Artificial Intelligence Review (2025) 58:218
https://doi.org/10.1007/s10462-025-11233-7

A comprehensive survey of specularity detection: state-of-the-art techniques and breakthroughs

Fengze Li[1,2] · Jieming Ma[1] · Hai-Ning Liang[3] · Zhongbei Tian[4] · Zhijing Wu[5] · Tianxi Wen[6] · Dawei Liu[1]

Accepted: 10 April 2025 / Published online: 23 April 2025

Abstract
Specularity poses significant challenges in computer vision (CV), often leading to performance degradation in various tasks. Despite its importance, the CV field lacks a comprehensive review of specularity detection techniques. This survey addresses this gap by synthesizing diverse definitions of specularity and providing a unified framework to enhance consistency. It also presents a systematic review of traditional and deep learning-based methods for detecting specularity. Comparative experiments on a standardized dataset enable in-depth evaluation of each method, highlighting their strengths and limitations. The survey further provides structured insights and guidance for selecting appropriate methods across diverse scenarios. Through this, it identifies key areas for future research, aiming to support the development of more advanced detection models. By integrating diverse methodologies and quantitative analyzes, this survey contributes to a deeper understanding of current advancements and potential innovations in specularity detection.

Keywords Computer vision · Specularity · Specularity detection

1 Introduction

Specularity refers to reflective phenomena that occur on various surfaces and objects. This includes reflective objects such as mirrors and polished metals, reflective surfaces like water, and specific reflective regions on objects, as shown in Fig. 1. These specular reflections introduce strong highlights and distortions, complicating image interpretation and leading to challenges in computer vision (CV) tasks (Tan et al. 2023; Anwer et al. 2023). Accurately detecting and interpreting these specularities is critical, as misinterpreting reflections can result in performance degradation in tasks such as object detection, segmentation, and scene understanding.

Despite advances in CV, the task of detecting and interpreting specularity remains inconsistent in different studies. Some researchers define specularity as specular objects, focusing on highly reflective entities such as mirrors (Yang et al. 2019; Lin et al. 2020; Mei et al.

Extended author information available on the last page of the article

Fig. 1 Specularity in life

2021; Tan et al. 2023; Lin et al. 2023; Lin and Lau 2023; Anwer et al. 2023; He et al. 2023), while others broaden the concept to include any reflective surface, typically referring to metallic or oily surfaces that exhibit specular reflections (Fu et al. 2020; Chen et al. 2023; Yu et al. 2014; Wu et al. 2021). This lack of a unified definition has resulted in various approaches to specularity detection, with no clear consensus on how to consistently address the issue in CV tasks.

This study first reviews various definitions of specularity in CV, which have evolved from geometrical, physical, and perceptual perspectives. The central research question is how to develop a consistent, unified definition of specularity that enhances the reliability and generalizability of specularity detection methods across diverse CV tasks. Thus, this review draws inspiration from multiple fields. For example, specularity in the field of computer graphics (CG) is often defined with precision using the Disney principled bidirectional reflectance distribution function (BRDF) (Burley and Studios 2012). From a geometric perspective, specularity is explained based on surface and light interactions, while the physical reflection model emphasizes material properties and light behavior. The perceptual angle considers how humans interpret reflective surfaces.

Thus, this review proposes a unified approach to specularity in CV, integrating these diverse perspectives into a mathematical framework, offering a more consistent solution for specularity detection. Specifically, this review suggests that specularity in CV should not be defined by the material or task, but by its fundamental physical and perceptual traits: any instance where light reflects off a surface in a concentrated manner, producing directional highlights. This study shows how this definition can be represented by a more precise mathematical model based on physical optics, the Fresnel equations (Lippincott and Stark 1982; Lvovsky 2013), which describe the behavior of light reflecting off smooth surfaces at the microscopic level. By unifying these multiple perspectives under a common mathematical framework, we not only deepen the theoretical understanding of specularity but also ensure that detection methods in CV are more coherent and adaptable across different contexts and applications.

With a clear definition of specularity established, this review summarizes the major contributions to specularity detection in CV, spanning from the earliest traditional methods to the most recent deep learning (DL)-based approaches. By examining both past and current research, this review aims to provide a comprehensive overview of how these methods have evolved over time. Traditional CV methods for specularity detection often rely on handcrafted features, physical principles, and mathematical algorithms, such as using color space transformations, intensity gradients, and reflection models to identify reflective regions. These methods, while effective in controlled environments, tend to struggle with generalization in more complex, real-world scenarios due to the variability of lighting conditions, surface materials, and reflection angles. Furthermore, their reliance on predefined rules limits their adaptability across diverse environments and surface types. On the other hand, DL-based approaches have demonstrated significant progress by automatically learning features that can differentiate reflective phenomena from actual scene content. Various DL models have contributed to improving the detection of specularities, leveraging large datasets and complex architectures. However, despite these advancements, DL methods still face challenges in accurately detecting and interpreting specularities in highly reflective or dynamically changing environments. There remains ample room for improvement in making these models more robust and generalizable across different scenarios.

Given the varying and often inconsistent definitions of specularity, the relative novelty of specularity research, and the field's strong focus on application-specific challenges, much of the existing work lacks generalizability across broader contexts. As a result, this paper provides a comprehensive overview of specularity detection, addressing the need for a unified and systematic review to consolidate diverse insights and guide future developments. Through extensive quantitative experiments on standardized datasets, this study critically evaluates the performance of various approaches, tracking their progress and highlighting improvements over time. Our analysis not only examines the evolution of methods but also details how datasets have been optimized, offering practical insights into the potential and limitations of each approach. This review has tested these methods across different specular datasets, providing a clear understanding of their applicability in various scenarios. This comprehensive evaluation aims to assist future researchers by offering concrete guidance on selecting the most appropriate methods for specific applications, as well as identifying areas where further improvement is necessary.

In addition to the experimental analysis, this survey identifies gaps in the current literature and suggests promising directions for future research. By uncovering under-explored areas and highlighting the limitations of existing methods, this review is expected to inspire the development of more advanced detection models. To the best of our knowledge, many of the analyses and insights presented here are offered for the first time. The Fig. 2 offers a structured overview of the core definitions, detection methods, and essential components reviewed in this paper, guiding readers through the main aspects of specularity detection. Therefore, the key contributions of this survey are as follows:

1. This paper provides the first comprehensive survey of specularity detection, establishing a unified and consistent definition for CV tasks by leveraging the precise theoretical foundations from CG. This definition serves as a guiding framework for understanding and detecting specularities across diverse CV applications.
2. This survey systematically reviews both traditional and deep learning-based methods for specularity detection, exploring the relationships between them. This survey highlights the unique strengths, weaknesses, and innovations of each approach, providing a comprehensive comparison of their evolution over time.
3. Through extensive quantitative experiments, this review evaluates the performance of various methods using unified and diverse metrics across multiple standardized datasets. These results offer clear, data-driven insights into the effectiveness of each approach, allowing for a more rigorous comparison of their practical applicability.
4. This paper explores a wide range of potential applications for specularity detection, identifies current challenges, and highlights complementary approaches such as salient

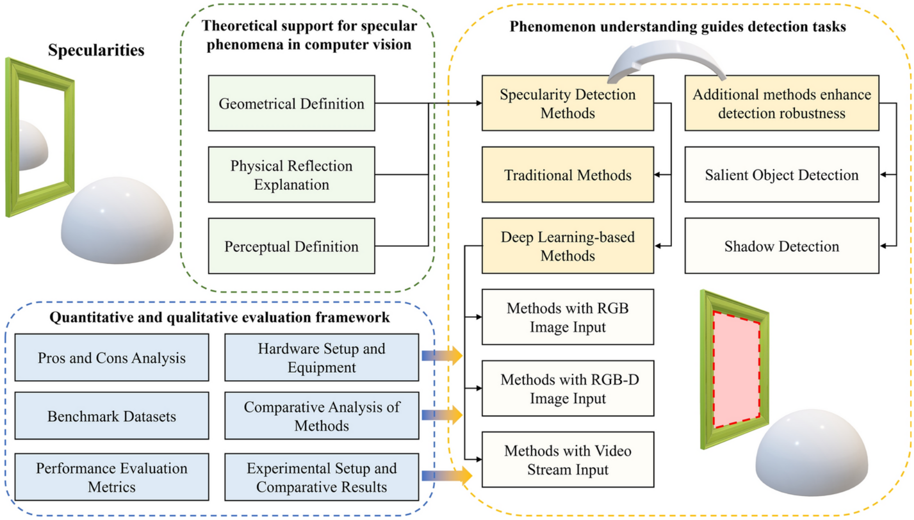

Fig. 2 Methodological framework for specularity detection. The diagram outlines key definitions, detection methods, and supporting components of this review

object and shadow detection. This study also provides targeted recommendations for integrating these techniques to improve specularity detection performance.

2 Defining specularity

This section establishes the foundational groundwork for unifying the understanding of specularity across both CG and CV, two fields where specularity plays a crucial role. By exploring the fundamental principles of specularity, as well as how it is defined in both CG and CV, this section reveals the consistency and overlap between the two domains. The convergence of these definitions forms the basis for a more cohesive and unified approach to specularity, which is essential for advancing research and improving methodologies in visual computing. This structure not only strengthens the theoretical framework but also provides a clearer path for handling specularity in practical applications, shown in Fig. 3.

This section begins with the Subsect. 2.1, which defines the core properties of specular reflections by examining how light interacts with smooth surfaces. The Subsect. 2.2 then outlines how specularity is characterized in CG, focusing on key models that describe reflective surfaces to achieve realism. Finally, the Subsect. 2.3 addresses how specularity is interpreted in CV, particularly with regard to its impact on the accuracy of vision systems. By highlighting the consistency in definitions across CG and CV, this section offers a unified conceptual framework that enhances both research and practical application of specularity in visual computing.

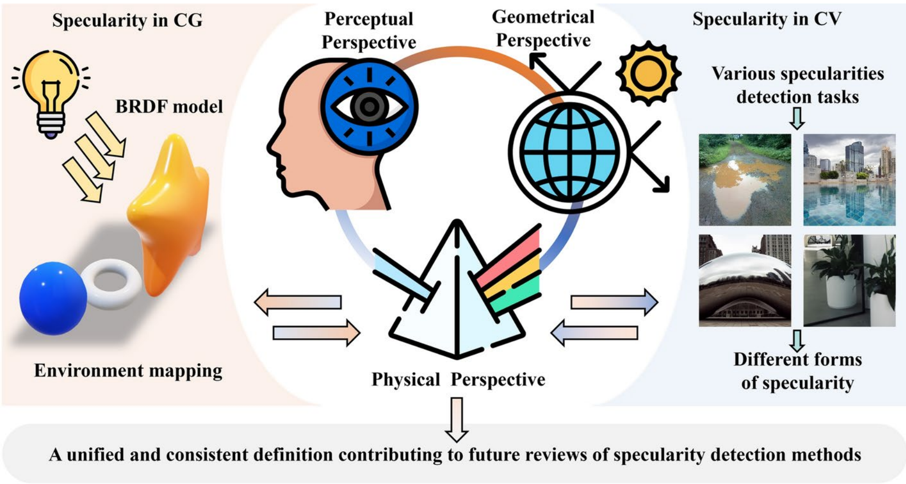

Fig. 3 The diagram of the unified specularity definition highlights the geometrical, physical, and perceptual perspectives, providing a clear foundation for future specularity detection tasks

2.1 Fundamentals of specularity

Specularity, often referred to as highlights or reflective properties of a surface, is a fundamental concept in both CG and CV. It arises from the interaction of light with smooth surfaces. When light hits a surface, it can either be absorbed, diffused, or reflected (Tan 2021). Specularity specifically refers to the reflection of light in a concentrated manner, as opposed to diffused reflection, where light is scattered in multiple directions. Specular reflections occur most prominently on smooth surfaces such as metals, ceramics, and glass, where light is reflected in a controlled, directional way, resulting in bright spots commonly known as specular highlights (Artusi et al. 2011). This survey defines specularity through three key perspectives: geometrical analysis, physical interaction, and perceptual interpretation.

From a geometrical perspective, specular highlights provide critical information about surface properties such as surface normals and curvature (Blake and Brelstaff 1988). To better understand this relationship, the microfacet model is often employed. The microfacet model assumes that a macroscopic surface is composed of a multitude of tiny reflective facets, each with its own surface normal (Cook and Torrance 1982; Oren and Nayar 1994). The distribution of these microfacets' orientations plays a significant role in determining how light reflects off the surface. For smoother surfaces, the microfacets tend to be aligned in a similar direction, whereas rougher surfaces exhibit a broader and more randomized distribution of microfacet normals. Specifically, the location of specular highlights is determined by the orientation of these surface normals at specific points. By analyzing the direction of reflected light, it becomes possible to infer the surface's overall orientation. Furthermore, specular reflections can also provide insights into the local curvature of the surface. A sharper, more focused specular highlight often corresponds to a smoother and flatter surface, while a diffused specular highlight suggests a rougher or more curved surface. By analyzing specular highlights from multiple viewpoints, variations in their positions provide valuable geometric information about the surface's structure. These insights contribute significantly to defining specularity in terms of surface normals and curvature, offering a clear geometrical framework for understanding and detecting specular reflections (Blake and Brelstaff 1988; Oren and Nayar 1997).

From a perceptual perspective, the human visual system uses specularity as a fundamental cue to interpret surface properties. Research has shown that the visual system processes specular highlights to distinguish between reflective and non-reflective surfaces, directly contributing to how specularity is defined in terms of material interpretation (Blake and Bülthoff 1990). The perception of these highlights, through their position, intensity, and contrast, gives important information about surface smoothness and reflectivity (Marlow et al. 2012). This perceptual interpretation plays a crucial role in shaping the overall understanding of specularity, as it reveals how humans naturally define and differentiate reflective characteristics in real-world environments (Fleming et al. 2004). This reliance on specularity plays a fundamental role in interpreting surface features and is crucial for accurate detection and analysis of reflective surfaces in CV.

In terms of physical interaction, specularity emerges from the way light interacts with surfaces, particularly through concentrated, directional reflections. When light encounters smooth surfaces such as metals, ceramics, or glass, it tends to reflect in a controlled, specular manner, producing bright highlights (Chen et al. 2006; Artusi et al. 2011). This behavior contrasts with diffuse reflection, where light scatters in multiple directions, typically seen on rough surfaces. The strength of this reflection is closely tied to both the angle of incidence and the material's physical properties. In optical physics, the Fresnel equations explain how light behaves at the interface between two media, governing both reflection and refraction (Angelopoulou and Poger 2003; Angelopoulou 2007). The intensity of specular reflection depends on factors such as the angle of incidence and the material's refractive index. For example, metals and dielectric materials exhibit distinct specular reflection characteristics, with metals reflecting more light over a broader range of angles (Zhang et al. 2015). These material-specific properties are essential for understanding specularity in CG, where the BRDF is employed to model light interactions with surfaces.

2.2 Specularity in computer graphics

The concept of specularity has long been integral to CG, as its accurate representation is crucial for rendering realistic images. Building on the physical principles of light reflection, including those described by the Fresnel equations (Moon 1940; Angelopoulou and Poger 2003, 2004; Angelopoulou 2007; Zhang et al. 2015), specularity models have evolved to capture the behavior of light as it interacts with different materials.

A significant advancement in CG was the introduction of the Disney BRDF (Burley and Studios 2012), which provided a more versatile model for simulating a wide range of material properties. Unlike earlier models, the Disney BRDF allows for the simulation of both rough, diffuse surfaces and highly reflective, mirror-like surfaces. This is achieved through parameters such as roughness, reflectance, and anisotropy, giving artists and researchers greater control over the appearance of specular highlights in rendered images (Ghosh et al. 2007). In addition to this, recent approaches in BRDF acquisition have further advanced the ability to capture realistic surface reflectance. For example, new methods have been developed to directly measure BRDFs using basis representations by projecting incident light as a sequence of basis functions from a spherical zone of directions. This technique allows for the reconstruction of specular and diffuse reflection components with high accuracy and efficiency, greatly enhancing the precision of specularity representation in CG. Furthermore, studies on the leaf BRDF, which analyze both spectral and directional variations, have identified the spectral invariance of the specular component in the visible spectrum, further contributing to a more comprehensive understanding of specular highlights across different materials (Bousquet et al. 2005).

Furthermore, techniques such as environment mapping and ray tracing have further improved the simulation of specular reflections in CG. Environment mapping captures the surrounding environment and applies it as a texture to reflective surfaces, while ray tracing simulates the actual paths of light rays as they interact with objects in a scene. Together with the Disney BRDF, these methods have enabled the creation of rendered images that are nearly indistinguishable from photographs (Amanatides 1992; Bajcsy et al. 1996; Feris et al. 2006; Morgand and Tamaazousti 2014; Aggarwal and Namboodiri 2016). These advancements have provided the field with unified tools and models to accurately simulate and define specularity. By combining mathematical models, such as those used for BRDF acquisition, with physically-based rendering techniques, the specularity in computer graphics has been more precisely defined, allowing for more realistic rendering of various material properties, from diffuse surfaces to highly reflective ones.

2.3 Specularity in computer vision

In the field of CV, specularity has been interpreted in various ways depending on the specific tasks and objectives (Mallick et al. 2006; Adato et al. 2007; Shen and Cai 2009; Lin and Shum 2001). Some CV tasks define specularity as reflective properties resembling mirror-like surfaces, focusing on entities like mirrors or polished metal that exhibit clear, well-defined reflections (Yang et al. 2019; Lin et al. 2020; Mei et al. 2021; Tan et al. 2023; Lin et al. 2023; Lin and Lau 2023; Anwer et al. 2023; He et al. 2023). Other tasks take a broader approach, defining specularity as any reflective characteristic present on surfaces, such as the irregular shine of smooth, metallic objects (Lippincott and Stark 1982; Shen and Zheng 2013; Morel et al. 2006, 2005; Fu et al. 2020; Son et al. 2020; Chen et al. 2023). Even more expansive definitions include reflections from less conventional materials, such as water surfaces or glass, where the reflective properties are less defined but still impactful on scene interpretation. These varying interpretations reflect the diverse challenges posed by specular reflections across different CV applications.

Traditionally, specularity in CV has been viewed as an artifact: something to be removed, particularly when it interferes with object recognition or surface detail recovery (Lippincott and Stark 1982; Lvovsky 2013; Shen and Zheng 2013; Morel et al. 2006, 2005, 2006). Early approaches treated specularity as a nuisance, especially in scenarios where reflections distorted the underlying material properties, and developed techniques to mitigate these effects by separating specular reflections from diffuse ones. Techniques such as specular flow have been proposed to detect and interpret how specular reflections distort based on surface curvature, revealing insights into surface geometry (Roth and Black 2006). However, as the field has advanced, researchers began to recognize that specular highlights can also provide valuable information about an object's surface properties, geometry, and material characteristics (Fu et al. 2020). Studies on specularity in static scenes have explored how reflections deviate from straight trajectories during camera movement, providing clues about the surface's orientation and geometry. This understanding has led to a more nuanced treatment of specularity, where reflections are considered informative rather than merely problematic. Specularity, in these contexts, is not only something to be detected and removed but also a feature that can be leveraged to improve scene interpretation.

Building on our insights from the Subsects. 2.2 and 2.3, it is evident that specularity in CV must be unified under a broader and more consistent definition. All forms of reflective behavior fall under the domain of specularity whether it involves mirror-like entities, metallic surfaces, or water reflections. These reflections share common physical and perceptual characteristics: they occur due to light interacting with smooth or semi-smooth surfaces and produce highlights that vary based on the surface material, geometry, and viewing angle. Thus, whether the task involves detecting specular reflections from mirrors, polished metal, or irregular surfaces such as water or glass, these are all valid manifestations of specularity.

