Surround-View Fisheye Optics in Computer Vision and Simulation: Survey and Challenges

Daniel Jakab[ID], Brian Michael Deegan[ID], Sushil Sharma, Eoin Martino Grua[ID], Jonathan Horgan, Enda Ward[ID], Pepijn van de Ven[ID], Anthony Scanlan[ID], and Ciarán Eising[ID], *Senior Member, IEEE*

Abstract— In this paper, we provide a survey on automotive surround-view fisheye optics, with an emphasis on the impact of optical artifacts on computer vision tasks in autonomous driving and ADAS. The automotive industry has advanced in applying state-of-the-art computer vision to enhance road safety and provide automated driving functionality. When using camera systems on vehicles, there is a particular need for a wide field of view to capture the entire vehicle's surroundings, in areas such as low-speed maneuvering, automated parking, and cocoon sensing. However, one crucial challenge in surround-view cameras is the strong optical aberrations of the fisheye camera, which is an area that has received little attention in the literature. Additionally, a comprehensive dataset is needed for testing safety-critical scenarios in vehicle automation. The industry has turned to simulation as a cost-effective strategy for creating synthetic datasets with surround-view camera imagery. We examine different simulation methods (such as model-driven and data-driven simulations) and discuss the simulators' ability (or lack thereof) to model real-world optical performance. Overall, this paper highlights the optical aberrations in automotive fisheye datasets, and the limitations of optical reality in simulated fisheye datasets, with a focus on computer vision in surround-view optical systems.

Index Terms— Surround-view, fisheye, field-of-view (FOV), optical effects, chromatic aberration, astigmatism, vignetting, computer vision, simulation, synthetic data, fisheye projection.

I. INTRODUCTION

COMPUTER Vision is a field of research where algorithms are designed to computationally interpret the scene from images of the real world that the camera is in. With the coming to the fore of neural networks in the last decade, computer vision research has grown exponentially.
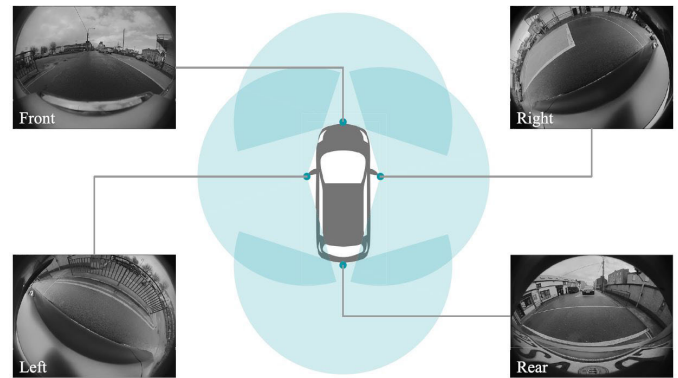
Fig. 1. **Illustration of a surround-view camera system for automated driving**. The cameras are FV (top left), MVR (top right), MVL (bottom left), and RV (bottom right) [1].

Automated driving is an industry where computer vision has begun to dominate [2]. In parallel, surround-view camera systems have received attention. Through a combination of several fisheye cameras located on a vehicle, such camera systems offer a full 360° view around the vehicle, as illustrated in Figure 1. Surround-view systems were traditionally employed for scene-viewing applications (such as blind-zone monitoring [3] and bird-eye view vision [4]). More recently, there has been particular interest in the computer vision tasks that can be undertaken by surround-view systems, developing from low-speed applications several years ago [5] to more complete perception tasks in vehicle autonomy [1], [6]. Part of this field of research is the creation of datasets of images from automotive scenes using specific camera and lens combinations. These datasets are the basis for developing, training, and deploying computer vision models in the real world. However, due to variations in optics and electronics (both in design and due to manufacturing tolerances), different cameras capture different scenes in different ways.

The impact of the electronic Image Signal Processing (ISP) on the performance of computer vision is discussed in detail in [7] and [8]. The impact of the variation in optics on computer vision performance is, perhaps, not well studied in a single place. In particular, wide fisheye cameras tend to suffer more from optical artifacts and are much more apparent in visual imagery than in other camera types [9, p.233]. Primarily, the large incidence angle of light due to field curvature in fisheye systems leads to much more serious astigmatism and a higher-order chromatic aberration, which is difficult to correct

especially as the FOV angle increases [10], [11]. Most papers only discuss strategies to compensate or correct for optical artifacts which is much more apparent in fisheye or wide FOV panoramic optical systems [11], [12], [13]. In contrast, in this paper, we discuss the optical artifacts associated with wide FOV systems and their impact on computer vision performance.

Fisheye cameras are unique and not well understood in the computer vision space as a result of the extreme optical artifacts present in the images [10], [11], [14]. For this reason, more experiments are needed to understand how computer vision performance is affected by the optical artifacts of this type of camera lens. One can consider that optical artifacts increase the further off-axis a light ray is. With the wide optics of fisheye, there is, therefore, a greater potential for optical artifacts. Thus, in this paper, we will discuss what is known about the impact of optical aberrations on computer vision performance in surround-view systems.

For safety-critical applications, such as vehicle autonomy, this lack of understanding of computer vision model behavior with different lenses is potentially dangerous, especially if all potential scenes cannot be investigated. Outcomes on the road can only be investigated by examining different traffic scenarios, including corner cases that rarely occur and are the most challenging to reproduce in test facilities. Simulation can be a powerful tool where any corner-case traffic scenario can be reconstructed as additional data for training computer vision models. In this paper, we will survey the popular vision simulation methods, and pay particular attention to their ability to simulate surround-view cameras accurately, including fisheye geometry and other optical artifacts. Not only should the optical artifacts in simulations be accurate for training, but if we can adapt the parameters of a simulated optical model, it would be possible to show how computer vision models behave using different camera lenses or cameras produced within varying manufacturing tolerances. However, despite its popularity in automated driving research, the use of simulation in this area is a relatively new development. For the particular use case of surround-view systems, significant further development is required for simulations to approach photorealism.

Since simulation is identified as a potential solution to the scarcity of well-sampled fisheye data, a problem exists where computer vision models should be trained with simulation data and deployed into the real world. However, as will be shown in this paper it is not as simple as applying fisheye distortion to simulation. The entire lens as well as its realistic effects on real-world images should be modelled in simulation. For example, Figure 2 shows a fisheye image with many readily apparent optical artifacts. For example, mechanical and optical vignetting is visible at the corners of the image where there is a slight shadow cast on the pixels. The effects of geometric distortion can be seen as the shape of the building is deformed from being rectangular to a convex-like shape. Lateral chromatic aberration is evident on the sides of buildings and trees. In particular, the left outline of the building exhibits a deep blue color, which contrasts with the light blue sky (see Section III-B Figure 5). The blurriness of the buildings and scenery constitutes both sagittal and tangential astigmatism.

The lack of optical artifacts in simulation is a shortcoming because the effects caused by optical models are a crucial part of the real world. If these are not simulated, there is a lack of information provided to computer vision models in training, leading to potentially unpredictable performance on the road. Therefore, in current circumstances, computer vision models cannot learn the full scope of the real world in simulation and arguably cannot be deployed completely safely as a result.

There have been other surveys in the field discussing fisheye cameras for automotive systems and simulation. None of them has a specific focus on the optical background for automotive surround-view systems and the optical artifacts that are created by the optical systems, other than the obvious geometric distortion. An early survey [3] focused on the scene-viewing application of blind-zone monitoring and discussed the effects of fisheye geometric distortion and light fall-off. In [15], a part-survey, part-positional argument is provided on how surround-view perception systems should be structured, based on what the authors coined the 4Rs of automotive surround-view computer vision (Reconstruction, Recognition, Relocalization, and Reorganization), itself based on earlier work known as the 3Rs of computer vision [16]. In Kumar et al. [1], a review of optical systems in automated driving for fisheye (i.e. fisheye projection models, the surround-view camera system, perception tasks, and popular automotive datasets) is conducted. In [17], a superficially similar survey to [1] is provided. However, in [17], the authors provide more detail on automotive fisheye datasets (both real and simulated), and the three tasks of object detection and tracking, semantic segmentation, mapping and localization, and surround-view monitoring. Other fisheye-related surveys such as Gao et al. [18], focus on the different types of wide field-of-view (FOV) systems for 360° panoramic, catadioptric as well as briefly discussing fisheye segmentation for autonomous driving. Finally, Mütsch et al [19] provide an insightful overview of the advantages and drawbacks of different simulator types for autonomous driving, where an important insight into the different types of simulation available, such as model-driven and data-driven simulation, is provided. A lot of crucial points on simulation are raised in Mutsch's work which will be investigated here from the surround-view camera perspective.

In all of the above surround-view surveys, only the obvious geometric distortion is surveyed in detail, and as such, we only mention this topic briefly in this paper. In none of the above are the other optical artifacts of surround-view cameras discussed (except for [3] that discusses light fall off). While simulations are discussed in these surveys as a means to augment datasets for automated driving development, there is also no discussion provided of how realistic the optics of the simulations are. In this paper, we aim to provide this discussion. Therefore, to complement the other surveys mentioned, this paper will (1) elaborate specifically on the types of image quality effects found in fisheye optical systems, (2) investigate how these optical artifacts impact computer vision, (3) establish the current trend in simulation frameworks, distinguishing between
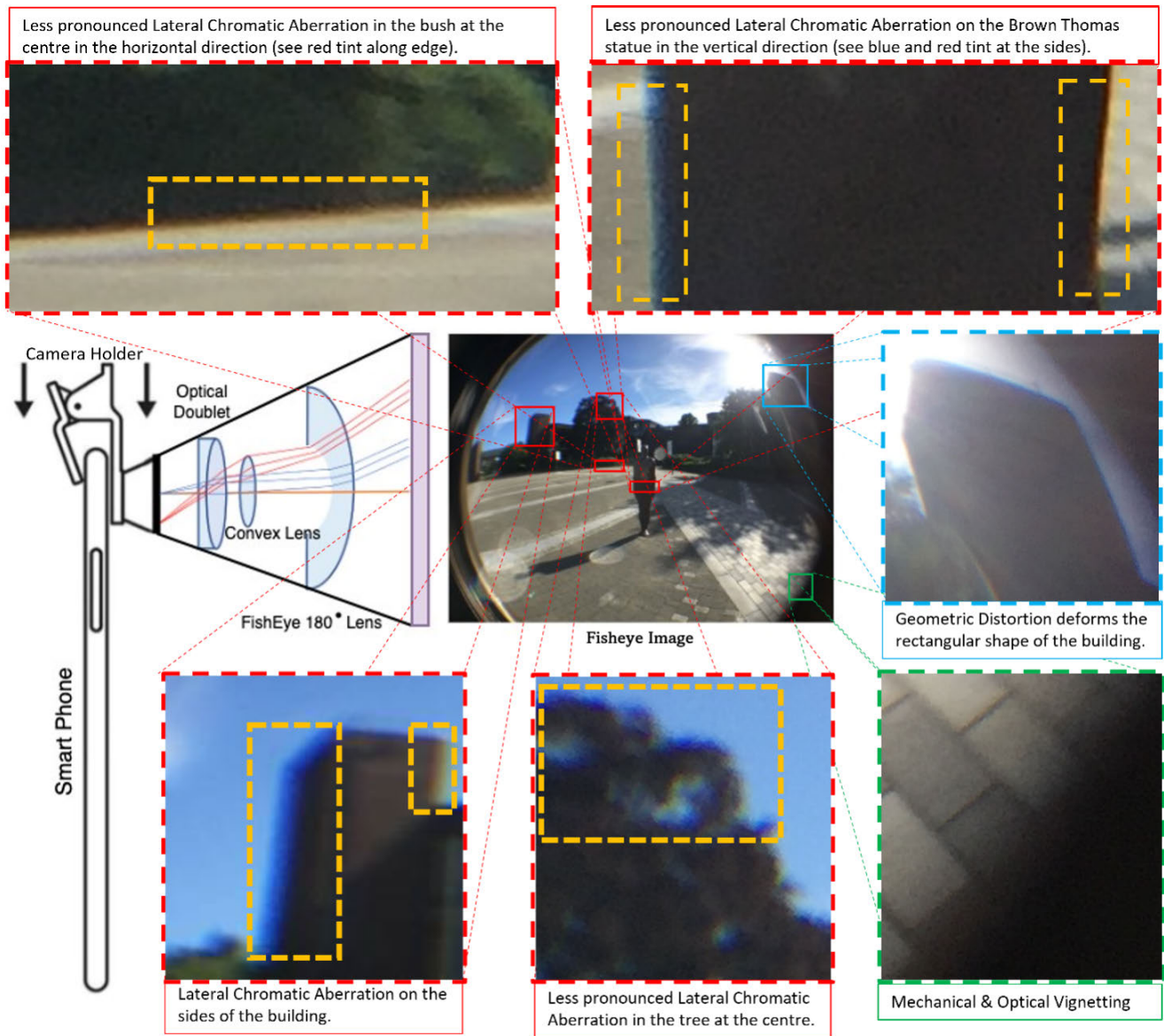
Fig. 2. **Illustration of 180° fisheye lens/camera combination**. The captured image shows many optical artifacts associated with fisheye cameras, which we will discuss later in the paper. Note: the marked dotted boxes (orange) represent lateral chromatic aberration (red) and strong astigmatism or optical blur, geometric distortion effects on a building (blue), and both mechanical and optical vignetting at the periphery with slight shadow (green).

different simulation approaches applicable for automated driving and considering the accuracy of optical artifacts in current simulations.

This paper is organized as follows. In Section II, we will briefly discuss what surround-view optical systems are and how they are designed. In Section III, we will discuss the main image quality factors including Chromatic Aberration, Astigmatism, Vignetting, and Geometric Distortion specifically for fisheye cameras and how these affect computer vision. Then we examine computer vision applied to all surround-view systems in Section IV, including surveying some recent new proposals on how to measure computer vision performance considering the spatially variant optics of surround-view cameras. Finally, in Section V, an in-depth discussion is made where the photorealism and image quality of simulations are discussed and how fisheye can become part of computer simulators (i.e. CARLA, Unity, etc.).

II. FISHEYE OPTICAL DESIGN

The optical design of a fisheye lens is a significant undertaking, and we do not intend to cover all the details here. This is the topic of papers (e.g., [20], [21]) and books (e.g., [9]) in itself. The aim here is to give a reader some concept of the complexity of the design and to highlight that, while much of the optical design is to combat optical artifacts, it is impossible to completely remove such artifacts.

