习算法3个方向介绍当前视觉增强技术。

1.1.1 视觉数据与降质建模

建立降质模型是构建视觉增强算法的基础之一，本节从噪声、降采样、光照以及雨雾特殊场景4个方面介绍视觉数据与降质模型。

1）噪声建模。噪声是图像视频中不希望出现的随机或失真性质的信号。这种信号可能来源于多种因素，例如摄像机传感器的电子元件、环境光线、电磁干扰或传输过程中的干扰等。常见的噪声分布包括高斯噪声、泊松噪声和椒盐噪声等。泊松—高斯噪声建模（Foi等，2008）是最为常用的噪声建模之一。它包括独立的两个部分，信号依赖的泊松噪声$\boldsymbol{n}_p$和信号独立的高斯噪声$\boldsymbol{n}_g$。即

$$\boldsymbol{y}=\boldsymbol{n}_p+\boldsymbol{n}_g \tag{1}$$

为了更精确地估计噪声，Abdelhamed等人（2019）提出了基于标准化流模型（normalizing flow）的噪声建模，展示了深度学习和数据驱动噪声分布的力量。

2）降采样建模。超分辨率任务认为，低分辨率图像（或视频）$\boldsymbol{y}$是由高分辨率图像$\boldsymbol{x}$经过某种退化作用得到的。双三次退化模型（Timofte等，2017）采用双立方插值生成低分辨率图像，传统降质建模（Liu和Sun，2014；Shocher等，2018）则将退化预设为一个下采样模糊核$\boldsymbol{k}$，具体为

$$\boldsymbol{y}=(\boldsymbol{x}\circledast\boldsymbol{k})\downarrow_s+\boldsymbol{n} \tag{2}$$

式中，$\circledast$表示卷积操作，n表示复杂噪声。该建模假设低分辨率图像首先通过一个高斯核或点扩散函数$\boldsymbol{k}$，获得模糊的图像$\boldsymbol{x}\circledast\boldsymbol{k}$，然后进行尺度系数为$s$的下采样操作以及额外的高斯噪声$n$。双三次退化可以看做是传统降质建模的一种特殊情况。考虑图像压缩损失，降质可进一步建模为

$$\boldsymbol{y}=[(\boldsymbol{x}\circledast\boldsymbol{k})\downarrow_s+\boldsymbol{n}]_{\mathrm{JPEG}} \tag{3}$$

退化模型一般由模糊核和噪声水平等因素刻画，根据是否事先知道这些因素，超分辨率方法可以大致分为非盲和盲超分辨率。

简单的退化模型难以有效覆盖真实图像的各种退化。为了解决这一问题，Zhang等人（2021a）提出一个由随机打乱的模糊、下采样和噪声退化组成的降质建模。其中，模糊由两个具有各向同性和各向异性高斯核的卷积近似；下采样随机选择最近邻插值、双线性插值和双三次插值；噪声则通过不同噪声等级的高斯噪声、不同质量系数的JPEG（joint photographic experts group）压缩，以及通过反向—前向的相机图像信号处理过程和RAW图像噪声模型模拟处理后的相机传感器噪声来生成。

3）光照建模。黑夜的低光照环境是最常见的恶劣场景之一，与之伴随的便是低光照环境下图像偏暗、噪声严重、存在色偏等成像挑战。Retinex模型有效建模了低光照图像、视频的降质，对于场景的纹理细节和光照分布进行解耦。Retinex理论的方法的基础是Land和McCann（1971）提出的视网膜大脑皮层理论（retinex theory），该理论认为人眼感知物体的亮度取决于环境的照明和物体表面对照射光的反射，因此可以根据该理论将图像分解为光照层和反射层。具体而言，对于观察图像$\boldsymbol{I}$，这一理论可以表示为

$$\boldsymbol{I}=\boldsymbol{R}\odot\boldsymbol{L} \tag{4}$$

式中，$\boldsymbol{R}$表示物体表面对照射光的反射，对环境光照独立，$\boldsymbol{L}$表示环境光照，⊙则表示逐元素乘法。

4）雨雾建模。在真实世界中，雨雾表现形式多样。为了应对这一多样性，现有研究提出了各种不同的雨雾建模方法，用于模拟各种类型的雨雾天气现象。考虑雨层和背景层的相互影响，Luo 等人（2015）提出了屏幕混合模型（screen blend model），具体为

$$\boldsymbol{R}=\boldsymbol{B}+\boldsymbol{S}-\boldsymbol{B}\odot\boldsymbol{S} \tag{5}$$

式中，⊙表示逐元素乘法操作，$\boldsymbol{R}$代表降质图像，$\boldsymbol{B}$和$\boldsymbol{S}$则分别代表背景层和雨层。

在强烈降雨的情况下，远处的雨纹和大气中水粒子的积聚形成了雨幕效应。这种雨水积聚产生的视觉效果在图像背景中呈现出类似雾的现象，从而诞生了大雨模型（heavy rain model）（Yang 等，2017）。对于附着在玻璃表面的雨滴，Qian 等人（2018）提出了雨滴遮罩模型，具体为

$$\boldsymbol{R}=(1-\boldsymbol{M})\odot\boldsymbol{B}+\boldsymbol{D} \tag{6}$$

式中，$\boldsymbol{D}$表示雨滴，$\boldsymbol{M}$是二进制掩码。

针对雨线和雨滴共同存在的场景，Quan 等人（2021）提出了混合雨模型，具体为

$$\boldsymbol{R}=(1-\boldsymbol{M})\odot(\boldsymbol{B}+\boldsymbol{S})+\alpha\boldsymbol{D} \tag{7}$$

式中，α表示大气照明系数。针对视频数据，采用类似于图像的建模方式，对每一帧进行雨雾建模。

1.1.2 传统恶劣场景视觉增强

针对恶劣图像视频信息降质场景，出现了大量的视觉增强算法。早期的算法常采用简单策略，如超分辨率方法主要基于插值技术，包括最近邻、双线性和双三次插值等。然而，这些方法受限于线性模型，难以还原高频细节。为了应对恶劣场景下复杂的降质问题，研究人员提出了更精细的算法。

直方图均衡化常用于图像处理中的亮度调整。直方图表示数字图像中像素的每一灰度级别的出现频次，即图像像素的概率分布函数，其描述了图像灰度级别的分布情况。利用直方图可以观察图像的灰度范围、每个灰度级别的频度、整幅图像的平均明暗程度和对比度等图像灰度分布特性。直方图均衡化算法能有效地对图像进行全局调整，改善画面亮度和对比度，然而其对于局部调整能力不佳。为克服这一弊端，Ketcham 等人（1974）引入了自适应的局部调整机制，提出自适应直方图均衡化。进一步提升速度和效率，Pizer 等人（1990）提出对窗口中心的像素均进行转换，降低时间开销，如统计 32 × 32 像素窗口内的直方图，而后将中心 16 × 16 像素按照该直方图进行转换。事实上，自适应直方图均衡化在取全图为窗口、调整窗口内所有像素时退化为全局的直方图均衡化算法。

不同于基于直方图均衡化的算法，基于视网膜大脑皮层理论的方法通过将图像拆解成反射层和光照层，并对光照层进行增亮来实现低光照图像增强。人工设计的基于 Retinex 模型的算法往往需要人为对光照层和反射层分量设计一个分解策略以及增强算法。Jobson 等人（1997）提出一种单尺度的基于 Retinex 模型的增强算法（single-scale retinex）。在原始 Retinex 模型的基础上，该工作认为对原始图像进行中心环绕滤波可以近似表示环境光照$\boldsymbol{L}$。

针对图像去雨问题，最初的技术主要是基于滤波方法。Xu 等人（2012）使用引导滤波来实现去雨效果。这一策略得到进一步增强。Zheng 等人（2013）结合了多种引导滤波技术，而 Kim 等人（2013）则利用雨滴独特的细长椭圆形状的性质。在此基础上，研究者尝试了基于先验知识的方法，利用了图像特性或学习到的知识。这类方法通常基于最大后验概率（maximum a posteriori，MAP）原理（Mu 等，2019），通过优化函数来获取无雨图像（Zhang 等，2017b）。对于视频去雨问题，Garg 和 Nayar（2007）在研究的早期阶段提出了雨滴的模型，并尝试通过调整摄像机设置来去除雨滴。他们的研究为理解雨滴在视频中的形态提供了重要的理论基础。随后的研究进一步提供了更加真实的雨滴建模，并强化了时空关联在视频去雨中的应用。Zhang 等人（2006）采用了 K 均值聚类方法，而 Park 和 Lee（2008）使用了卡尔曼滤波器。这些方法利用了视频帧间的时间连续性和空间相关性，以有效地分离雨滴和背景。Barnum 等人（2010）引入了在频域中处理雨和雪的策略，关注如何从时空和频率角度来分析和去除视频中的降雨和降雪现象。