Therefore, this review suggests that specularity in computer vision (CV) should be defined not by the specific material or task, but by its fundamental physical and perceptual traits. Specularity can be characterized as any instance where light reflects off a surface in a concentrated manner, producing directional highlights. This broader definition encompasses various manifestations of specularity, including those observed on mirrors, polished metals, and irregular surfaces like water and glass, aligning with both the physical properties discussed in the Subsect. 2.1 and the models used in CG. Adopting this consistent definition across CV tasks will allow researchers to develop more coherent methods for detecting, segmenting, and even leveraging specular reflections in various applications.

Furthermore, this review suggests refining the definition of specular reflection in this work by incorporating a more precise mathematical model based on physical optics. Specifically, the Fresnel equations (Lippincott and Stark 1982; Lvovsky 2013) model the behavior of light reflecting off smooth surfaces at the microscopic level. In practical scenarios, this model is often combined with diffuse reflections, where surface irregularities larger than the wavelength of light influence the overall reflection. To provide a consistent and unified framework, the specularity can be mathematically described by the following integral, based on the Fresnel equations:

$$L_s = \int_{\Omega} F(\theta_i, \theta_r) \cdot (\mathrm{n} \cdot \mathrm{l})^p \; d\Omega,$$

where L_s represents the specular reflection intensity, and Ω denotes the range of possible incident angles. The term $F(\theta_i, \theta_r)$ is the reflection coefficient derived from the Fresnel equations, which depends on both the incident angle θ_i and the reflection angle θ_r. Additionally, n is the surface normal vector, and represents the direction of the incident light. The exponent p controls the sharpness of the specular highlight, which is related to the surface roughness or smoothness.

This mathematical model-based definition provides a unified approach to specularity that applies across different surface types and reflection scenarios, offering a more consistent and coherent model for specularity detection in CV. The application of this model across different tasks, such as detecting specular reflections from mirrors, polished metals, or water surfaces, involves selecting specific subsets of this general model depending on the physical properties of the surface and the task requirements.

3 Specularity detection methods

With a well-defined understanding of specularity established in the previous section from geometrical, physical, and perceptual perspectives, this section now turns to how these insights guide the development of specularity detection methods in computer vision tasks. The theoretical foundation laid by these definitions provides clarity and structure, enabling a more systematic exploration of detection techniques. This section examines both traditional and deep learning-based approaches to specularity detection, each contributing unique advancements to the field. These methods are assessed for their capabilities, limitations, and the foundations they provide for future work.

The traditional approaches described in Subsect. 3.1 are reviewed chronologically, focusing on their dependency on external hardware, primary advantages, and contributions to the field. These methods utilized mathematical optimization and control and laid the groundwork for the subsequent research, particularly in their application to early CV systems. By evaluating traditional methods in terms of their hardware dependencies and overall impact, this section better demonstrates how they have shaped current approaches and what potential they hold for future advancements. Also, this section highlights the critical milestones in traditional methods and evaluates their strengths in enabling accurate specularity detection while acknowledging their limitations in terms of generalizability and scalability.

After exploring traditional methods, this section will transition to the review of deep learning-based approaches in Subsect. 3.2. With the advent of deep learning techniques, specularity detection has seen rapid progress, with significant advancements in network architectures and data-driven models. In this section, deep learning approaches are presented chronologically, but the primary focus is placed on the type of input data they utilize, resulting in a classification based on the following categories: methods that rely on RGB image input, methods that incorporate RGB-D image input, and those that leverage video stream input. Methods utilizing RGB-D input, in particular, offer promising potential, as they leverage depth information to enhance specularity detection in complex scenes. Recent advancements in three-dimensional (3D) vision, such as improved depth estimation and surface reconstruction techniques, have enriched the capabilities of RGB-D-based models, highlighting the growing intersection between specularity detection and 3D vision tasks. This categorization allows for a deeper understanding of how different data modalities impact model performance and design, as well as how deep learning techniques have evolved to accommodate various challenges associated with these input types.

3.1 Traditional approaches

Early methods for detecting and removing specularity laid a critical foundation for the later development of artificial intelligence (AI)-based techniques by addressing the complex challenges posed by reflective surfaces (Son et al. 2020; Fu et al. 2024). These traditional approaches, developed during the 1990s and early 2000s, emphasized the analysis of both the geometric and physical properties of reflections. This included factors such as surface normals, the positioning of light sources, and the interaction between diffuse and specular reflections. Understanding these core properties is essential for identifying and processing specular highlights, which often obscure important details in scenes and complicate feature extraction or surface reconstruction. This subsection will provide an in-depth analysis and comparison of representative traditional methods. Each method's key contributions and innovations at the time of their introduction will be discussed, as illustrated in Fig. 4.

In the early 1990 s, Wolff (1990) pioneered the use of polarization to distinguish between metallic and dielectric surfaces based on their reflection characteristics. His method, grounded in the Fresnel reflection model, demonstrated that specular reflections from dielectric materials are strongly polarized, while diffuse reflections are generally unpolarized. This polarization difference provided a reliable means for material classification, particularly in industrial machine vision tasks such as circuit board inspection. Wolff's work laid the foundation for later polarization-based approaches to specularity detection, emphasizing the importance of the polarization state of light in accurately isolating specular reflections.

Building upon these early insights, Umeyama and Godin (2004) introduced a more advanced polarization-based technique in 2004, which employed independent component analysis to separate diffuse and specular components in images. By capturing images through a rotating polarizer, their method exploited the polarization properties of specular highlights to distinguish them from the underlying diffuse reflection. This approach proved effective in real-world scenarios, where complex lighting and material interactions often complicate specularity detection. Their work expanded the use of polarization to not only classify materials but also to address the challenges of separating reflection components in natural scenes.

Another one of the earliest approaches, proposed in 1992 by Amanatides (1992), focused on detecting and eliminating specular aliasing. Their method incorporated geometrical properties such as surface normals, gaze direction, and light source positioning. By analyzing these parameters, they managed to remove specular aliasing without increasing the sampling rate, thereby optimizing the detection of specular highlights. Their work paved

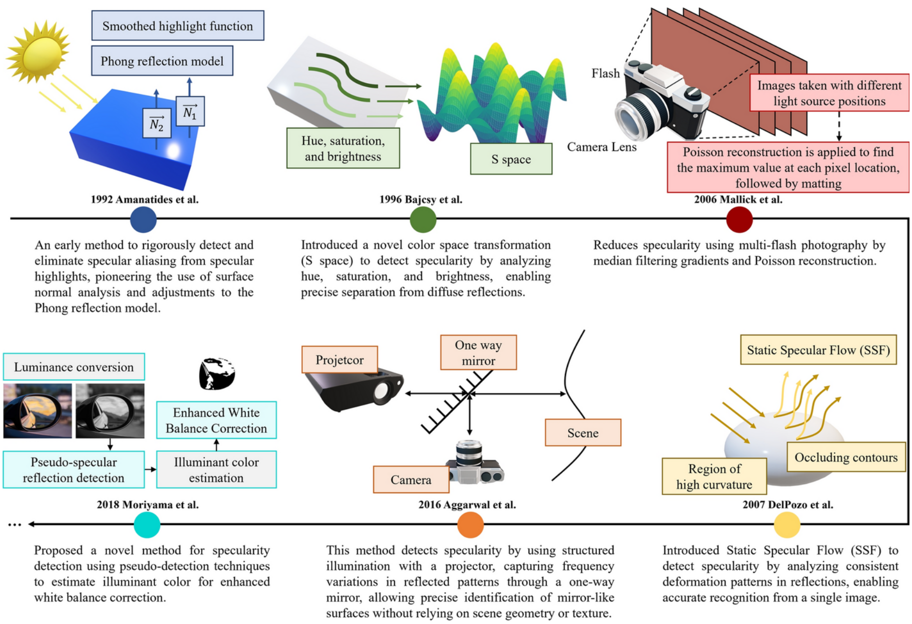

Fig. 4 Timeline illustrating key milestones in traditional specularity detection methods

the way for future geometric-based approaches to tackle specularity in scenes with complex light interactions.

Ching et al. (1993) proposed a specular highlights detection method utilizing multiple gray-scale images from different viewpoints. By leveraging key characteristics of specular highlights, such as strong intensity peaks and viewpoint dependence, the method reduced the ambiguities faced by traditional gray-scale techniques. Through de-emphasizing the brightest regions in intensity histograms during the correlation process, their approach improved the reliability of specularity detection. This method, while computationally efficient due to its reliance on gray-scale images, introduced challenges related to computational load when multiple views were needed, particularly in the presence of occlusions.

Following this, in 1996, Bajcsy et al. (1996) advanced the field by introducing a method that employed color space transformations to separate diffuse and specular reflections. Using the hue, saturation, and brightness (HSB) color space, they enhanced the contrast between specular and inter-reflections, significantly improving segmentation accuracy. Their innovative approach underscored the growing recognition of color space's role in distinguishing reflection types, marking an important step forward in specularity detection techniques.

By 2006, specularity detection was further advanced by Feris et al. (2006), who introduced a technique based on multi-flash photography. Their method used images captured from different light source positions to isolate specular highlights. Through the use of Poisson equations to model the gradient field, they were able to remove or reduce specular reflections. This approach shifted the focus from detecting specularity in a single image to using multiple images, enhancing accuracy in complex scenes. In 2007, DelPozo and Savarese (2007) extended this research by developing the static specular flow (SSF) method. Their technique analyzed consistent deformation patterns in reflections caused by specular surfaces with high curvature or occluding contours. By identifying these static flows, their method allowed for the detection of specularity using a single image. This advancement was significant for simplifying the detection process in natural scenes without relying on complex multi-image setups.

Entering the 2010s, new research efforts began to explore novel computational approaches for separating specularity in images, leading to significant advancements in the field. Among these, Shen and Zheng (2013) introduced a real-time specular highlight removal algorithm based on intensity ratios, marking an important contribution to specularity detection. The method is grounded in the observation that, for diffuse pixels, the intensity ratio between the maximum and range values (maximum minus minimum) remains independent of surface geometry. By leveraging this property, the algorithm efficiently computes the specular fractions of image pixels using intensity ratios.

For textured surfaces, pixels are grouped into clusters, and the intensity ratios of these clusters are used to classify reflections. This pixel-wise approach operates without the need for explicit specular pixel identification or local interactions, offering both computational efficiency and simplicity. Running four times faster than previous methods, the algorithm delivers improved accuracy in removing specular highlights. Shen et al.'s work offers a critical step forward in advancing real-time specularity detection, providing a scalable solution for real-world applications in dynamic environments.

In 2014, Morgand and Tamaazousti (2014) expanded on the work in color space analysis, specifically focusing on the hue-saturation-value (HSV) color space to process multiple illumination sources with varying intensities. Their approach focused on real-time processing of these illumination changes with fast detectors, allowing the method to deal with lighting jumps. By adjusting contrast and utilizing automatic thresholding techniques, their method effectively detected specular reflections, continuing the trend of leveraging color space transformations for enhanced accuracy.

Further advancements in specularity detection came in 2016, when Aggarwal and Namboodiri (2016) introduced a projector-camera system for mirror-like surface segmentation. By projecting structured sinusoidal patterns and analyzing the frequency variations in the reflected patterns, their method enabled precise identification of mirror-like surfaces. This approach leveraged the relationship between the captured local frequencies and the projected pattern, eliminating the need for complex scene geometry or texture data.

Furthermore, the 2018 study by Moriyama et al. (2018) proposed a method for illuminant color estimation that relies on the pseudo-detection of specular reflection, which is achieved by comparing the average pixel values of brighter and darker regions in an image. This approach stands out for its simplicity and its ability to effectively approximate specular highlights without the need for complex hardware or structured lighting setups.

In recent years, Wen et al. (2021) further advanced polarization-guided specular reflection separation by developing a unified optimization strategy that combines both RGB and polarimetric information. Their method introduces the concept of a polarization chromaticity image, which accurately clusters pixels with similar diffuse colors, enabling precise separation of specular and diffuse components. This approach addresses the challenges posed by varying illumination conditions and offers robust results in diverse environments. The integration of polarization with chromaticity in Wen's work represents a significant step forward in the refinement of polarization-based techniques for specularity detection, building on the foundational work of Wolff (1990) and Umeyama and Godin (2004).

One of its strengths is its capacity to improve white balance correction by isolating specular reflections from diffuse reflections, leading to more accurate color estimation. However, the method can struggle with highly complex scenes where specular and diffuse components are difficult to distinguish. The key advancement of their method is its novel use of localized pixel brightness analysis, which laid groundwork for further refinement in specularity detection techniques.

Table 1 compares the strengths, weaknesses, and method characteristics of all the traditional models referenced above by including a method description and exploring whether the method exploits additional hardware support or modifications. The differences between traditional methods are explicitly compared. Of particular note is the increasing number of methods that rely on external hardware support, modify existing imaging devices or propose new hardware-algorithm systems. The results suggest that these are a trend for traditional methods in the field of specular detection, where the combination with hardware can lead to better model performance during a period of slow breakthroughs in mathematical algorithms.

In addition to these observations, the foundational contributions of traditional specularity detection methods to AI-based techniques should be emphasized. These earlier approaches provided crucial insights into the geometric and physical behavior of reflective surfaces, such as surface normals and light interactions, which are now integral to AI models. By thoroughly analyzing these factors, traditional methods laid the groundwork for the development of AI models capable of recognizing and simulating the complex properties of specular reflections. The interaction between diffuse and specular reflections, long addressed

Table 1 Summary of traditional specularity detection methods: key features, contributions, and future directions for improvement

Method	Description	Hardware dependency	Advantages and contributions	Future directions and implications
Amanatides (1992)	Auxiliary algorithm	✗	1. One of the earliest works addressing specular aliasing, establishing a key foundation for specularity detection research. 2. Clamping specular function parameters improved detection efficiency without increasing sampling rates.	1. This method highlighted the need for future techniques to handle more complex surface reflectance models. 2. It suggested improvements for detecting specularity in high-curvature regions, encouraging more adaptive approaches.
Ching et al. (1993)	Detection algorithm	✗	1. Uses color image segmentation to distinguish specular and diffuse reflections effectively. 2. Efficiently handles dielectric materials under uniform illumination.	1. Could improve specularity detection in complex scenes with varied lighting. 2. Potential for real-time specularity detection in dynamic environments.
Bajcsy et al. (1996)	Detection algorithm	✗	1. Introduces a color image segmentation approach to distinguish specular and diffuse reflections effectively. 2. Handles dielectric materials under uniformly colored illumination, advancing specularity detection through the dichromatic model.	1. Further enhancement to handle more complex scenes with more diverse material properties. 2. Offers potential for handling occlusion alongside specular highlights in future work.
Feris et al. (2006)	Multi-flash camera system	✓	1. Innovatively uses a multi-flash camera setup to detect specularity by capturing images under varying lighting conditions. 2. It solves a Poisson equation, effectively minimizing specularities while preserving image texture.	1. Expanding the number of light sources could enhance specularity detection performance, particularly for high-curvature surfaces. 2. Sets the foundation for real-time applications in photography and vision systems, with a potential in dynamic scenes or textured regions.
DelPozo and Savarese (2007)	Detection algorithm	✗	1. Introduces static specular flow (SSF) for detecting specular surfaces, advancing low-level specularity detection. 2. Achieves strong classification accuracy, identifying reflective objects without prior shape assumptions.	1. Enhancing detection in complex scenes with intricate geometries and varied lighting could improve performance. 2. Integrating SSF with other cues could lead to more comprehensive real-world specularity detection systems.
Shen and Zheng (2013)	Removal algorithm	✗	1. Introduces a real-time method to separate specularity using intensity ratios. 2. Achieves superior computational efficiency without the need for iterative frameworks or specular pixel identification.	1. Further optimization could enhance performance on dark and achromatic surfaces. 2. Increasing robustness could allow for wider use in general-purpose real-time applications.

Table 1 (continued)

Method	Description	Hardware dependency	Advantages and contributions	Future directions and implications
Morgand and Tamaazousti (2014)	Threshold-based detection algorithm	✗	1. Introduces a threshold-based, real-time method using hue-saturation-value (HSV) color space to detect specularity under varying lighting conditions. 2. Handles over-saturation and lighting variations through contrast equalization and automatic thresholding.	1. Refining thresholding techniques could improve detection in highly reflective or complex environments. 2. Integrating multi-view data could further enhance robustness.
Aggarwal and Namboodiri (2016)	Structured illumination detection algorithm	✓	1. Utilizes a Projector-Camera system with structured light to segment specularity without needing prior environment information. 2. Detects specularity by analyzing local frequency changes, advancing in scenes with multiple reflections.	1. Further refinement of the Projector-Camera system could improve detection accuracy in complex scenes with inter-reflections.
Moriyama et al. (2018)	White balance estimation algorithm	✗	1. Proposes a white balance method based on pseudo-detection of specularities, improving the accuracy of illuminant color estimation. 2. Provides a non-iterative approach by the difference between brighter and darker pixel averages.	1. Optimizing parameter tuning could enhance the method's performance across diverse lighting conditions.

by traditional geometric and color space analysis, remains a challenge for AI-based vision systems. Furthermore, multi-modal data sources generated by traditional methods, such as multi-flash photography or structured lighting techniques, enriched the datasets that AI techniques now rely on, thereby enhancing the robustness and accuracy of modern specularity detection.

3.2 DL-based approaches

This subsection examines how recent DL-based models have more effectively extracted specularity features and computed the specularity color space inspired by traditional methods. These advancements in DL-based architectures have led to significant improvements in specularity detection compared to earlier techniques.

Fig. 5 provides a comprehensive overview of the evolution of DL-based models for specularity detection, illustrating key advancements since 2019. The models are organized according to input data types, RGB images, RGB-D data, and video stream showcasing the progression in both data handling and architectural complexity. The figure highlights how early models focused on feature extraction and context enhancement. In contrast, more recent models have introduced advanced techniques such as depth information integration, self-attention mechanisms, and uncertainty modeling. This timeline-based diagram demonstrates a clear shift from simpler architectures that emphasize basic feature extraction and fusion to more sophisticated frameworks incorporating multiscale features, attention-based modules, and methods for handling ambiguous specular reflections. Additionally, the figure

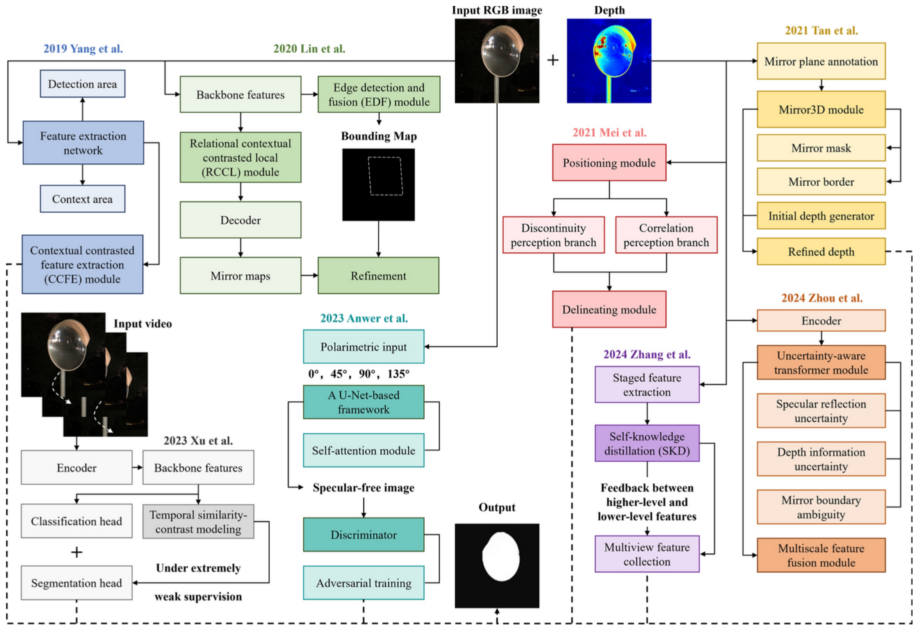

Fig. 5 Key milestones in deep-learning based specularity detection networks

captures the growing emphasis on robustness and generalization, with many recent models leveraging polarimetric input, depth uncertainty, and multiview feature aggregation to better handle complex scenes and challenging specular conditions.