The goal of an optical camera system is to form a sharp image on the image sensor. We can geometrically consider this as light rays converging on an image plane, as demonstrated in Figure 2, where we show a fisheye image captured by a fisheye lens. The optical lens illustrated in Figure 2 shows a simple three-lens system, though fisheye lenses commonly have more elements. The system shown consists of (1) a fisheye lens, i.e., a lens with a very wide FOV; (2) a convex lens where light rays converge upon exiting; and (3) an

optical doublet which is a lens consisting of two different optical media and is typically found in fisheye camera systems. In an optical doublet, the two different media are typically two different types of glass material fused together, which aids in the reduction of chromatic aberration in lenses (discussed in further detail in Section III-B). Naturally, fisheye optical systems can become more complex when additional lenses are added to this system depending on its application. In [9, 259-266], further detail is introduced about ray tracing by modeling optical surfaces as matrices.

Ray tracing is a method for modeling the projection of a virtual ray (representing the light ray) through an optical system, analyzing the refraction of light as it enters different lens media (OpticStudio[1] is an example of such a ray-tracing software tool). An optical system has virtual light rays that should ideally converge at a specific point at the end of an optical system. Both the thickness and the refractive index of the medium affect the direction of virtual rays. In an optical system different types of media can be placed (i.e., lenses with different refractive indices) together using Snell's law [9, 259]. Each lens has an individual power associated with it which is the reciprocal of its focal length as described by the equation:

$$\phi_i = \frac{1}{f_i} \tag{1}$$

where f_i is the ith focal length of a lens in an optical system where there can be n optical lenses. A typical fisheye lens (which is the large disc-shaped lens in Fig. 2) has very high optical power due to its very small focal length leading to the extreme convergence of light. Abnormally high optical powers lead to extreme optical aberrations as light bends at a wide angle through media. It is difficult to remove the optical artifacts caused by extreme lens powers. Optical Artifacts that directly affect optical systems such as Chromatic Aberration, Astigmatism, and Vignetting are discussed in Section III. The power tends to increase as more convex lenses are added to an optical system. However, not all lenses in an optical system are positive (e.g., the optical doublet shown in Fig. 2, in which the convex part of the lens is positive (converging) whereas the concave part is negative (diverging)). The overall lens can be positive if the convex part has higher power than the diverging part. The total power of an optical system (reciprocal of the focal length), can be described as:

$$\Phi = \phi_1 + \phi_2 + \ldots + \phi_i + \ldots + \phi_n \tag{2}$$

where, Φ is the total power of the optical system, ϕ_i is the power of the ith individual lens, where there are n lenses in the optical system.

The design of fisheye optics is an ongoing topic of research, with some recent advances in the use of neural networks for lens design [22]. An optical system with more lenses does not necessarily imply a better-performing system. In one sense, a more efficient optical design implies using the optimum number of lenses. An eight-element wide FOV (120°) lens was proposed in [23]. However, this system suffers from optical artifacts above 120° [21]. In [21], a simplified fisheye design

---

that optimizes optical artifacts is presented. Optical aberrations will be discussed in the following sections, where we will also survey the impact of the particular aberrations on the performance of computer vision.

III. FISHEYE OPTICAL ABERRATIONS

In the previous section, we discussed the complexity of lens design and how aberrations are not possible to completely negate through optics. In this section, we look at what the major optical aberrations are and survey their impact on computer vision: (A) Spatially Variant Optical Quality, (B) Chromatic Aberration, (C) Astigmatism, (D) Vignetting, and (D) Geometric Distortions. Specifically, the spatial variation on the image of the optical aberrations influences how objects are perceived by the computational observer.

A. Spatially Variant Optical Quality

The spatially variant optical quality of a camera system can be measured by examining the image spots generated by rays propagating through the optical system (known as a spot diagram) [27]. The spots of light created by a camera on the image plane are known as the Point Spread Functions (PSFs) of a camera system and are essentially the impulse responses of the camera system. In normal optics, the PSF can be approximately Gaussian (e.g., Figure 3a) [28], [29]. However, in fisheye optics, the PSF can be highly non-Gaussian (Figure 3b). The spread of PSFs in the image is shown in Figure 4. Ideally, a camera system is spatially invariant where the same PSF describes the spatial impulse response of the camera at any point on the image sensor. However, fisheye cameras are spatially variant optical systems where optical quality can differ greatly across the image plane.

Therefore, spatially variant optical quality should be considered when processing fisheye images. While the measurement metrics for optical quality are of significant relevance, they will only be briefly discussed here and are not the main focus of this review. Traditional techniques exist for measuring the quality of cameras in the industry such as the Modulation Transfer Function (MTF) [30], [31] and signal-to-noise ratio (SNR) (peak SNR or PSNR is a studied metric for image quality assessment alongside MTF [32], [33]). Lin et al. [34] considered evaluating image quality for motion blur and camera exposure based on both MTF and SNR techniques. However, the MTF has been considered insufficient for image applications in certain cases. For example, a study by Wolf et al. [35] shows that MTF and Refractive Power metrics are not sufficient for measuring windscreen sharpness alone for artificial intelligence algorithms especially, where the camera is located behind the windscreen. As a result, other traditional metrics such as Shannon Information Capacity (SIC) [36] and Noise Equivalent Quanta (NEQ) [37] are now being reconsidered as a way to evaluate computer vision performance and could potentially form part of a research direction in terms of computer vision for optical systems [34].

As a result of the challenge of measuring and fine-tuning optical systems for artificial intelligence, recent work has investigated the impact of modeling and varying the optical
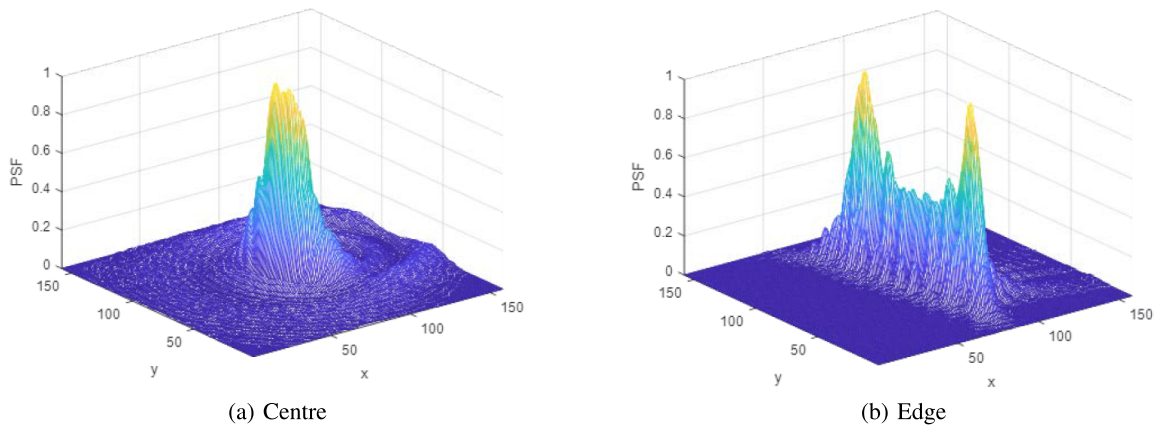
(a) Centre

(b) Edge

Fig. 3. **Measurement of PSF for a simulated fisheye lens at the center and edge locations** [24]. Note that in the center, the PSF is approximately Gaussian, and has a relatively narrower peak indicating a sharper image. The Edge PSF is highly non-Gaussian and non-isotropic.
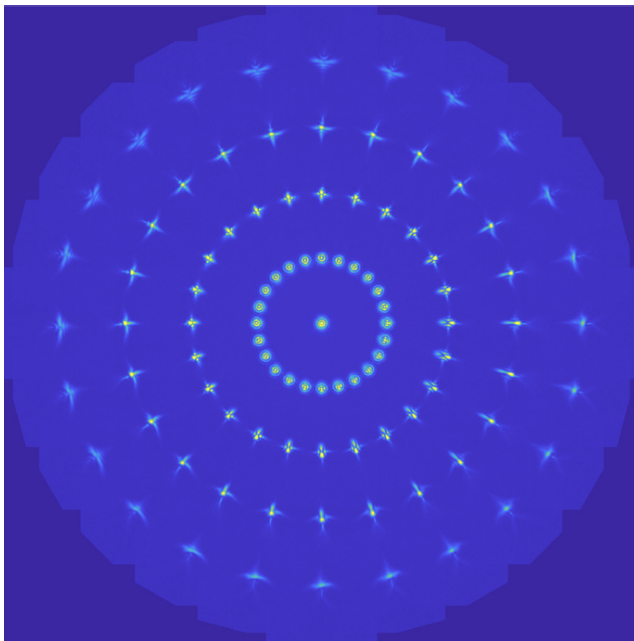

Fig. 4. **PSF model of a full image** [25], [26]. Note the highly varying individual PSFs, from approximately isotropic in the center to highly non-isotropic at the periphery.

parameters of narrow FOV optical designs using Neural Networks (NNs). In the work of Wittpahl et al. [25], for a 60° standard FOV camera, an artificial neural network is trained to predict the spatially variant point spread function, given three input (1) defocus, (2) field and (3) azimuth. This can then be used to artificially model the impact of these three parameters on the optical quality of an image, whereby the PSF model is convolved with every pixel of a target image using PSFs determined by the ANN-based prediction, with downsampling and interpolation techniques using weighted kernels.

Furthermore, Lehmann et al. [26] performed an investigation by examining the spatial resolution and accuracy of the model in [25]. The optimal topology of the ANN was found in this process by using Mean Square Error (MSE) as a decisive measure of modeling the 60° FOV model. It was found by a variation of between 8 to 448 neurons; the mean performance reached a plateau optimum of around 3 MSE within 50 to 100 neurons, with overfitting occurring when more neurons are added to the topology. For each topology configuration (i.e. variation in ANN) training was started 100 times with different randomized bias [26].

In principle, there is no reason that the models described in [25] and [26] cannot be generalized to any optical model by design choice. For example, if a wide FOV camera model is to replace the narrow FOV model, in principle it is possible to reach a custom ANN topology for that specific model using the same approach and would be interesting future work.

A caveat for the work of [25] and [26] should be highlighted. The images used were captured from a real camera. High-end optics were used in which the optical aberrations were small but non-zero. As such, the ANN-modeled PSF was applied on top of an existing PSF of the image lens, which will lead to distortion of the output image. As mentioned in [26], to apply the optical model to the pre-recorded data, the data taken by the original image lens would need to be deconvolved and then convolved with the optical model. A natural question arises as to whether it would be possible to apply the model directly to data collection in simulation environments (see Section V), which would be a fruitful direction of research.

B. Chromatic Aberration

Chromatic aberration occurs when long, medium, and short wavelengths of the visible spectrum (corresponding to red, green, and blue light) focus at different points in the image plane (i.e., do not converge at one specific point). This is an undesirable effect causing "color fringes" to appear at discontinuities (i.e., edges) in an image.

There are both lateral (i.e., light dispersing on the y-axis) and longitudinal (i.e., light dispersing on the x-axis) versions of chromatic aberration. Lateral chromatic aberration is demonstrated in Figure 5, where the further the rays of light fall from the center of the fisheye lens or optical axis, the wider the dispersion of light appears on the image plane. Lateral chromatic aberration typically increases towards the periphery of the FOV. As it is, essentially, a function of light ray incident angle, it is particularly noticeable and is very difficult to completely compensate for in fisheye lenses [38].
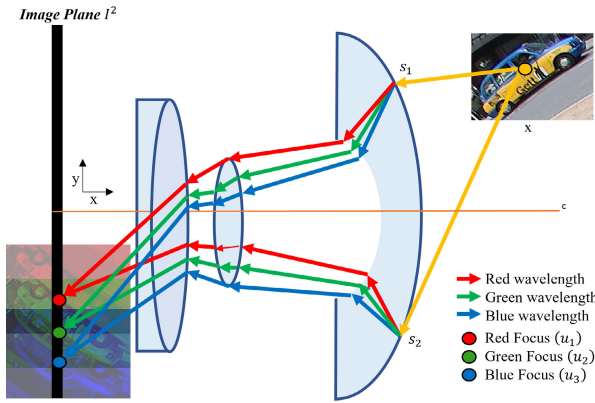
Fig. 5. **Lateral chromatic aberration on a fisheye lens**. Light splits into its components red, green, and blue (RGB) along the y-axis. We are only interested in the red, green, and blue components of the light, as an image sensor typically has only red, green, and blue pixels.

Fig. 6. **Zoomed-in segment of a Woodscape dataset image showing evidence of lateral chromatic aberration**. Note the slight red/purple and green tint in the pixels along the outline of the car and the road marking.

In Figure 6, it can be seen that lateral chromatic aberration exists in images in the fisheye Woodscape dataset [39].

Conversely, longitudinal chromatic aberration occurs across the entire image regardless of the position from which light intersects the fisheye lens. In this case, the wavelengths of light which are longer (i.e., red and infrared) tend to converge past the image plane whereas shorter wavelengths (i.e., blue) converge before the image plane. Longitudinal chromatic aberration is easier to correct where stopping down a lens (i.e., increasing the f-number) compensates for this effect. Referring to Figure 5, there is no evidence of significantly visible longitudinal chromatic aberration in Woodscape.

Chromatic aberration is typically compensated for either through achromatic doublets in the optics (such as the optical doublet seen for the fisheye lens design in Figure 2) or in the Image Signal Processing (ISP) stage before demosaicing of a camera system [40], [41]. Demosaicing is the stage at which a Bayer RAW signal (having one color per pixel) is converted into the standard RGB format. Images with lateral chromatic aberration perform poorly in demosaicing as the color effects shift the expected pixel detail in the resulting RGB images. Lateral chromatic aberration is usually corrected using Look-Up Tables calculated for a given lens type as a

post-processing step, but this is often not feasible. Recent work has examined blind methods for correcting lateral chromatic aberration [41]. It should be noted that the post-processing to compensate for chromatic aberration does not fix the problem. However, it is quite evident that it is not feasible to completely correct chromatic aberration in fisheye cameras.

Pontinen et al [14] compared chromatic aberration between a Sigma 8mm 1:4 D EX Fisheye [42] and AF DX Fisheye-Nikkor 10.5mm f/2.8G ED [43] by applying different correction models to identify which is more suitable for 180° wide FOV lenses. The four correction models used in this study were shifting and scaling, affine, projective, and polynomial. These models were applied on the red and blue color channels where the green color channel was the reference channel for correction. It was found that the shifting and scaling correction model is adequate whereas the other models didn't improve chromatic aberration correction overall. What was interesting from this study was the analysis of the camera behaviors, where the change in aperture size and focus settings on the cameras influenced chromatic aberration unique to each camera design. For example, with increasing aperture size from f/4 to f/32, the red channel aberration increased by 0.5 pixels for the Sigma lens. With the Nikon lens, there was very little variation for the red channel.