尽管传统的恶劣场景视觉增强算法取得了不错的效果，但它们依赖手工设计的规则，缺乏对真实和未知降质场景的鲁棒性。

1.1.3 基于深度学习模型的恶劣场景视觉增强

深度神经网络在图像分类、目标检测和人脸识别等领域表现出色，同时在低层次计算机视觉问题

上，如超分辨率、风格迁移、颜色转换和纹理迁移等任务中也展现出优异性能。随着深度神经网络框架的更新换代，研究者们结合卷积神经网络、Transformer结构和扩散模型提出了多样化的视觉增强方法。本节将从模型架构创新的角度介绍基于深度学习模型的恶劣场景视觉增强方法。

1）卷积神经网络。早期的深度学习视觉增强模型采用朴素的架构设计。面向超分辨率任务，SRCNN（super-resolution convolutional neural network）（Dong等，2014）通过构建一个三层卷积神经网络，从双三次降采样的低分辨率图像重建出高分辨率图像，证明了卷积神经网络提取图像局部相关性的能力。面向图像去噪任务，DnCNN（denoising convolutional neural network）（Zhang等，2017a）采用了残差策略，模型预测噪声而不是预测去噪后的图像，从而降低训练难度。低光照增强方面，最早的深度学习模型由Lore等人（2017）提出。其在使用随机伽马变换和高斯噪声合成的成对低光照—正常光照数据上训练，模型则采用深度自解码器（auto-encoder）结构，采用了监督学习的方式完成梯度反串和自编码器的参数更新。雨雾去除方面，Eigen等人（2013）首次使用卷积神经网络（convolutional neural network，CNN）进行图像去雨处理。

随着深度学习技术的发展，更多框架设计相继提出以改进视觉增强效果并适配目标降质场景。例如针对去雨任务，Qian等人（2018）引入注意力模块，使其更适用于处理雨滴。Li等人（2018b）提出了一种包括对称编码器和解码器部分的非局部增强网络，以更准确地提取用于描述雨带的有效特征，然后精细地保存图像细节。Zhang等人（2020）使用条件生成对抗网络。Jiang等人（2020）进一步设计了一个多尺度渐进融合网络，利用相关信息跨尺度的雨层执行单幅图像去噪任务。

视觉增强技术不仅可以处理图像数据，还可以处理视频数据。Chen等人（2018a）将超像素引入了视频去雨卷积网络。面向暗光增强，Triantafyllidou等人（2020）提出了一种低光照视频合成框架SIDGAN（see-in-the-dark GAN），其可以通过具有中间域映射的半监督双CycleGAN生成动态视频数据。为了训练这个框架，Triantafyllidou等人（2020）从Vimeo-90K数据集中收集真实世界的视频，低光照原始视频数据和对应的长曝光图像则来自DRV（dark raw video）数据集。自监督学习技术也应用于合成降质条件下的视频重建（Yang等，2020a），为大规模训练模型提供了有效的技术方案。

2）Transformer模型。Transformer模型是一种基于自注意力机制的神经网络架构，它利用自注意力机制和位置编码来捕捉输入序列的上下文信息。得益于此，Transformer模型在图像增强和恢复领域的各种任务上表现出更为卓越的性能。SwinIR（image restoration using swin transformer）模型（Liang等，2021）利用Transformer模型捕捉图像中的上下文信息，从而更好地在超分辨率任务上恢复图像的细节和纹理。Liang等人（2022）提出了一个轻量级的单图像递归去雨Transformer网络。Dudhane等人（2023）提出使用Transformer结构对多幅低光照图像的输入进行融合及增强。其框架结合了邻域及全局特征，并设计了多尺度分层对齐模块来对多幅图像进行预处理，并且提出了无参考特征增强模块来渐进地融合特征。在完成特征对齐和特征增强后，使用循环采样机制来进一步加强帧间的信息融合，其框架在多帧超分辨率、去噪以及低光照增强领域均获得了性能增益。面对多种降质场景，Zamir等人（2022）提出了一种基于Transformer的架构，对于去雨、去模糊等多种降质场景的恢复重建均有良好效果。

3）扩散模型。扩散模型（diffusion models）基于多步小尺度的加噪去噪过程完成分布间的转换，研究者通常使用扩散模型完成复杂分布（如自然图像分布）与简单分布（如标准高斯分布）之间的转换，从而构建从高斯噪声到自然图像的生成模型。得益于扩散模型展现的强大生成能力，其生成的图像往往符合自然图像分布且具有较高的质量。Kawar等人（2022）提出将扩散模型应用于图像重建领域，使用其先验生成能力来辅助模型产生质量更好的预测结果。谷歌团队将低分辨率图像视为扩散模型的条件，并通过低分辨率条件生成高分辨率图像，提出了基于扩散模型的超分辨率方法SR3（super-resolution via repeated refinement）（Saharia等，2023）。Nguyen等人（2023）则将扩散模型应用于低光照场景下的文字图像增强，借由扩散模型的预测能力和高频信号生成能力还原文字的细节。与领域先进方法相比，该模型有效地提升了图像的质量以及文字可识别度，挖掘了扩散模型在病态逆问题求解方面的潜力。

1.2 恶劣场景下视觉理解技术

不同于传统视觉增强的目标为全面提升人眼对图像视频的视觉感知效果，新一代视觉增强以及分析方法考虑降质场景下机器视觉对图像视频的理解性能。本文将从数据集构建、传统算法以及深度学习算法3个方向介绍当前恶劣场景下的视觉理解技术。

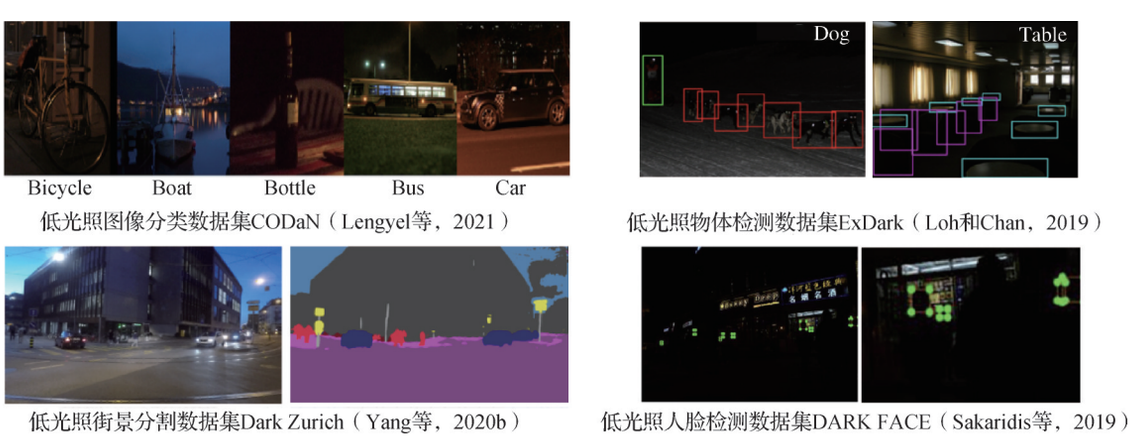
1.2.1 恶劣场景下视觉理解数据集

数据集是开发机器学习算法的基础。降质场景下视觉分析理解数据集的建设带动了一系列恶劣场景视觉任务的发展。其中，低光照降质场景的代表性数据集如图1所示。

针对最基础的图像分类任务，CODaN（common objects day and night）数据集（Lengyel等，2021）包含1.5万幅分辨率为224×224像素的彩色暗光图像，分为10个类，每个类有1 550幅图像。其中，所有图像均具有相同的尺寸，且相互排斥，既不包含不同类的物体对象，也不属于多个类的物体对象。

针对更为复杂的物体检测任务，Loh和Chan（2019）构建了ExDark数据集，并同时使用手工设计和深度学习的特征分析了低光照图像。该数据集包含了7 363幅从极低光环境到黄昏的低光图像，包括12个图像类和局部对象边界框标注的对象类。针对恶劣场景下特定物体的检测任务，在人脸检测方面，Nada等人（2018）构建了基准人脸检测数据集UFDD（unconstrained face detection dataset）-Haze，该数据集包含在不同天气条件下（包括雨、雪、雾等）捕获的真实世界图像。DARK FACE（Yang等，2020b）是一个大规模低光照人脸数据集，包含10 000幅低光照人脸图像，其中6 000幅为标注可获得的公开训练集，每一幅图像标注了人脸的边界框。以该数据集为基础的UG^{2}竞赛诞生了一系列的暗光人脸检测模型。行人检测被用于许多视觉应用，例如视频监控和自动驾驶。恶劣场景下的行人检测任务也备受关注。KAIST（Korea advanced institute of science and technology multispectral pedestrian benchmark）多光谱行人检测数据集（Hwang等，2015）包含9.5万对颜色—热量图，数据从车辆上拍摄，包含夜晚和白天的标签。ECP（eurocity persons）数据集（Braun等，2019）拍摄自欧洲12个国家的31个不同的城市，包括白天和晚上的图像，带标注的边界框总数超过20万。NightOwl数据集（Neumann等，2019）由40个行业标准相机记录的夜间视频序列组成，包括27.9万个视频帧，涵盖了3个国家，包括不同的季节和天气条件。所有的帧都是完全标注的，并标注了附加的对象属性，如遮挡、姿势和难度，以及跟踪信息，以识别跨多个帧的相同对象。NightOwl数据集包含了用于验证检测器鲁棒性的大量背景帧，提供了用于本地超参数调优的验证集，以及用于提交服务器上评估的测试集。