3.2.1 Methods with RGB image input

In the context of specularity detection, RGB image input has been widely utilized due to the simplicity and direct access to color information from standard cameras (Süsstrunk et al. 1999). Many early models in specularity detection relied heavily on the RGB color space to analyze specular highlights, separating them from diffuse reflections. These methods focus on leveraging the color and intensity variations in the RGB channels to isolate and detect specularity in scenes. Although effective in many controlled scenarios, these early methods were often limited in generalization and robustness, especially in complex environments where lighting conditions varied significantly.

Contextual contrast feature extraction. A key milestone in using RGB image input for specularity detection is MirrorNet, introduced by Yang et al. in 2019 (Yang et al. 2019). This model represents a significant advancement in specularity detection, particularly in mirror detection tasks, and laid the foundation for much of the subsequent work in this field. MirrorNet's architecture successfully addressed the challenges associated with detecting mirrors, which often involve complex reflections and boundary ambiguities. For a considerable period, MirrorNet was considered the baseline for specularity detection, especially in the context of mirror detection and segmentation. Additionally, Yang et al. provided the first large-scale open-source mirror dataset (MSD), which became an essential resource for evaluating various object detection and specularity detection models. This section explores how MirrorNet and similar RGB-based models have evolved to meet the demands of specularity detection, leveraging their strengths and addressing their limitations.

Overall, Yang et al. developed an innovative contextual contrast feature extraction (CCFE) module inspired by how humans distinguish mirrors or highly reflective objects by detecting discontinuities. Humans typically focus on slight interruptions in texture or subtle color changes, which is mirrored in CV systems by targeting weak contrasts between mirrored and non-mirrored regions. The CCFE module is designed to aggregate long-range contextual contrast information hierarchically, allowing for the detection of mirrors of different sizes. As shown in Fig. 6, MirrorNet starts by passing a single input image through a feature extraction network (FEN) to extract multi-layer features. The deepest features, rich in semantic information, are sent to the CCFE module to enhance contrast features and generate a coarse mirror map. This attention-based map suppresses non-mirror noise, enabling better localization of the mirror region. Through a process of coarse-to-fine refinement, the model up-samples and refines the mirror map to output a high-resolution image.

However, MirrorNet has its limitations. Its heavy reliance on the boundary contrast between mirror and non-mirror regions significantly reduces the accuracy of mirror detection in cases where the contrast between the interior and exterior of the mirror is not distinct. This boundary-based approach struggles in low-contrast scenarios, making it difficult for the model to detect mirrors in real-world environments consistently.

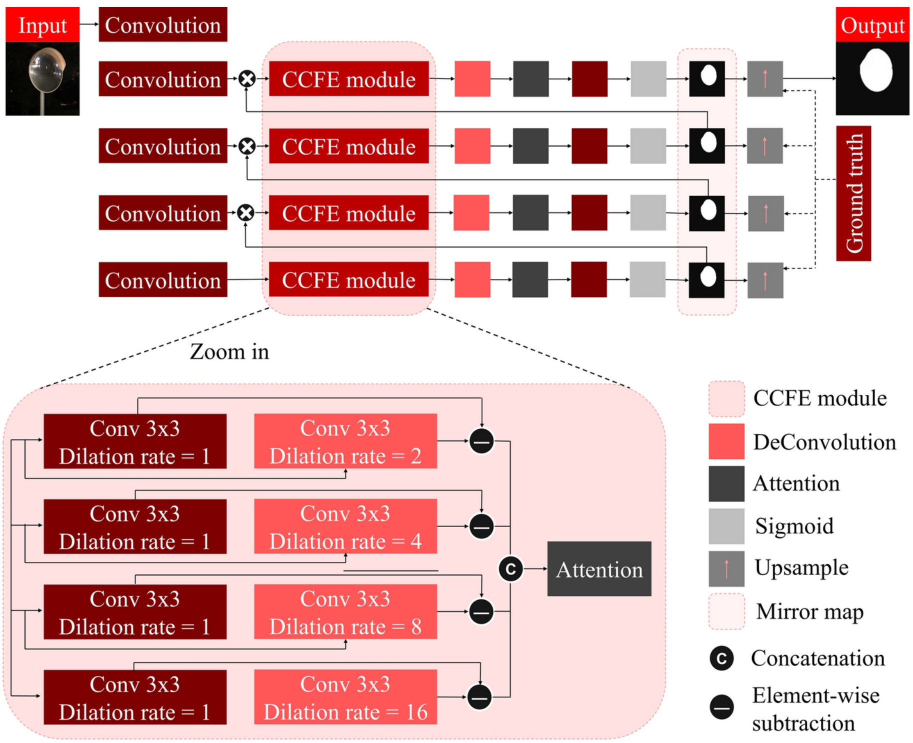

Fig. 6 The structure of MirrorNet (Yang et al. 2019), which consists of a feature extraction network and an embedded CCFE module

Relational context and edge fusion. To address these limitations, Lin et al. proposed the progressive mirror detection (PMD) model in 2020 (Lin et al. 2020), which delivers improved results. The PMD model combines two innovative modules: the relational contextual contrasted local (RCCL) module and the edge detection and fusion (EDF) module. The key idea behind PMD is to progressively explore the relationship between objects inside and outside the mirror, looking for correspondences that can help identify potential mirror regions. Once the model learns these relationships, it can infer the mirror regions more effectively. These regions are then refined through explicit edge detection, utilizing local features to enhance precision. By integrating relational context through the RCCL module and refining mirror boundaries with the EDF module, PMD resolves the contrast dependency issue of MirrorNet and demonstrates greater robustness in handling challenging, low-contrast scenes.

The proposed progressive mirror detection model is based on two novel modules. RCCL is used to extract all the contextual contrasts in the image and the relational features therein to locate all potential mirror regions as far as possible. EDF, on the other hand, targets multi-scale mirror edges. Combining these two modules, the model extracts the final exact mirror regions through a refinement network. The workflow will be introduced in Fig. 7, the multi-scale features of the input image are extracted by the baseline backbone network from (Xie et al. 2017), which are then extracted by RCCL and turned into relational contextual features. The role of the decoders in the network model structure is to output a mirror map using the features generated by the RCCL. While the EDF module produces a boundary map based on the input features from the previous modules and the mirror edge features. Finally the process moves towards the refinement module, which takes all the predictions and outputs a mirror map.

YOLO-based method. In 2022, Mirror-YOLO (Li et al. 2022), based on the YOLOv4 backbone (Bochkovskiy et al. 2020), introduced a novel approach to mirror detection by integrating advanced attention mechanisms and instance segmentation capabilities. The core innovation lies in its attention focus mechanism using the convolutional block attention module (CBAM) (Woo et al. 2018). This mechanism enables the model to selectively highlight essential channel and spatial features, allowing for enhanced detection of low-contrast mirror boundaries, which had been a significant challenge in previous models. Additionally, Mirror-YOLO incorporates a hypercolumn-stairstep feature fusion strategy in its feature

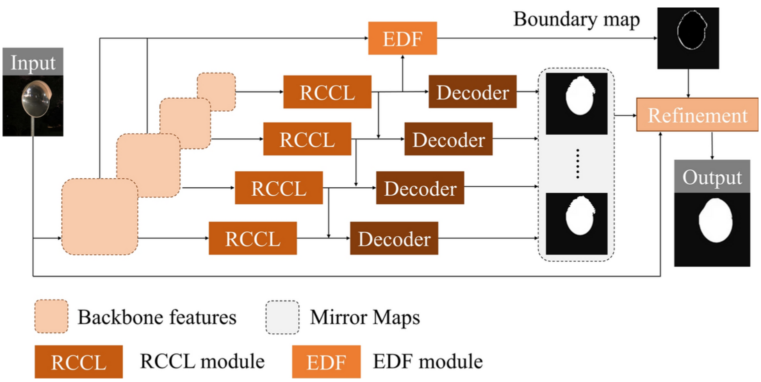

Fig. 7 Overview of the PMD model (Lin et al. 2020)

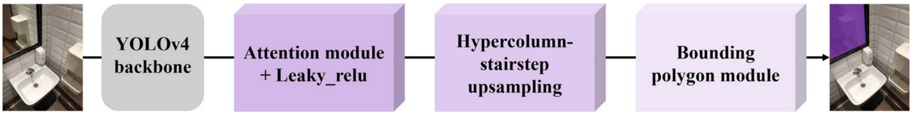

Fig. 8 The overall structure of Mirror-YOLO (Li et al. 2022)

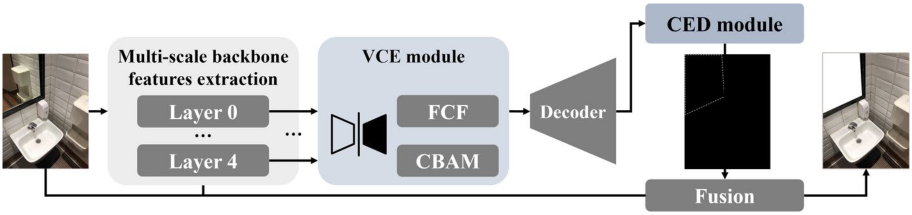

Fig. 9 The pipeline of VCNet (Tan et al. 2023)

extraction layers. Unlike standard fusion techniques, this approach incrementally scales feature maps to improve spatial consistency across layers, resulting in smoother and more coherent features that bolster detection accuracy. Mirror-YOLO further introduces mirror bounding polygons, such as segmentation, enhancing its ability to delineate mirrors with irregular contours more accurately. The overall pipeline is shown in Fig. 8.

Despite its considerable results, Mirror-YOLO has several limitations. One primary drawback is the increased computational cost due to the integration of the CBAM and its feature fusion strategy. These components significantly raise the model's complexity, making it less suitable for real-time applications, especially in high-resolution scenarios. Additionally, the reliance on attention mechanisms can lead to potential performance bottlenecks in resource-constrained environments or edge devices, which are becoming increasingly common in applications requiring mirror detection.

Chirality-guided boundary detection. To address the persistent challenge of accurately identifying mirror boundaries in low-contrast and visually ambiguous settings, Tan et al. introduced an innovative approach in the form of visual chirality called VCNet in 2023 (Tan et al. 2023). Unlike previous methods that rely on contextual cues or relational correspondences between mirror and non-mirror regions, this method leverages the inherent left-right asymmetry, i.e., chirality, found in mirrored reflections. This novel strategy allows the model to discern mirrored regions even when other boundary cues are weak or absent, significantly enhancing its robustness and accuracy in complex scenes. The proposed VCNet achieves this by incorporating two primary modules as shown in Fig. 9: the visual chirality embedding (VCE) module, which captures chirality features through a flipping-convolution-flipping (FCF) operation, and the chirality-guided edge detection (CED) module, which integrates multi-scale chirality information to refine the detection of mirror boundaries.

By focusing on these chirality-based features, VCNet offers a marked improvement in handling intricate, poorly defined mirror boundaries compared to earlier models, showcasing the potential of visual chirality as a powerful cue in specularity detection. However, while visual chirality introduces a unique advantage, it also presents limitations when addressing occlusions or highly complex mirror surfaces where chirality cues may be less effective. This dependency highlights areas for further exploration, such as refining chirality feature extraction to bolster mirror detection in diverse, real-world environments.

Unsupervised representation learning. Following the VCNet (Tan et al. 2023), a self-supervised approach to specularity detection emerged, with Lin et al. introducing a new self-supervised learning (SSL) pre-training framework (Lin and Lau 2023) that addresses several limitations seen in earlier models. As shown in Fig. 10, this SSL framework presents a new paradigm by shifting the focus from supervised pre-training to task-specific, unsupervised learning. Specifically, it operates through a three-stage pre-training strategy. The image-level pre-training stage equips the model with the foundational knowledge of reflection properties by distinguishing between original and flipped images. This stage builds essential global understanding, capturing fundamental geometrical cues associated with mirrored surfaces. Next, the patch-level pre-training stage refines the model's sensitivity to local reflection patterns, enhancing its capacity to identify spatial coherence between mirrored and non-mirrored regions. Finally, the pixel-level pre-training stage enables the model to reconstruct masked areas within mirror images, honing its ability to recognize fine-grained details crucial for accurately identifying mirror boundaries.

Compared to models like MirrorNet (Yang et al. 2019) and PMD (Lin et al. 2020), which rely heavily on contextual contrast and progressive boundary refinement, Lin et al.'s SSL framework reduces dependency on labeled data by training the model to understand intrinsic mirror properties through task-specific cues. This unsupervised learning strategy provides notable advantages in generalization and adaptability. However, the model may still encounter challenges when handling low-contrast mirrors or surfaces that lack distinct reflective features. Despite these limitations, this innovative approach highlights the potential of self-supervised learning in specularity detection, laying the groundwork for future models that aim to reduce reliance on extensive labeled datasets while improving detection robustness in complex environments.

Polarimetric data for specularity-free images. In the same year, Anwer et al. introduced another distinctive approach that leverages polarimetric imaging to dynamically model illumination variations (Anwer et al. 2023). Their proposed multi-domain specular highlight mitigation generative adversarial network (SHMGAN) employs polarimetric data as the foundation for both specular highlight detection and mitigation. Unlike earlier models, which rely on RGB or contextual cues, SHMGAN captures the nuanced interactions between light and material surfaces, using polarimetric input to create specularity-free images. This feature makes it particularly powerful for applications in environments with complex lighting conditions or materials exhibiting irregular reflective properties.

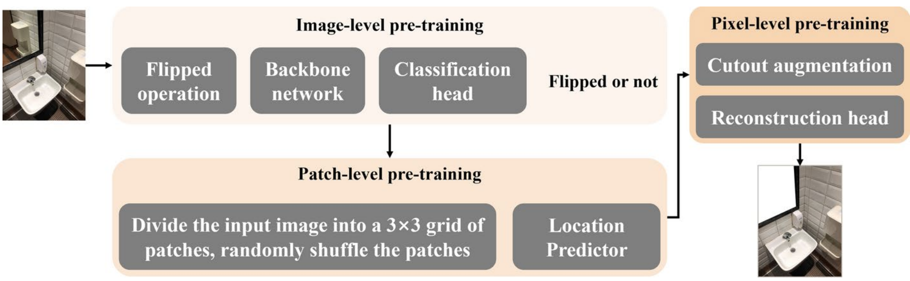

Fig. 10 The new SSL pre-training framework (Lin and Lau 2023)

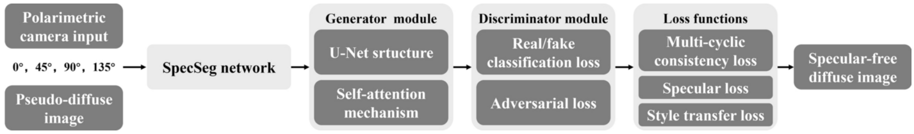

Fig. 11 The illustration of SHMGAN's architecture (Anwer et al. 2023)

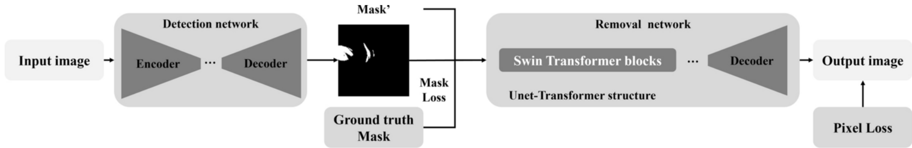

Fig. 12 The UNet-transformer-based specularity detection framework (Wu et al. 2023)

SHMGAN's architecture integrates a specular segmentation (SpecSeg) network with a GAN-based mitigation model, creating a joint framework capable of detecting and removing specular highlights, shown in Fig. 11. The SpecSeg network isolates reflective regions by leveraging information from polarimetric variations, while the GAN component generates high-quality images devoid of these highlights. Compared to earlier works, polarimetric data input grants it a unique advantage, particularly in handling extreme lighting contrasts and challenging, reflective surfaces. In addition, this polarimetric approach broadens the scope of specularity detection, allowing it to operate in scenarios where traditional methods may be less effective.

The reliance on polarimetric data requires specialized equipment, limiting its feasibility in general-use applications. Additionally, the robustness of SHMGAN in highly dynamic or occluded scenes remains an area for further refinement. However, its use of polarimetric imaging introduces a promising direction for future research. It suggests potential applications beyond computer vision, such as in medical imaging or material inspection, where specularity control is essential for clarity and precision.

UNet-transformer-based method. A notable advancement in the same year was achieved by Wu et al., who proposed an UNet-transformer architecture designed for joint specular highlight detection and removal (Wu et al. 2023). This model combines the strengths of UNet's encoder-decoder framework with the global feature-capturing capabilities of transformers, particularly the Swin transformer (Liu et al. 2021). By leveraging transformers, the model improves contextual understanding, especially in complex and dynamic specular highlight areas, where capturing relationships between highlight and diffuse regions is critical. This integration allows the network to model the mapping between specular highlights and their surroundings, which reduces uncertainty in highlight-dense regions and enhances removal accuracy. This UNet-transformer-based approach incorporates a two-branch system which is shown in Fig. 12: the first branch detects specular highlights using UNet, generating a highlight mask, while the second branch performs highlight removal, guided by the detected mask. This guidance mechanism minimizes chromatic aberrations and color inconsistencies, which are common challenges in specular removal tasks. By structuring the network this way, Wu et al. address the limitations of earlier detection and removal models, where detection errors or lack of guidance often lead to noticeable artifacts. The integration of the transformer into the removal branch enables the model to capture both local details and global contextual cues, making it effective in handling both small and large specular regions.

However, this model's reliance on the Swin transformer's self-attention mechanism introduces potential computational challenges, mainly when processing high-resolution images or scenarios that require real-time responsiveness. Furthermore, while the model demonstrates robust performance on individual images, its ability to generalize to video sequences or multi-view inputs has not yet been thoroughly investigated. This limitation opens avenues for future research, as an adequate adaptation to sequential or multi-angle data could significantly enhance its applicability. Thus, the subsequent sub-subsection 3.2.3 will introduce models specifically designed for handling video sequence inputs, exploring their advantages in dynamic scenes.

Spatial-frequency fusion. In 2024, Xie et al. introduced the cross-space-frequency window transformer model (CSFwinformer), a novel framework that enhances mirror detection by effectively leveraging spatial and frequency domain features (Xie et al. 2024). The key innovation in CSFwinformer lies in its cross-space-frequency window alignment module (SFWA), which aligns spatial and frequency features, enabling the model to capture nuanced mirror textures across various video frames. This alignment is particularly beneficial for detecting mirrors in dynamic scenes, where spatial cues alone may be insufficient. CSFwinformer integrates a dilated window attention module (DWA) to address the need for a global context, which is crucial in distinguishing mirror regions from complex backgrounds while preserving local details. Additionally, the model incorporates a cross-modality context contrast module (CMCC), aligning spatial and frequency modalities, allowing for a comprehensive understanding of reflective properties in different scenarios. The overall pipeline is shown in Fig. 13.

An essential component of CSFwinformer is its spatial attention module (SAM) (Kirillov et al. 2023), which enhances the model's ability to focus on relevant regions by dynamically assigning attention weights to spatial features based on their importance. The SAM plays a critical role in refining the mirror detection process by directing the model's focus toward significant areas within each frame, thereby enhancing the accuracy of segmentation results, particularly in complex backgrounds. By concentrating on high-impact areas, SAM reduces the influence of irrelevant regions, improving the overall efficiency and effectiveness of the model.