It is intuitive to think that chromatic aberration will negatively impact the performance of computer vision tasks, as it is generally considered an unwanted optical artifact. However, this is not completely clear (and there is relatively little work examining this). According to a 2019 study by Chang et al [44], it has been shown both qualitatively and quantitatively that images with chromatic aberration outperform 'all-in-focus' images (i.e., images without any optical aberrations) both in Depth Estimation and 3D Object Detection applications using the NYU Depth v2 dataset [45] and a subset of the KITTI dataset [46]. Table I gives a detailed insight into the effects of different optical artifacts according to literature. Chang et al. [44] is the only paper to date that gives evidence of a chromatic aberrated lens and a freeform lens performing better than the baseline or original data. Further evidence of this behaviour can be seen in recent work where robustness is significant against coma (a form of chromatic aberration) for a ResNeXt50 network [32]. Additionally, Depth Estimation for defocused NYUv2 images performs better than the baseline (*all-in-focus* images) for a D3-Net architecture [47] (see Table I). Whereas in all other literature concerning chromatic aberration a negative or degrading performance is observed for Object Detection and Instance Segmentation applications. However, since the optical designs used in these experiments have evidence of both chromatic aberration and astigmatism, it cannot be conclusively determined that the performance degradation is caused by any one artifact. Also in recent work [48], [49], longitudinal chromatic aberration was explicitly exploited for depth estimation on a bi-convex lens using the different spectral channels of unsharpness. The utility of chromatic aberration in tasks involving depth (e.g., depth estimation and 3D bounding box tasks) may be explained by the fact that the amount of chromatic aberration present is dependent on the depth of the imaged object,

TABLE I

COMPARISON OF OPTICAL ARTIFACTS ON COMPUTER VISION

Category	Year	Author	Dataset	Tasks	Impact	Computer Vision Performance
Chromatic Aberration & Astigmatism	2018	[47]	NYU Depth v2	Depth Estimation	positive	Using the D3-Net architecture for the Depth from Defocus approach proposed outperforms 'all-in-focus' for NYU Depth v2 RGB images with additional blur where root mean squared error (RMSE) improved from 0.706 to 0.144 which is about 79.6% error reduction.
	2019	[44]	KITTI, NYU Depth v2,	3D Object Detection, Depth Estimation,	positive	The second lowest linear MSE for Chromatic Aberration of 0.4496 (NYU Depth v2), and 2.2566 (KITTI). The third lowest linear MSE for Astigmatism of 0.4561(NYU Depth v2) and 2.3634 (KITTI). Significant improvements can also be seen for 3D Object Detection where the freeform lens improved by +8.34% AP(Easy mode), +3.2% AP (moderate mode) and +0.17% AP (hard mode) respectively. Perhaps surprisingly, this work demonstrates that a simple lens with chromatic aberration performs almost as well as a freeform lens for the tasks discussed. There is, however, some further work that backs this up [76].
	2021	[57]	BDD100k, Cityscapes,	Object Detection, Instance Segmentation,	negative	Cooke Triplet model applied to datasets for a defocus range between Δz = {-1.25, -0.75, -0.5, 0, +0.5, +0.75, +1.25}. For car detection, the baseline car 78% AP degraded to between 67%-76% AP for BDD100k. For car instance segmentation, a baseline of 68.9% AP degraded to between 58.9%-65.69% AP for Cityscapes.
	2021	[53]	BDD100k	Object Detection,	negative	Applying defocus parameters to BDD100k images shows a clear drop in spatial performance of 7% AP at the periphery of the spatial domain for Δz = -1.25 and 5.4% AP for Δz = +1.25. Overall results for degraded BDD100k show an improvement from a baseline of 90.72% AP to between 83.758% and 89.24% AP.
	2022	[59]	BDD100k	Object Detection	negative	10k images of BDD100k for the baseline (original images) 61% AP degraded by a Cooke Triplet model to between 54%-55% AP where little variations in results were observed for approximation models.
	2022	[60]	BDD100k,	Instance Segmentation	negative	Cooke Triplet model was applied to BDD100k from [53]. SPI and SRI metrics LAMR* increased from 0.6367 (baseline) to between 0.64361 and 0.68354 (degraded).
	2024	[77]	Primary Research	Object Detection	positive	The first major comprehensive analysis of primary research data for pedestrians and cars collected using a BlackFly-S 8.9MP camera and degraded between Δz = {-1.5, -1.0, -0.5, 0, +0.5, +0.1, +1.5}. Results show for a degraded dataset of -0.5λ, object detection performance for Average Precision between a 50% and 95% IoU threshold was in most cases higher than the baseline across six different deep learning models. Out of all six models Faster RCNN Swin Tiny FPN was the most robust against degradation achieving a performance difference of 7.22% (large objects) and 20.72% (small objects).
Chromatic Aberration	2018	[75]	KITTI,	Object Detection	neutral	Modeling optical artifacts by degrading synthetic data (i.e., VKITTI and GTA) to improve performance on real-life data (i.e., KITTI). For 21k VKITTI images, 58.25% AP (baseline KITTI) improved to 62.52% (KITTI with degraded VKITTI). Similarly, for 50k GTA, 53.26% AP (baseline KITTI) improved to 55.85% AP (KITTI with degraded GTA) on the same held-out set of KITTI. While introducing Chromatic Aberration into simulation shows improvements; it does not prove it has either a positive or negative impact on real-life data.
Vignetting	2019	[70]	KITTI,	Object Detection	positive	Applying vignetting to KITTI images using an integration sphere improved precision from 83.12% to 88.96% which (somewhat surprisingly) gives a gain of 5.84% for the Faster RCNN with ResNet50.
	2021	[62]	DEV, CIFAR-10, Tiny ImageNet,	Adversarial Attack & Classification	negative	Radial Anisotropic AVA (RA-AVA) degraded images from DEV, CIFAR-10, and Tiny ImageNet datasets producing either the highest or second highest transfer success rates for all. Neural Network architectures such as ResNet50 gives degraded performance of 46.34% compared to the baseline (original) for Tiny ImageNet.
Distortion	2019	[6]	Cityscapes, SynWoodscape, SVScape	Semantic Segmentation	negative	Restricted deformable convolution improved the performance on real-world data to as high as 74.2% mIoU using the ERFNET-RDC-8 which is a VGG-based architecture neural network [78]. Proposing a new architecture for distortion gives an alternative over the traditional Computer Vision algorithms and is likely the only way to remove performance degradation of this optical artifact.
	2023	[79]	Woodscape, SynWoodscape,	Optical Flow	negative	RAFT network is finetuned using SynWoodscape and gives the lowest End Point Error (EPE) on Woodscape (1.40) compared to both Sintel (5.12) and Jointly trained (1.48) versions. However, results signify that Distortion does not have a good effect on traditional optical flow.
	2023	[80]	Woodscape,	Semantic Segmentation	negative	Viewpoint Augmentation gives the highest result (66.91% mIoU) compared to both images without augmentation (63.80% mIoU) and with affine augmentation (66.60% mIoU). Using augmentation to improve results further shows evidence that traditional Computer Vision algorithms are not able to handle distortion well

Note: The Impact column signifies any quantitative difference in statistical metrics considering optical artifacts where a positive impact implies a better performance in metrics which includes but is not limited to Average Precision (% AP), Mean Squared Error (MSE), Precision, End-point Error (EPE), mean Intersection over Union (mIoU).
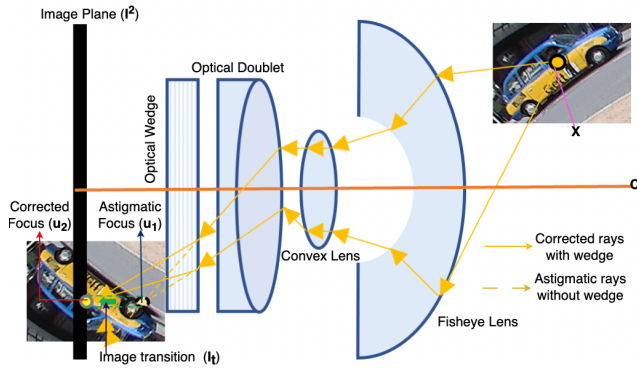
Fig. 7. **An optical wedge or prism used for astigmatism correction**. As shown in the diagram by adding the prism in front of an optical system; the rays are refracted to focus the light rays on the image plane producing a sharp image.
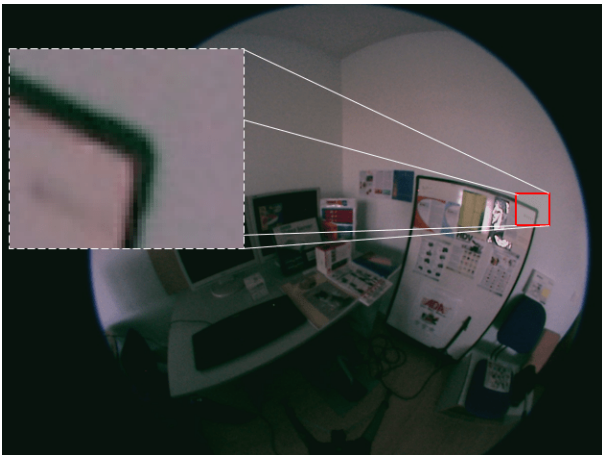

Fig. 8. **An example of a fisheye image displaying astigmatism**. Note the increased 'blur' in the horizontal direction compared to the vertical. The Woodscape dataset does not exhibit significant astigmatism - this sample is taken from the non-automotive hybrid structure-from-motion and omnidirectional camera calibration dataset [50]. Note that some chromatic aberration and vignetting are also visible.

and this relationship can potentially be exploited. However, these experiments were performed on narrow FOV cameras. As discussed, chromatic aberration can be particularly severe in fisheye images, and no work (to the authors' knowledge) explicitly examines the impact of fisheye lateral chromatic aberration on computer vision performance. Additionally, it is not clear if other tasks, such as object detection or instance segmentation, are impacted negatively by the presence of significant chromatic aberration.

C. Astigmatism

Astigmatism is an optical artifact that impacts an image by causing a directionally dependent defocus (Figure 8). It is caused by an effect whereby rays in two perpendicular planes have two different foci, as shown in Figure 9. It is introduced into the optical system either due to a lens not being rotationally symmetric or due to misalignment in lens elements and can degrade the performance of an optical system. Concerning Figure 9, the chief ray is a light ray passing through the center of the aperture of the system. From

this ray, two different planes can be defined horizontally and vertically (also known as the sagittal and tangential planes). Likewise, there exist sagittal and tangential rays that form two unique foci (i.e., F_t, F_s). Hence two different images are formed (i.e., I_t, I_s). There can be many variations of sagittal and tangential planes by varying the angle of incidence of the chief ray [9, 274-276]. In an astigmatic system, it is impossible to focus both tangential and sagittal rays on one plane to create a sharp image. It has been noted that in an astigmatic system, the orientation of the object affects how imaging occurs [51]. For example, the greater the angle between the object and the optical axis, the greater the astigmatic difference is between F_t and F_s, hence blurriness increases in the image. Notice in Figure 9, two elliptical images are formed where blurring is vertical for I_t and horizontal for I_s. As astigmatism increases, the divergence between sagittal and tangential rays increases depending on manufacturing defects. The *circle of least confusion* is the smallest circle or spot that a lens system can make. The presence of astigmatism can change depending on lens manufacturing imperfections. A typical PSF in these conditions is shown in Section II (Figure 3b) (located at the periphery of the PSF lens model in Figure 4), which represents an astigmatic system occurring at the periphery of the lens where strong radial distortion is present. Astigmatism impacts the PSF of a given camera system directly (though it is not the only impact). There have been many attempts over the past few decades to correct astigmatism in optical systems. For example, the rear half of a classical symmetric Celor lens was used to reduce spherical aberration and astigmatism in a fisheye lens design [21], [52]. As described by Muller et al [53], the Cooke Triplet is an optical system with the ability to compensate for astigmatism. In that work, the Cooke Triplet was applied on a subset of the Berkely Deep Drive (BDD100k) [54] automotive dataset for blurring between a defocus range of $\Delta z = \pm 1.25$, where a Mask RCNN [55] model was evaluated across this range of defocus. Within this range, it was shown that different Regions of Interest (ROI) from the center to the edge of an image have noticeably varying degrees of blurriness proving that the spatial domain of an image has inconsistent image quality and causes a significant drop in both statistical and spatial performance metrics for the Mask RCNN model. From these examples, it is clear that optical astigmatism has an undesired effect but is very difficult to control in complex optical systems. For wide-angle lenses such as fisheye, astigmatism is particularly problematic where the radially varying PSFs (discussed in Section II) affect the degree of blurring around the periphery of the lens.

Due to the mass production and time constraints of producing camera systems, this optical aberration cannot be eliminated. To compensate for astigmatism, optical systems could sometimes end up being rebuilt several times, but this is industrially undesirable [56], [57]. To correct linear astigmatism, a manufactured lens can be rolled on its mechanical seat so that one surface becomes tilted by an angle of α measured in arc minutes which can be estimated by (3) [56]:
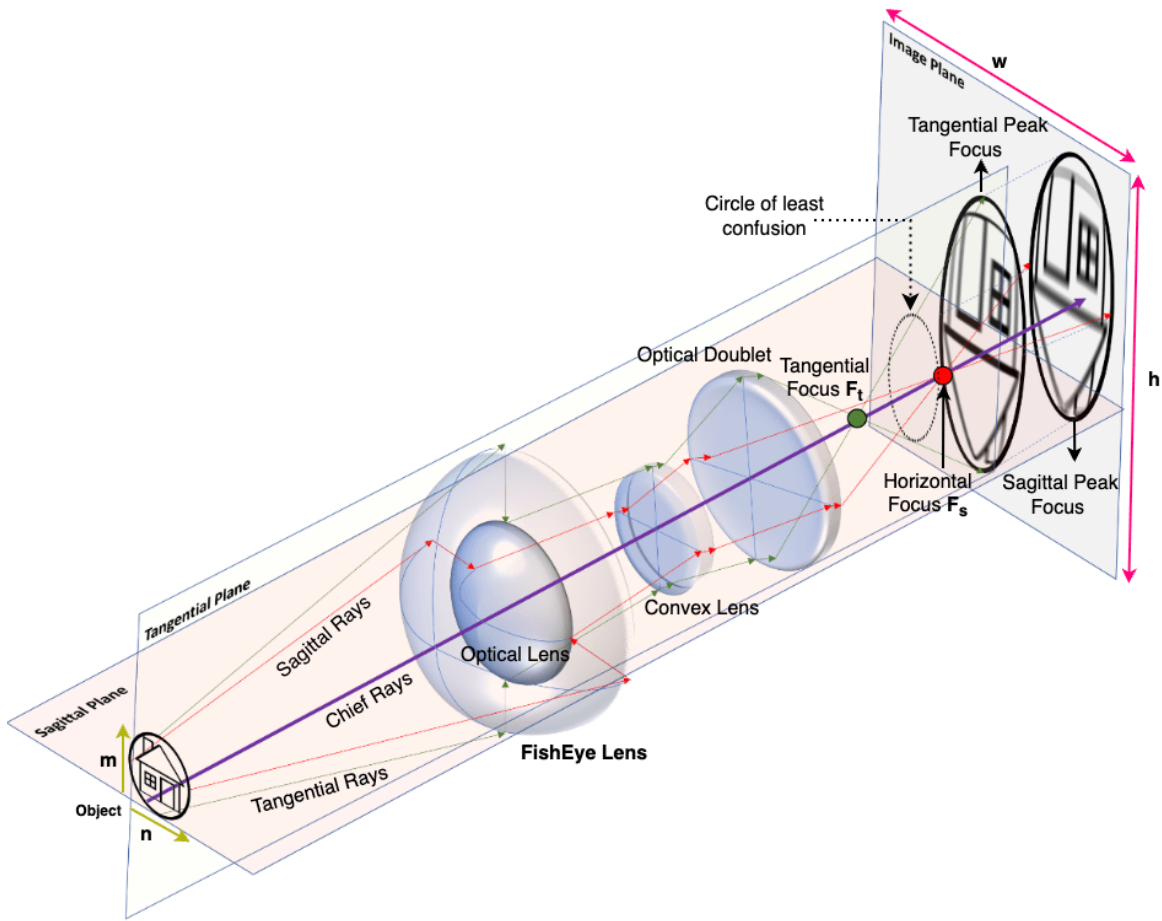
$$\alpha = \frac{d}{R} \qquad (3)$$