自动驾驶是恶劣场景视觉分析理解最常见且极具挑战性的应用。对于在雨雾、暗光等降质场景下实现高准确率的自动驾驶需求，推动了一系列街景理解尤其是语义分割数据集的建立。早期数据标注较为有限，Mapillary Vistas数据集（Neuhold等，2017）包含白天和夜间的街景数据，但不包含相应的标签。Dai和Van Gool（2018）构建的Nighttime Driving街景分割数据集包含了3.5万幅从正常光照到微光到夜



低光照图像分类数据集CODaN（Lengyel等，2021）　低光照物体检测数据集ExDark（Loh和Chan，2019）

低光照街景分割数据集Dark Zurich（Yang等，2020b）　低光照人脸检测数据集DARK FACE（Sakaridis等，2019）

图1　低光照降质场景代表性视觉分析理解数据集

Fig. 1　Low-light degradation scenarios representative visual understanding datasets

间的街景图像，但仅标注了50幅夜间图像用于通测。Dark Zurich街景分割数据集（Sakaridis等，2019）包含约9 000幅不同时段的街景图像，此外还包含了相机的GPS信息，以粗略地建立不同时段下图像的映射，例如将夜间或者微光图像与相应的白天图像相匹配。Sakaridis等人（2018）提出了Foggy-Cityscapes数据集，旨在研究雾天条件下的场景理解问题。但在该数据集中，雾天图像不是实拍获得而是通过在Cityscapes数据集中的晴朗天气图像上模拟雾效应来生成。类似地，Sindagi等人（2020）在Cityscapes数据集上合成了Rainy- Cityscapes数据集，用来评估雨场景下的目标检测性能。真实雾天任务驱动数据集最广泛使用的是Li等人（2019）提出的RTTS（real-world task-driven testing set）数据集，相较于Foggy-Cityscape具有更大规模。作为RESIDE（realistic single image dehazing）数据集的一个子集，包含5 000幅已标注和未标注的真实世界雾天图像，图像主要涵盖交通和驾驶场景。进一步扩大数据集的规模与包含更多的降质类型，驾驶视频数据集BDD-100K（Berkeley deep drive）（Yu等，2020）包含超过1 100 h 10 000个视频的驾驶视频，涵盖了一天中不同的时间、天气条件和驾驶场景，同时包含了10种任务标注。该数据集同样包含了车辆和其他机动车的标注框。

受限于高昂的标注成本，大多语义分割数据集中含标注的数据量较少。对此，Sakaridis等人（2021）提出了ACDC（adverse conditions dataset with correspondences），涵盖不同的视觉降质条件，可用于训练和测试不利视觉条件下的语义分割方法，如图2所示。ACDC由4 006幅图像组成，这些图像平均分布在4种常见的降质条件下：雾、夜间、雨和雪。每个降质条件图像都带有高质量的细像素级语义标注、正常条件下拍摄的相同场景的对应图像，以及用于区分图像内清晰和不确定语义内容区域的二进制掩码。因此，ACDC既支持标准语义分割，也支持不确定性感知语义分割。此外，专注于夜间语义分割任务，NightCity（Tan等，2021）是截至目前规模最大的有标注夜间街景数据集，其包含了4 297幅包含密集标注的夜间街景图像。

1.2.2 基于深度学习模型的恶劣场景视觉理解

从零开始构筑一个降质数据集费时费力，相对地，正常场景下数据集资源非常丰富。因此，如何将模型从正常场景迁移至恶劣场景受到了大量关注。

以摆脱对标注的依赖为任务目标，无监督域适应（unsupervised domain adaptation）是一个简单直接的解决方案。经典方法包括特征对齐（Ben-David等，2010；Long等，2015，2017；Sun和Saenko，2016）、对抗学习（Tzeng等，2017；Ganin等，2016；Kim等，2019a）、像素迁移（Russo等，2018；Hoffman等，2018；Inoue等，2018）和伪标签（Khodabandeh等，2019；D'Innocente等，2020）等。在目标检测中，模型通常不仅要考虑全局域间隙，还需要考虑局部域间隙

图2 ACDC数据集样例（Sakaridis等，2021）

Fig. 2 Examples of the ACDC dataset (Sakaridis et al., 2021)

(Chen等,2018b;Saito等,2019)。然而,上述方法仅为正常场景下两个域的域迁移任务所设计,没有考虑到视觉信息降质的恶劣场景的特性,因此在正常场景至恶劣场景的域迁移任务上性能有限。

为了摆脱恶劣场景应用环境下,机器分析模型对重新构筑训练数据集的依赖,面向恶劣场景的无监督域迁移技术孕育而生。不同于考虑人眼视觉的视觉增强工作,面向恶劣场景的域迁移工作考虑机器视觉,尤其是基于深度学习的下游模型。早期工作往往忽视了降质条件下的机器视觉,直到最近几年,面向恶劣场景的无监督域迁移技术才渐渐受到关注。

针对RAW图像上的物体检测任务,YOLO-in-the-Dark(Sasagawa和Nagahara,2020)最早提出使用粘合层(glue layer)和对抗式生成模型组合不同域的预训练模型。首先在域A和域B上分别预训练模型,其中域A的模型为从RAW图像预测RGB图像,而域B上的模型在RGB图像上进行物体检测,根据两个模型的特征边界提取一部分模型,然后通过粘合层进行接合。训练过程使用了基于生成模型的知识蒸馏,通过将特征表示作为输入来训练粘合层,不需要借助其他的数据集。

针对夜间街景识别,DANNet(domain adaptation network)(Wu等,2021)采用对抗学习来调整模型,并借助了暗光、微光和正常光照的不同亮度图像。DANNet的任务为将街景分割模型从仅包含白天数据的数据集Cityscapes迁移到同时包含白天、暗光和夜间数据的Dark Zurich数据集,其中Dark Zurich数据集包含了无标签的利用GPS(global positioning system)记录信息大致对齐的不同光照图像对。研究者们提出首先将模型从Cityscapes迁移到Dark Zurich中的白天数据,然后将在Dark Zurich白天数据上的预测值作为Dark Zurich夜间数据的伪标签监督信息,用于训练域迁移模型。此外,DANNet采用了权重共享的语义分割网络结构,还使用了提亮网络、对抗学习和权重重新赋值策略。提亮网络让不同光照的图像趋于一致,让语义分割网络对光照不要过于敏感。判别器通过输出空间的对抗学习,判断语义预测结果是来自源域还是目标域。小物体的标注更少,为了弥补这种偏差,使用了权重重新赋值策略。

在传统域适应方法之外,零样本昼夜域适应方法考虑了一个更严格的条件,即无法获取真实的夜间图像。针对这一更严格的假设,CIConv(color invariant convolution)(Lengyel等,2021)基于Kubelka-Munk理论的图像信息模型,推导出一种用来提取光照不变表征的边缘检测器,将提取光照不变因子的过程嵌入神经网络。一些研究工作也关注了低光照图像搜索(Jenicek和Chum,2019)夜间深度估计(Wang等,2021a)、低光照图像匹配(Song等,2021)等任务。

对于恶劣场景下的视觉理解,一些去雨去雾算法(Yang等,2017)也长期应用于直接优化检测结果。Chen等人(2018b)假设恶劣天气条件导致域的偏移,并通过提出一个域适应的Faster-RCNN方法来克服这一问题。Shan等人(2019)提出了使用Cycle-GAN框架在图像级和特征级进行联合的场景转换,试图将恶劣场景转换为一般场景进行视觉理解。Halder等人(2019)提出了一个基于物理渲染的雨天场景轻量级语义分割框架,提升了雨天场景下的语义分割性能。Qian等人(2021)提出了一个多模态车辆检测网络,利用激光雷达和雷达信号获得建议,然后将这两个传感器的区域特征融合在一起,得到最终的检测结果。

1.3 恶劣场景下视觉增强与理解协同计算

近年来,越来越多的研究开始专注于在恶劣场景下的视觉增强与理解协同计算,即在提高视觉呈现质量的同时提升下游机器分析模型的场景理解性能,使得恶劣场景下的人眼视觉与机器视觉两种需求能够同时被满足。Sun等人(2019)提出了一个联合优化轻量化模型,可以与现有的基于图像的高级感知算法无缝集成。Yu等人(2022)从人类感知和机器分析任务的角度,分析了去雨任务下的对抗攻击,对不同级别的攻击和使用各种损失/目标生成扰动的各种方法进行全面的实证评估。Zhang等人(2022)同时解决了雨场景的语义分割和去雨任务,并且开发了一个语义融合网络和视图融合网络,分别融合语义信息和多视图信息,使得语义分割和去雨可以互相促进。