One of CSFwinformer's significant advantages is its ability to enhance detection robustness without relying on depth information, unlike many other models that incorporate RGB-D data. However, this transformer-based structure and the computational complexity associated with cross-modality alignment impose high resource demands, potentially limiting the model's applicability in real-time and resource-constrained scenarios. Future

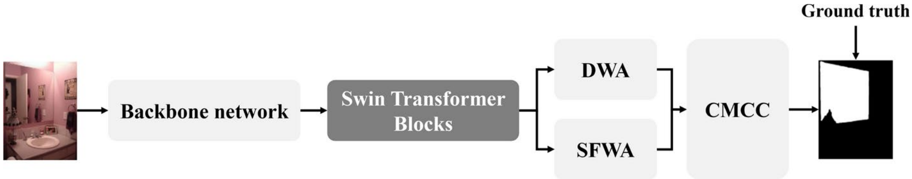

Fig. 13 The architecture of CSFwinformer, incorporating modules like SFWA, DWA, and CMCC to enable effective spatial-frequency alignment and robust mirror detection across complex specular scenes (Xie et al. 2024)

research could explore optimizing CSFwinformer's computational efficiency through techniques such as model pruning or designing lightweight transformer architectures, which would make it feasible for real-time mirror detection applications. Additionally, introducing adaptive frame sampling strategies may further improve efficiency by balancing computational load while maintaining accuracy.

Given the effectiveness of CSFwinformer's spatial-frequency approach, future extensions could consider integrating depth information. This inclusion would allow the model to address challenges associated with depth discontinuities often found in reflective and specular surfaces. Incorporating RGB-D data would provide the model with enhanced spatial structure, thereby improving its segmentation capabilities in complex environments. This potential expansion sets the stage for exploring models that combine spatial, frequency, and depth features, which could deliver more robust performance in specularity detection across diverse conditions. Therefore, this review will continue to discuss in depth the specularity detection model with RGB-D images as input in the following sub-subsection 3.2.2.

3.2.2 Methods with RGB-D image input

In specularity detection, RGB-D image input introduces a significant advancement by incorporating depth information alongside traditional RGB channels (Song et al. 2019). Unlike standard RGB data, which captures only color and intensity, RGB-D data provides an additional dimension of spatial structure, allowing models to differentiate between specular and diffuse regions with greater accuracy (Shaikh and Chai 2021). The depth channel enables a more detailed understanding of the scene's geometry, which is particularly beneficial in detecting specular boundaries and handling ambiguities in reflective surfaces. This depth information proves invaluable in challenging environments, where variations in lighting and reflections can complicate detection. By using RGB-D input, recent methods have developed more robust models that can better generalize across diverse and complex scenes, enabling improved accuracy in identifying specularities and reflective objects across various spatial conditions.

Depth-aware boundary enhancement. In 2021, Mei et al. introduced an innovative positioning and delineating approach called PDNet to specularity detection with depth-aware solution shown in Fig. 14, marking the first method that combines RGB and depth information for enhanced mirror boundary detection (Mei et al. 2021). Building upon the

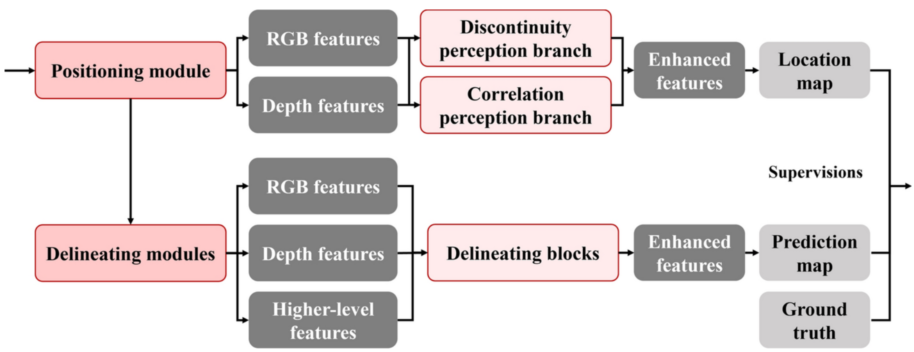

Fig. 14 The structure of proposed positioning and delineating modules in PDNet (Mei et al. 2021)

foundation laid by earlier works such as MirrorNet (Yang et al. 2019), PDNet utilizes depth cues to address limitations associated with RGB-only methods. By capturing depth discontinuities, PDNet significantly improves boundary accuracy, especially in scenarios where RGB data alone does not clearly delineate between mirrored and non-mirrored regions. This integration of depth information enables the model to overcome ambiguities in reflection-heavy environments, establishing a more reliable framework for specularity detection.

Another notable contribution is the RGBD-Mirror dataset, designed for rigorous evaluation in RGB-D mirror segmentation tasks. This pioneering dataset, combined with PDNet's depth-aware detection, establishes a new standard for leveraging depth in specularity detection, addressing challenges that RGB-only approaches face.

However, the reliance on high-quality depth data introduces potential limitations, particularly in environments where depth information may be noisy or incomplete. This dependency may reduce the PDNet's adaptability in diverse settings with variable depth quality. Despite these constraints, Mei et al. paves the way for future research that integrates RGB with depth, setting a precedent for combining multi-modal data to tackle complex visual ambiguities in specularity detection.

Depth-enhanced 3D plane. Tan et al.'s work introduced Mirror3DNet (Tan et al. 2021), pioneering the use of 3D mirror plane prediction to advance specularity detection. This approach addresses a critical aspect of specular surfaces: the inability of conventional depth sensors to capture accurate depth information for mirrors. In the context of specularity detection, predicting the 3D plane of mirrors is essential, as specular surfaces disrupt depth continuity and create ambiguity in scene interpretation. Using RGB-D data, Mirror3DNet refines depth perception within specular regions, thereby enhancing segmentation accuracy and robustness in detecting mirrors and reflective surfaces. Mirror3DNet's primary focus is its dual-stage prediction process shown in Fig. 15, integrating Mask R-CNN (He et al. 2017) for initial mirror region detection and PlaneRCNN (Liu et al. 2019) for 3D plane estimation within those identified areas. The model first uses RGB data to identify potential mirror areas, then applies depth data to predict the 3D mirror plane, effectively correcting for erroneous depth readings in reflective regions. This method significantly strengthens

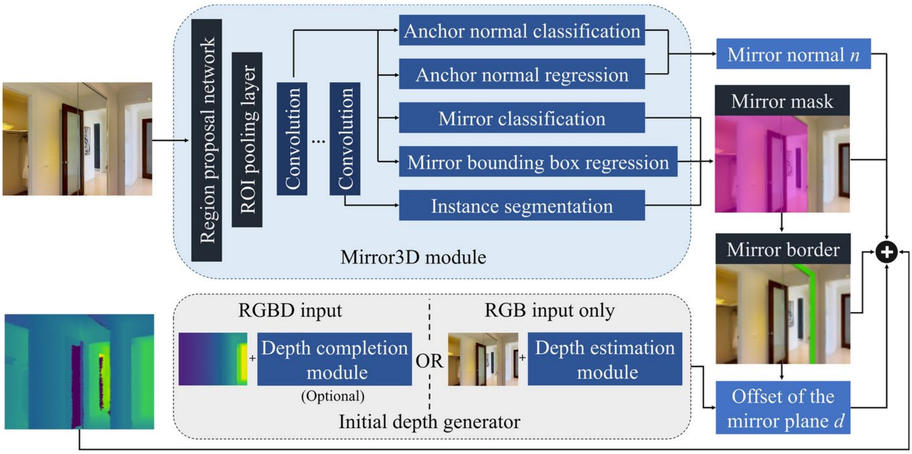

Fig. 15 The general diagram showing Mirror3DNet architecture as applied to two types of input (Tan et al. 2021)

specularity detection, as it reconstructs a reliable depth map that accurately detects specular planes from the surrounding environment. Additionally, the introduction of the Mirror3D dataset, which includes annotated 3D mirror planes, provides a rich training resource that captures the unique reflective characteristics of mirrors, enhancing depth-based segmentation in complex indoor scenes. Furthermore, Tan et al. were the first to address the scarcity of datasets for 3D specularity detection by providing 3D mirror surface annotations for three widely used RGB-D datasets: NYUv2 (Silberman et al. 2012), Matterport3D (Chang et al. 2017), and ScanNet (Dai et al. 2017). These datasets were optimized and enhanced through a process of data fusion, creating a more comprehensive resource tailored to the unique challenges of 3D specularity detection.

While Mirror3DNet achieves notable depth refinement for mirrors, the accuracy of its 3D plane predictions can be affected by the initial segmentation quality and the completeness of surrounding depth information. In complex environments with disrupted depth continuity, such as highly reflective settings, the model may face challenges in consistently identifying mirror boundaries and accurately estimating depth. Future research could explore integrating adaptive depth correction methods to handle variable data quality, as well as multi-view approaches to strengthen depth estimation in reflective regions. Additionally, incorporating techniques that analyze reflected objects within mirrors could enhance depth perception and overall scene interpretation, addressing some of the limitations observed in single-view scenarios.

Uncertainty-guided localization. In 2024, Zhou et al. introduced UTLNet, a pioneering model designed to handle the ambiguities in mirror boundaries using RGB-D data through a novel uncertainty-aware transformer framework (Zhou et al. 2024). UTLNet incorporates a specialized uncertainty-aware module that works within the transformer's attention mechanism, effectively managing low-confidence regions in both RGB and depth channels. By leveraging the unique collapse-and-expand uncertainty aggregation (CEUA) module, UTLNet enhances the precision of mirror boundary segmentation by dynamically adjusting feature fusion based on input uncertainties. This structured approach shown in Fig. 16 directly addresses challenges in boundary delineation, marking a significant advancement in managing visual ambiguity for specularity detection. UTLNet's architecture, by integrating uncertainty estimation with transformer-based processing, ensures robust segmentation performance across complex reflective environments.

The advantage of UTLNet lies in its capacity to capture long-range dependencies within reflective scenes, thanks to the CEUA module, which selectively merges RGB and depth features. This architecture positions UTLNet to manage complex scenarios with inconsistent or ambiguous depth information, making it particularly effective in challenging indoor environments. Future work could focus on enhancing UTLNet's efficiency, potentially through lightweight adaptations of the transformer structure. Additionally, exploring adaptive resolution and sampling techniques could help optimize the model's computational demands while preserving its accuracy in real-time applications. This would allow UTLNet

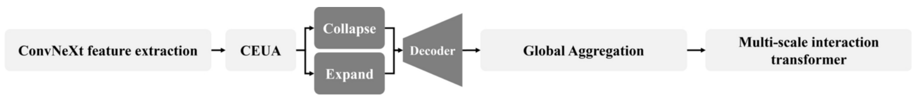

Fig. 16 The general framework of UTLNet and the corresponding CEUA architecture (Zhou et al. 2024)

to maintain high segmentation quality under varying resource constraints, broadening its applicability in practical, dynamic environments.

Knowledge distillation-based strategy. In 2024, two significant approaches using knowledge distillation were introduced for RGB-D mirror segmentation, each with distinct strategies and contributions to specularity detection. The first model, SEMCNet, utilizes a self-knowledge distillation strategy with a staged extraction and multiview collection framework (Zhang et al. 2024). SEMCNet employs both CNNs and transformers to iteratively refine feature extraction across different perspectives, such as normal, expanded, and remote views. This multiview strategy enhances the model's ability to capture subtle specular details by leveraging depth data alongside RGB inputs, making it particularly effective in handling ambiguous or complex mirror regions. By utilizing self-knowledge distillation, SEMCNet achieves improved segmentation accuracy through continuous learning from its predictions, thereby enriching the detection of specular features in challenging environments.

In contrast, the second model introduces MGNet, a morphology-guided network that focuses on efficiency and adaptability for resource-constrained applications (Zhou et al. 2024). MGNet consists of a teacher-student setup, where the teacher network (MGNet-T) incorporates an erosion-dilation fusion module (EDFM) to enhance the detection of mirror edges and maintain essential textures. Through knowledge distillation, the smaller student model (MGNet-S) inherits key insights from MGNet-T, enabling it to achieve high accuracy while maintaining a lightweight structure. This design prioritizes the purification of RGB-D features by filtering noise in depth data and optimizing texture clarity in RGB images, making MGNet-S suitable for real-time applications without sacrificing segmentation quality.

As shown in Fig. 17, the primary difference between SEMCNet and MGNet lies in their design goals: SEMCNet emphasizes robust feature extraction through multiview collection for high accuracy in complex environments, while MGNet focuses on creating a compact and efficient model optimized for edge and texture preservation. Both approaches underscore the versatility of knowledge distillation in RGB-D mirror segmentation, showing potential pathways for further enhancing specularity detection through adaptive feature refinement and resource-efficient designs.

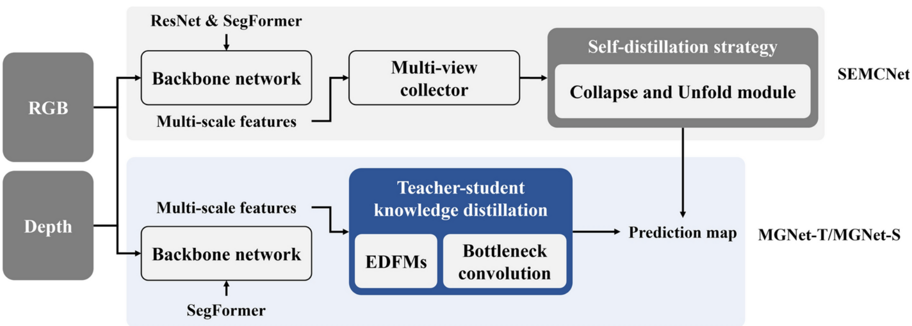

Fig. 17 Comparative architectures of SEMCNet (top) (Zhang et al. 2024) and MGNet (bottom) (Zhou et al. 2024), both utilizing knowledge distillation for RGB-D mirror segmentation. SEMCNet emphasizes feature refinement through self-distillation, while MGNet employs a teacher-student framework for efficiency

3.2.3 Methods with video stream input

The use of video stream input in specularity detection provides a unique advantage by incorporating temporal information from consecutive frames, adding depth to the analysis beyond what is possible with individual RGB or RGB-D images. Unlike static frames, video data captures the dynamic evolution of lighting, reflections, and object movements, enabling models to track and interpret specular properties over time (Jiao et al. 2021). This temporal context assists in differentiating persistent specular highlights from transient reflections or environmental noise, especially in complex, real-world scenarios with fluctuating lighting and reflective conditions. The sequential nature of video allows models to aggregate information about the consistency and behavior of reflective surfaces, uncovering specularities that might otherwise be missed in single-frame analysis (Zhu et al. 2020). By leveraging the continuity provided by video input, recent approaches have introduced advanced models capable of achieving greater robustness and accuracy in detecting specularities, particularly in dynamic and constantly changing environments.

Dual correspondence modeling. In 2023, Lin et al. introduced VMD-Net (Lin et al. 2023), a model for video mirror detection that utilizes dual correspondence (DC) through spatial and temporal cues. The DC module operates in two stages: first, it captures intra-frame correspondences with a global relation (GR) block for spatial analysis within frames and a cross-attention (CA) module to establish short-term temporal correspondences between frames. The second stage extends this to long-term temporal correspondences across the video sequence, using reverse cross-attention to enhance the relationship between reflections and real objects. As shown in Fig. 18, this dual-stage approach enables VMD-Net to achieve high accuracy in detecting mirrors within dynamic scenes.

Yet, VMD-Net's emphasis on spatial correspondences reduces its precision in defining mirror edges, particularly in scenes with ambiguous boundaries. Furthermore, the absence of explicit edge supervision can lead to coarse mirror boundaries in complex environments. Future work could enhance this model by integrating edge-specific supervision or refining the cross-attention mechanism to capture finer structural details. Additionally, incorporating multi-scale temporal correspondences could strengthen its robustness across varied specular scenarios.

Temporal similarity modeling. Next year, Xu et al. introduced the ZOOM model, which uses extremely weak supervision to enhance video-based mirror detection by leveraging temporal similarity-contrast across frames (Xu et al. 2024). Unlike conventional methods that rely on dense pixel annotations, ZOOM requires only binary indicators to identify the presence of mirror regions, making it a cost-effective solution. The model's core innovation lies in its ability to identify mirror reflections by analyzing the consistency of temporal features across adjacent frames. This approach highlights regions with high inter-frame similarity as potential mirrors while contrasting them against non-mirror regions that display more significant temporal variability. Such a design addresses the challenge of detecting

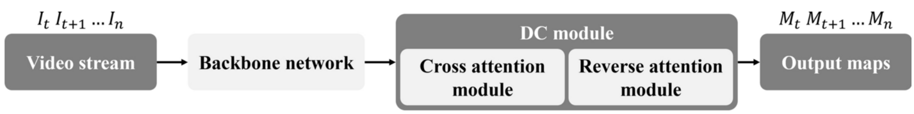

Fig. 18 The framework of VMD-Net with the proposed DC module (Lin et al. 2023)

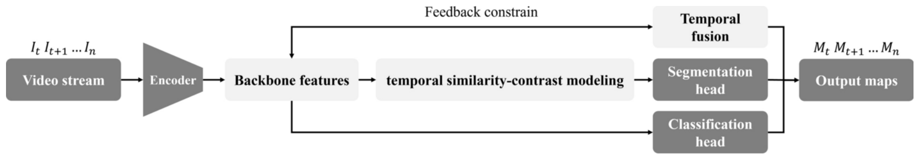

Fig. 19 The overview pipeline of ZOOM model (Xu et al. 2024)

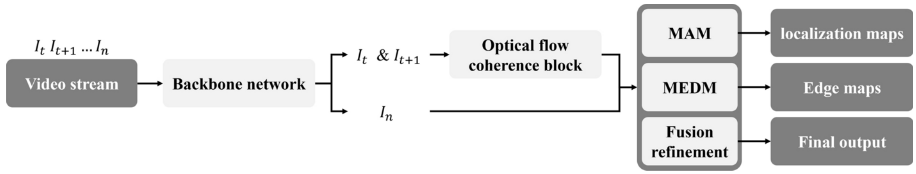

Fig. 20 MG-VMD model pipeline for video-based specularity detection, using MAM, MEDM, and fusion refinement modules to produce localization and edge maps (Warren et al. 2024)

mirrors in dynamic video environments by maximizing feature consistency across frames, thus improving specularity detection accuracy without intensive manual labeling.

Additionally, ZOOM introduces a feedback constraint mechanism shown in Fig. 19 that refines localization through temporal fusion, generating pseudo-labels that improve mirror segmentation. However, ZOOM encounters challenges in highly reflective or low-contrast scenarios where binary indicators may fail to capture fine-grained specularity details. Future work could explore the integration of attention mechanisms or multi-frame aggregation to improve its sensitivity to complex reflective patterns, while also investigating self-supervised techniques to further minimize dependency on labeled data.

Motion cue exploitation. In the same year, Warren et al. introduced the motion-guided video mirror detection (MG-VMD) model, leveraging the unique cue of motion inconsistency to enhance mirror detection in video streams (Warren et al. 2024). The core innovation of MG-VMD lies in its two key modules: the motion attention module (MAM) and the motion-guided edge detection module (MEDM). MAM identifies mirror regions by focusing on motion discrepancies in optical flow fields, enabling it to discern reflective surfaces even in complex scenes. MEDM complements this by guiding edge feature learning based on motion inconsistencies, allowing the model to delineate mirror boundaries with greater precision. Together, these modules form a robust system for mirror detection in dynamic environments as shown in Fig. 20, where the motion inside mirror reflections often differs from that outside.

While MG-VMD demonstrates substantial improvements in specularity detection, it still faces challenges, particularly in differentiating mirrors from windows or other reflective surfaces with inconsistent motion patterns. This limitation suggests that MG-VMD may over-predict mirrors in scenes where windows exhibit similar motion cues. Future work could explore integrating additional contextual cues, such as depth or material properties, to better distinguish between mirrors and other reflective objects. Additionally, enhancing the model's adaptability to low-contrast environments could further strengthen its robustness in a wider array of real-world scenarios.