Fig. 9. **Demonstration of astigmatism in a fisheye lens**. Note the mismatching foci where the astigmatic images represent varying divergences of both sagittal and tangential rays.

where d is the lateral displacement of the lens due to rolling and R is the radius of curvature of the surface on the seat. Finally, a thin optical wedge is added to the optical system just before the light reaches the image plane as described by [56], [58]:

$$W_{LA} = -\frac{1 - n^2}{n} * u * \bar{u} * \alpha * y \tag{4}$$

where n is the index of refraction of the wedge, u is the slope of the first order marginal ray before entering the wedge, $\bar{u}$ is the slope of the first order chief ray before entering the wedge in image space respectively, and y is the marginal ray height at the wedge [56]. Figure 7 demonstrates these parameters on a sample optical wedge-shaped as a rectangular prism.

It has been demonstrated that astigmatism can be undesirable for computer vision algorithms as it can influence the performance of object detection and segmentation algorithms due to the blurriness of the target objects [57, p. 43]. Substantial evidence in literature all demonstrate this for computer vision [47], [53], [57], [59], [60] (see Table I). An example of a negative impact on performance can be seen in [53, p. 60-61], where it is demonstrated that using a Mask-RCNN model the Average Precision (AP) worsens in performance on the Precision-Recall (PR) curves by applying a defocus of $\Delta z = \pm 1.25$ using the Cooke Triplet

lens. Contrarily, a positive impact can be seen in 3D Object Detection results from Chang and Wetzstein [44], using an optimized mask for a freeform lens initialized with astigmatism. For example, when trained on 5500 KITTI [46] images the optimized lens model produced 37.51% Average Precision (AP) which is 10.8% higher for easy 3D object localization when compared to the all-in-focus (original) KITTI dataset. Similarly, there is an 8.34% AP improvement for 'easy' 3D object detection ('easy' being a difficulty parameter of the KITTI performance metrics). When considering astigmatism for the application of Depth Estimation, a popular term was coined in the 20th century called Depth-from-Defocus (DfD) [61]. Using DfD, Carvalho et al. [47] synthetically defocused datasets (such as NYUv2 [45]) and depth was learned from the defocused blur showing a performance improvement when compared to 'all-in-focus' (unaugmented) images. As with the PSF, machine learning tasks that require depth estimation may benefit from the depth-dependent defocus.

D. Vignetting

Vignetting (otherwise known as 'light fall off') is an optical phenomenon that causes gradual darkening at the periphery of an image. Vignetting occurs due to the varying angle of light across a sensor, the intrinsic lens characteristics, and physical occlusions [62]. It is most apparent in extremely wide
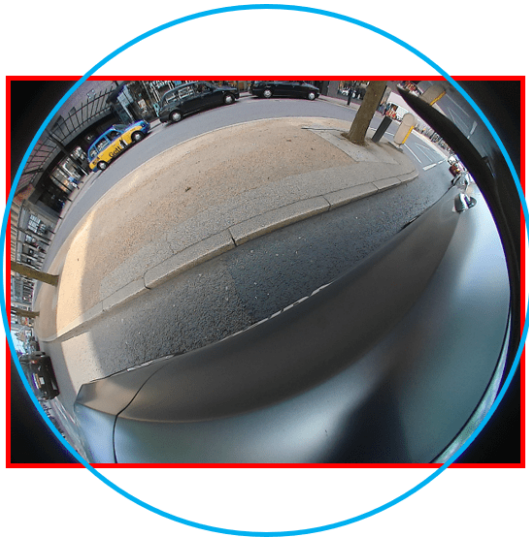
Fig. 10. **Mechanical vignetting in the Woodscape dataset**. Note the presence of black corners due to a mismatch in the sensor and lens aperture.

FOV cameras, such as fisheye. There are four main causes of vignetting [63]:

1) **Mechanical vignetting** is the physical occlusion of light to a camera sensor which gives a loss of information in the resulting image [63]. This is generally a trade-off in the design of fisheye cameras. Reducing the mechanical vignetting would necessitate the sacrifice of camera FOV, with most fisheye cameras designed either to be full-circle (ref Figure 8) or cropped-circle (ref Figure 6).

2) The **cosine-fourth-power-law** describes the illumination fall off due to the apparent elliptic eccentricity of the entrance pupil from the point of view of the incident light ray [64]. The greater the off-axis angle of the object being imaged the more elliptic the entrance pupil becomes. This increased eccentricity of the entrance pupil from the point of view of the object means less light propagates through the pupil, and hence we have an illumination fall off.

3) **Optical vignetting** occurs due to physical occlusion between lens elements (literally, rear elements are shaded by those in front). It can often be reduced by reducing the diameter of the aperture (otherwise known as 'stopping down the lens') [21].

4) **Pixel vignetting** is caused by the physical structure of the pixels in an image sensor [65]. In an image sensor, the top layer of the sensor is not (typically) the light-sensitive layer. When light is incident on a pixel, it must travel through a 'tunnel' (layers of CMOS) before it hits the photodiode. The deeper this 'tunnel' is relative to the pixel area, the more a shadow is cast by the pixel walls, causing increased light fall off with increased incident angle.

Both optical and pixel vignetting can be compensated for at the image signal processing stage (ISP) of a camera system [7], [63]. For the Woodscape dataset, vignetting is largely compensated with radial digital gain correction using a lens shading block [39]. This post-processing compensation may fix the visual appearance of the vignetting, but the signal-to-noise

ratio will not be improved. There is also mechanical vignetting along the aperture, as seen in Figure 6. The "black corners" that are distinctly visible are due to the lens hood that prevents the image sensor from seeing the complete FOV. The on-automotive hybrid structure-from-motion and omni-directional camera calibration dataset [50] shows some level of optical/pixel/cosine-fourth vignetting (ref Figure 8).

The ISO12233 standard [66] has introduced polynomial fitting and non-uniform compensation to correct for vignetting and geometric distortion effects in Edge Spatial Frequency Response (e-SFR) measurements [67]. When high (1 side) and additional low-signal (2 sides) corrections are applied, the SFR decreases in both stages, hence reducing vignetting in measurements. ISO12233 is used in measurements for surround-view cameras in autonomous vehicles [68]. In terms of computer vision performance, yet again, somewhat counterintuitively, vignetting has been shown to have some positive impact on statistical metrics for object detection. For instance, it has been shown that applying vignetting to the KITTI [46] and Virtual KITTI [69] datasets for 2D object detection on cars [70], where a Faster-RCNN [71] model with a region proposal network (RPN) (pre-trained on ImageNet [72]) was used, precision increased from 83.12% to 88.96% when vignetting was applied to the images [70] (see Table I). Also, the number of car detections increased at the periphery of the images, which is unusual as one would assume that vignetting would cause a performance drop at the periphery. However, precision alone is not sufficient for completely determining the apparent positive impact of vignetting on performance, as it is well-known that precision is biased towards the positive class [73]. As a result of insufficient statistical metrics in these experiments, more investigation is needed on the impact of vignetting on computer vision performance. On the contrary, an Adversarial Vignetting Attack (AVA) model confused a ResNet50 [74] network which demonstrates that vignetting can have an unpredictable effect on computer vision performance [62] (see Table I). More experiments are needed to understand these behaviors. Table I provides a detailed comparison of experiments in the literature aimed at assessing optical artifacts in computer vision. It is reasonable to deduce that, among the identified papers, both negative and positive effects of optical artifacts can occur depending on the nature of the examined tasks. It cannot be conclusively assumed that introducing optical artifacts into training data results in degraded performance. For example, Depth Estimation experiments show a positive impact, and augmenting synthetic data improves performance on real-world data [75]. It is also evident that individual metrics, such as precision, improve when vignetting is applied to KITTI (see Table I). However, degrading real-world images with vignetting as an adversarial attack on algorithms gives degrading performance [62]. This demonstrates that image degradation can lead to positive results for datasets with simulation but negative on real-world data.

E. Geometric Distortion

The geometry of fisheye cameras has been covered in several previous surveys [1], [3], and as such we only give a brief discussion here on the topic.
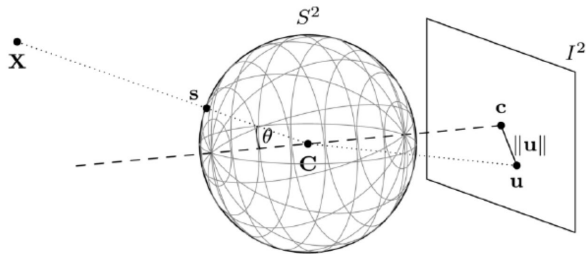
Fig. 11. **Geometric Relationship between a fisheye point at** u **on the two-dimensional image and an equivalent point at** s **on a unit sphere** [1]. The projective sphere is a more natural geometric object for modeling the rays of an omnidirectional/fisheye camera, in the same manner that the projective plane is used for narrower FOV cameras [87].

The very strong radial distortion in fisheye optics is, perhaps, the most obvious undesirable artifact. In [1], there is an in-depth discussion on different fisheye mathematical models. It is especially interesting to note a fisheye camera's relationship with the 2D image of size I^2 from a geometrical perspective. Here we give a very brief overview but would encourage readers to examine the previous surveys for more details. In Figure 11, θ is the angle at which the ray of light from an object X maps from a unit sphere onto an image plane. C is the center of the unit sphere. The radial distance $||u||$ of the projected image point **u** to the image center **c** is proportional to this incident angle θ. Some of the classic geometric projection models that represent fisheye distortion include (1) equidistant, (2) stereographic, and (3) orthographic projections. Each projection model has a slight variation of the general mapping strategy illustrated by Figure 11 [1]. Each model has its own set of unique functions for both projection and unprojection and these models in theory can be applied to fisheye images to either remove the distortion or add distortion to rectilinear images. The polynomial Brown-Conrady model (including calibration methods), is implemented in many commonly used software packages and libraries (e.g., MATLAB [88], [89]). The popular Woodscape datasets [39], [90], examined closely in this paper, use a full fourth-order polynomial model for fisheye projection:

$$||\mathbf{u}|| = r\left(\theta\right) = a_1\theta + a_2\theta^2 + a_3\theta^3 + a_4\theta^4 \quad (5)$$

Neural networks are not specifically built for distorted images which lead to an expected degrading performance [15]. For example, due to the distortion in fisheye imagery, the implicit assumption of translation invariance in CNNs does not hold true. Furthermore, the use of augmentation methods does not necessarily imply more data equates to better performance. Cho et al. [80] present viewpoint augmentation as the method that gives the least degrading performance for fisheye (see Table I). More recent developments show that deformable convolution networks [91] would effectively solve the distortion problem when considering the use of offsets and geometric transformations in images. Deng et al. [6] propose a restricted version of this strategy where the model was shown to be robust for large radial distortions in surround-view systems (see Table I).

F. Discussion

What is perhaps of most interest is that, when surveying the admittedly sparse literature, it is not clear that all optical aberrations impact the performance of all automotive computer vision tasks negatively (see Table I). Several pieces of work have demonstrated that certain tasks have been improved by optical aberrations. For certain tasks, such as those that extract depth from the image, this can make some sense due to the depth-dependent nature of certain optical aberrations. Considering the sparsity of the literature, it is evident that further work in this space is required.

It should be noted that there are some elements related to optics that we haven't discussed completely in this section. For example, images taken by cameras are also affected by natural elements of the scene, such as sunlight causing sun flare (as is the case in Fig 2). False artifacts appear near the aperture of the lens in scenes with sun flare [92]. This produces unwanted blurs at the periphery of the image and can be mistaken by computer vision systems when performing object detection. Actors such as pedestrians and cars can be occluded by the sun flare, leading to missed detections. This is especially critical when the sun is low in the sky in which case sun flare can be unavoidable despite efforts to remove it in camera production.

Additionally, there is some work ongoing on designing optical systems for computer vision. Designing optical lenses for automotive computer vision tasks is especially challenging given the requirement of using high-quality and standard high-resolution input images for the training of DNNs and other recent approaches (e.g., Vision Transformers [93] and Bidirectional Encoder representation from Image Transformers [94]). Research on lens design optimization concerning computer vision is ongoing. In Yang et al [86], a Task-Driven End-to-End Lens design approach has been introduced as a more feasible option over conventional lens designs. End-to-end optical design is a relatively new area of research that has shown promising results when applied to common computer vision tasks, including object detection. However, these lens designs rely on well-developed End-to-End optical design solutions [86], [95], [96], [97]. However, for example, the TaskLens model [86] shows a lot of potential and unexplored areas in optical design outperforming the image classification of three traditional lens designs developed using conventional lens design methods (i.e., the doublet, triplet, and quadruplet each having a 68.8° FOV and f-number of f/2.8). There is some very recent work showing a trend in the adoption of optical lens modeling using artificial intelligence as a strategy to design for computer vision performance as shown in Table II.

IV. COMPUTER VISION IN SURROUND-VIEW CAMERAS

In the previous section, we briefly discussed the impact of individual optical artifacts. In this section, we examine computer vision performance in surround-view cameras as a whole. This is illustrated using an application of the YOLOv7 object detection algorithm on the Woodscape dataset. What is evident from the previous section is that for fisheye, the optical artifacts are strongly spatially variant (in narrow FOV cameras, this spatial variance can typically be safely ignored).