2 国内研究现状

2.1 恶劣场景下视觉增强技术

国内研究者对于视觉增强亦进行了较为深入的

研究。本节同样从视觉数据与降质建模、传统算法以及深度学习算法的3个方向介绍国内相关的视觉增强技术，并最后进行国内外研究进展的分析比较与总结。

2.1.1 视觉数据与降质建模

为了更好地解决真实场景超分辨率问题，腾讯的Wang等人(2021b)在降质建模中引入了一个高阶退化建模过程，即结合数个重复的降质过程，每个过程都是经典的退化模型，包括模糊、重采样、噪声和JPEG压缩等，还考虑了合成过程中常见的振铃和过度锐利伪影。在暗光的极端场景下，传统手工设计的噪声建模失效。针对这一问题，哈尔滨工业大学的Cao等人(2023)提出了基于学习的传感器噪声建模，基于对典型成像过程中噪声形成的研究，提出了一种新的物理引导的ISO(International Organization for Standardization)相关噪声建模方法，建立了一个基于标准化流模型的框架表征CMOS(Complementary Metal Oxide Semiconductor)相机复杂的噪声特性。

对于雨造成的降质，目前最广泛使用的线性叠加模型由天津大学团队(Ren等，2019a)提出，该模型实现了雨层和背景层的分离，具体为

$$\boldsymbol{R}=\boldsymbol{B}+\boldsymbol{S} \tag{8}$$

式中，$\boldsymbol{R}$代表降质图像，$\boldsymbol{B}$和$\boldsymbol{S}$分别代表背景层和雨层。

香港中文大学的Hu等人(2019)进一步根据雨天视觉效果与场景深度信息的关系，提出了一种深度引导雨模型。相机距离较远的物体的可见性随深度变化，雾比雨更明显地遮挡这些远处的物体，建模为

$$\boldsymbol{R}=(\boldsymbol{1}-\boldsymbol{S}-\boldsymbol{F})\odot\boldsymbol{B}+\boldsymbol{S}+\boldsymbol{F}\odot\boldsymbol{A} \tag{9}$$

式中，$\boldsymbol{F}$代表雾层。

基于对雨纹和蒸汽相互交织的观察，电子科技大学的Wang等人(2020)提出了传输介质模型，通过将雨和雾都视为传输介质来重新模拟雨天成像，具体为

$$\boldsymbol{R}=(\boldsymbol{T}_s+\boldsymbol{T}_v)\odot\boldsymbol{B}+[1-(\boldsymbol{T}_s+\boldsymbol{T}_v)]\odot\boldsymbol{A} \tag{10}$$

式中，$\boldsymbol{T}_s$和$\boldsymbol{T}_v$分别代表雨和雾的传输图。

2.1.2 传统恶劣场景视觉增强

国内学者同样基于Retinex模型针对在恶劣场景下的视觉增强展开了研究。中国科学院信息工程研究所的Guo等人(2017)提出通过在RGB三通道中分别寻找最大值来迭代地估计每一个像素的光照。此外，原始的光照图通过一个结构先验的约束进行精细化，以得到最终的光照图。结合光照估计，进行颜色与光照的分解，并对光照进行固定系数的伽马变换，重建后得到最终的增强结果。厦门大学的Fu等人(2016)提出了一个加权的变分模型来同时估计反射和光照，该方法称为SRIE(simultaneous reflectance and illumination estimation)。Fu等人(2016)认为对数变换无法取得很好的效果，因此结合对数变换的研究采用了一个加权的变分模型来获得更好的表征。这个表征被加入到正则化项中，以获得更好的先验表示。相较于传统的模型，SRIE能够更好地保持反射中的细节，在增强光照的同时还可以一定程度上抑制噪声。在求解中，Fu等人(2016)采用了交替极小化的策略。

传统的去雨方法依赖基于图像统计的手工设计先验知识来恢复雨天图像。台湾师范大学的Kang等人(2012)使用字典学习和稀疏编码将雨天图像分解成不同的组成部分，然后重构干净的图像。研究者探索了从图像分解(Fu等，2011)到复杂的正则化先验(Chen等，2014)的多种策略。Kang等人(2012)利用雨滴的方向梯度直方图(histogram of oriented gradient，HOG)特征，将图像中的雨和非雨部分进行分类。在此之后，台湾国立清华大学的Chen和Hsu(2013)开发了一种广义低秩模型来去除图像中的雨。类似地，华中科技大学的Chang等人(2017)利用低秩图像分解模型从雨天图像中恢复无雨图像。香港中文大学的Zhu等人(2017)提出一个带有3种图像先验的联合优化框架，用于从雨天图像中去除雨条纹。

2.1.3 基于深度学习模型的恶劣场景视觉增强

国内学者同样借助深度学习技术构建恶劣场景视觉增强模型。北京航空航天大学的Lyu等人(2018)提出了一个端到端的多分支提亮网络MBLLEN(multi-branch low-light enhancement network)，该模型通过特征提取模块、增强模块和融合模块提取有效的特征表示，提高了低光照增强的性能。中国科学院信息工程研究所的Ren等人(2019b)设计了一个更复杂的端到端网络，包括用于图像内容增强的编码器—解码器网络和用于图像边缘增强的循环神经网络。与之类似，多曝光融合网络TBEFN(two-branch exposure-fusion network)(Lu

和Zhang，2021）在两个分支中估计一个迁移函数，从中可以得到两个增强结果。最后，采用一种简单的平均机制对两幅图像进行融合，并通过细化单元对结果进一步的细化。此外，金字塔网络、残差网络和拉普拉斯金字塔（Laplacian pyramid）也用于低光照增强。这些方法通过常用的端到端网络结构有效地学习特征表示。

自编码器的网络结构在低光照任务上缺乏可解释性。华中科技大学的Shen等人（2017）利用深度卷积网络代替多尺度Retinex算法中的高斯核，提出了一个低光照增强网络MSR-net（multi-scale retinex convolutional neural network）。Shen等人（2017）认为，经典多尺度视网膜增强方法可以看做数个高斯卷积核与1×1卷积操作的组合，于是提出使用权重可以学习的深度神经网络的卷积操作代替固定参数的高斯卷积核。MSR-net首先对输入图像进行多尺度的对数转换操作，然后经过主体网络，最后进行颜色重建。MSR-net虽然是基于多尺度Retinex算法的增强方法，但是并没有充分分析和利用低光照图像的光学性质。同样基于Retinex理论，北京大学的Wei等人（2018）为了在缺乏Retinex分解真值的条件下训练提亮网络，设计了基于成对图像反射相同、光照不同但仅有一通道的无监督学习策略。此外，Wei等人（2018）拍摄了一组成对的正常光照—低光照真实数据集。这种结合Retinex理论的深度学习框架衍生了一系列模型。为了减少计算量，天津大学的Li等人（2018a）提出了一种用于低光照图像增强的轻量级网络LightenNet，它只有4层，以低光照图像作为输入，然后估计其光照图，并基于Retinex理论，将输入图像除以光照图得到增强图像。Zhang等人（2019）设计了一个包含3个子网络的KinD模型，分别用于层分解、反射重建和亮度调整。之后，作者通过多尺度亮度注意力模块改善了KinD的视觉缺陷，这个改进版本称为KinD++（Zhang等，2021b）。

在去雨方面，中国科学技术大学Fu等人（2017）提出了深度细节网络，其他研究如多流密集网络（Zhang和Patel，2018）则专注于雨滴密度。对雨滴大气过程的理解促进了新方法的发展。北京大学Yang等人（2017）提出用于逐步去雨的循环神经网络（recurrent neural network，RNN）架构。中山大学Li等人（2018b）加入了上下文聚合网络。北京大学Liu等人（2018）设计了J4R-Net模型，专注于视频中动态雨滴模式的识别和去除。哈尔滨工业大学Ren等人（2019a）在渐进式循环去雨网络中，提出了一个简洁的递归展开残差网络。

2.2 恶劣场景下视觉理解技术

国内相关研究者对于基于深度学习模型的恶劣场景视觉理解进行了较为深入的研究。

针对低光照场景下的无监督域适应任务，上海交通大学的Cui等人（2021）提出MAET（multitask auto encoding transformation），将相机拍摄技术与机器学习融合，通过将RAW图像到RGB图像的变换与物体检测预测共同编码来学习特征，使模型能够捕获光照变换中的隐式特征。具体而言，MAET框架假设图像信号处理过程（image signal processing，ISP）中的参数分布已知，依此从正常光照的RGB图像出发，通过反向ISP过程和RAW域的线性亮度降低操作获取低光照RGB图像。模型训练时，MAET采用了多任务训练范式，即在优化物体检测任务的同时，要求模型解耦图像降质中使用的参数，以学习光照相关的信息学习视觉结构。为了避免两个任务特征的过度纠缠，MAET采用了正交切线正则化约束特征，最大化不同任务输出的切线之间的正交性。MAET框架可以广泛地应用于各类目标检测模型上，并取得显著的性能提升。

针对低光照场景下的零样本域迁移任务，北京大学的Luo等人（2023）构建了一个涉及图像和模型两个层面的相似性最小—最大框架。在图像层面上，该方法通过暗化过程生成了一个合成的夜间域，该域与白天域的特征相似性最小，以放大域间隙。在模型层面上，该方法通过最大化两个域中图像的特征相似性，以实现昼夜域适应。