4 Complementary approaches in specularity detection

In this section, salient object detection (SOD) and shadow detection, while not originally intended for specularity detection, have made notable contributions to this field. Many current specularity detection models draw inspiration from SOD, as both approaches aim to identify visually prominent regions within an image (Wang et al. 2021). Given that specularities often manifest as high-contrast features, SOD models provide a valuable framework for isolating these reflective areas through multi-scale feature extraction. This capacity to capture prominent regions aligns with specularity detection's goals, where diverse specular surfaces need precise handling across scales and intensities.

Similarly, shadow detection shares technical foundations with specularity detection, especially in feature extraction and boundary delineation (Vasluianu et al. 2024). Both shadows and specular reflections arise from complex interactions between light and surfaces, requiring nuanced edge and boundary recognition techniques. Methods from shadow detection that emphasize boundary precision prove advantageous in specularity detection by enabling the accurate delineation of reflective areas in visually intricate scenes. These shared technical methodologies bolster the accuracy and adaptability of specularity detection models across challenging lighting conditions.

By integrating advancements from SOD and shadow detection, specularity detection gains access to refined feature extraction and boundary recognition techniques, resulting in more robust detection performance. Therefore, this section will review these complementary approaches, highlighting their technical relevance and contributions to specularity detection advancements.

4.1 Salient object detection

4.1.1 Why SOD is foundational for specularity detection

SOD and specularity detection share fundamental principles rooted in a DL-based model's capacity to identify visually distinct regions within images. In both domains, the objective is to isolate prominent features, though their contexts differ: SOD focuses on objects that stand out based on perceptual cues, while specularity detection emphasizes the capture of reflective, high-intensity regions within scenes. The convergence of these fields lies in their reliance on spatial contrast, boundary precision, and feature differentiation.

In our reviewed DL-based models, both SOD and specularity detection benefit from architectures designed to balance high-level semantic understanding with low-level boundary refinement. Techniques used in SOD, such as multi-scale feature extraction, attention mechanisms, and boundary enhancement, directly influence the robustness of specularity detection by capturing the subtle transitions and high-contrast edges characteristic of reflective regions. Models developed in SOD that integrate global contextual awareness, such as pyramid pooling and transformer-based architectures, have proven especially beneficial in specularity detection, where reflections vary in shape, intensity, and spatial context (Borji et al. 2019; Zhou et al. 2024). Furthermore, the incorporation of uncertainty modeling in recent SOD advancements aligns well with the need for specularity detection to handle varying degrees of reflectance and occlusion in dynamic environments. Thus, SOD frameworks provide valuable insights and structural adaptations for specularity detection tasks, especially as they push the boundaries of segmentation accuracy through advanced techniques in attention, multi-scale contextualization, and edge precision. These shared technical foundations allow for a seamless application of SOD advancements in refining and enhancing specularity detection models, addressing complexities inherent in reflective surface analysis.

4.1.2 Influential SOD models for advancing specularity detection

In 2017, Zhao et al. introduced the Pyramid Scene Parsing Network (PSPNet) (Zhao et al. 2017), designed to overcome the limitations of fully convolutional networks (FCNs) (Long et al. 2015) by incorporating a pyramid pooling module to capture multi-scale contextual information. This module enables PSPNet to extract global features, essential for managing complex scenes, making it especially valuable for specularity detection where detailed spatial context aids in distinguishing reflective regions from non-reflective surroundings.

Deng et al. proposed R^3Net in 2018 (Deng et al. 2018), integrating recursive residual refinement blocks (RRBs) to learn the residuals between saliency predictions and ground truth, refining the model's precision in capturing prominent features. This approach is particularly adaptable for specularity detection, as it can improve the delineation of specular regions by enhancing boundary accuracy and adjusting to subtle variations in reflectance.

In 2019, EGNet by Zhao et al. (2019) introduced an edge-guidance mechanism, significantly improving boundary precision-a critical aspect for accurately detecting specular boundaries in mirror or reflective object detection tasks. Liu et al.'s PoolNet (Liu et al. 2019) further contributed by implementing a global guidance module (GGM) and feature aggregation module (FAM), which together enhance salient detail sharpening. These modules are beneficial for specularity detection where edge differentiation is key to distinguishing reflective from non-reflective areas. Similarly, BASNet (Qin et al. 2019) leveraged a novel residual refinement module (RRM) and a hybrid loss that integrates pixel-, region-, and feature map-level supervision, creating a robust multi-level feature framework suitable for handling diverse reflective surfaces in specularity tasks. Wu et al.'s CPDNet (Wu et al. 2019) emphasized computational efficiency with a cascaded partial decoder, a fast and accurate approach advantageous for real-time specularity detection applications.

In 2022, Wu et al. introduced the Extremely-Downsampled Network (EDN) (Wu et al. 2022), which applies extreme downsampling to achieve a global view of the entire image, efficiently locating salient objects. By using Scale-Correlated Pyramid Convolution (SCPC) for multi-level feature integration in the decoder, EDN improves boundary alignment, which is crucial for detecting specular highlights, especially in complex environments.

Ma et al. in 2023 proposed boosting broader receptive fields (BBRF) (Ma et al. 2023), addressing multi-scale challenges with a bilateral extreme stripping (BES) encoder and dynamic complementary attention module (DCAM). These components increase the model's ability to handle extreme scale variations, a feature essential for specularity detection where reflections may appear across diverse spatial scales. The loop compensation strategy in BBRF further enhances the detection of scale-specific features, reducing errors when identifying specular regions across different object sizes.

In 2023, Wang et al. introduced the multiple enhancement network (MENet) (Wang et al. 2023), a multi-level model that refines salient regions through pixel, region, and object-level feature enhancement. MENet's use of a multi-scale feature enhancement module allows it to flexibly manage detail across scales, directly benefiting specularity detection where reflections can vary in scale and shape within a single scene. In the same year, Tian et al.'s modeling distributional uncertainty (Tian et al. 2023) tackles out-of-distribution (OOD) uncertainties in SOD, a significant advancement for specularity detection. By explicitly addressing distributional uncertainty, this model enhances robustness in dynamic visual conditions, enabling more reliable detection of specular regions in images with inconsistent lighting or complex textures.

Finally, in 2024, Mao et al. proposed the generative transformer with inferential GAN (iGAN) (Mao et al. 2024), a transformer-based approach that includes a latent variable for predictive uncertainty. This feature is particularly beneficial for specularity detection in high-contrast scenes where reflections introduce prediction challenges. The iGAN's ability to model prediction uncertainty enhances its adaptability to the contextual nuances of specular surfaces, improving accuracy and reliability in complex environments.

4.2 Shadow detection

4.2.1 Context-aware boundary precision for specularity detection

Shadow detection contributes to specularity detection through its robust methodologies for feature extraction, edge delineation, and context-aware segmentation. Both shadow and specularity detection are rooted in the identification of visually distinct regions resulting from light-surface interactions, making DL-based approaches highly effective in these tasks. By leveraging shared properties, such as the dependency on lighting context and boundary precision, shadow detection models offer valuable insights for specularity detection, particularly in handling intricate reflections and edge transitions.

Furthermore, from a DL-based perspective, shadow detection models are engineered to distinguish subtle changes in light intensity and texture, which parallels the needs of specularity detection where reflective boundaries must be precisely identified. Techniques like multi-scale feature extraction, common in shadow detection, are highly applicable to specularity tasks, as they capture both global context and localized intensity shifts, essential for accurately identifying reflective surfaces within complex scenes. Additionally, shadow detection models often incorporate spatial attention mechanisms that focus on low-intensity regions with context-dependent cues; this approach can be adapted to detect specular reflections by refining model attention to high-intensity, reflective areas in variable lighting conditions.

Therefore, in essence, shadow detection's advanced boundary recognition and context-awareness techniques provide a foundational framework for specularity detection, enabling more accurate, context-sensitive segmentation of reflective regions. By adopting these methodologies, specularity detection models can achieve enhanced performance in delineating reflective boundaries, even in environments with complex lighting or high visual clutter.

4.2.2 Technical insights from shadow detection models for specularity detection

In 2020, Wang et al. introduced the light-guided instance shadow-object association (LISA) model. This dual-branch framework jointly detects shadow instances and their casting objects while estimating light direction (Wang et al. 2020). By linking shadows to their sources, LISA enables precise boundary delineation, which is essential for specularity detection where reflections require careful edge handling. The model's structured approach to shadow-object association provides valuable techniques for refining specular region boundaries, especially in scenes with complex object-shadow interactions.

In 2022, Ding et al. proposed shadow-consistent correspondence (SC-Cor) for video shadow detection, focusing on temporal stability by learning pixel-to-set correspondences across frames (Ding et al. 2022). SC-Cor's weakly-supervised alignment maintains shadow feature consistency even under varying textures and illumination, directly benefiting specularity detection in dynamic scenes. This stability is crucial for reliably capturing shifting reflective surfaces, making SC-Cor's temporal coherence strategies adaptable to tracking specular highlights across frames.

Also in 2022, Zhu et al. introduced single image shadow detection via complementary mechanism, which employs dual interactive branches to detect shadow and non-shadow regions simultaneously, enhancing boundary precision through mutual feature exchange (Zhu et al. 2022). This model's complementary structure prevents misclassification between shadowed and non-shadowed areas, a principle that can be applied to specularity detection to improve boundary clarity and accurately isolate reflective regions from non-reflective backgrounds.

In 2024, Wang et al. advanced shadow segmentation in dynamic contexts with the model of detect-any-shadow (ShadowSAM) (Wang et al. 2024), refining SAM with a long short-term attention mechanism for video consistency. ShadowSAM's attention-driven temporal coherence is instrumental in specularity detection, where consistent tracking of reflective boundaries across frames is essential. By adapting SAM to focus on high-contrast areas, ShadowSAM offers a robust approach for capturing specular reflections that change with motion or lighting.

Collectively, these models contribute to specularity detection by enhancing boundary precision, temporal stability, and context-aware segmentation: core aspects necessary for accurately handling reflective surfaces in complex scenes. Their innovations in light-guided association, temporal alignment, and complementary feature processing extend the technical capabilities of specularity detection in both static and dynamic environments.

5 Specularity datasets

Specularity detection relies on high-quality datasets that capture a wide range of reflective surfaces under diverse environmental conditions. To comprehensively evaluate different approaches, the datasets used in specularity detection research can be categorized based on the dimensionality of their input data, aligning with the structure of this review. Specifically, datasets are grouped into 2D datasets, which primarily consist of RGB images; RGB-D datasets, which include depth information alongside color images; and video datasets, which provide temporal sequences for analyzing specular reflection dynamics.

Each category of datasets presents distinct advantages and challenges. 2D datasets are widely used for training and benchmarking models due to their accessibility and large-scale availability. However, they often struggle with ambiguous cases where depth information would aid in distinguishing specular reflections from actual objects. RGB-D datasets mitigate this issue by integrating depth cues, allowing for a more accurate separation of specular highlights from object textures. Finally, video datasets extend the analysis to temporal domains, offering a unique perspective on how specular reflections change over time and enabling the development of methods that leverage motion cues.

5.1 Widely-used 2D dataset

The mirror segmentation dataset (MSD), introduced by Yang et al. in 2019 (Yang et al. 2019), is one of the most widely used and specialized 2D datasets for specularity detection, focusing on mirrors as reflective surfaces. It consists of 4,018 annotated mirror images, of which 3,677 were captured indoors and 341 outdoors. To ensure diverse and balanced training and testing sets, the dataset is segmented by mirror type, resulting in 3,063 images for training and 955 for testing. MSD offers comprehensive annotation of mirror-specific attributes, such as location, shape, position, and global color contrast, which are critical for analyzing specular regions within complex scenes. This level of detailed annotation provides a robust benchmark for specularity detection tasks, particularly in distinguishing mirrors from surrounding objects in diverse environments. However, MSD's high similarity across images has been noted as a limitation, potentially reducing the robustness of models trained on this dataset due to limited scene diversity.

In response to the limitations of MSD, the progressive mirror detection dataset (PMD) (Lin et al. 2020) provides a more diverse and comprehensive benchmark for mirror-based specularity detection. PMD significantly enhances diversity by including 6,461 images that cover a broader array of indoor and outdoor scenes. This dataset was constructed by selectively aggregating all images containing mirrors from six large-scale public datasets. By drawing from these sources, PMD achieves a high degree of image-scene variety, addressing the homogeneity issue in MSD and providing a more realistic distribution of everyday settings that include both flat and concave mirrors. As shown in Table 3, both MSD and PMD datasets are valuable for advancing specularity detection. MSD provides detailed annotations of mirror-specific attributes, while PMD enhances diversity by incorporating varied scenes and mirror types. Together, these datasets serve as benchmarks that capture the complexities of reflective surfaces, supporting the development of more robust and generalizable specularity detection models.

To further explore the availability of specular-related images beyond dedicated datasets like MSD and PMD, this review analyzes several large-scale, open-source datasets that cover diverse real-world scenes. As shown in Table 2, this study identifies and quantifies specular-related images across six commonly used public datasets, including ADE20K (Zhou et al. 2019, 2017), NYUD-V2 (Silberman et al. 2012), MINC (Bell et al. 2015), Pascal-Context (Mottaghi et al. 2014), SUNRGBD (Song et al. 2015), and COCO-Stuff

Table 2 Specular-related image statistics from different public datasets

Dataset	Total number of specular-related image
ADE20K Zhou et al. (2019, 2017)	1340
COCO-Stuff Caesar et al. (2018)	3700
MINC Bell et al. (2015)	390
NYUD-V2 Silberman et al. (2012)	155
SUNRGBD Song et al. (2015)	720
Pascal-Context Mottaghi et al. (2014)	102

Table 3 Comparison of statistics and features between MSD (Yang et al. 2019) and PMD (Lin et al. 2020)

Attribute	MSD Yang et al. (2019)	PMD Lin et al. (2020)
Total images	4018	6,461
Indoor/outdoor split	3677/341	Mixed indoor and outdoor scenes
Image resolution	512 x 640	Various resolutions
Capturing method	Smartphones, Real-world scenes	Synthetic, Camera types unspecified
Training/testing split	3063/955	Leave-one-out cross-validation
Attributes	Shape, location, color contrast	Shape, size, diverse backgrounds
Annotation diversity	Limited, high similarity	Diverse, six public sources

(Caesar et al. 2018). Although the number of specular-related images within these datasets is relatively limited compared to specialized benchmarks, they still provide valuable supplementary data for integrating specularity detection into broader vision tasks. The additional sources contribute to future dataset fusion efforts.

5.2 RGB-D datasets

The RGB-D mirror segmentation dataset (RGBD-Mirror), introduced in PDNet by Mei et al. (2021), comprises 3,049 RGB-D images collected from Matterport3D (Chang et al. 2017), SUNRGBD (Song et al. 2015), and ScanNet (Dai et al. 2017). Each image is annotated with mirror masks and depth maps, enabling models to utilize depth discontinuities for precise mirror boundary detection. This dataset's integration of depth cues is especially beneficial for specularity detection, as it enhances the ability to identify and segment reflective surfaces within complex scenes.

The Mirror3D dataset, proposed by Tan et al. in 2021 (Tan et al. 2021), focuses on 3D mirror plane prediction and depth refinement. It includes 7,011 annotated mirror planes across 5,894 frames sourced from Matterport3D (Chang et al. 2017), ScanNet (Dai et al. 2017), and NYUv2 (Silberman et al. 2012). This dataset addresses depth inaccuracies caused by mirrors in RGB-D images, providing corrected depth data on mirror surfaces to support reliable 3D scene reconstruction. The Mirror3D dataset is essential for specularity detection applications that require accurate depth information for reflective surfaces.

Both of them offer crucial depth information that enhances specularity detection capabilities beyond what 2D datasets provide, as shown in Table 4. While RGBD-Mirror focuses on pixel-level mirror masks and leverages depth discontinuities for precise boundary detection, Mirror3D addresses depth inaccuracies and supports 3D mirror plane refinement. Together, these RGB-D datasets provide a more comprehensive foundation for developing models that require accurate spatial understanding of reflective surfaces, enabling more robust specularity detection in complex 3D environments.

5.3 Video datasets

The VMD-D dataset, introduced by Lin et al. (2023), is one of the first large-scale datasets specifically created for video mirror detection. It comprises 14,987 frames across 269 video

Table 4 Comparison of statistics and features between RGBD-Mirror (Mei et al. 2021) and Mirror3D (Tan et al. 2021)

Attribute	RGBD-mirror Mei et al. (2021)	Mirror3D Tan et al. (2021)
Total images	3,049	5,894
Source datasets	Matterport3D, SUN-RGBD, ScanNet	Matterport3D, Scan-Net, NYUv2
Mirror instances	Pixel-level mirror masks	7,011 3D mirror plane annotations
Depth data	RGB-D depth provided	Refined depth on surfaces
Background	Indoor, outdoor	Indoor, outdoor, mixed
Focus	Depth cues for mirror boundary detection	3D mirror plane estimation and depth refinement
Camera information	Kinect	Kinect, RealSense
Annotation method	Pixel-level masks cre-ated by annotators	3D mirror plane an-notations with depth refinement
Application	Enhances boundary precision in 2D images	Refines depth accu-racy in 3D surfaces

clips, collected from diverse real-world scenes. Each frame in VMD-D is annotated with precise mirror masks, enabling the development of models that leverage both intra-frame and inter-frame correspondences for mirror detection. The dataset facilitates video-based specularity detection by providing high-diversity and long-temporal video sequences, which are essential for evaluating the robustness of detection models under varied scenarios.

The ZOOM dataset, proposed by Xu et al. (2024), focuses on weakly-supervised video mirror detection. It contains 200 video clips (totaling 12,490 frames) sourced from Charades and Charades-Ego datasets and captured in indoor scenes using smartphones. Unlike fully supervised datasets, ZOOM employs frame-level binary mirror indicators instead of pixel-wise annotations, reducing the annotation burden. This design enables the exploration of temporal consistency for mirror detection, leveraging the contrast between mirrored and non-mirrored frames to enhance specularity detection under weak supervision conditions.

The motion mirror dataset (MMD), introduced by Warren et al. (2024), includes 37 nine-second video clips, each with pixel-level annotations for mirrors and surrounding edges. It captures a range of mirror sizes, shapes, and positions under diverse lighting conditions, offering a more challenging benchmark than previous datasets. MMD is instrumental for specularity detection, especially in scenarios with motion inconsistencies, as it emphasizes real-world conditions where mirrors display varying motion patterns relative to the scene.

As shown in Table 5, these three datasets provide support for specularity detection by leveraging temporal information. Video datasets offer distinct advantages over static 2D datasets, capturing dynamic scenes and enabling models to use motion cues, temporal consistency, and inter-frame correlations. The temporal dimension is crucial for detecting specular surfaces in complex environments, where mirrors and reflections vary with changes in lighting, motion, and perspective. These datasets thus promote the development of robust and generalizable models for specularity detection in real-world applications.

Table 5 Comparison of statistics and features between VMD-D (Lin et al. 2023), ZOOM dataset (Xu et al. 2024), and MMD (Warren et al. 2024)

Attribute	VMD-D Lin et al. (2023)	ZOOM Xu et al. (2024)	MMD Warren et al. (2024)
Total frames	14,988	12,490	9,727
Total videos	269	200	37
Video resolution	1920 x 1080	Various	1920 x 1080
Frame rate	30 fps	Not specified	30 fps
Source	Real-world daily-life scenes	Specific video dataset	Real-world daily-life scenes
Annotation type	Pixel mask	Frame-level binary	Pixel mask and edges
Background	Diverse scenes	Indoor only	Mixed indoor and outdoor
Focus	Temporal consistency	Weak supervision	Motion, lighting diversity
Application	Long-term analysis	Minimal supervision	Dynamic scenes
Dataset split	143 training, 126 test videos	150 training, 50 test videos	18 training, 19 test videos

6 Evaluation metrics

Accurate evaluation metrics are essential for assessing specularity detection performance, where models need to capture reflective regions with precision and structural fidelity. This section reviews multiple common metrics, underscoring their relevance for specularity detection, as shown in Table 6. Each metric builds upon fundamental precision and recall (PR) principles:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{1}$$

where TP, FP, TN, and FN denote true positives, false positives, true negatives, and false negatives. These metrics are foundational in measuring detection accuracy, balancing false positives with true reflective region identification-a core challenge in specularity detection.