TABLE II
OPTICAL LENS MODELLING WITH ARTIFICIAL INTELLIGENCE

Year	Paper	FOV	Type of Model	Computer Vision Performance
2017	[81]	160°	Multi-Layer Perceptron	Neurolens first proposed and simulated for four different lenses including fisheye where an acyclic four-layer perceptron network is trained ((1x) input layer, (2x)hidden layers and (1x) output layer). The number of layers is increased if the prediction and testing error (ϵ) exceeds (0.2% - 0.8%).
2018	[25], [26], [82]	60°	Neural Network	The first proposal of realistic image degradation using a spatially variant distribution of the PSF of an optical lens trained using non-linear regression with an ANN. Results show that training the model with the three input parameters defocus (Δz), image height (R) and azimuth (ϕ) gives a qualitatively realistic degradation. Lehmann et al. [26], [82] optimizes the same model with Mean Squared Error (MSE) reaching an optimal of 3 within 50 to 100 neurons.
2023	[83]	OC*(130° - 266°), UZH [84](124° - 166°)	Neural Network	A second Neurolens is proposed for distortion and vignetting for the Lensfun database and is extended to OC and UZH generalizing well across different lenses.
2023	[85]	68.8°	Curriculum Learning (Reinforcement Learning)	Autolens is proposed in this work where Curriculum Learning is used to design a pinhole camera model based on geometrical optics and is compatible with Zemax OpticStudio.
2023	[86]	68.8°	Neural Network	Task-Driven Lens modeling from scratch uses supervised lens design to outperform conventional lens systems in image classification where for an optical doublet ResNet50 classification accuracy improved by 1.54% for the TaskLens model and for a Large Patch vision transformer (ViT-L/16) by 0.57% respectively.

Note: OC* = Omnidirectional Camera Calibration (OCamCalib), DGauss = Double Gaussian lens

Therefore, perhaps, global metrics of performance do not give us the full picture of the performance of computer vision from fisheye cameras. In this section, we therefore also briefly discuss recent proposals on spatially dependent performance metrics.

A. Computer Vision Performance

Computer Vision research is a consistently developing field where new algorithms are being developed or optimized to solve the perception problem in images. To get an idea of the speed of development, You-Only-Look-Once version 8 (YOLOv8) [98] was released approximately 5 months after the previous version, i.e., YOLOv7 [99]. To give an indication of the rate performance improvement for YOLO, using the baseline models comparison, YOLOv7, when trained on the MS COCO [100] dataset with 75% fewer parameters than YOLOv4 [101], is capable of reaching 66.7% AP which is 1.5% better than YOLOv4 [99]. The general trend for each YOLO version release is based on performance optimization such as speeding up non-maximum suppression (NMS) and maintaining previous structural improvements such as being an anchor-free network [102].

As we have already seen, despite the highly dynamic nature of computer vision research at the moment, there is relatively little research done into how optical characteristics, discussed in Section III, affect computer vision statistical metrics. As discussed by Brummel et al. [60], spatial metrics are a more insightful approach than statistical metrics (we shall discuss these in the next subsection). To our knowledge, to date, there hasn't been a more rigorous analysis comparing both metrics across the FOV of fisheye. This remains an open challenge for the community. However, here we will give some indication of the complexity of the problem.

Figure 12 shows the output of YOLOv7 inference run on a WoodScape sample. In the Regions of Interest (ROIs) chosen from the image both training from scratch and transfer learning on MS COCO [100] were performed for YOLOv7. Comparing both results, there are fewer false positives (FPs) in the ROIs for transfer learning such as the omission of

the bin and scooter. Higher IoUs are evident for transfer learning out of the True Positives (TPs) identified suggesting a better performance. Most IoU results average between 0.49-0.83 which is a better performance for smaller objects than for typical behavior where smaller objects tend to give lower results in images [103]. However, the current statistics of YOLO are not meaningful, especially concerning the optical artifacts discussed in this article. Going forward we need an accessible platform on which specific metrics are broken down for each optical artifact, especially in fisheye images. There are multiple instances of objects being mislabeled by YOLOv7. For example, two bicycles placed next to each other or a green bin with four wheels are labeled as 'vehicles' in the right background. A rack of bicycles is also detected as a vehicle in the left background and the scooter in the right background is incorrectly labeled as 'person'. These False Positive errors are likely (in the terminology of Hoiem [104]) due to similarity or background confusion. Woodscape does not have a 'scooter' or a 'bin' label as part of the ground truth labeling. Object detection errors of fisheye images are inherently affected by the optical artifacts where the objects in the scenes are deformed as shown in recent literature [6], [44], [53], [62]. Not only must the model deal with optical artifacts discussed previously, but fisheye also introduces rotations of the objects dependent on the scene positions. Examine the bicycles in Figure 12. All are approximately vertical in the scene (i.e., they are standing on the ground), but in the fisheye image, they appear in a range of different orientations. Any model must learn rotational invariance to perform well on fisheye surround-view cameras.

In addition to optical and geometric effects, a practical limitation of using fisheye cameras on a vehicle is that much of the image will be composed of the bodywork of the ego-vehicle itself, potentially causing image reflections to be mistaken as objects in the scene (e.g., see Figure 13). Recently, a simple U-Net [105] architecture with a Res-Net50 [74] encoder was proposed to extract areas of the vehicle ego-body in the surround-view images of Woodscape to prevent unnecessary computation and detection of false positives [106].
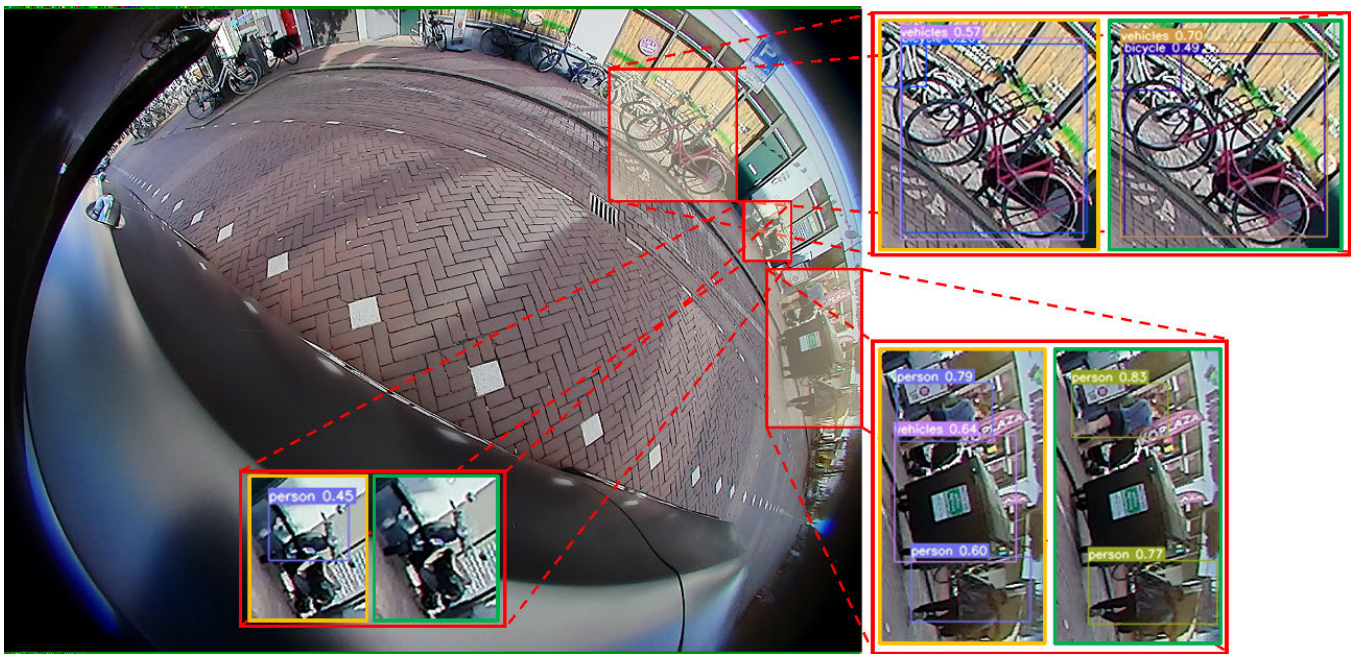
Fig. 12. **YOLOv7 inference regions of interest (ROIs) (red box) on the Woodscape left camera image was performed. Qualitative results show both training from scratch (orange box) and transfer learning (green box)**. Notice the misclassification of object types in both instances. The orientation of the object in the image due to its location can confound an object detection network. Note: the pre-trained model used in transfer learning was pre-trained on the MS COCO dataset.

Fig. 13. **False positive detected on the vehicle bodywork** due to reflections visible in the surround-view camera [106].

Braun et al. [107] discuss the difficulty in accounting for all possible scenarios with automotive cameras. Feasibility is one of the main obstacles to building robust computer vision models for the industry where data needs to be collected in real-time to account for the consistently changing environmental conditions and traffic scenarios that have not been tested during development. There are large datasets including BDD100k [54] and KITTI [46] covering a sample of potential road traffic events on roads but the data in these represent just a sample of possible events in the automotive context. There may be completely different conditions or exceptions such as the time of year when festivals disrupt the expected flow of traffic or unplanned roadworks which adds to the unpredictability of traffic. In short, it is not possible to test all road traffic scenarios using only real-world datasets that

have a limited size. For this reason, in Section V, we will discuss the current state of surround-view simulation systems. First, however, we will briefly discuss some newly developed metrics for computer vision that take into account the spatial variance of image and scene.

B. New Spatially Variant Performance Metrics

As mentioned previously, PSFs are used to model the degree of blurring in images created by an optical system. The PSFs can represent the way the optical resolution of the camera system changes from the center to the periphery of an image. However, measuring PSFs is not straightforward. As mentioned, the PSF is essentially the spatial impulse response of the camera. To measure the PSF, you need a scene with specific hardware (such as lasers or specific scene targets) [108]. A simpler and more practical measure of spatial resolution is to use the modulation transfer function (MTF) [109], which is a measure of the spatial frequency response of the camera. Indeed, there are mechanisms to extract the MTF of natural scenes without the need for any specific hardware or target [110].

The correlation between MTF and computer vision performance has been investigated in [59], where it is concluded that MTF is not a suitable predictor for the performance of computer vision. The test results show very little difference in optical performance characteristics using a Cooke Triplet despite the use of completely different sampling methods (i.e., superposition (SP) and isoplanar (ISO) with blocks of area 320 pixels^2) [59].

A potential strategy identified in [59] is the use of spatial metrics [53], [60]. Spatial metrics are a useful approach to quantifying computer vision algorithms, as they report

performance measures in different regions within the image. This is especially interesting given the spatially variant optical aberrations exhibited in fisheye cameras. The two key metrics proposed are the Spatial Recall Index (SRI) and the Spatial Precision Index (SPI) as described by (6) and (7) respectively. These are spatial variants that generalize the statistical metrics of Recall and Precision, as the names indicate. In [59], both metrics were adapted to quantify the spatial performance of a Cascade Mask RCNN model [111] for instance segmentation in automotive scenes. In each case, the statistical metrics were adapted by assigning the equivalent statistical value (i.e., both Precision and Recall) to each pixel occupied by the object in the bounding box. Using the metrics from (6) and (7) a correlation was observed between the spatially varying optical performance and the spatial performance of the Cascade Mask RCNN model. The SRI is given as:

$$SRI = \left[\sum_{n=1}^{N} \begin{cases} 1 & (x, y) \in [TP_n \cap P_n] \\ 0 & else \end{cases} \right]$$
$$\oslash \left[\sum_{k=1}^{K} \begin{cases} 1 & (x, y) \in [P_k] \\ 0 & else \end{cases} \right] \tag{6}$$

where, $(x, y) \in [TP_n \cap P_n]$ are all the pixels belonging to the n-th true positive mask and the k-th ground-truth mask denoted by P_k. The element-wise division denoted by $\oslash$ of the sum of all TPs and the sum of all ground truths results in the SRI metric.

Similarly, the SPI metric is found by element-wise division of N true positive instances and the sum of N true positive and J false positive instances.

$$SPI = \left[\sum_{n=1}^{N} \begin{cases} 1 & (x, y) \in [TP_n \cap P_n] \\ 0 & else \end{cases} \right]$$
$$\oslash \left[\sum_{n=1}^{N} \begin{cases} 1 & (x, y) \in [TP_n] \\ 0 & else \end{cases} + \sum_{j=1}^{J} \begin{cases} 1 & (x, y) \in [FP_j] \\ 0 & else \end{cases} \right] \tag{7}$$

where, $(x, y) \in [FP_j]$ denotes the j-th false positive instance and $(x, y) \in [TP_n]$ denotes the n-th true positive instance. Notice the same numerators are used in both (6) and (7). For more information please refer to the original work [53], [60].

These spatial metrics were used on degraded versions of the BDD100k datasets where a defocus of $\Delta z = \pm 1.25$ was applied. Differences in spatial performance were examined by comparing the results from the degraded datasets to the baseline BDD100k dataset. This is known as the spatial performance drop (SRI_{drop} and SPI_{drop}):

$$SRI_{drop} = SRI_{base} - SRI_{\Delta z} \tag{8}$$
$$SPI_{drop} = SPI_{base} - SPI_{\Delta z} \tag{9}$$

where SRI_{base} and SPI_{base} denote the spatial results of the baseline dataset and $SRI_{\Delta z}$ and $SPI_{\Delta z}$ denote the spatial results of the degraded datasets.

Four observations can be made from the experiments in [53] and [60]:

1) It was found that AP dropped from 87.21% (i.e. baseline) to 84.78% for a defocus of $\Delta z = +1.25$ and 82.72% for $\Delta z = -1.25$, respectively.
2) Spatial performance drop increases towards the periphery or edges of the images for a negative defocus (i.e. $\Delta z = -1.25$). Conversely, performance drop increases towards the center of the images for a defocus of $\Delta z = +1.25$ suggesting that statistical metrics vary across the FOV.
3) Ground truth instances were missed due to the worsening image quality between $\Delta z = \pm 1.25$.
4) Finally, the number of FPs was not influenced negatively by the degradations.

Both of these works demonstrate the additional insight that can be obtained from these metrics and should become part of the analysis when considering optical artifacts in fisheye images.