对于雾天条件下的目标检测，台北科技大学的Huang等人（2021）提出了一个双子网网络，包括一个检测子网和一个恢复子网。这个网络可以通过结合可见性增强任务和目标检测任务进行多任务学习训练，从而超越纯目标检测器。温州肯恩大学的Zheng等人（2022）提出了语义引导交互网络，该网络通过三阶段去雨方式考虑了SID和语义分割之间的充分交互，即粗略去雨、语义信息提取、语义引导去雨3阶段，得到了更有效的雨天环境下语义分割模型。

2.3 恶劣场景下视觉感知与理解协同计算

面向视觉感知与理解的协同计算，国内的研究者们也取得了一系列进展。北京大学Guo等人(2020)提出了任务驱动的模式，设计了一个包含语义精细化残差网络和新颖的两阶段语义感知联合训练方法的新型去雨网络，可以联合测试去雨和分割结果。华中科技大学的Li等人(2022)通过一个创新的自上而下和自下而上的统一范式，交替执行去雨和图像分割任务，这种框架允许两个子任务相互协作，从而提升彼此的性能。温州肯恩大学Zheng等人(2022)使用去雨网络结合感知对比学习，同时增强图像的去雨效果和语义分割能力。

暗光增强方面，大连理工大学的Liu等人(2021)提出了一种轻量级且高效的低光照图像增强网络RUAS(retinex-inspired unrolling with architecture search)。该方法以Retinex理论为基础，首先建模了低光照图像的内在欠曝光结构，并将其优化过程展开以构建整体传播结构。随后，RUAS通过神经网络搜索技术搜索低光照图像的先验架构，以减少网络的参数量和计算代价。同样来自大连理工大学的Ma等人(2022)则提出了一种自校正照明(self-calibrated illumination，SCI)学习框架，旨在针对现实世界中复杂、未知的低光照场景，实现快速、灵活且鲁棒的图像增强。SCI框架的核心在于建立一个级联的照明学习过程，采用权重共享机制，有效地处理低光照图像增强任务。考虑到级联模式可能带来的高计算负担，研究者们设计了一个自校正模块，使各阶段的结果收敛到同一状态，以减少计算成本。此外，SCI框架通过定义无监督训练损失，增强了模型适应一般场景的能力。通过全面的实验，研究团队展示了SCI框架对相机操作设置的稳定性和对模型无关的通用性。这些属性在现有的研究中通常缺失，但对于提高技术的适应性和通用性至关重要。SCI框架在低光照人脸检测和夜间语义分割等实际夜间视觉分析场景中具有较高的实用价值，对于在夜间环境下进行高效且准确的视觉分析具有重大意义。

2.4 国内外研究进展比较

长期以来，恶劣场景下的视觉感知与理解一直备受学术界和工业界关注。在国外，研究者早期建立了完善的数字图像处理学科体系，并提出了许多经典的降质建模与视觉增强算法。随着深度学习技术的兴起，国内的研究者在恶劣场景下的视觉增强方面开始贡献更多高质量、高影响力的工作。近年来，针对自动驾驶、智慧城市等具有挑战性的恶劣场景应用，国内外开始关注视觉理解技术。国外研究者基于其技术积累，为学术界提供了许多基础性思路和解决方案，尤其是大规模数据集的构建。国内学者逐渐加大力度，相关研究逐渐成为主流并提出了大量优秀方法，部分国内团队已达到国际领先水平。

3 发展趋势与展望

尽管恶劣场景下视觉感知与理解技术快速发展，在面对复杂降质场景、未知下游任务时，现有算法仍然鲁棒性受限。此外，人工智能大模型为视觉感知与理解技术带来了新的机遇与挑战。本节基于前文对国内外技术发展的梳理，展望恶劣场景下视觉感知与理解技术的未来方向。

1)面向复杂降质场景。在面对复杂降质场景时，现实世界的图像可能会同时受到多种因素的影响，如强烈的雨雾、光照动态变化、低照度环境和图像损坏，这使得模型需要处理未知和多样化的图像特征。目前的挑战在于大部分现有模型仅针对特定降质场景设计，引入了大量先验知识，难以适用于其他降质场景。例如，雨雾去除模型在网络框架中嵌入了雨层—背景层的建模，虽然在雨雾去除任务上表现出良好效果，但无法适用于超分辨率、暗光增强等其他任务。开发能够鲁棒应对多种复杂降质的高效增强方法是图像增强领域极具挑战性的任务。

2)面向未知下游任务。现有恶劣场景视觉理解模型的构建依赖下游任务信息，包括目标域数据分布、降质先验和预训练的下游任务模型等，难以做到对任意任务及分析模型鲁棒。然而，在现实应用中，下游任务信息难以准确获取。此外，多数方法局限于某个单一的机器分析下游任务，无法泛化至新的下游任务场景。解决这一挑战需要研究者不断提升恶劣场景视觉感知方法的泛化性和适应性，从而为现实机器分析应用提供更好的性能与可靠性。

3)大模型的机遇与挑战。大模型在各领域取得了显著的成就，如语言模型BERT(Devlin等，2019)在自然语言处理中的广泛应用。目前，不少工作已证明大模型在增强重建等低层次计算视觉任务上具

有巨大的潜力。然而，大模型的高复杂性也带来了挑战，包括计算资源需求大、训练时间长和模型优化困难等问题。同时，恶劣场景下模型泛化能力也是亟待解决的挑战，需要更全面的数据构建策略和更有效的模型优化方法来提升视觉大模型在恶劣场景视觉感知与理解任务上的性能和可靠性。

4 结　语

本文系统地分析了国内外恶劣场景下视觉感知与理解领域的重要研究进展，包括图像视频与降质建模、恶劣场景视觉增强、恶劣场景下视觉分析理解等技术，以及恶劣场景下视觉感知与理解的协同计算。此外，本文比较了国内外研究进展，并展望恶劣场景下视觉感知与理解的未来发展趋势。

致　谢　本文由中国图象图形学学会多媒体专委会组织撰写，该专委会链接为 https://www.csig.org.cn/16/201612/49319.html。

参考文献（References）

Abdelhamed A, Brubaker M and Brown M. 2019. Noise flow: noise modeling with conditional normalizing flows//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 3165-3173 [DOI: 10.1109/ICCV.2019.00326]

Barnum P C, Narasimhan S and Kanade T. 2010. Analysis of rain and snow in frequency space. International Journal of Computer Vision, 86(2/3): 256-274 [DOI: 10.1007/s11263-008-0200-2]

Ben-David S, Blitzer J, Crammer K, Kulesza A, Pereira F and Vaughan J W. 2010. A theory of learning from different domains. Machine Learning, 79(1/2): 151-175 [DOI: 10.1007/s10994-009-5152-4]

Braun M, Krebs S, Flohr F and Gavrila D M. 2019. EuroCity persons: a novel benchmark for person detection in traffic scenes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(8): 1844-1861 [DOI: 10.1109/TPAMI.2019.2897684]

Cao Y, Liu M, Liu S, Wang X T, Lei L and Zuo W M. 2023. Physics-guided ISO-dependent sensor noise modeling for extreme low-light photography//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 5744-5753 [DOI: 10.1109/CVPR52729.2023.00556]

Chang Y, Yan L X and Zhong S. 2017. Transformed low-rank model for line pattern noise removal//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 1735-1743 [DOI: 10.1109/ICCV.2017.191]

Chen D Y, Chen C C and Kang L W. 2014. Visual depth guided color image rain streaks removal using sparse coding. IEEE Transactions on Circuits and Systems for Video Technology, 24(8): 1430-1455 [DOI: 10.1109/TCSVT.2014.2308627]

Chen J, Tan C H, Hou J H, Chau L P and Li H. 2018a. Robust video content alignment and compensation for rain removal in a CNN framework//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 6286-6295 [DOI: 10.1109/CVPR.2018.00658]

Chen Y H, Li W, Sakaridis C, Dai D X and Van Gool L. 2018b. Domain adaptive faster R-CNN for object detection in the wild//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 3339-3348 [DOI: 10.1109/CVPR.2018.00352]

Chen Y L and Hsu C T. 2013. A generalized low-rank appearance model for spatio-temporally correlated rain streaks//Proceedings of 2013 IEEE International Conference on Computer Vision. Sydney, Australia: IEEE: 1968-1975 [DOI: 10.1109/ICCV.2013.247]

Cui Z T, Qi G J, Gu L, You S D, Zhang Z H and Harada T. 2021. Multitask AET with orthogonal tangent regularity for dark object detection//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 2533-2542 [DOI: 10.1109/ICCV48922.2021.00255]

Dai D X and Van Gool L. 2018. Dark model adaptation: semantic image segmentation from daytime to nighttime//Proceedings of the 21st International Conference on Intelligent Transportation Systems. Maui, USA: IEEE: 3819-3824 [DOI: 10.1109/ITSC.2018.8569387]

Devlin J, Chang M W, Lee K and Toutanova K. 2019. BERT: pre-training of deep bidirectional Transformers for language understanding//Proceedings of 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis, USA: Association for Computational Linguistics: 4171-4186 [DOI: 10.18653/V1/N19-1423]