Mean square error (MSE) and **root mean square error (RMSE)** evaluate the average squared differences between predictions and ground truth, with RMSE providing a normalized error measure:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2, \tag{2}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^{n} (y_j - \hat{y}_j)^2}. \tag{3}$$

where low MSE and RMSE values indicate precise boundary detection of specular reflections, critical for accurate localization in complex scenes.

Table 6 Evaluation metrics for specularity detection: suitability and comparative analysis

Metric	Mathematical expressions	Suitability for specularity detection
MSE	$\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$	1. Sensitive to large errors, highlighting discrepancies in specular regions. 2. Limited for structural detail, impacting reflective surface assessment.
RMSE	$\sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$	1. Similar to MSE, emphasizes large deviations. 2. Useful where significant errors disrupt detection accuracy.
MAE	$\frac{1}{n}\sum_{i=1}^{n}\lvert\hat{y}_i - y_i\rvert$	1. Stable against outliers, suitable for intensity variations. 2. Good for evaluating brightness in specular reflections.
F-Score	$\left(1+\beta^2\right)\cdot\frac{\text{Precision}\cdot\text{Recall}}{\beta^2\cdot\text{Precision}+\text{Recall}}$	1. Balances Precision and Recall, adaptable via β. 2. Suitable for reducing false positives in complex scenes.
PSNR	$20\cdot\log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right)$	1. Measures signal fidelity, though overlooks structure. 2. Limited precision in assessing fine specular details.
SSIM	$[l(\mathrm{x},\mathrm{y})]^{\alpha}[c(\mathrm{x},\mathrm{y})]^{\beta}[s(\mathrm{x},\mathrm{y})]^{\gamma}$	1. Captures structural similarity in brightness and contrast. 2. Ideal for maintaining structure in specular regions.
S-Measure	$\alpha\times S_o + (1-\alpha)\times S_r$	1. Extends SSIM, focusing on region similarity. 2. Effective for coherent specular object detection.
E-Measure	$\frac{1}{W\times H}\sum_{i=1}^{W}\sum_{j=1}^{H}\phi_S(i,j)$	1. Integrates pixel and image-level alignment. 2. Enhances assessment in complex reflective scenes.

Mean absolute error (MAE) measures the average absolute difference, suitable for evaluating brightness variations in specular regions:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\lvert\hat{y}_i - y_i\rvert\,. \tag{4}$$

The **F-score** balances Precision and Recall, with F_β allowing for emphasis on Precision (often prioritized in specularity detection):

$$F_1 = 2\cdot\frac{\text{Precision}\cdot\text{Recall}}{\text{Precision}+\text{Recall}}, \tag{5}$$

$$F_\beta = \frac{\left(1+\beta^2\right)\text{Precision}\cdot\text{Recall}}{\beta^2\text{Precision}+\text{Recall}}. \tag{6}$$

Peak signal-to-noise ratio (PSNR) quantifies image quality by comparing pixel intensities, where higher values indicate reduced noise and better preservation of reflective detail:

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right) = 20 \cdot \log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right), \tag{7}$$

where MAX_I represents maximum pixel intensity (usually 255 for 8-bit images). High PSNR indicates accurate, artifact-free representation of specular regions.

The **structural similarity index (SSIM)** assesses structural similarity, capturing brightness, contrast, and structural alignment essential for reproducing complex specular textures:

$$\begin{gathered} SSIM(\mathrm{x},\mathrm{y}) = [l(\mathrm{x},\mathrm{y})]^{\alpha}[c(\mathrm{x},\mathrm{y})]^{\beta}[s(\mathrm{x},\mathrm{y})]^{\gamma}, \\ l(\mathrm{x},\mathrm{y}) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \quad c(\mathrm{x},\mathrm{y}) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \quad s(\mathrm{x},\mathrm{y}) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}. \end{gathered} \tag{8}$$

Higher SSIM values reflect better preservation of specular structures, crucial for visual quality in detection results.

The **S-measure** focuses on structural similarity by combining object- and region-based assessments, crucial for capturing coherent shapes of specular objects:

$$S = \alpha \cdot S_o + (1 - \alpha) \cdot S_r, \tag{9}$$

where S_o and S_r denote object- and region-based similarities.

The **enhanced-alignment measure (E-measure)** combines pixel and global image values, offering improved alignment assessment for specular regions:

$$E = \frac{1}{W \cdot H} \sum_{i=1}^{W} \sum_{j=1}^{H} \phi_S(i,j), \tag{10}$$

where ϕ_S is the enhanced symmetry matrix, capturing the correlation between predicted and true specular regions. High E-measure values indicate strong alignment in complex, reflective scenes.

In this review, the majority of the studies rely on RGB images along with their corresponding specular masks, where the masks are binary images used to differentiate specular regions from non-specular regions. The RGB images capture the color information of the scene, while the specular masks emphasize the reflective surfaces, marking the areas that exhibit specular highlights and excluding the rest of the scene. To illustrate the application of the evaluation metrics, this section presents an example below in Fig. 21.

This example demonstrates the process of evaluating specularity detection performance using RGB images and specular masks. The RGB image displays a scene containing specular highlights, where reflective surfaces are captured in color. The ground truth mask indicates the true specular regions, marked in white, effectively delineating the reflective areas in the scene. On the other hand, the predicted specular mask highlights the regions detected by the model, showing the areas the algorithm identifies as specular. By comparing the predicted and ground truth specular masks, various evaluation metrics can be computed.

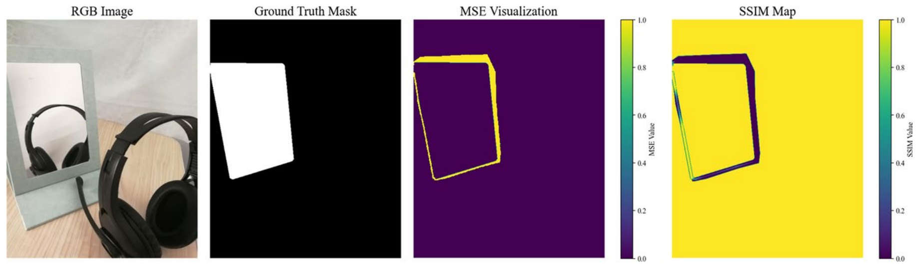

Fig. 21 Example of specularity detection metrics

7 Quantitative experiments and empirical analysis

This section presents an empirical analysis to identify key challenges in specularity detection. Specifically, it conducts a comprehensive benchmark evaluation, representing the first large-scale, input-type-based quantitative comparison of specularity detection models. This approach enables a deeper understanding of the strengths and limitations inherent in current models. Additionally, this section evaluates the performance of complementary approaches introduced in Sect. 4, evaluating their effectiveness across various specularity datasets. Finally, the experimental results offer a quantitative assessment of mainstream models, providing insights into their relative efficacy and potential areas for improvement.

7.1 Comparative experiments

The comparative experiments were conducted in a Python 3.8 environment with CUDA 11.8 on an NVIDIA RTX 4090 GPU. Hyperparameters, including learning rate and batch size, were set following the guidelines from each referenced paper to ensure consistency across comparisons. Adjustments were made only when necessary to align with the parameters specified in the original implementations.

Our proposed structured experiments aim to deliver an in-depth understanding of model performance across different input types, providing valuable insights into their applicability and limitations for specularity detection.

RGB image input. This experiment evaluates specularity detection models along with complementary approaches, specifically SOD and shadow detection models. Given the previously discussed architectural and methodological similarities among these models, a unified assessment enables a direct performance comparison. Such an arrangement provides insight into the state-of-the-art (SOTA) capabilities of specularity detection, SOD, and shadow detection methods under standardized experimental conditions, allowing for a clearer analysis of their comparative performance.

RGB-D image input. The second segment focuses on models that utilize RGB-D data, incorporating depth information alongside color to improve detection accuracy, especially in scenes with complex spatial structures and diverse lighting. The objective is to assess the strengths and limitations of RGB-D input for specularity detection, identifying cases where depth cues either enhance or limit model effectiveness in capturing specular regions.

Video stream input. The final experiment examines models that process video streams, which are designed to maintain temporal consistency in detecting specular regions across frames. Video-based methods are especially relevant for real-time applications in dynamic environments, where reflective surfaces may vary in appearance due to movement or lighting changes. This evaluation measures the robustness of video stream models and their ability to accurately track specular regions over time, offering a comprehensive comparison of performance in temporally varying scenarios.

7.2 Results snalysis

RGB image input. The quantitative results presented in Table 7 demonstrate a clear performance advantage for models specifically designed for specularity detection when tested on the MSD and PMD datasets. Notably, CSFwinformer (Xie et al. 2024) achieves the best results with an MAE of 0.030 on MSD and 0.039 on PMD, alongside an impressive F_β score of 0.865 and 0.836, respectively, establishing it as the SOTA in this experimental setup. Similarly, the PSNR values of 29.50 on MSD and 28.02 on PMD further confirm CSFwinformer's robustness in handling reflective surfaces, underscoring its precision in both pixel-level and structural fidelity metrics. Other specularity detection models, such as SSL (Lin and Lau 2023) and VCNet (Tan et al. 2023), also demonstrate competitive performance, with SSL achieving an MAE of 0.038 on MSD and a high F_β score of 0.847, highlighting the ability of these models to accurately localize and delineate specular regions. The specularity detection models outperform the SOD and shadow detection models in all metrics, reflecting their specialized design for handling the complexities of reflective surfaces.

The shadow detection models, including ShadowSAM (Wang et al. 2024) and SC-Cor (Ding et al. 2022), show considerably lower performance on the specular datasets. This is primarily due to the fundamental differences in how shadows and specular reflections are perceived and represented in images. Shadows are characterized by their diffuse nature, lacking sharp boundaries and high-frequency detail, which shadow detection models are tuned to capture. In contrast, specular surfaces exhibit distinct high-frequency reflections and abrupt boundaries that require a precise understanding of local intensity and structure. The lack of adaptation in shadow detection models to handle these sharp reflective characteristics explains their relatively poor performance, with ShadowSAM achieving only an MAE of 0.080 and an F_β score of 0.700 on MSD, considerably lower than the specialized specularity models.

While SOD models, such as EDN (Wu et al. 2022) and BBRF (Ma et al. 2023), perform moderately well on the specular datasets, they still fall short of the specularity detection models. SOD techniques are inherently designed to detect prominent objects based on salient features like brightness, color, and edge contrast, which partially aligns with the requirements of specularity detection. For example, EDN achieves an MAE of 0.092 on MSD and an F_β score of 0.740, showing a reasonable capacity to localize reflective regions. However, the primary limitation of SOD models on specular datasets is their lack of depth in modeling the high-frequency textures and precise boundary delineation needed to capture specular details accurately. Although SOD methods provide a decent baseline, they lack the specific adaptations, such as feature refinement and structural alignment modules, that make specularity detection models superior in this context.

A key consideration beyond accuracy is the computational efficiency of these models, as measured by speed using frames per second (FPS) and model complexity using parameter number (Params). Notably, SC-Cor (Ding et al. 2022), despite showing competitive accu-

Table 7 Quantitative comparison of methods with RGB image input on MSD and PMD datasets using MAE, F_β, and PSNR as evaluation metrics

Category	Method	MAE ↓		F_β ↑		PSNR ↑		Speed		Params
		MSD	PMD	MSD	PMD	MSD	PMD	MSD	PMD	
SOD	PSPNet Zhao et al. (2017)	0.138	0.145	0.600	0.580	22.30	21.90	10.7	9.6	68.07
	R^3Net Deng et al. (2018)	0.125	0.138	0.630	0.610	23.50	22.80	–	–	–
	PoolNet Liu et al. (2019)	0.120	0.133	0.650	0.625	24.00	23.40	40.5	36.4	68.26
	BASNet Qin et al. (2019)	0.110	0.126	0.680	0.645	24.50	24.10	36.2	34.6	87.06
	EGNet Zhao et al. (2019)	0.105	0.122	0.700	0.670	25.00	24.50	10.7	9.6	108.07
	CPDNet Wu et al. (2019)	0.098	0.115	0.720	0.695	25.50	24.90	32.4	30.5	47.85
	EDN Wu et al. (2022)	0.092	0.110	0.740	0.710	26.00	25.40	51.7	46.5	42.85
	BBRF Ma et al. (2023)	0.088	0.105	0.755	0.730	26.50	26.00	39.3	35.5	74.4
Shadow detection	LISA Wang et al. (2020)	0.090	0.104	0.660	0.635	25.50	24.80	–	–	–
	SC-Cor Ding et al. (2022)	0.085	0.100	0.670	0.650	26.00	25.20	74.14	74.02	232.63
	Mutual complementary model Zhu et al. (2022)	0.083	0.098	0.685	0.665	26.20	25.50	35.3	31.7	10.95
	ShadowSAM Wang et al. (2024)	0.080	0.095	0.700	0.685	26.50	26.00	183.3	164.9	93.73
Specularity detection	MirrorNet Yang et al. (2019)	0.066	0.073	0.815	0.784	27.32	26.71	36.83	33.48	–
	PMD Lin et al. (2020)	0.061	0.069	0.828	0.803	27.88	27.20	–	–	–
	Mirror-Yolo Li et al. (2022)	0.064	0.071	0.821	0.792	27.45	26.89	27.00	25.50	55.30
	UNet-transformer specular model Wu et al. (2023)	0.055	0.065	***0.842***	***0.815***	28.04	27.42	–	–	–
	SHMGAN Anwer et al. (2023)	0.058	0.068	0.831	0.804	27.91	27.31	39.2	37.6	80.10
	VCNet Tan et al. (2023)	***0.054***	**0.059**	0.834	0.805	**28.83**	***27.48***	–	–	–
	SSL Lin and Lau (2023)	**0.038**	***0.059***	**0.847**	**0.826**	***28.76***	*28.10*	–	–	–
	CSFwinformer Xie et al. (2024)	*0.030*	*0.039*	*0.865*	*0.836*	*29.50*	**28.02**	100.6	88.5	84.75

The three best scores are marked in italic, bold, and bolditalic, respectively

racy, has an extremely high parameter count of 232.63 million (M), making it computationally expensive for real-time applications. In contrast, ShadowSAM (Wang et al. 2024), which leverages the SAM for segmentation, also has a significant computational footprint, operating at 183.3 FPS on MSD while maintaining a moderate parameter count of 93.73 M. On the other hand, the best-performing specularity detection model, CSFwinformer (Xie et al. 2024), balances accuracy and efficiency with 100.6 FPS on MSD and 88.5 FPS on

PMD, with 84.75 M parameters, demonstrating its ability to handle reflective surfaces efficiently. Among the traditional SOD models, CPDNet (Wu et al. 2019) and BBRF (Ma et al. 2023) exhibit moderate parameter sizes of 47.85 M and 74.4 M, respectively, while running at 32.4 FPS and 39.3 FPS on MSD. Some models could not be evaluated in terms of computational efficiency due to limitations in open-source availability or insufficient details in the original publications, and these cases are marked as '-' in the Table 7. Overall, while certain models exhibit superior accuracy, their practical deployment largely depends on computational efficiency. Therefore, in real-world applications, balancing performance and real-time usability remains a crucial factor.

RGB image input. The results in Table 8 highlight the effectiveness of RGB-D input-based specularity detection models, particularly MGNet-T (Zhou et al. 2024), which demonstrates SOTA performance on both the RGBD-Mirror and Mirror3D datasets. MGNet-T achieves an MAE of 0.029 on RGBD-Mirror and 0.036 on Mirror3D, indicating its high accuracy in capturing reflective surfaces with minimal error. In terms of F_β, MGNet-T also leads with scores of 0.850 on RGBD-Mirror and 0.710 on Mirror3D, further solidifying its capability to precisely identify specular regions across varying spatial structures. Its PSNR values of 28.50 on RGBD-Mirror and 23.00 on Mirror3D underscore its robustness in preserving image quality and detail, positioning it as the top-performing model in this evaluation.

One of the primary reasons for the high performance of MGNet-T and other recent models, such as SEMCNet (Zhang et al. 2024) and MGNet-S, is the incorporation of knowledge distillation mechanisms. Knowledge distillation enables these models to leverage feature representations from a teacher model, enhancing the student model's learning process with refined, high-level information. In the case of MGNet-T and MGNet-S, the teacher-student setup allows for the extraction of critical texture and depth features essential for accurately detecting specular boundaries. This setup improves model efficiency and accuracy by transferring vital insights from the teacher to the student model, thereby enabling MGNet-S to achieve strong performance with a lighter computational footprint. For instance, MGNet-S

Table 8 Quantitative comparison of methods with RGB-D image input on RGBD-Mirror and Mirror3D datasets using MAE, F_β, and PSNR as evaluation metrics

Method	MAE ↓		F_β ↑		PSNR ↑	
	RGBD-Mirror	Mirror3D	RGBD-Mirror	Mirror3D	RGBD-Mirror	Mirror3D
PDNet Mei et al. (2021)	0.045	0.058	0.780	0.620	25.60	20.40
Mirror3DNet Tan et al. (2021)	0.040	0.052	0.800	0.640	26.10	21.10
UTLNet Zhou et al. (2024)	0.035	0.049	0.815	***0.663***	26.50	21.50
SEMCNet Zhang et al. (2024)	**0.030**	***0.043***	**0.840**	**0.680**	**27.80**	***22.10***
MGNet-S (Student) Zhou et al. (2024)	***0.034***	**0.042**	***0.825***	0.661	***27.10***	**22.80**
MGNet-T (Teacher) Zhou et al. (2024)	*0.029*	*0.036*	*0.850*	*0.710*	*28.50*	*23.00*

The three best scores are marked in italic, bold, and bolditalic, respectively

Table 9 Quantitative comparison of methods with video stream input on VMD-D and MMD datasets using MAE, F_β, and PSNR as evaluation metrics

Method	MAE ↓		F_β ↑		PSNR ↑	
	VMD-D	MMD	VMD-D	MMD	VMD-D	MMD
VMD-Net Lin et al. (2023)	0.152	0.186	0.871	0.845	27.39	27.22
ZOOM Xu et al. (2024)	0.145	0.179	0.866	0.840	27.64	27.10
MG-VMD Warren et al. (2024)	*0.134*	*0.155*	*0.880*	*0.865*	*28.03*	*27.86*

The best scores are marked in italic

attains an MAE of 0.034 and F_β of 0.825 on RGBD-Mirror, showing a balanced trade-off between efficiency and accuracy.

The use of multiview feature refinement in SEMCNet also contributes to its competitive performance, especially on RGBD-Mirror, where it achieves an MAE of 0.030 and an F_β score of 0.840. By incorporating self-knowledge distillation within its architecture, SEMCNet enables iterative learning across different views, enhancing its robustness in specular detection. This feature refinement allows SEMCNet to capture subtle details in reflective regions, which are challenging to detect due to depth ambiguities. Overall, the success of these models highlights the benefits of knowledge distillation and depth-enhanced learning in improving accuracy and efficiency for RGB-D-based specularity detection.