V. SURROUND-VIEW SIMULATIONS

Now we will turn our attention to surround-view simulation. As a way of resolving the lack of understanding of combining fisheye cameras with computer vision (especially due to the lack of public datasets available); automotive simulation can be utilized. However, as will be discussed in this section, automotive simulation lacks the optical properties of the camera and there is no standardized process of perception algorithms for automotive simulation. As we have already discussed, datasets will, no matter how large they are, suffer from a limited set of scenarios in which they are captured. Additionally, the collection of real-life data for training AI applications in the automotive industry (including autonomous driving) is a time-consuming and expensive process [1], [75]. The BDD100k dataset [54] has over 100k driving videos collected from more than 50k rides capturing New York and the San Francisco Bay area (among others) and is one of the largest automotive datasets released to date. Woodscape [39] has 10k publicly released images from three distinct locations (USA, Europe, and China) and is the first dataset of its kind to address surround-view cameras for automated driving. These datasets, while giving a reasonable amount of data for training computer vision algorithms for research, have their limitations where they only show a fraction of what can happen in the real world. Thus, it is fruitful to provide an overview (and some discussions) of current trends in automotive scene simulations, including model-driven and data-driven simulators, which are proposed to be used for addressing corner cases in autonomous driving. In the context of simulation, we also revisit the optical artifacts from Section III and discuss how they should be combined with the fisheye distortion models for photorealism.

A. The Argument for Simulations

As stated previously simulation lacks the optical properties of the camera but can potentially be considered as a direction of research for integrating the camera design with the simulation tools available. It is quite intuitive why simulations are needed for the development of automated driving solutions. However, it is still worth taking the time to make

this explicit (perhaps for the uninitiated reader). Every time a new traffic scenario is found that is not covered by the public datasets discussed (or other datasets), the new information should be added and re-split into the given ratios of training, validation, and testing. For example, if new data was added to the Woodscape dataset in the experiment from Section IV-A Figure 12, the data would need to be re-split into the (80%-10%-10%) ratio and YOLOv7 would need to be retrained. This would be very inefficient in the long term, especially if a diverse set of camera models with different lens calibrations would need to be incorporated into the datasets. Therefore, by adding simulation data to real-world data, the performance of computer vision can improve on real-world data. This can also be possible where there is a lack of ground-truth data in real-world images for training computer vision. For example, as of the writing of this survey, SynWoodscape [90] is the only publicly available synthetic fisheye dataset with pixel-wise optical flow ground truth. For the original Woodscape, it was not possible to obtain optical flow ground truth in the real scenes, and synthetic data was identified as the only feasible solution [39]. As a logical next step, Shen's experiments [79], demonstrate using the ground truth from SynWoodscape on the Recurrent All-Pairs Field Transforms (RAFT) [112] optical flow algorithm to improve performance and correct erroneous optical flow in the original Woodscape. Table I shows that synthetic data can be considered to simulate optical artifacts due to the realism of automotive simulators. However, as clearly stated in this section, there are quantitative performance differences between degraded simulation data and real-world data. Therefore, a more comprehensive analysis of computer vision algorithms is required and any simulation results shown here by no means represents a definitive analysis. Simulation must develop in the direction of combining the optical properties of the camera with simulation for there to be any comprehensive analysis.

However, the lack of real-world optical artifacts inside the simulation presents a disadvantage when it comes to optimizing algorithms for the real world. In the context of model-driven simulators, the only approach to solving this problem is by implementing the lens design inside the simulator which would lead to the inherent addition of optical artifacts and a more realistic synthetic image than before. One such strategy would be combining the well-known OpticStudio[1] with CARLA [113] or Unity where the ability to consistently alter lens models gives flexibility to the user as a way of varying optical artifacts inside the simulator.

As of this moment, data augmentation with pixel unprojection (to a ray), rotation, and re-projection (of the ray to pixel) of the fisheye data can be used as a means of creating augmented fisheye images for training computer vision algorithms [80], [114]. However, creating augmented fisheye data affects the distribution of optical artifacts in the images in a possibly unrealistic manner. Additionally, by using the original images for augmentation, no new lens configuration can be learned from the augmented data, as the augmentation just models a warped version of the lens configuration of the original system. Therefore, augmented fisheye data is not necessarily a solution to examining optical performance behavior

on computer vision. If the original data were to be synthesized (or augmented) for training, both the original and altered datasets would only ever have one specific lens configuration. This could potentially lead to overfitting computer vision on one specific lens configuration, which is not useful for scaling to different cameras.

The naive approach to solving the data problem would be to iterate over an entire set of camera lenses. This is impossible for three reasons:

1) Firstly, using only one camera is not enough to test all possible combinations. Every time a new cycle of experiments begins the camera would need to be disassembled and reassembled with new lens calibrations, which would be a time-consuming process with potentially irreproducible results.

2) Secondly, it is infeasible to test all possible lens calibration combinations for computer vision where data is collected with a group of cameras with different optical configurations. This is mainly because it is impossible to determine the number of cameras needed for this experiment without making mistakes in assumptions, for example, about the optical artifacts discussed in Section III.

3) Finally, it would likely only be possible to estimate this number via trial-and-error experiments. This could potentially lead to an unpredictable number of iterations in experiments to demonstrate a smooth transition of results iterating over the different lens designs.

A solution to this problem is to use simulation by varying optical parameters of lenses and examine the effect on computer vision performance. As of today's research, automotive simulation can be broken down into the hierarchy presented in Figure 14, where simulation can be both model-driven and data-driven.

B. Model-Driven Simulators

Model-driven simulation is where simulation is created using computer programs to design a scene and landscape using blueprint assets and mesh vectors. As described by Mutsche et al. [19], model-driven simulation is more comparable to modern video games. As shown in Figure 14, there is a wide range of model-driven simulators available. Each simulator is designed with sensors or multi-modal outputs (i.e., camera, radar, Light Detection and Ranging (LiDAR), etc.) applicable to autonomous driving. Model-driven simulation is challenging in automated driving mainly because it represents an artificial and near-perfect version of real life, which does not actually exist. There is a visible difference in texture used on assets, buildings, and the environment. For example, the asphalt on the road as depicted in Figure 15 is unnaturally brighter than the road surface in Figure 12, clearly indicating that, while objects may resemble real life, current automotive sensor simulators struggle with texture. Despite these limitations, simulation is improving rapidly in terms of visual quality. The recent release of Unreal Engine 5.2 (UE5) has much more control over the look and feel of objects in scenes, such as applying dust on the clear coating of vehicles. These
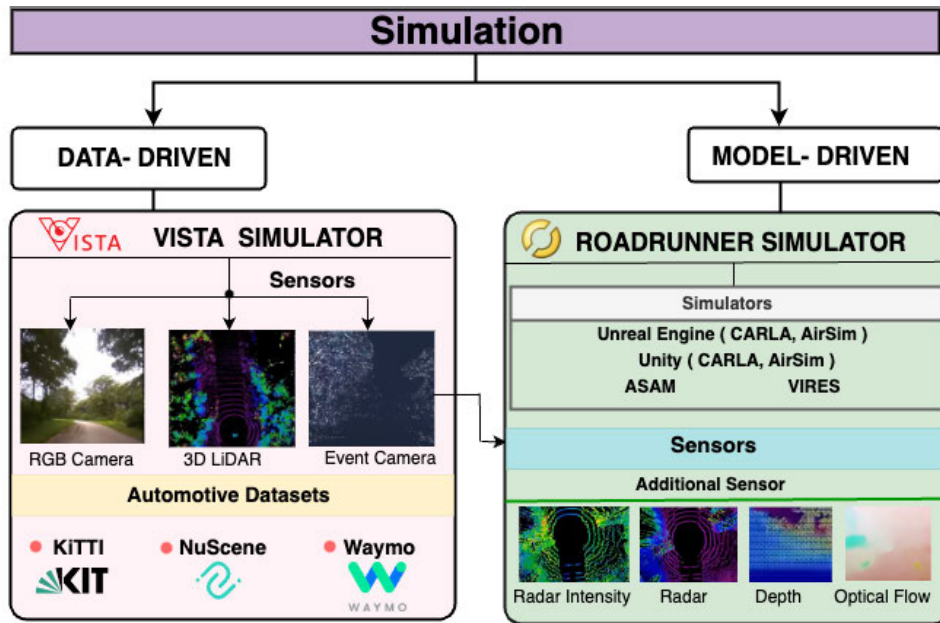
Fig. 14.　**Simulation Hierarchy**. MATLAB RoadRunner custom maps can be exported to any model-driven simulator. Unlike model-driven simulators, VISTA uses real-life datasets (e.g., KITTI, NuScenes, and Waymo) to generate simulated images or additional viewpoints of the same environment. Also, the sensors available in the VISTA simulator (i.e., RGB Camera, 3D LiDAR, and Event Camera) can also be found in the model-driven simulators.
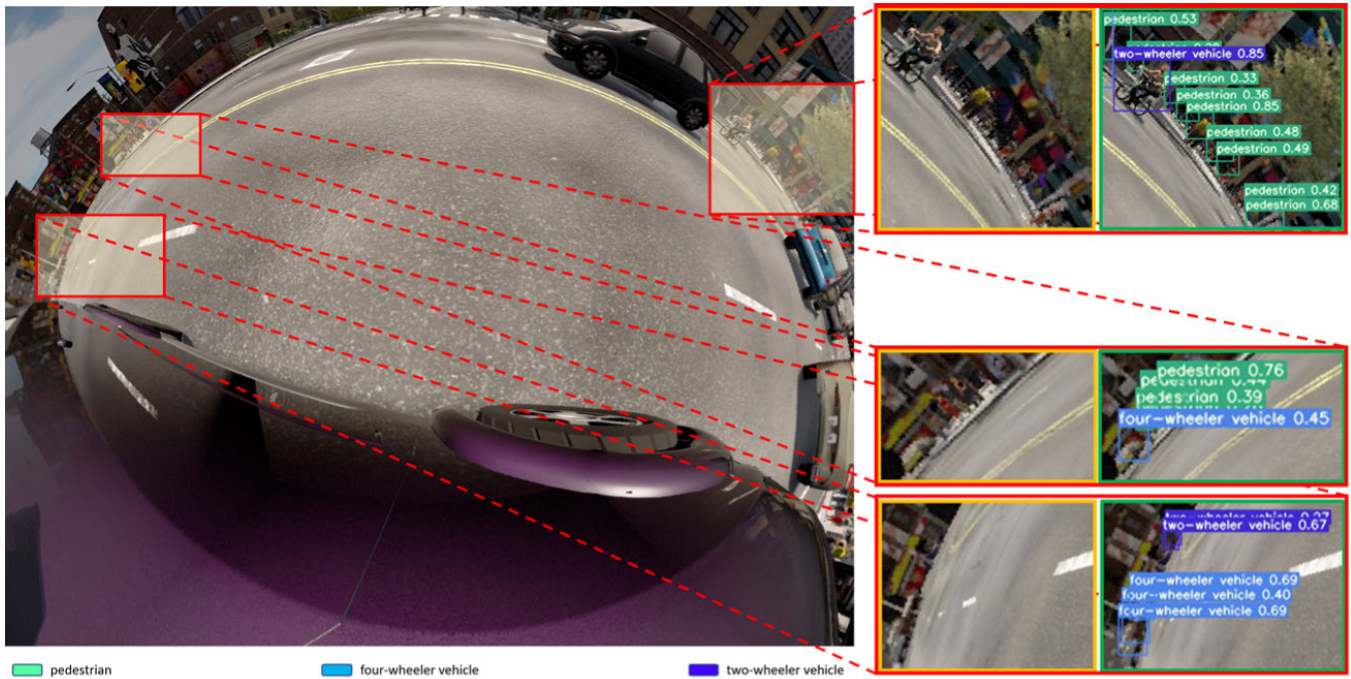

Fig. 15.　**YOLOv7 trained from scratch on SynWoodscape and inference on an MVL sample (green box)**. Zoomed areas of the image show regions of interest (ROIs) for object detection (red box). See Figure 12 for comparison. Note the absence of mechanical vignetting (dark corners), artificial lighting (uniform colors), and the presence of the outline of the cube map fisheye projection model used for distortion (a dark outline of a cube in the center of the image.

visual improvements suggest that model-driven simulation is steadily advancing towards photorealism, at least superficially. By utilizing automotive simulators, such as CARLA [113], scarcity of data issues are addressed in datasets for supervised autonomous driving. Therefore, if there were a way to combine real and synthetic data, it would be possible to measure optical performance across a range of lens calibrations to understand which lens combination would be ideal for which

computer vision model. Especially in the space of surround-view cameras, a simulator such as CARLA with Unreal Engine [113] would be useful to test out these different scenarios in combination with fisheye distortion. However, CARLA cannot natively model fisheye cameras.

To overcome this, to the best of the authors' knowledge, all fisheye models created in CARLA use an intermediate cube map for the creation of the fisheye image [90], [115],
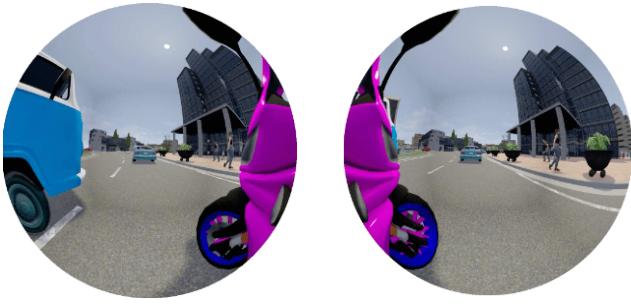
Fig. 16. **Synthetic fisheye images from the OmniScape dataset left and right views** [115]. Note the visibility of the projection cube as shadows in the image.
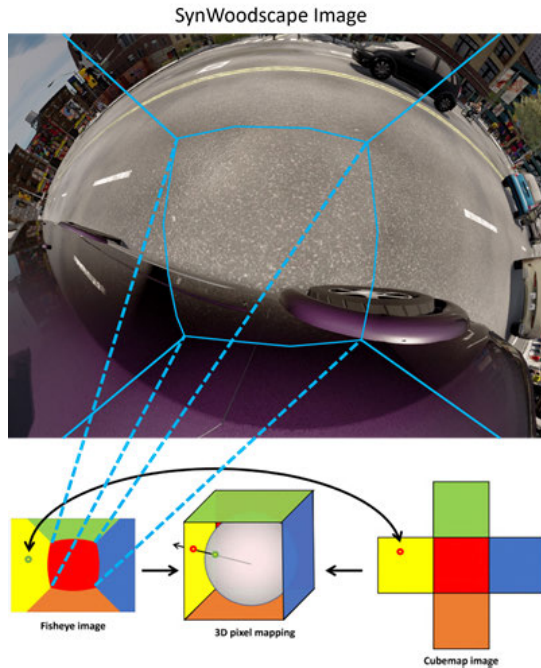

Fig. 17. **Mapping of the cube map image's pixels to the fisheye image** [90]. Five or six perspective cameras are created in the simulator. Each side of the cube map has a 90° FOV. These are then post-processed into a fisheye image.