D'Innocente A, Borlino F C, Bucci S, Caputo B and Tommasi T. 2020. One-shot unsupervised cross-domain detection//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: IEEE: 732-748 [DOI: 10.1007/978-3-030-58517-4_43]

Dong C, Loy C C, He K and Tang X. 2014. Learning a deep convolutional network for image super-resolution//Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland: Springer: 184-199 [DOI: 10.1007/978-3-319-10593-2_13]

Dudhane A, Zamir S W, Khan S, Khan F S and Yang M H. 2023. Burstormer: burst image restoration and enhancement Transformer//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 5703-5712 [DOI: 10.1109/CVPR52729.2023.00552]

Eigen D, Krishnan D and Fergus R. 2013. Restoring an image taken

through a window covered with dirt or rain//Proceedings of 2013 IEEE International Conference on Computer Vision. Sydney, Australia: IEEE: 633-640 [DOI: 10.1109/ICCV.2013.84]

Foi A, Trimeche M, Katkovnik V and Egiazarian K. 2008. Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data. IEEE Transactions on Image Processing, 17(10): 1737-1754 [DOI: 10.1109/TIP.2008.2001399]

Fu X Y, Huang J B, Zeng D L, Huang Y, Ding X H and Paisley J. 2017. Removing rain from single images via a deep detail network//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 1715-1723 [DOI: 10.1109/CVPR.2017.186]

Fu X Y, Zeng D L, Huang Y, Zhang X P and Ding X H. 2016. A weighted variational model for simultaneous reflectance and illumination estimation//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 2782-2790 [DOI: 10.1109/CVPR.2016.304]

Fu Y H, Kang L W, Lin C W and Hsu C T. 2011. Single-frame-based rain removal via image decomposition//Proceedings of 2011 IEEE International Conference on Acoustics, Speech and Signal Processing. Prague, Czech Republic: IEEE: 1453-1456 [DOI: 10.1109/ICASSP.2011.5946766]

Ganin Y, Ustinova E, Ajakan H, Germain P, Larochelle H, Laviolette F, Marchand, M and Lempitsky V S. 2016. Domain-adversarial training of neural networks. Journal of Machine Learning Research, 17(1): 2096-2030 [DOI: 10.5555/2946645.2946704]

Garg K and Nayar S K. 2007. Vision and rain. International Journal of Computer Vision, 75 (1): 3-27 [DOI: 10.1007/s11263-006-0028-6]

Guo M X, Chen M T, Ma C, Li Y, Li X F and Xie X D. 2020. High-level task-driven single image deraining: segmentation in rainy days//Proceedings of the Neural Information Processing. Bangkok, Thailand: Springer: 350-362 [DOI: 10.1007/978-3-030-63830-6_30]

Guo X J, Li Y and Ling H B. 2017. LIME: low-light image enhancement via illumination map estimation. IEEE Transactions on Image Processing, 26(2): 982-993 [DOI: 10.1109/TIP.2016.2639450]

Halder S, Lalonde J F and De Charette R. 2019. Physics-based rendering for improving robustness to rain//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 10202-10211 [DOI: 10.1109/ICCV.2019.01030]

He K M, Sun J and Tang X O. 2009. Single image haze removal using dark channel prior//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE: 1956-1963 [DOI: 10.1109/CVPR.2009.5206515]

Hoffman J, Tzeng E, Park T, Zhu J Y, Isola P, Saenko K, Efros A A and Darrell T. 2018. CyCADA: cycle-consistent adversarial domain adaptation//Proceedings of the 35th International Conference on Machine Learning. Stockholm, Sweden: PMLR: 1989-1998

Hu X W, Fu C W, Zhu L and Heng P A. 2019. Depth-attentional features for single-image rain removal//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 8014-8023 [DOI: 10.1109/CVPR. 2019.00821]

Huang S C, Le T H and Jaw D W. 2021. DSNet: joint semantic learning for object detection in inclement weather conditions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(8): 2623-2633 [DOI: 10.1109/TPAMI.2020.2977911]

Hwang S, Park J, Kim N, Choi Y and Kweon I S. 2015. Multispectral pedestrian detection: benchmark dataset and baseline//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE: 1037-1045 [DOI: 10.1109/CVPR.2015.7298706]

Inoue N, Furuta R, Yamasaki T and Aizawa K. 2018. Cross-domain weakly-supervised object detection through progressive domain adaptation//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 5001-5009 [DOI: 10.1109/CVPR.2018.00525]

Jenicek T and Chum O. 2019. No fear of the dark: image retrieval under varying illumination conditions//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 9695-9703 [DOI: 10.1109/ICCV.2019.00979]

Jiang K, Wang Z Y, Yi P, Chen C, Huang B J, Luo Y M, Ma J Y and Jiang J J. 2020. Multi-scale progressive fusion network for single image deraining//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 8343-8352 [DOI: 10.1109/CVPR42600.2020.00837]

Jobson D J, Rahman Z and Woodell G A. 1997. Properties and performance of a center/surround retinex. IEEE Transactions on Image Processing, 6(3): 451-462 [DOI: 10.1109/83.557356]

Kang L W, Lin C W and Fu Y H. 2012. Automatic single-image-based rain streaks removal via image decomposition. IEEE Transactions on Image Processing, 21 (4): 1742-1755 [DOI: 10.1109/TIP.2011.2179057]

Kawar B, Elad M, Ermon S and Song J M. 2022. Denoising diffusion restoration models [EB/OL]. [2024-01-01]. https://arxiv.org/pdf/2201.11793.pdf

Ketcham D J, Lowe R W and Weber J W. 1974. Image Enhancement Techniques for Cockpit Displays. Hughes Aircraft Co Culver City Calif Display Systems Lab

Khodabandeh M, Vahdat A, Ranjbar M and Macready W. 2019. A robust learning approach to domain adaptive object detection//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 480-490 [DOI: 10.1109/ICCV.2019.00057]

Kim J H, Lee C, Sim J Y and Kim C S. 2013. Single-image deraining using an adaptive nonlocal means filter//Proceedings of 2013 IEEE International Conference on Image Processing. Melbourne, Austra-

lia: IEEE: 914-917 [DOI: 10.1109/ICIP.2013.6738189]

Kim S, Choi J, Kim T and Kim C. 2019a. Self-training and adversarial background regularization for unsupervised domain adaptive one-stage object detection//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 6091-6100 [DOI: 10.1109/ICCV.2019.00619]

Land E H and McCann J J. 1971. Lightness and retinex theory. Journal of the Optical Society of America, 61(1): 1-11 [DOI: 10.1364/josa.61.000001]

Lengyel A, Garg S, Milford M and van Gemert J C. 2021. Zero-shot day-night domain adaptation with a physics prior//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 4379-4389 [DOI: 10.1109/ICCV48922.2021.00436]

Li B Y, Ren W Q, Fu D P, Tao D C, Feng D, Zeng W J and Wang Z Y. 2019. Benchmarking single-image dehazing and beyond. IEEE Transactions on Image Processing, 28(1): 492-505 [DOI: 10.1109/TIP.2018.2867951]

Li C Y, Guo J C, Porikli F and Pang Y W. 2018a. LightenNet: a convolutional neural network for weakly illuminated image enhancement. Pattern Recognition Letters, 104: 15-22 [DOI: 10.1016/j.patrec.2018.01.010]

Li G B, He X, Zhang W, Chang H Y, Dong L and Lin L. 2018b. Non-locally enhanced encoder-decoder network for single image de-raining//Proceedings of the 26th ACM International Conference on Multimedia. Seoul Korea (South): ACM: 1056-1064 [DOI: 10.1145/3240508.3240636]

Li Y, Chang Y, Yu C F and Yan L X. 2022. Close the loop: a unified bottom-up and top-down paradigm for joint image deraining and segmentation//Proceedings of the 36th AAAI Conference on Artificial Intelligence. Virtual: AAAI: 1438-1446 [DOI: 10.1609/aaai.v36i2.20033]

Liang J Y, Cao J Z, Sun G L, Zhang K, Van Gool L and Timofte R. 2021. SwinIR: image restoration using Swin Transformer//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision Workshops. Montreal, Canada: IEEE: 1833-1844 [DOI: 10.1109/ICCVW54120.2021.00210]

Liang Y C, Anwar S and Liu Y. 2022. DRT: a lightweight single image deraining recursive Transformer//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. New Orleans, USA: IEEE: 588-597 [DOI: 10.1109/CVPRW56347.2022.00074]

Liu C and Sun D Q. 2014. On Bayesian adaptive video super resolution. IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(2): 346-360 [DOI: 10.1109/TPAMI.2013.127]

Liu J Y, Yang W H, Yang S and Guo Z M. 2018. Erase or fill? Deep joint recurrent rain removal and reconstruction in videos//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 3233-3242 [DOI: 10.1109/CVPR.2018.00341]

Liu R S, Ma L, Zhang J A, Fan X and Luo Z X. 2021. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 10556-10565 [DOI: 10.1109/CVPR46437.2021.01042]