Video stream input. The quantitative results in Table 9 illustrate the effectiveness of video stream input for specularity detection, with MG-VMD (Warren et al. 2024) achieving the best performance across all metrics. On the VMD-D dataset, MG-VMD reached an MAE of 0.134, an F_β score of 0.880, and a PSNR of 28.03, indicating superior accuracy and robustness in capturing specular highlights within dynamic video environments. Similarly, on the MMD dataset, MG-VMD maintained its advantage with an MAE of 0.155, F_β of 0.865, and PSNR of 27.86, establishing it as the leading model under the tested conditions.

The outstanding performance of MG-VMD can be attributed to its innovative use of motion cues through the MAM and MEDM. These modules enable MG-VMD to effectively exploit motion inconsistencies in video frames, which is crucial for distinguishing persistent specular reflections from transient changes in the scene. By focusing on regions with differing motion patterns, MG-VMD accurately delineates mirror boundaries and reflective surfaces, even in complex settings. This approach leverages temporal coherence and spatial motion differences to enhance detection precision, making MG-VMD particularly suited for real-world applications involving fluctuating lighting and reflective conditions.

8 Applications and future prospects of specularity detection

8.1 Applications

Specularity detection plays a crucial role in enhancing the accuracy and robustness of various applications across fields. The ability to effectively identify and manage specular reflections is becoming increasingly essential. This subsection explores several key application domains where specularity detection offers significant benefits, as well as future research directions to further advance its impact.

8.1.1 Medical imaging and surgery

In clinical ophthalmology, specular microscopy is a non-invasive diagnostic tool used to examine the corneal endothelium (Chaurasia and Vanathi 2021). The detection and management of specular reflections during imaging are essential to achieving high-quality endothelial cell images, as reflections can obscure critical details. Accurate specularity detection improves the clinical utility of this method, aiding in the assessment of corneal health and the diagnosis of conditions such as endothelial dystrophy (Monkam et al. 2021). Further, in nanophotonics, specular-reflection photonic nanojets (s-PNJs) are generated when light reflects off dielectric microstructures near a flat mirror (Minin et al. 2020). These nanojets enable the manipulation of nanoparticles or atoms, which is critical for applications in biophysics, optical trapping, and nanotechnology. In this context, the ability to control specular reflections allows for more stable and effective optical trapping of particles, enhancing medical or research applications in biotechnology and nanotechnology.

Moreover, for minimally invasive surgery (MIS), specular reflections in endoscopic images can severely disrupt the surgeon's ability to view anatomical structures clearly. Specularity detection and removal are thus crucial for ensuring accurate feature extraction and image registration, particularly in surgical navigation systems. A typical example is the Adaptive-RPCA decomposition, which provides an advanced solution for removing specular highlights in endoscopic image sequences (Li et al. 2020). Such application examples demonstrate that the ability to accurately detect and remove specular reflections improves the reliability of surgical navigation systems, leading to better decision-making during complex surgical procedures.

8.1.2 Industrial intelligence

Specularity detection enhances the accuracy and efficiency of automated inspection systems, which is critical for defect detection, material classification, and quality control in industrial applications. By addressing the challenges posed by reflective surfaces, these methods improve the precision of measurements and ensure consistent high-quality manufacturing (He et al. 2024). In the automotive sector, robotic inspection systems detect defects such as scratches and dents on semi-specular painted surfaces by analyzing specular reflections (Akhtar et al. 2020). These systems reduce the time and variability associated with human inspections, improving overall process efficiency and ensuring the consistent quality of automotive parts.

In addition, for optical components in industries like aerospace and electronics, specularity detection in this case ensures that high-precision optical components are thoroughly inspected, improving the reliability of critical parts (Huang et al. 2019). Furthermore, in the field of automated defect detection for car body surfaces, deflectometry analyzes specular reflections to detect small surface deformations like dents and scratches. This method improves quality control by providing a rapid, precise way to assess surface integrity (Molina et al. 2017). The ability to detect and interpret specular reflections in highly reflective surfaces enhances the overall speed and accuracy of inspections in automotive manufacturing. Moreover, polarization-based material classification can leverage specular reflections to differentiate between materials like metals and dielectrics (Wolff 1990), which useful in applications that require accurate material identification, such as in automated material handling systems.

8.1.3 Autonomous driving

Specularity detection is essential in autonomous driving, where specular surfaces such as road signs, wet roads, and water puddles pose significant challenges for vision-based perception systems. Current autonomous driving approaches can be broadly categorized into pure vision-based methods and multi-modal fusion strategies that incorporate radar, LiDAR, and other sensors. In both cases, detecting and managing specular reflections enhances scene understanding, improves object recognition, and mitigates perception errors caused by misleading high-intensity reflections.

In vision-centric autonomous systems, specular reflections can obscure critical information in images, leading to misinterpretations in object detection and scene segmentation. Advanced 2D and 3D specularity detection algorithms mentioned earlier in this review can be applied to handle these issues by suppressing reflections in input images or explicitly modeling them to enhance perception accuracy. For example, detecting and filtering specularities on road surfaces allows for more accurate lane detection, particularly in adverse weather conditions where water puddles create strong reflections that distort edge-based segmentation algorithms. For multi-modal sensor fusion approaches, detecting specular reflections enables better integration of vision and LiDAR data. Since LiDAR relies on active illumination, reflective surfaces can introduce high-intensity return signals, leading to incorrect depth estimations. By identifying specular regions in images, LiDAR processing pipelines can apply reflection-aware filtering techniques to refine depth estimation and improve sensor fusion consistency.

Beyond open-road scenarios, tunnel environments present unique challenges due to complex lighting conditions and multi-dimensional reflection effects. Traditional tunnel lighting design considers road surface reflections but often neglects the influence of diffuse reflections from tunnel walls and ceilings. By incorporating specularity detection into tunnel perception models, autonomous vehicles can adjust their perception algorithms to account for varying reflection properties, reducing false detections and enhancing overall driving safety (Shen et al. 2022).

8.1.4 Virtual & augmented reality

Realistic reflections and accurate lighting interactions are essential for immersive virtual reality and augmented reality (VR/AR) experiences. Misaligned specular highlights can cause artifacts, disrupt depth perception, and reduce visual coherence between virtual and real elements. In the VR environment, temporal reprojection has been studied to assess its impact on the subjective perception of specular reflections. Results indicate that accurate reprojection of specular highlights improves visual comfort and reduces artifacts, especially under low sampling conditions (Mišiak et al. 2023). The ability to detect and reproject specular reflections effectively ensures that materials retain their intended reflective properties while maintaining user immersion in immersive multimedia (Le et al. 2023).

For AR applications, researchers have explored how specular reflections contribute to realism and perceptual consistency. Some studies have focused on reproducing reflections in specular surfaces, demonstrating that maintaining material-specific reflectance properties enhances the coherence of digital overlays (Zhang et al. 2021). Similarly, in AR-based surgical navigation, specular reflections in endoscopic images have been investigated for their impact on feature extraction and overlay accuracy (Li et al. 2020).

In summary, advancements in the research of specularity promote VR/AR to more realistic rendering, enhance user immersion, and facilitate applications such as medical AR navigation and reflective aerial imaging. By accurately detecting and integrating specular effects, these methods contribute to seamless virtual-physical interactions and improved perceptual consistency in immersive environments.

8.2 Future prospects

The field of specularity detection has witnessed significant advancements, yet there remain several open questions and promising directions for future research. These can be divided into two main areas: further development of review articles on specularity detection techniques, and potential advancements in the algorithms.

8.2.1 Future review articles on specularity detection

One direction for future reviews is to expand the discussion of non-DL-based models, particularly those grounded in physics or mathematics. For example, traditional models such as the Blinn-Phong reflection model and those based on Fresnel equations could be revisited and explored in more depth. These models provide valuable insights into the physical interactions between light and reflective surfaces and can offer more interpretable and efficient solutions, particularly in controlled environments where DL models may not be the most effective. Additionally, polarization techniques deserve more attention in future reviews. By exploring the mathematical principles behind polarization and its application in specularity detection, further insights can be gained into its potential to inform future algorithms. These traditional techniques could be combined with more modern algorithms or even physical equipments to create hybrid models that leverage the strengths of both.

For DL-based models, future reviews should pay more focus toward practical deployment considerations, specifically comparing models based on memory usage, training time, or rendering efficiency. While many recent models exhibit comparable performance in terms of detection accuracy, there is a growing need to evaluate models from the perspective of real-world applications, where memory efficiency and processing time are crucial. Particularly in time-sensitive and resource-constrained environments, such as autonomous driving and real-time AR applications, the efficiency of models in terms of computational resources will become as important as their detection accuracy. Therefore, future research could explore trade-offs between accuracy and efficiency, providing clearer guidance for selecting models that balance performance with resource requirements.

8.2.2 Advancing specularity detection algorithms

Several opportunities for the further development of specularity detection lie ahead. One promising area is the exploration of self-supervised or even unsupervised learning techniques. The reliance on large, labeled datasets is a significant limitation in the current landscape of specularity detection, and reducing this dependence could open up new opportunities for more flexible and generalizable models. This challenge is not unique to specularity detection but reflects a broader trend in CV, where approaches borrowed from natural language processing (NLP) often ignore valuable perceptual information. Self-supervised and unsupervised learning techniques, if developed and applied effectively, could minimize the need for large-scale labeled datasets, offering more adaptable and scalable solutions that could be applied across a wide range of environments.

Another key area of research is the fusion of multi-modal data. While multi-modal fusion, such as combining LiDAR and RGB images has shown promise in autonomous driving and robotics, it remains under-explored when it comes to dealing with specular reflections. Specularity can distort both LiDAR and image data, leading to inaccuracies in object detection and scene reconstruction. Further investigation into the fusion of data from multiple sensors, such as radar, LiDAR, and thermal cameras, could greatly improve specularity detection in challenging environments. By using the complementary strengths of different sensors, future algorithms could be made more robust, offering improved accuracy in complex and dynamic environments like autonomous driving or industrial automation.

Furthermore, knowledge distillation combined with Large Language Models (LLMs) presents a novel direction for enhancing specularity detection capability. Recent advances in LLMs, particularly their ability to process multi-modal data and transfer knowledge across domains, could significantly benefit detection systems. LLMs can integrate diverse sources of information, including conventional image data, depth maps, weather conditions, and lighting information, which can provide valuable context for improving specularity detection. By fine-tuning and incorporating domain-specific LLMs, specularity detection systems can become more context-aware, adapting to real-world scenarios with greater flexibility and robustness. This approach shows the path to leveraging a wider range of data, enhancing detection accuracy and enabling more reliable real-time models.

In conclusion, the future of specularity detection will be shaped by both the refinement of existing methods and the introduction of innovative techniques that push beyond current capabilities. Whether through revisiting traditional models or optimizing state-of-the-art approaches, there is immense potential to enhance the performance, efficiency, and applicability of specularity detection across diverse domains. Successfully addressing challenges in complex scenarios like specularity detection will not only advance this field but also provide valuable insights and methodologies applicable to other similar challenges in computer vision and beyond.

9 Conclusion

This survey has provided the first comprehensive review of specularity detection in computer vision, establishing a structured, unified framework that integrates diverse definitions and methodologies from both traditional and deep learning approaches. By synthesizing perspectives from geometrical, physical, and perceptual domains, this review has presented a coherent definition of specularity that offers a mathematics-based foundation for consistent analysis across diverse CV applications. This unified definition has been crucial in addressing the ambiguities surrounding specularity, which have historically led to fragmented approaches and limited generalizability in CV tasks.

In reviewing the evolution of specularity detection techniques, this study has systematically analyzed and categorized both traditional and modern deep learning-based methods. Traditional approaches, though foundational, have often struggled to generalize in complex environments due to their reliance on handcrafted features and predefined reflection models. Conversely, deep learning models have demonstrated remarkable potential in detecting and interpreting specular regions, especially with the availability of large datasets and advanced architectures. However, challenges remain, particularly in highly reflective or dynamically changing environments, where model robustness and adaptability are critical. The quantitative evaluation of these methods across multiple datasets has highlighted both the strengths and limitations of each approach, providing valuable insights for future research and application.

Furthermore, this survey has identified emerging applications for specularity detection in fields such as autonomous driving, industrial inspection, and medical imaging, where accurate specularity handling is increasingly essential. This review has explored the potential of complementary techniques, such as salient object and shadow detection, to enhance specularity detection performance in complex scenarios. By proposing an integration of these techniques, along with recommendations for incorporating depth and temporal information, this study aims to guide future advancements in specularity detection models.

The findings of this survey underscore the necessity of continued research to improve the robustness and accuracy of specularity detection. This review encourages further exploration of knowledge distillation, multimodal data fusion, and self-supervised learning as promising directions to enhance generalization across diverse environments. Additionally, based on our experimental results, models such as CSFwinformer (Xie et al. 2024), SSL (Lin and Lau 2023), MG-VMD (Warren et al. 2024) and MGNet (Zhou et al. 2024) have shown outstanding performance in terms of MAE, F_β, and PSNR, with the best overall performance in specularity detection tasks across different datasets collected in this survey, within their respective input categories. Ultimately, this review is expected to serve as a foundational resource for researchers and practitioners, fostering a deeper understanding of specularity detection and inspiring innovations that drive the field forward.

Acknowledgements This research is supported by the Natural Science Foundation of China (62472361), the Suzhou Science and Technology Project-Key Industrial Technology Innovation (SYG202122), 2024 Suzhou Innovation Consortium Construction Project (LHT202406), Suzhou Municipal Key Laboratory for Intelligent Virtual Engineering (SZS2022004), the XJTLU Postgraduate Research Scholarship (Grand No. PGRS1906004), the XJTLU AI University Research Centre, Zooming New Energy-XJTLU Smart Energy Joint Laboratory, Jiangsu Province Engineering Research Centre of Data Science and Cognitive Computation, Suzhou Data Innovation Application Laboratory and SIP High level innovation platform.

Author contributions Fengze Li developed the main idea, designed the experiments, and wrote the primary text of the manuscript. All other authors provided guidance and support throughout the research experiments and writing process.

Data availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

References

Adato Y, Vasilyev Y, Ben-Shahar O, Zickler T (2007) Toward a theory of shape from specular flow. In: 2007 IEEE 11th international conference on computer vision. IEEE, pp. 1–8. https://doi.org/10.1109/ICCV.2007.4408883

Aggarwal R, Namboodiri AM (2016) Detection and segmentation of mirror-like surfaces using structured illumination. In: Proceedings of the tenth indian conference on computer vision, graphics and image processing, pp. 1–8. https://doi.org/10.1145/3009977.3010020

Akhtar S, Tandiya A, Moussa., Tarry C (2020) A robotics inspection system for detecting defects on semi-specular painted automotive surfaces. In: 2020 IEEE international conference on robotics and automation (ICRA), pp. 8928–8934. https://doi.org/10.1109/ICRA40945.2020.9196980

Amanatides J (1992) Algorithms for the detection and elimination of specular aliasing. Proc Graph Interface 92:86–93. https://doi.org/10.20380/GI1992.11

Angelopoulou E (2007) Specular highlight detection based on the fresnel reflection coefficient. In: 2007 IEEE 11th international conference on computer vision. IEEE, pp. 1–8. https://doi.org/10.1109/ICCV.2007.4409097

Angelopoulou E, Poger S (2003) Color of specular highlights. In: Human vision and electronic imaging VIII. SPIE 5007:298–309. https://doi.org/10.1117/12.473905

Angelopoulou E, Poger S (2004) Specular highlights of plastic surfaces and the fresnel coefficient. In: Photonics North 2004: photonic applications in astronomy, biomedicine, imaging, materials processing, and education. SPIE, 5578:465–475. https://doi.org/10.1117/12.566776

Anwer A, Ainouz S, Saad NM, Ali SSA, Meriaudeau F (2023) Joint network for specular highlight detection and adversarial generation of specular-free images trained with polarimetric data. Neurocomputing 559:126769. https://doi.org/10.1016/j.neucom.2023.126769

Artusi A, Banterle F, Chetverikov D (2011) A survey of specularity removal methods. Comput Graph Forum 30:2208–2230. https://doi.org/10.1111/j.1467-8659.2011.01971.x

Bajcsy R, Lee SW, Leonardis A (1996) Detection of diffuse and specular interface reflections and inter-reflections by color image segmentation. Int J Comput Vision 17(3):241–272. https://doi.org/10.1007/BF00128233

Bell S, Upchurch P, Snavely N, Bala K (2015) Material recognition in the wild with the materials in context database. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3479–3487. https://doi.org/10.1109/CVPR.2015.7298970

Blake A, Bülthoff H (1990) Does the brain know the physics of specular reflection? Nature 343(6254):165–168. https://doi.org/10.1038/343165a0

Blake A, Brelstaff G (1988) Geometry from specularities. In: 1988 Second international conference on computer vision. IEEE computer society, pp. 394–395. https://doi.org/10.1109/CCV.1988.590016

Bochkovskiy A, Wang C-Y, Liao H-YM (2020) Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934

Borji A, Cheng M-M, Hou Q, Jiang H, Li J (2019) Salient object detection: a survey. Comput Visual Media 5(2):117–150. https://doi.org/10.1007/s41095-019-0149-9

Bousquet L, Lachérade S, Jacquemoud S, Moya I (2005) Leaf brdf measurements and model for specular and diffuse components differentiation. Remote Sens Environ 98(2–3):201–211. https://doi.org/10.1016/j.rse.2005.07.005

Burley B, Studios WDA (2012) Physically-based shading at disney. In: Acm Siggraph, vol. 2012, pp. 1–7. https://disneyanimation.com/publications/physically-based-shading-at-disney/

Caesar H, Uijlings J, Ferrari V (2018) Coco-stuff: Thing and stuff classes in context. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1209–1218. https://doi.org/10.1109/CVPR.2018.00132