[116], [117], shown in Figure 17. The cube map can then be mapped to a fisheye image using any fisheye model. For example, in [115], a 10,000 image dataset of fisheye from CARLA was converted using a six cubic map model, as shown in Figure 16. The cube map images were converted to fisheye using the model from [118]. The SynWoodscape dataset took a similar approach, except using the fourth-order polynomial model from [39]. However, a major drawback of this two-step fisheye image creation is that, due to the initial cubic mapping, there are faint dark outlines of the cube map projection model itself, such as can be seen in OmniScape (Figure 16, and in SynWoodscape (Figure 17).

Despite the lack of realism in synthetic data, YOLOv7 does surprisingly well at identifying objects in SynWoodscape (split into an 80%:10%:10% ratio) which has the same degree of distortion as Woodscape (see Figure 15). Most notably, even with a relatively small dataset of 2.5k images, it is capable of identifying very small pedestrians with the IoU threshold set to 0.25. A recently released synthetic automotive

surround-view dataset, called Fisheye camera images and BEV maps from Simulated Sequences of Ego car Motion (FB-SSEM) [119], was built using the Unity game engine. This dataset is interesting, as it has both Depth BEV and fisheye segmentation views, which can be used for a wide range of computer vision algorithms such as object detection, segmentation, and trajectory prediction (i.e., predicting the position of the ego-vehicle in a scene). As the Unity game engine has built-in fisheye support, it is interesting to note that we don't have the same artifacts that occur with the cube map approach in CARLA. However, this adds a restriction that, in the Unity game engine, one can only model a fisheye camera using the projection model implemented in the Unity source code. As previously discussed, there are many models of fisheye projection that one may be interested in [1].

A diverse set of maps is required for an algorithm to learn more efficiently, as information collected from only one map or scene has a bias for this specific traffic scenario. RoadRunner[2] offers a solution that is a Map Editor tool designed to create automotive maps for simulators like CARLA or Unity. With this tool, researchers can design their custom maps to suit their applications in autonomous driving. Using maps with different props, actors, and scenery can also widen the possibilities of fisheye dataset creation.

C. Data-Driven Simulators

Data-driven simulation is a relatively unexplored area in the simulation field where instead of designing computer models that imitate the realism of automotive scenarios, publicly gathered datasets (such as the KITTI [46], NuScenes [120], and Waymo [121] datasets) can be synthesized for simulation which would have a more photorealistic appearance. A publicly available interface is being constructed to interface with these three datasets.[3] As of the writing of this paper, the interface for these public datasets has not yet been released on the official website. The second version of Virtual Image Synthesis and Transformation for Autonomy (VISTA 2.0) [122] has the current capabilities to accommodate different sensor systems such as RGB, LiDAR, and Event-based cameras.

The lack of photorealism in simulators like CARLA raises the question as to whether model-driven simulation is reliable for development and testing in the long term, despite its practicality and feasibility. This is where photorealistic simulation, such as VISTA 2.0 can potentially help [122], [123]. VISTA is a data-driven simulator, which can adapt its actors and virtual environment to that seen in real-life datasets. Data-driven simulation is a relatively new area of simulation development which, as yet, is unproven in surround-view imaging for automated driving.

The adaptation of public datasets such as KITTI, will provide an advantage in creating custom synthetic datasets following these benchmark implementations. For example, creating a synthetic fisheye dataset using Woodscape with

---

VISTA would be a convenient way of generating synthetic fisheye data without learning the complexity of model-driven simulators. This would be especially useful to target scenarios not covered by the traditional simulators and improve the overall quality of testing. However, for this paper, it is important to note that, at the time of publication, there seem to be no plans to convert any surround-view dataset with VISTA.

The benefit of VISTA is that it can generate an infinite set of local viewpoints of trajectories from only a sparsely sampled set of real-world viewpoints of a scene. Each newly generated viewpoint is a different view of the same scene. In Amini's experiments [123], VISTA outperformed three other self-driving techniques with viewpoint augmentation: (1) Domain Randomization [124], (2) Sim-to-Real Domain Adaption [125], and (3) Imitation Learning [126]. Unlike the other self-driving strategies, VISTA required no lane interventions, and the near-crash recovery rate improved by 30% when compared to Imitation Learning.

The VISTA algorithm works in the following manner [122]:

1) In the ego-vehicle's state, a continuous kinematic model is updated consisting of the virtual agent's dynamics and those of the human driver.
2) The closest frame concerning the vehicle's pose is taken and a scene is recreated based on the ego-vehicle's perspective.
3) The frame is projected into 3D space to reconstruct the scene.
4) From the 3D space a coordinate transformation is performed using the virtual agent's viewpoint.
5) A new 2D simulated observation predicting the next frame of the scene is reprojected from the 3D transformation back as a new viewpoint.
6) The newly generated observation is sent back to the agent which then decides the ego-vehicle's next course of action in the scene. For example, if the road veers to the right, the driving wheel is turned to the right.

While data-driven simulation looks powerful, even if in the early stages, there are still some obvious artifacts. For example, if one carefully examines the examples provided in [122], under a modeled camera translation, the resulting images have unrealistic perspective distortions introduced.

D. Optical Artifacts in Simulation

While simulators produce data that can enhance performance for computer vision, questions remain as to how effectively lens models can translate to a simulation where there is no lens present in the virtual world. Of course, data-driven simulators pose an interesting question as to whether lens configuration is transferable from real datasets, such as KITTI. However, as clearly indicated in SynWoodscape Figure 18, there is no evidence of optical artifacts such as chromatic aberration, astigmatism (or, in fact, any significant PSF), or vignetting. These simulators are not designed to incorporate real-world optics and can often bring their own artifacts (like the visibility of the projection cube).

In Carlson et al. [75] optical artifacts are accounted for in computer vision experiments by applying them using filters

Fig. 18. **SynWoodscape left camera image showing no evidence of lateral or longitudinal chromatic aberration**. See Figure 6 for comparison. Note also that the projection cube is visible as a shadow structure in the image.

Fig. 19. **GTA dataset with optical artifacts applied** [75].

(such as chromatic aberration) to the Virtual KITTI [69] and Grand Theft Auto (GTA) [127] datasets. A sample illustration of the resulting image from GTA can be seen in Figure 19. It was shown in [75], from the results with baseline (unaugmented) and augmented (i.e., synthetic images with chromatic aberration, blur, noise, etc.), that by adding the optical artifacts uniformly across the synthetic images in training, Faster Region Convolution Neural Network (RCNN) models performed better with slightly improved mean Average Precision (mAP) compared to the unaugmented synthetic datasets (see Table I). Questions remain unanswered as to how to control the application of these optical artifacts in the spatial domain, especially for fisheye lenses where strong optical aberrations are not uniform across the image.

In summary, while graphical simulation may seem like a feasible solution to cover unforeseen scenarios on roads, it is not the overall solution as currently, simulation alone cannot completely replace the realism of actual camera footage [107].

VI. CONCLUSION

Automotive surround view cameras have become a niche research area for computer vision. This survey has identified the gaps and limitations of current strategies to train computer vision systems for wide FOV cameras, both in real life and in simulation. In particular, this survey has discussed fisheye optical systems, the optical effects that inherently affect

these systems, and the concept of adding optical effects in simulation.

Concerning computer vision performance in surround-view systems, there is little work in the space that examines the impact of optical artifacts. Some computer vision tasks, such as those that require depth extraction, appear to be improved when certain artifacts are present. One might think, therefore, that for these tasks designing a camera system with a significantly greater presence of these artifacts may be beneficial. However, this is not indicated by the results in this area, in which only residual artifacts are examined. Likely, the imperfections due to these optical artifacts would quickly damage the performance of these tasks, were they too great. Regardless, other tasks are almost certainly negatively affected by optical artifacts. What is clear is that significantly more work in this space is required for the community to understand completely the impact of optical artifacts on computer vision performance, particularly for surround-view fisheye cameras, where design challenges mean these artifacts can be significant.

From this survey, two limitations in surround-view simulators can be identified. Firstly, **popular simulators lack the simulation of lenses from the real world**. Geometric distortion models have been applied in simulation to introduce fisheye distortion. However, real fisheye images are not just affected by distortion but by a wide range of optical effects, as has been discussed. Also, it is only possible to apply optical effects to simulated images external to simulators, which is a problem as there is little to no chance of reproducibility for projects even if it can be achieved indirectly outside of simulators. Secondly, as identified in the discussions from Section V, there is a **scarcity of both real and simulated fisheye datasets for automated driving**. This hinders the development of surround-view perception systems, where current simulation data may prove unsuitable for training, partially due to the lack of optical effects.

References

[1] V. R. Kumar, C. Eising, C. Witt, and S. K. Yogamani, "Surround-view fisheye camera perception for automated driving: Overview, survey & challenges," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 4, pp. 3638–3659, Apr. 2023.

[2] J. Janai, F. Güney, A. Behl, and A. Geiger, "Computer vision for autonomous vehicles: Problems, datasets and state of the art," *Found. Trends Comput. Graph. Vis.*, vol. 12, nos. 1–3, pp. 1–308, 2020.

[3] C. Hughes, M. Glavin, E. Jones, and P. Denny, "Wide-angle camera technology for automotive applications: A review," *IET Intell. Transp. Syst.*, vol. 3, no. 1, p. 19, 2009.

[4] Y.-C. Liu, K.-Y. Lin, and Y.-S. Chen, "Bird's-eye view vision system for vehicle surrounding monitoring," in *Robot Vision*, G. Sommer and R. Klette, Eds. Berlin, Germany: Springer, 2008, pp. 207–218.

[5] M. Heimberger, J. Horgan, C. Hughes, J. McDonald, and S. Yogamani, "Computer vision in automated parking systems: Design, implementation and challenges," *Image Vis. Comput.*, vol. 68, pp. 88–101, Dec. 2017.

[6] L. Deng, M. Yang, H. Li, T. Li, B. Hu, and C. Wang, "Restricted deformable convolution-based road scene semantic segmentation using surround view cameras," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 10, pp. 4350–4362, Oct. 2020.

[7] L. Yahiaoui, J. Horgan, B. Deegan, S. Yogamani, C. Hughes, and P. Denny, "Overview and empirical analysis of ISP parameter tuning for visual perception in autonomous driving," *J. Imag.*, vol. 5, no. 10, p. 78, Sep. 2019.

[8] D. Molloy et al., "Impact of ISP tuning on object detection," *J. Imag.*, vol. 9, no. 12, p. 260, Nov. 2023.

[9] E. Hecht, *Optics*. London, U.K.: Pearson, 2017.

[10] L. Fan and L. Lu, "Calculation of the wave aberration of field curvature and color aberrations of an ultrawide-angle optical system," *Opt. Commun.*, vol. 479, Jan. 2021, Art. no. 126414.

[11] C. B. Martin, "Design issues of a hyperfield fisheye lens," *Proc. SPIE*, vol. 5524, pp. 84–92, Mar. 2004.

[12] S. Gao et al., "Compact and lightweight panoramic annular lens for computer vision tasks," *Opt. Exp.*, vol. 30, no. 17, p. 29940, 2022.

[13] Y.-H. Huang, Z.-Y. Liu, Y.-G. Fu, and H. Zhang, "Design of a compact two-channel panoramic optical system," *Opt. Exp.*, vol. 25, no. 22, p. 27691, 2017.

[14] P. Pöntinen, "Study on chromatic aberration of two fisheye lenses," in *International Society for Photogrammetry and Remote Sensing*, vol. 37. Beijing, China: ISPRS Congress, 2012, p. 27.

[15] C. Eising, J. Horgan, and S. Yogamani, "Near-field perception for low-speed vehicle automation using surround-view fisheye cameras," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 13976–13993, Sep. 2022.

[16] J. Malik et al., "The three R's of computer vision: Recognition, reconstruction and reorganization," *Pattern Recognit. Lett.*, vol. 72, pp. 4–14, Mar. 2016.

[17] Y. Qian, M. Yang, and J. M. Dolan, "Survey on fish-eye cameras and their applications in intelligent vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 12, pp. 22755–22771, Dec. 2022.

[18] S. Gao, K. Yang, H. Shi, K. Wang, and J. Bai, "Review on panoramic imaging and its applications in scene understanding," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–34, 2022.

[19] F. Mütsch, H. Gremmelmaier, N. Becker, D. Bogdoll, M. R. Zofka, and J. M. Zöllner, "From model-based to data-driven simulation: Challenges and trends in autonomous driving," 2023, *arXiv:2305.13960*.

[20] L. Fan and L. Lu, "Design of a simple fisheye lens," *Appl. Opt.*, vol. 58, no. 19, p. 5311, 2019.

[21] A. M. Samy and Z. Gao, "Simplified compact fisheye lens challenges and design," *J. Opt.*, vol. 44, no. 4, pp. 409–416, Dec. 2015.

[22] C.-L. Tien, C.-Y. Chiang, and W.-S. Sun, "Design of a miniaturized wide-angle fisheye lens based on deep learning and optimization techniques," *Micromachines*, vol. 13, no. 9, p. 1409, Aug. 2022.

[23] S.-N. Lee. (2014). *Fisheye Lens System and Photographing Apparatus*. [Online]. Available: https://patents.google.com/patent/US8638507B2/ko

[24] P. D. Burns, D. Williams, J. Griffith, H. Hall, and S. Cahall, "Application of ISO standard methods to optical design for image capture," *Electron. Imag.*, vol. 32, no. 9, pp. 240-1–240-7, Jan. 2020.

[25] C. Wittpahl, H. B. Zakour, M. Lehmann, and A. Braun, "Realistic image degradation with measured PSF," *Electron. Imag.*, vol. 30, no. 17, pp. 149-1–149-6, Jan. 2018.

[26] M. Lehmann, C. Wittpahl, H. B. Zakour, and A. Braun, "Resolution and accuracy of nonlinear regression of point spread function with artificial neural networks," *Opt. Eng.*, vol. 58, no. 4, p. 1, Apr. 2019.

[27] G. Krishnan and S. K. Nayar, "Cata-fisheye camera for panoramic imaging," in *Proc. IEEE Workshop Appl. Comput. Vis.*, Jan. 2008, pp. 1–8.

[28] L. Yang, X. Zhang, and J. Ren, "Adaptive Wiener filtering with Gaussian fitted point spread function in image restoration," in *Proc. IEEE 2nd Int. Conf. Softw. Eng. Service Sci.*, Jul. 2011, pp. 208–212.

[29] T. Eboli, J.-M. Morel, and G. Facciolo, "Fast two-step blind optical aberration correction," in *Proc. CVF Eur. Conf. Comput. Vis.*, S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, Eds., 2022, pp. 693–708.

[30] *Photography—Electronic Still Picture Imaging—Resolution and Spatial Frequency Responses*, Standard BS ISO 12233:2017, Brit. Standard Inst., 2017.