Loh Y P and Chan C S. 2019. Getting to know low-light images with the exclusively dark dataset. Computer Vision and Image Understanding, 178: 30-42 [DOI: 10.1016/j.cviu.2018.10.010]

Long M S, Cao Y, Wang J M and Jordan M I. 2015. Learning transferable features with deep adaptation networks//Proceedings of the 32nd International Conference on Machine Learning. Lille, France: JMLR.org: 97-105

Long M S, Zhu H, Wang J M and Jordan M I. 2017. Deep transfer learning with joint adaptation networks//Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia: JMLR.org: 2208-2217

Lore K G, Akintayo A and Sarkar S. 2017. LLNet: a deep autoencoder approach to natural low-light image enhancement. Pattern Recognition, 61: 650-662 [DOI: 10.1016/j.patcog.2016.06.008]

Lu K and Zhang L H. 2021. TBEFN: a two-branch exposure-fusion network for low-light image enhancement. IEEE Transactions on Multimedia, 23: 4093-4105 [DOI: 10.1109/TMM.2020.3037526]

Luo R D, Wang W J, Yang W H and Liu J Y. 2023. Similarity min-max: zero-shot day-night domain adaptation//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 8070-8080 [DOI: 10.1109/ICCV51070.2023.00744]

Luo Y, Xu Y and Ji H. 2015. Removing rain from a single image via discriminative sparse coding//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE: 3397-3405 [DOI: 10.1109/ICCV.2015.388]

Lyu F F, Lu F, Wu J H and Lim C. 2018. MBLLEN: low-light image/video enhancement using CNNs//Proceedings of the British Machine Vision Conference. Newcastle, UK: BMVA: #220

Ma L, Ma T Y, Liu R S, Fan X and Luo Z X. 2022. Toward fast, flexible, and robust low-light image enhancement//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 5627-5636 [DOI: 10.1109/CVPR52688.2022.00555]

Mu P, Chen J, Liu R S, Fan X and Luo Z X. 2019. Learning bilevel layer priors for single image rain streaks removal. IEEE Signal Processing Letters, 26(2): 307-311 [DOI: 10.1109/LSP.2018.2889277]

Nada H, Sindagi V A, Zhang H and Patel V M. 2018. Pushing the limits of unconstrained face detection: a challenge dataset and baseline results//Proceedings of the 9th IEEE International Conference on Biometrics Theory, Applications and Systems. Redondo Beach, USA: IEEE: 1-10 [DOI: 10.1109/BTAS.2018.8698561]

Neuhold G, Ollmann T, Bulò S R and Kontschieder P. 2017. The mapillary vistas dataset for semantic understanding of street scenes//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 5000-5009 [DOI: 10.1109/ICCV.2017.534]

Neumann L, Karg M, Zhang S S, Scharfenberger C, Piegert E, Mistr S, Prokofyeva O, Thiel R, Vedaldi A, Zisserman A and Schiele B. 2019. NightOwls: a pedestrians at night dataset//Proceedings of the 14th Asian Conference on Computer Vision. Perth, Australia: Springer: 691-705 [DOI: 10.1007/978-3-030-20887-5_43]

Nguyen C M, Chan E R, Bergman A W and Wetzstein G. 2023. Diffusion in the dark: a diffusion model for low-light text recognition [EB/OL]. [2024-01-01]. https://arxiv.org/pdf/2303.04291.pdf

O'Handley D A and Green W B. 1972. Recent developments in digital image processing at the image processing laboratory at the jet propulsion laboratory. Proceedings of the IEEE, 60(7): 821-828 [DOI: 10.1109/PROC.1972.8781]

Park W J and Lee K H. 2008. Rain removal using Kalman filter in video//Proceedings of 2008 International Conference on Smart Manufacturing Application. Goyangi, Korea (South): IEEE: 494-497 [DOI: 10.1109/ICSMA.2008.4505573]

Pizer S M, Johnston R E, Ericksen J P, Yankaskas B C and Muller K E. 1990. Contrast-limited adaptive histogram equalization: speed and effectiveness//Proceedings of the 1st Conference on Visualization in Biomedical Computing. Atlanta, USA: IEEE: 337-345 [DOI: 10.1109/VBC.1990.109340]

Qian K, Zhu S L, Zhang X Y and Li L E. 2021. Robust multimodal vehicle detection in foggy weather using complementary lidar and radar signals//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 444-453 [DOI: 10.1109/CVPR46437.2021.00051]

Qian R, Tan R T, Yang W H, Su J J and Liu J Y. 2018. Attentive generative adversarial network for raindrop removal from a single image//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 2482-2491 [DOI: 10.1109/CVPR.2018.00263]

Quan R J, Yu X, Liang Y Z and Yang Y. 2021. Removing raindrops and rain streaks in one go//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 9143-9152 [DOI: 10.1109/CVPR46437.2021.00903]

Ren D W, Zuo W M, Hu Q H, Zhu P F and Meng D Y. 2019a. Progressive image deraining networks: a better and simpler baseline//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 3932-3941 [DOI: 10.1109/CVPR.2019.00406]

Ren W Q, Liu S F, Ma L, Xu Q Q, Xu X Y, Cao X C, Du J P and Yang M H. 2019b. Low-light image enhancement via a deep hybrid network. IEEE Transactions on Image Processing, 28(9): 4364-4375 [DOI: 10.1109/TIP.2019.2910412]

Russo P, Carlucci F M, Tommasi T and Caputo B. 2018. From source to target and back: symmetric bi-directional adaptive GAN//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 8099-8108 [DOI: 10.1109/CVPR.2018.00845]

Saharia C, Ho J, Chan W, Salimans T, Fleet D and Norouzi M. 2023. Image super-resolution via iterative refinement. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(4): 4713-4726 [DOI: 10.1109/TPAMI.2022.3204461]

Saito K, Ushiku Y, Harada T and Saenko K. 2019. Strong-weak distribution alignment for adaptive object detection//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 6949-6958 [DOI: 10.1109/CVPR.2019.00712]

Sakaridis C, Dai D X and Van Gool L. 2018. Semantic foggy scene understanding with synthetic data. International Journal of Computer Vision, 126(9): 973-992 [DOI: 10.1007/s11263-018-1072-8]

Sakaridis C, Dai D X and Van Gool L. 2019. Guided curriculum model adaptation and uncertainty-aware evaluation for semantic nighttime image segmentation//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 7373-7382 [DOI: 10.1109/ICCV.2019.00747]

Sakaridis C, Dai D X and Van Gool L. 2021. ACDC: the adverse conditions dataset with correspondences for semantic driving scene understanding//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 10745-10755 [DOI: 10.1109/ICCV48922.2021.01059]

Sasagawa Y and Nagahara H. 2020. YOLO in the dark - domain adaptation method for merging multiple models//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: IEEE: 345-359 [DOI: 10.1007/978-3-030-58589-1_21]

Shan Y H, Lu W F and Chew C M. 2019. Pixel and feature level based domain adaptation for object detection in autonomous driving. Neurocomputing, 367: 31-38 [DOI: 10.1016/j.neucom.2019.08.022]

Shen L, Yue Z H, Feng F, Chen Q, Liu S H and Ma J. 2017. MSR-net: low-light image enhancement using deep convolutional network [EB/OL]. [2024-01-01]. https://arxiv.org/pdf/1711.02488.pdf

Shocher A, Cohen N and Irani M. 2018. Zero-shot super-resolution using deep internal learning//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 3118-3126 [DOI: 10.1109/CVPR.2018.00329]

Sindagi V A, Oza P, Yasarla R and Patel V M. 2020. Prior-based domain adaptive object detection for hazy and rainy conditions//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: IEEE: 763-780 [DOI: 10.1007/978-3-030-58568-6_45]

Song W Z, Suganuma M, Liu X, Shimobayashi N, Maruta D and Okatani T. 2021. Matching in the dark: a dataset for matching image

pairs of low-light scenes//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 6009-6018 [DOI: 10.1109/ICCV48922.2021.00597]

Sun B C and Saenko K. 2016. Deep CORAL: correlation alignment for deep domain adaptation//Proceedings of 2016 European Conference on Computer Vision Workshops. Amsterdam, the Netherlands: IEEE: 443-450 [DOI: 10.1007/978-3-319-49409-8_35]

Sun H, Ang M H and Rus D. 2019. A convolutional network for joint deraining and dehazing from a single image for autonomous driving in rain//Proceedings of 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems. Macau, China: IEEE: 962-969 [DOI: 10.1109/IROS40897.2019.8967644]

Tan X, Xu K, Cao Y, Zhang Y H, Ma L Z and Lau R W H. 2021. Night-time scene parsing with a large real dataset. IEEE Transactions on Image Processing, 30: 9085-9098 [DOI: 10.1109/TIP.2021.3122004]