Chang A, Dai A, Funkhouser T, Halber M, Niessner M, Savva M, Song S, Zeng A, Zhang Y (2017) Matterport3d: Learning from rgb-d data in indoor environments. arXiv preprint arXiv:1709.06158
Chaurasia S, Vanathi M (2021) Specular microscopy in clinical practice. Indian J Ophthalmol 69(3):517–524. https://doi.org/10.4103/ijo.IJO_574_20
Chen J, Song Y, Li D, Lin X, Zhou S, Xu W (2023) Specular removal of industrial metal objects without changing lighting configuration. IEEE Trans Ind Inf. https://doi.org/10.1109/TII.2023.3297613
Chen T, Goesele M, Seidel H-P (2006) Mesostructure from specularity. In: 2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06), 2, 1825–1832. https://doi.org/10.1109/CVPR.2006.182. IEEE
Ching W-S, Toh P-S, Er M-H (1993) A new specular highlights detection algorithm using multiple views. In: Proceedings of IEEE singapore international conference on networks/international conference on information engineering'93. IEEE, vol. 1, pp. 474–478. https://doi.org/10.1109/SICON.1993.515809
Cook RL, Torrance KE (1982) A reflectance model for computer graphics. ACM Trans Graph 1(1):7–24. https://doi.org/10.1145/357290.357293
Dai A, Chang AX, Savva M, Halber M, Funkhouser T, Nießner M (2017) Scannet: Richly-annotated 3d reconstructions of indoor scenes. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 5828–5839. https://doi.org/10.1109/CVPR.2017.261
DelPozo A, Savarese S (2007) Detecting specular surfaces on natural images. In: 2007 IEEE conference on computer vision and pattern recognition. IEEE, pp. 1–8. https://doi.org/10.1109/CVPR.2007.383215
Deng Z, Hu X, Zhu L, Xu X, Qin J, Han G, Heng P-A (2018) R3net: recurrent residual refinement network for saliency detection. In: Proceedings of the 27th international joint conference on artificial intelligence. AAAI Press, pp. 684–690. https://doi.org/10.24963/ijcai.2018/95
Ding X, Yang J, Hu X, Li X (2022) Learning shadow correspondence for video shadow detection. In: European conference on computer vision. Springer, pp. 705–722. https://doi.org/10.1007/978-3-031-19790-1_42
Feris R, Raskar R, Tan K-H, Turk M (2006) Specular highlights detection and reduction with multi-flash photography. J Braz Comput Soc 12(1):35–42. https://doi.org/10.1007/BF03192386
Fleming RW, Torralba A, Adelson EH (2004) Specular reflections and the perception of shape. J Vis 4(9):10–10. https://doi.org/10.1167/4.9.10
Fu G, Zhang Q, Zhu L, Lin Q, Wang Y, Fan S, Xiao C (2024) Towards high-resolution specular highlight detection. Int J Comput Vision 132(1):95–117. https://doi.org/10.1007/s11263-023-01845-3
Fu G, Zhang Q, Lin Q, Zhu L, Xiao C (2020) Learning to detect specular highlights from real-world images. In: Proceedings of the 28th ACM international conference on multimedia, pp. 1873–1881. https://doi.org/10.1145/3394171.3413586
Ghosh A, Achutha S, Heidrich W, O'Toole M (2007) Brdf acquisition with basis illumination. In: 2007 IEEE 11th international conference on computer vision. IEEE, pp. 1–8. https://doi.org/10.1109/ICCV.2007.4408935
He R, Lin J, Lau RW (2023) Efficient mirror detection via multi-level heterogeneous learning. Proc AAAI Conf Artif Intell 37:790–798. https://doi.org/10.1609/aaai.v37i1.25157
He Y, Wu B, Mao J, Jiang W, Fu J, Hu S (2024) An effective mid-based visual defect detection method for specular car body surface. J Manuf Syst 72:154–162. https://doi.org/10.1016/j.jmsy.2023.11.014
He K, Gkioxari G, Dollár P, Girshick R (2017) Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision, pp. 2961–2969. https://doi.org/10.1109/ICCV.2017.322
Huang Y, Yue H, Fang Y, Wang W, Liu Y (2019) Structured-light modulation analysis technique for contamination and defect detection of specular surfaces and transparent objects. Opt Express 27(26):37721–37735. https://doi.org/10.1364/OE.27.037721
Jiao L, Zhang R, Liu F, Yang S, Hou B, Li L, Tang X (2021) New generation deep learning for video object detection: a survey. IEEE Trans Neural Netw Learn Syst 33(8):3195–3215. https://doi.org/10.1109/TNNLS.2021.3053249
Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg AC, Lo W-Y, *et al*(2023) Segment anything. In: Proceedings of the IEEE/CVF international conference on computer vision, pp. 4015–4026. https://doi.org/10.1109/ICCV51070.2023.00371
Le TV, Choi Y, Lee JY (2023) Specular detection and rendering for immersive multimedia. IEEE Multimedia 30(2):38–47. https://doi.org/10.1109/MMUL.2023.3262195
Li R, Pan J, Si Y, Yan B, Hu Y, Qin H (2020) Specular reflections removal for endoscopic image sequences with adaptive-rpca decomposition. IEEE Trans Med Imaging 39(2):328–340. https://doi.org/10.1109/TMI.2019.2926501
Li F, Ma J, Tian Z, Ge J, Liang H-N, Zhang Y, Wen T (2022) Mirror-yolo: a novel attention focus, instance segmentation and mirror detection model. In: 2022 7th International conference on frontiers of signal processing (ICFSP). IEEE, pp. 76–80. https://doi.org/10.1109/ICFSP55781.2022.9925001

Lin J, Lau RWH (2023) Self-supervised pre-training for mirror detection. In: 2023 IEEE/CVF international conference on computer vision (ICCV), pp. 12193–12202. https://doi.org/10.1109/ICCV51070.2023.01123
Lin S, Shum H-Y (2001) Separation of diffuse and specular reflection in color images. In: Proceedings of the 2001 Ieee computer society conference on computer vision and pattern recognition. Cvpr 2001. IEEE, vol. 1,. https://doi.org/10.1109/CVPR.2001.990495
Lin J, Tan X, Lau RWH (2023) Learning to detect mirrors from videos via dual correspondences. In: 2023 IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp. 9109–9118. https://doi.org/10.1109/CVPR52729.2023.00879
Lin J, Wang G, Lau RWH (2020) Progressive mirror detection. In: 2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp. 3694–3702. https://doi.org/10.1109/CVPR42600.2020.00375
Lippincott HW, Stark H (1982) Optical-digital detection of dents and scratches on specular metal surfaces. Appl Opt 21(16):2875–2881. https://doi.org/10.1364/AO.21.002875
Liu J-J, Hou Q, Cheng M-M, Feng J, Jiang J (2019) A simple pooling-based design for real-time salient object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 3917–3926. https://doi.org/10.1109/CVPR.2019.00404
Liu C, Kim K, Gu J, Furukawa Y, Kautz J (2019) Planercnn: 3d plane detection and reconstruction from a single image. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 4450–4459. https://doi.org/10.1109/CVPR.2019.00458
Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S, Guo B (2021) Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF international conference on computer vision, pp. 10012–10022. https://doi.org/10.1109/iccv48922.2021.00986
Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3431–3440. https://doi.org/10.1109/TPAMI.2016.2572683
Lvovsky AI (2013) Fresnel equations. Encyclopedia Opt Eng 27:1–6
Ma M, Xia C, Xie C, Chen X, Li J (2023) Boosting broader receptive fields for salient object detection. IEEE Trans Image Process 32:1026–1038. https://doi.org/10.1109/TIP.2022.3232209
Mallick SP, Zickler T, Belhumeur PN, Kriegman DJ (2006) Specularity removal in images and videos: A pde approach. In: computer vision–ECCV 2006: 9th European conference on computer vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I 9. Springer, pp. 550–563. https://doi.org/10.1007/11744023_43
Mao Y, Zhang J, Wan Z, Tian X, Li A, Lv Y, Dai Y (2024) Generative transformer for accurate and reliable salient object detection. IEEE transactions on circuits and systems for video technology, 1–1 https://doi.org/10.1109/TCSVT.2024.3469286
Marlow PJ, Kim J, Anderson BL (2012) The perception and misperception of specular surface reflectance. Curr Biol 22(20):1909–1913. https://doi.org/10.1016/j.cub.2012.08.009
Mei H, Dong B, Dong W, Peers P, Yang X, Zhang Q, Wei X (2021) Depth-aware mirror segmentation. In: 2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp. 3043–3052. https://doi.org/10.1109/CVPR46437.2021.00306
Minin I, Geints YE, Zemlyanov A, Minin O (2020) Specular-reflection photonic nanojet: physical basis and optical trapping application. Opt Express 28(15):22690–22704. https://doi.org/10.1364/OE.400460
Mišiak M, Fuhrmann A, Latoschik ME (2023) A subjective quality assessment of temporally reprojected specular reflections in virtual reality. In: 2023 IEEE conference on virtual reality and 3D user interfaces abstracts and workshops (VRW), pp. 825–826. https://doi.org/10.1109/VRW58643.2023.00255
Molina J, Solanes JE, Arnal L, Tornero J (2017) On the detection of defects on specular car body surfaces. Robot Comput-Integrated Manuf 48:263–278. https://doi.org/10.1016/j.rcim.2017.04.009
Monkam P, Wu J, Lu W, Shan W, Chen H, Zhai Y (2021) Easyspec: automatic specular reflection detection and suppression from endoscopic images. IEEE Trans Comput Imaging 7:1031–1043. https://doi.org/10.1109/TCI.2021.3112117
Moon P (1940) A table of fresnel reflections. J Math Phys 19(1–4):1–33
Morel O, Stolz C, Meriaudeau F, Gorria P (2006) Active lighting applied to three-dimensional reconstruction of specular metallic surfaces by polarization imaging. Appl Opt 45(17):4062–4068. https://doi.org/10.1364/AO.45.004062
Morel O, Meriaudeau F, Stolz C, Gorria P (2005) Polarization imaging applied to 3d reconstruction of specular metallic surfaces. In: Machine vision applications in industrial inspection XIII 5679:178–186. https://doi.org/10.1117/12.586815. SPIE
Morgand A, Tamaazousti M (2014) Generic and real-time detection of specular reflections in images. In: 2014 International conference on computer vision theory and applications (VISAPP). IEEE 1:274–282. https://doi.org/10.5220/0004680102740282

Moriyama D, Ueda Y, Koga T, Suetake N, Uchino E (2018) Illuminant color estimation method based on pseudo-detection of specular reflection. In: 2018 25th IEEE international conference on image processing (ICIP). IEEE, pp. 2262–2266. https://doi.org/10.1109/ICIP.2018.8451423
Oren M, Nayar SK (1997) A theory of specular surface geometry. Int J Comput Vision 24(2):105–124. https://doi.org/10.1023/A:1007954719939
Oren M, Nayar SK (1994) Generalization of lambert's reflectance model. In: Proceedings of the 21st annual conference on computer graphics and interactive techniques. SIGGRAPH '94, pp. 239–246. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/192161.192213
Qin X, Zhang Z, Huang C, Gao C, Dehghan M, Jagersand M (2019) Basnet: Boundary-aware salient object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 7479–7489. https://doi.org/10.1109/CVPR.2019.00766
Roth S, Black MJ (2006) Specular flow and the recovery of surface structure. 2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06) 2:1869–1876. https://doi.org/10.1109/CVPR.2006.290
Shaikh MB, Chai D (2021) Rgb-d data-based action recognition: a review. Sensors 21(12):4246. https://doi.org/10.3390/s21124246
Shen H-L, Cai Q-Y (2009) Simple and efficient method for specularity removal in an image. Appl Opt 48(14):2711–2719. https://doi.org/10.1364/ao.48.002711
Shen H-L, Zheng Z-H (2013) Real-time highlight removal using intensity ratio. Appl Opt 52(19):4483–4493. https://doi.org/10.1364/AO.52.004483
Shen Y, Deng Y, Li T, Zhou L, Feng S, Zhu H (2022) Determining multidimensional diffuse reflection effects in city tunnel lighting environment. Build Environ 212:108796. https://doi.org/10.1016/j.buildenv.2022.108796
Silberman N, Hoiem D, Kohli P, Fergus R (2012) Indoor segmentation and support inference from rgbd images. In: European conference on computer vision. Springer, pp. 746–760. https://doi.org/10.1007/978-3-642-33715-4_54
Son M, Lee Y, Chang HS (2020) Toward specular removal from natural images based on statistical reflection models. IEEE Trans Image Process 29:4204–4218. https://doi.org/10.1109/TIP.2020.2967857
Song X, Jiang S, Wang B, Chen C, Chen G (2019) Image representations with spatial object-to-object relations for rgb-d scene recognition. IEEE Trans Image Process 29:525–537. https://doi.org/10.1109/TIP.2019.2933728
Song S, Lichtenberg SP, Xiao J (2015) Sun rgb-d: A rgb-d scene understanding benchmark suite. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 567–576. https://doi.org/10.1109/CVPR.2015.7298655
Süsstrunk S, Buckley R, Swen S (1999) Standard rgb color spaces. In: Color and imaging conference, vol. 7, pp. 127–134. Society of imaging science and technology. https://infoscience.epfl.ch/handle/20.500.14299/213105
Tan RT (2021) In: Ikeuchi, K. (ed.) Specularity, Specular Reflectance, pp. 1185–1188. Springer, Cham. https://doi.org/10.1007/978-3-030-63416-2_538
Tan X, Lin J, Xu K, Chen P, Ma L, Lau RWH (2023) Mirror detection with the visual chirality cue. IEEE Trans Pattern Anal Mach Intell 45(3):3492–3504. https://doi.org/10.1109/TPAMI.2022.3181030
Tan J, Lin W, Chang AX, Savva M (2021) Mirror3d: Depth refinement for mirror surfaces. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 15990–15999. https://doi.org/10.1109/CVPR46437.2021.01573
Tian X, Zhang J, Xiang, M, Dai Y (2023) Modeling the distributional uncertainty for salient object detection models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 19660–19670. https://doi.org/10.1109/CVPR52729.2023.01883
Umeyama S, Godin G (2004) Separation of diffuse and specular components of surface reflection by use of polarization and statistical analysis of images. IEEE Trans Pattern Anal Mach Intell 26(5):639–647. https://doi.org/10.1109/TPAMI.2004.1273960
Vasluianu F-A, Seizinger T, Zhou Z, Wu Z, Chen C, Timofte R, Dong W, Zhou H, Tian Y, Chen J, *et al* (2024) Ntire 2024 image shadow removal challenge report. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 6547–6570. https://doi.org/10.1109/CVPRW63382.2024.00654
Wang W, Lai Q, Fu H, Shen J, Ling H, Yang R (2021) Salient object detection in the deep learning era: an in-depth survey. IEEE Trans Pattern Anal Mach Intell. https://doi.org/10.1109/TPAMI.2021.3051099
Wang Y, Zhou W, Mao Y, Li H (2024) Detect any shadow: segment anything for video shadow detection. IEEE Trans Circ Syst Video Technol 34(5):3782–3794. https://doi.org/10.1109/TCSVT.2023.3320688
Wang T, Hu X, Wang Q, Heng P-A, Fu C-W (2020) Instance shadow detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 1880–1889. https://doi.org/10.1109/CVPR42600.2020.00195

Wang Y, Wang R, Fan X, Wang T, He X (2023) Pixels, regions, and objects: multiple enhancement for salient object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 10031–10040. https://doi.org/10.1109/CVPR52729.2023.00967

Warren A, Xu K, Lin J, Tam GK, Lau RW (2024) Effective video mirror detection with inconsistent motion cues. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 17244–17252. https://doi.org/10.1109/CVPR52733.2024.01632

Wen S, Zheng Y, Lu F (2021) Polarization guided specular reflection separation. IEEE Trans Image Process 30:7280–7291. https://doi.org/10.1109/TIP.2021.3104188

Wolff LB (1990) Polarization-based material classification from specular reflection. IEEE Trans Pattern Anal Mach Intell 12(11):1059–1071. https://doi.org/10.1109/34.61705

Woo S, Park J, Lee J-Y, Kweon IS (2018) Cbam: Convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV), pp. 3–19. https://doi.org/10.1007/978-3-030-01234-2_1

Wu Z, Zhuang C, Shi J, Guo J, Xiao J, Zhang X, Yan D-M (2021) Single-image specular highlight removal via real-world dataset construction. IEEE Trans Multimedia. https://doi.org/10.1109/TMM.2021.3107688

Wu Y-H, Liu Y, Zhang L, Cheng M-M, Ren B (2022) Edn: Salient object detection via extremely-downsampled network. IEEE Trans Image Process 31:3125–3136. https://doi.org/10.1109/TIP.2022.3164550

Wu Z, Guo J, Zhuang C, Xiao J, Yan D-M, Zhang X (2023) Joint specular highlight detection and removal in single images via unet-transformer. Comput Visual Media 9(1):141–154. https://doi.org/10.1007/s41095-022-0273-9

Wu Z, Su L, Huang Q (2019) Cascaded partial decoder for fast and accurate salient object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 3907–3916. https://doi.org/10.1109/CVPR.2019.00403

Xie Z, Wang S, Yu Q, Tan X, Xie Y (2024) Csfwinformer: Cross-space-frequency window transformer for mirror detection. IEEE Trans Image Process. https://doi.org/10.1109/TIP.2024.3372468

Xie S, Girshick R, Dollár P, Tu Z, He K (2017) Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1492–1500. https://doi.org/10.1109/CVPR.2017.634

Xu K, Siu TW, Lau RW (2024) Zoom: learning video mirror detection with extremely-weak supervision. Proc AAAI Conf Artif Intell 38:6315–6323. https://doi.org/10.1609/aaai.v38i6.28450

Yang X, Mei H, Xu K, Wei X, Yin B, Lau R (2019) Where is my mirror? In: 2019 IEEE/CVF international conference on computer vision (iccv), pp. 8808–8817. https://doi.org/10.1109/iccv.2019.00890

Yu D, Han J, Jin X, Han J (2014) Efficient highlight removal of metal surfaces. Signal Process 103:367–379. https://doi.org/10.1016/j.sigpro.2013.11.021

Zhang W, Qiu J, Liu L (2015) Deviation characteristics of specular reflectivity of micro-rough surface from fresnel's equation. J Quant Spectrosc Radiat Transfer 160:50–62. https://doi.org/10.1016/j.jqsrt.2015.03.023

Zhang A, Zhao Y, Wang S (2021) An improved augmented-reality framework for differential rendering beyond the Lambertian-world assumption. IEEE Trans Visual Comput Graph 27(12):4374–4386. https://doi.org/10.1109/TVCG.2020.3004195

Zhang H, Ran X, Zhou W (2024) Self-knowledge distillation-based staged extraction and multiview collection network for rgb-d mirror segmentation. IEEE Signal Process Lett. https://doi.org/10.1109/LSP.2024.3386470

Zhao J-X, Liu J-J, Fan D-P, Cao Y, Yang J, Cheng M-M (2019) Egnet: Edge guidance network for salient object detection. In: Proceedings of the IEEE/CVF international conference on computer vision, pp. 8779–8788. https://doi.org/10.1109/ICCV.2019.00887

Zhao H, Shi J, Qi X, Wang X, Jia J (2017) Pyramid scene parsing network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2881–2890. https://doi.org/10.1109/CVPR.2017.660

Zhou B, Zhao H, Puig X, Xiao T, Fidler S, Barriuso A, Torralba A (2019) Semantic understanding of scenes through the ade20k dataset. Int J Comput Vision 127(3):302–321. https://doi.org/10.1007/s11263-018-1140-0

Zhou W, Cai Y, Zhang L, Yan W, Yu L (2024) Utlnet: Uncertainty-aware transformer localization network for rgb-depth mirror segmentation. IEEE Trans Multimedia 26:4564–4574. https://doi.org/10.1109/TMM.2023.3323890

Zhou W, Cai Y, Qiang F (2024) Morphology-guided network via knowledge distillation for rgb-d mirror segmentation. IEEE Trans Intell Transp Syst. https://doi.org/10.1109/TITS.2024.3404654

Zhou H, Lin Y, Yang L, Lai J, Xie X (2024) Benchmarking deep models on salient object detection. Pattern Recogn 145:109951. https://doi.org/10.1016/j.patcog.2023.109951

Zhou B, Zhao H, Puig X, Fidler S, Barriuso A, Torralba A (2017) Scene parsing through ade20k dataset. In: Proceedings of the IEEE conference on computer vision and pattern recognition. https://doi.org/10.1109/CVPR.2017.544

Zhu H, Wei H, Li B, Yuan X, Kehtarnavaz N (2020) A review of video object detection: datasets, metrics and methods. Appl Sci 10(21):7834. https://doi.org/10.3390/app10217834

Zhu Y, Fu X, Cao C, Wang X, Sun Q, Zha Z-J (2022) Single image shadow detection via complementary mechanism. In: Proceedings of the 30th ACM international conference on multimedia, pp. 6717–6726. https://doi.org/10.1145/3503161.3547904

Authors and Affiliations

Fengze Li[1,2] · Jieming Ma[1] · Hai-Ning Liang[3] · Zhongbei Tian[4] · Zhijing Wu[5] · Tianxi Wen[6] · Dawei Liu[1]

✉ Jieming Ma
jieming.ma@xjtlu.edu.cn

Fengze Li
fengzeli@liverpool.ac.uk

Hai-Ning Liang
hainingliang@hkust-gz.edu.cn

Zhongbei Tian
z.tian@bham.ac.uk

Zhijing Wu
zw436@cam.ac.uk

Tianxi Wen
ucapwen@ucl.ac.uk

Dawei Liu
dawei.liu@xjtlu.edu.cn

1 School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou, China

2 Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool, UK

3 Computation Media and Arts Thrust, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China

4 Department of Electronic, Electrical and Systems Engineering, University of Birmingham, Birmingham, UK

5 Department of Physics, University of Cambridge, Cambridge, UK

6 Department of Physics & Astronomy, UCL, London, UK