Timofte R, Agustsson E, Gool L V, Yang M H, Zhang L, Lim B, Son S, Kim H, Nah S, Lee K M, Wang X T, Tian Y P, Yu K, Zhang Y L, Wu S X, Dong C, Lin L, Qiao Y, Loy C C, Bae W, Yoo J, Han Y, Ye J C, Choi J S, Kim M, Fan Y C, Yu J H, Han W, Liu D, Yu H C, Wang Z Y, Shi H H, Wang X C, Huang T S, Chen Y J, Zhang K, Zuo W M, Tang Z M, Luo L K, Li S H, Fu M, Cao L, Heng W, Bui G, Le T, Duan Y, Tao D C, Wang R X, Lin X, Pang J X, Xu J C, Zhao Y, Xu X Y, Pan J S, Sun D Q, Zhang Y J, Song X B, Dai Y C, Qin X Y, Huynh X P, Guo T T, Mousavi H S, Vu T H, Monga V, Cruz C, Egiazarian K, Katkovnik V, Mehta R, Jain A K, Agarwalla A, Praveen C V S, Zhou R F, Wen H D, Zhu C, Xia Z Q, Wang Z T and Guo Q. 2017. NTIRE 2017 challenge on single image super-resolution: methods and results//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu, USA: IEEE: 1110-1121 [DOI: 10.1109/CVPRW.2017.149]

Triantafyllidou D, Moran S, McDonagh S, Parisot S and Slabaugh G. 2020. Low light video enhancement using synthetic data produced with an intermediate domain mapping//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: IEEE: 103-119 [DOI: 10.1007/978-3-030-58601-0_7]

Tzeng E, Hoffman J, Saenko K and Darrell T. 2017. Adversarial discriminative domain adaptation//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 2962-2971 [DOI: 10.1109/CVPR.2017.316]

Wang K, Zhang Z Y, Yan Z Q, Li X, Xu B B, Li J and Yang J. 2021a. Regularizing nighttime weirdness: efficient self-supervised monocular depth estimation in the dark//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 16035-16044 [DOI: 10.1109/ICCV48922.2021.01575]

Wang X T, Xie L B, Dong C and Shan Y. 2021b. Real-ESRGAN: training real-world blind super-resolution with pure synthetic data//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision Workshops. Montreal, Canada: IEEE: 1905-1914 [DOI: 10.1109/ICCVW54120.2021.00217]

Wang Y L, Song Y B, Ma C and Zeng B. 2020. Rethinking image deraining via rain streaks and vapors//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: IEEE: 367-382 [DOI: 10.1007/978-3-030-58520-4_22]

Wei C, Wang W J, Yang W H and Liu J Y. 2018. Deep retinex decomposition for low-light enhancement [EB/OL]. [2024-01-01]. https://arxiv.org/pdf/1808.04560.pdf

Wu X Y, Wu Z Y, Guo H, Ju L L and Wang S. 2021. DANNet: a one-stage domain adaptation network for unsupervised nighttime semantic segmentation//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 15764-15773 [DOI: 10.1109/CVPR46437.2021.01551]

Xu J, Zhao W, Liu P and Tang X L. 2012. Removing rain and snow in a single image using guided filter//Proceedings of 2021 IEEE International Conference on Computer Science and Automation Engineering. Zhangjiajie, China: IEEE: 304-307 [DOI: 10.1109/CSAE.2012.6272780]

Yang W H, Tan R T, Feng J S, Liu J Y, Guo Z M and Yan S C. 2017. Deep joint rain detection and removal from a single image//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 1685-1694 [DOI: 10.1109/CVPR.2017.183]

Yang W H, Tan R T, Wang S Q and Liu J Y. 2020a. Self-learning video rain streak removal: when cyclic consistency meets temporal correspondence//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 1717-1726 [DOI: 10.1109/CVPR42600.2020.00179]

Yang W H, Yuan Y, Ren W Q, Liu J Y, Scheirer W J, Wang Z Y, Zhang T H, Zhong Q Y, Xie D, Pu S L, Zheng Y Q, Qu Y Y, Xie Y H, Chen L, Li Z H, Hong C, Jiang H, Yang S Y, Liu Y, Qu X C, Wan P F, Zheng S, Zhong M H, Su T Y, He L Z, Guo Y D, Zhao Y, Zhu Z F, Liang J X, Wang J W, Chen T Y, Quan Y H, Xu Y, Liu B, Liu X, Sun Q, Lin T Y, Li X C, Lu F, Gu L, Zhou S D, Cao C, Zhang S F, Chi C, Zhuang C B, Lei Z, Li S Z, Wang S Z, Liu R Z, Yi D, Zuo Z M, Chi J N, Wang H, Wang K, Liu Y X, Gao X Y, Chen Z Y, Guo C, Li Y Z, Zhong H C, Huang J, Guo H, Yang J F, Liao W J, Yang J G, Zhou L G, Feng M Y and Qin L K. 2020b. Advancing image understanding in poor visibility environments: a collective benchmark study. IEEE Transactions on Image Processing, 29: 5737-5752 [DOI: 10.1109/TIP.2020.2981922]

Yu F, Chen H F, Wang X, Xian W Q, Chen Y Y, Liu F C, Madhavan V and Darrell T. 2020. BDD100K: a diverse driving dataset for heterogeneous multitask learning//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 2633-2642 [DOI: 10.1109/CVPR42600.2020.00271]

Yu Y, Yang W H, Tan Y P and Kot A C. 2022. Towards robust rain removal against adversarial attacks: a comprehensive benchmark analysis and beyond//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 6003-6012 [DOI: 10.1109/CVPR52688.2022.00592]

Zamir S W, Arora A, Khan S, Hayat M, Khan F S and Yang M H. 2022. Restormer: efficient Transformer for high-resolution image restoration//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 5718-5729 [DOI: 10.1109/CVPR52688.2022.00564]

Zhang H and Patel V M. 2018. Density-aware single image de-raining using a multi-stream dense network//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 695-704 [DOI: 10.1109/CVPR.2018.00079]

Zhang H, Sindagi V and Patel V M. 2020. Image de-raining using a conditional generative adversarial network. IEEE Transactions on Circuits and Systems for Video Technology, 30(11): 3943-3956 [DOI: 10.1109/TCSVT.2019.2920407]

Zhang K, Liang J Y, Van Gool L and Timofte R. 2021a. Designing a practical degradation model for deep blind image super-resolution//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 4771-4780 [DOI: 10.1109/ICCV48922.2021.00475]

Zhang K, Zuo W M, Chen Y J, Meng D Y and Zhang L. 2017a. Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising. IEEE Transactions on Image Processing, 26(7): 3142-3155 [DOI: 10.1109/TIP.2017.2662206]

Zhang K, Zuo W M, Gu S H and Zhang L. 2017b. Learning deep CNN denoiser prior for image restoration//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 2808-2817 [DOI: 10.1109/CVPR.2017.300]

Zhang K H, Luo W H, Yu Y J, Ren W Q, Zhao F, Li C S, Ma L, Liu W and Li H D. 2022. Beyond monocular deraining: parallel stereo deraining network via semantic prior. International Journal of Computer Vision, 130(7): 1754-1769 [DOI: 10.1007/s11263-022-01620-w]

Zhang X P, Li H, Qi Y Y, Leow W K and Ng T K. 2006. Rain removal in video by combining temporal and chromatic properties//Proceedings of 2006 IEEE International Conference on Multimedia and Expo. Toronto, Canada: IEEE: 461-464 [DOI: 10.1109/ICME.2006.262572]

Zhang Y, Guo X, Ma J, Liu W and Zhang J. 2021. Beyond brightening low-light images. International Journal of Computer Vision, 129(4): 1013-1037 [DOI: 10.1007/s11263-020-01407-x]

Zhang Y H, Zhang J W and Guo X J. 2019. Kindling the darkness: a practical low-light image enhancer//Proceedings of the 27th ACM International Conference on Multimedia. Nice, France: ACM: 1632-1640 [DOI: 10.1145/3343031.3350926]

Zheng S, Lu C J, Wu Y X and Gupta G. 2022. SAPNet: segmentation-aware progressive network for perceptual contrastive deraining//Proceedings of 2022 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops. Waikoloa, USA: IEEE: 52-62 [DOI: 10.1109/WACVW54805.2022.00011]

Zheng X H, Liao Y H, Guo W, Fu X Y and Ding X H. 2013. Single-image-based rain and snow removal using multi-guided filter//Proceedings of the 20th Neural Information Processing. Daegu, Korea (South): Springer: 258-265 [DOI: 10.1007/978-3-642-42051-1_33]

Zhu L, Fu C W, Lischinski D and Heng P A. 2017. Joint bi-layer optimization for single-image rain streak removal//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 2545-2553 [DOI: 10.1109/ICCV.2017.276]

作者简介

汪文靖,女,博士研究生,主要研究方向为视觉增强和视觉生成。E-mail:daooshee@pku.edu.cn

刘家瑛,通信作者,女,副教授,主要研究方向为智能媒体计算与视觉理解。E-mail:liujiaying@pku.edu.cn

杨文瀚,男,副研究员,主要研究方向为恶劣环境下的图像/视频增强、视觉信号编码表征。E-mail:yangwh@pcl.ac.cn

方玉明,男,教授,主要研究方向为计算机视觉、多媒体信号处理和视觉质量评估。E-mail:leo.fangyuming@foxmail.com

黄华,男,教授,主要研究方向为可视媒体计算。E-mail:huahuang@bnu.edu.cn