[31] *Photography—Electronic Still Picture Imaging—Resolution and Spatial Frequency Responses*, Standard BS ISO 12233:2023, Brit. Standard Inst., 2023.

[32] P. Müller, A. Braun, and M. Keuper, "Classification robustness to common optical aberrations," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2023, pp. 3632–3643.

[33] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[34] H. Lin et al., "Simulating motion blur and exposure time and evaluating its effect on image quality," *Electron. Imag.*, vol. 35, no. 16, pp. 117-1–117-5, Jan. 2023.

[35] D. Werner Wolf, M. Ulrich, and A. Braun, "Windscreen optical quality for AI algorithms: Refractive power and MTF not sufficient," 2023, *arXiv:2305.14513*.

[36] N. L. Koren, "Measuring camera Shannon information capacity with a Siemens star image," *Electron. Imag.*, vol. 32, no. 9, pp. 347-1–347-10, Jan. 2020.

[37] B. W. Keelan, "Imaging applications of noise equivalent quanta," *Electron. Imag.*, vol. 28, no. 13, pp. 1–7, Feb. 2016.

[38] T. Sun, Y. Peng, and W. Heidrich, "Revisiting cross-channel information transfer for chromatic aberration correction," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 3268–3276.

[39] S. Yogamani et al., "WoodScape: A multi-task, multi-camera fisheye dataset for autonomous driving," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9308–9318.

[40] A. Lluis-Gomez and E. A. Edirisinghe, "Chromatic aberration correction in RAW domain for image quality enhancement in image sensor processors," in *Proc. IEEE 8th Int. Conf. Intell. Comput. Commun. Process.*, Aug. 2012, pp. 241–244.

[41] S. Petersson, H. Grahn, and J. Rasmusson, "Blind correction of lateral chromatic aberration in raw Bayer data," *IEEE Access*, vol. 9, pp. 99455–99466, 2021.

[42] Sigma Corporation—Wayback Machine. (2007). *8 mm F4 Ex Dg Circular Fisheye*. Accessed: Apr. 7, 2023. [Online]. Available: https://web.archive.org/web/20070324051136/

[43] Nikon Incorporated. (2003). *Fun and Compact Fisheye Lens for Dx-Format Cameras*. Accessed: Apr. 7, 2023. [Online]. Available: https://www.nikonusa.com/en/nikon-products/product/camera-lenses/af-dx-fisheye-nikkor-10.5mm-f%252f2.8g-ed.html

[44] J. Chang and G. Wetzstein, "Deep optics for monocular depth estimation and 3D object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2019, pp. 10193–10202.

[45] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from RGBD images," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, vol. 7576, 2012, pp. 746–760.

[46] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *Int. J. Robot. Res.*, vol. 32, no. 11, pp. 1231–1237, Sep. 2013.

[47] M. Carvalho, B. Le Saux, P. Trouvé-Peloux, A. Almansa, and F. Champagnat, "Deep depth from defocus: How can defocus blur improve 3D estimation using dense neural networks?" in *Proc. Eur. Conf. Comput. Vis. Workshops*, 2018.

[48] R. Horn et al., "Investigations on the potential application of machine vision lenses for depth measurement by exploiting chromatic aberrations," *Meas., Sensors*, vol. 23, Oct. 2022, Art. no. 100410.

[49] R. Horn, M. Rosenberger, G. Notni, A. Golomoz, R. Fütterer, and P.-G. Dittrich, "Investigation on chromatic aberrations and its potential for application in depth measurement," *Meas., Sensors*, vol. 18, Dec. 2021, Art. no. 100113.

[50] Y. Bastanlar, A. Temizel, Y. Yardimci, and P. Sturm, "Multi-view structure-from-motion for hybrid camera scenarios," *Image Vis. Comput.*, vol. 30, no. 8, pp. 557–572, Aug. 2012.

[51] B. M. Deegan, P. E. Denny, V. Zlokolica, B. Dever, and L. Russell, "Addressing challenges of modulation transfer function measurement with fisheye lens cameras," *Proc. SPIE*, vol. 9403, pp. 96–105, May 2015.

[52] J. M. Geary, *Introduction to Lens Design: With Practical ZEMAX Examples*. Richmond, VA, USA: Willmann-Bell Richmond, 2002.

[53] P. Müller, M. Brummel, and A. Braun, "Spatial recall index for machine learning algorithms," in *Proc. London Imag. Meeting, Imag. Deep Learn.*, 2021, pp. 58–62.

[54] F. Yu et al., "BDD100K: A diverse driving dataset for heterogeneous multitask learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2636–2645.

[55] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.

[56] J. Sasián, "Control of linear astigmatism aberration in a perturbed axially symmetric optical system and tolerancing," *Appl. Sci.*, vol. 11, no. 9, p. 3928, Apr. 2021.

[57] M. Brummel, "Performance evaluation of state-of-the-art computer vision systems in the field of autonomous driving in dependency of optical parameters," Ph.D. dissertation, Fachbereich - Elektro- & Informationstechnik, Hochschule Düsseldorf, Düsseldorf, Münsterstraße, 2021.

[58] J. M. Sasián, "Aberrations from a prism and a grating," *Appl. Opt.*, vol. 39, no. 1, p. 34, 2000.

[59] P. Müller and A. Braun, "Simulating optical properties to access novel metrological parameter ranges and the impact of different model approximations," in *Proc. IEEE Int. Workshop Metrol. Automot. (MetroAutomotive)*, Jul. 2022, pp. 133–138.

[60] M. Brummel, P. Müller, and A. Braun, "Spatial precision and recall indices to assess the performance of instance segmentation algorithms," *Electron. Imag.*, vol. 34, no. 16, pp. 101-1–101-6, Jan. 2022.

[61] M. Subbarao and G. Surya, "Depth from defocus: A spatial domain approach," *Int. J. Comput. Vis.*, vol. 13, no. 3, pp. 271–294, Dec. 1994.

[62] B. Tian, F. Juefei-Xu, Q. Guo, X. Xie, X. Li, and Y. Liu, "AVA: Adversarial vignetting attack against visual recognition," in *Proc. 30th Int. Joint Conf. Artif. Intell.*, Aug. 2021, pp. 1046–1053.

[63] D. Potoc and D. Petrinovic, "Creating a synthetic image for evaluation of vignetting modeling and estimation," in *Proc. 45th Jubilee Int. Conv. Inf., Commun. Electron. Technol. (MIPRO)*, May 2022, pp. 918–923.

[64] D. A. Kerr. (2007). *Derivation of the Cosine Fourth Law for Falloff of Illuminance Across a Camera Image*. [Online]. Available: http://www.Cosine_Fourth_Falloff.pdf

[65] P. B. Catrysse, X. Liu, and A. E. Gamal, "QE reduction due to pixel vignetting in CMOS image sensors," *Proc. SPIE*, vol. 3965, pp. 420–430, Aug. 1117.

[66] *Photography—Electronic Still Picture Imaging—Resolution and Spatial Frequency Responses*, ISO Standard 2233:2023, 2023.

[67] P. D. Burns, K. Masaoka, K. Parulski, and D. Wueller, "Updated camera spatial frequency response for ISO 12233," *Electron. Imag.*, vol. 34, no. 9, pp. 357-1–357-6, Jan. 2022.

[68] K. Parulski, D. Wueller, P. Burns, and H. Yoshida, "Creation and evolution of ISO 12233, the international standard for measuring digital camera resolution," *Electron. Imag.*, vol. 34, no. 9, pp. 347-1–347-7, Jan. 2022.

[69] A. Gaidon, Q. Wang, Y. Cabon, and E. Vig, "Virtual worlds as proxy for multi-object tracking analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2016, pp. 4340–4349.

[70] K. Saad and S.-A. Schneider, "Camera vignetting model and its effects on deep neural networks for object detection," in *Proc. IEEE Int. Conf. Connected Vehicles Expo (ICCVE)*, Nov. 2019, pp. 1–5.

[71] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2015.

[72] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[73] D. M. W. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," 2020, *arXiv:2010.16061*.

[74] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[75] A. Carlson, K. A. Skinner, R. Vasudevan, and M. Johnson-Roberson, "Modeling camera effects to improve visual learning from synthetic data," in *Proc. Eur. Conf. Comput. Vis. Workshops*, 2018.

[76] G. S. Jung and Y. H. Won, "Compact and fast depth sensor based on a liquid lens using chromatic aberration to improve accuracy," *Opt. Exp.*, vol. 29, no. 10, p. 15786, 2021.

[77] D. Molloy et al., "Analysis of the impact of lens blur on safety-critical automotive object detection," *IEEE Access*, vol. 12, pp. 3554–3569, 2024.

[78] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[79] S. Shen, L. Kerofsky, and S. Yogamani, "Optical flow for autonomous driving: Applications, challenges and improvements," 2023, *arXiv:2301.04422*.

[80] J. Cho, J. Lee, J. Ha, P. Resende, B. Bradai, and K. Jo, "Surround-view fisheye camera viewpoint augmentation for image semantic segmentation," *IEEE Access*, vol. 11, pp. 48480–48492, 2023.

[81] Q. Zheng and C. Zheng, "NeuroLens: Data-driven camera lens simulation using neural networks," *Comput. Graph. Forum*, vol. 36, no. 8, pp. 390–401, Dec. 2017.

[82] M. Lehmann, C. Wittpahl, H. B. Zakour, and A. Braun, "Resolution and accuracy of non-linear regression of psf with artificial neural networks," *Proc. SPIE*, vol. 10695, May 2018, Art. no. 106950C.

[83] W. Xian, A. Božič, N. Snavely, and C. Lassner, "Neural lens modeling," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 8435–8445.

[84] J. Delmerico, T. Cieslewski, H. Rebecq, M. Faessler, and D. Scaramuzza, "Are we ready for autonomous drone racing? The UZH-FPV drone racing dataset," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2019, pp. 6713–6719.

[85] X. Yang, Q. Fu, and W. Heidrich, "Curriculum learning for ab initio deep learned refractive optics," 2023, *arXiv:2302.01089*.

[86] X. Yang, Q. Fu, Y. Nie, and W. Heidrich, "Image quality is not all you want: Task-driven lens design for image classification," 2023, *arXiv:2305.17185*.

[87] L. Mariotti and C. Eising, "Spherical formulation of geometric motion segmentation constraints in fisheye cameras," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 5, pp. 4201–4211, May 2022.

[88] MATLAB. (2022). *Correct Fisheye Image for Lens Distortion—MATLAB Undistortfisheyeimage—Mathworks United Kingdom*. [Online]. Available: https://uk.mathworks.com/help/vision/ref/undistortfisheyeimage.html

[89] MATLAB. (2022). *What is Camera Calibration?—MATLAB & Simulink—Mathworks United Kingdom*. [Online]. Available: https://uk.mathworks.com/help/vision/ug/camera-calibration.html

[90] A. R. Sekkat et al., "SynWoodScape: Synthetic surround-view fisheye camera dataset for autonomous driving," *IEEE Robot. Autom. Lett.*, vol. 7, no. 3, pp. 8502–8509, Jul. 2022.

[91] J. Dai et al., "Deformable convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 764–773.

[92] J. A. Jakubiec, K. Van Den Wymelenberg, M. Inanici, and A. Mahic, "Accurate measurement of daylit interior scenes using high dynamic range photography," in *Proc. CIE Lighting Quality Energy Efficiency Conf.*, 2016, pp. 42–52.

[93] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

[94] H. Bao, L. Dong, S. Piao, and F. Wei, "BEiT: BERT pre-training of image transformers," 2021, *arXiv:2106.08254*.

[95] Q. Sun, C. Wang, F. Qiang, D. Xiong, and H. Wolfgang, "End-to-end complex lens design with differentiable Ray tracing," *ACM Trans. Graph.*, vol. 40, no. 4, pp. 1–13, 2021.

[96] Z. Li, Q. Hou, Z. Wang, F. Tan, J. Liu, and W. Zhang, "End-to-end learned single lens design using fast differentiable ray tracing," *Opt. Lett.*, vol. 46, no. 21, p. 5453, 2021.

[97] E. Tseng et al., "Differentiable compound optics and processing pipeline optimization for end-to-end camera design," *ACM Trans. Graph.*, vol. 40, no. 2, pp. 1–19, 2021.

[98] G. Jocher, A. Chaurasia, and J. Qiu. (2023). *YOLO by Ultralytics*. [Online]. Available: https://github.com/ultralytics/ultralytics

[99] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 7464–7475.

[100] T. Lin et al., "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 740–755.

[101] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.

[102] J. Terven and D. Cordova-Esparza, "A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS," 2023, *arXiv:2304.00501*.

[103] Z. Zheng et al., "Enhancing geometric factors in model learning and inference for object detection and instance segmentation," *IEEE Trans. Cybern.*, vol. 52, no. 8, pp. 8574–8586, Aug. 2022.

[104] D. Hoiem, Y. Chodpathumwan, and Q. Dai, "Diagnosing error in object detectors," in *Proc. 12th Eur. Conf. Comput. Vis.*, 2012, pp. 340–353.

[105] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*. Munich, Germany: Springer, 2015, pp. 234–241.

[106] C. Hogan and G. Sistu, "Automatic vehicle ego body extraction for reducing false detections in automated driving applications," in *Proc. Irish Conf. Artif. Intell. Cogn. Sci.* Cham, Switzerland: Springer, 2022, pp. 264–275.

[107] A. Braun, "Automotive mass production of camera systems: Linking image quality to AI performance," *Technisches Messen*, vol. 90, no. 3, pp. 205–218, Mar. 2023.

[108] A. Mosleh, P. Green, E. Onzon, I. Begin, and J. M. P. Langlois, "Camera intrinsic blur kernel estimation: A reliable framework," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 4961–4968.

[109] M. R. Schroeder, "Modulation transfer functions: Definition and measurement," *Acta Acustica United Acustica*, vol. 49, no. 3, pp. 179–182, 1981.

[110] O. van Zwanenberg, S. Triantaphillidou, R. B. Jenkin, and A. Psarrou, "Estimation of ISO12233 edge spatial frequency response from natural scene derived step-edge data," *J. Imag. Sci. Technol.*, vol. 65, no. 6, pp. 060402-1–060402-16, Nov. 2021.

[111] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving into high quality object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6154–6162.

[112] Z. Teed and J. Deng, "RAFT: Recurrent all-pairs field transforms for optical flow," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, Aug. 2020, pp. 402–419.

[113] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *Proc. Conf. Robot Learn.*, 2017, pp. 1–16.

[114] A. Konrad, C. Eising, G. Sistu, J. McDonald, R. Villing, and S. Yogamani, "FisheyeSuperPoint: Keypoint detection and description network for fisheye images," in *Proc. 17th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, 2022, pp. 340–347.

[115] A. R. Sekkat, Y. Dupuis, P. Vasseur, and P. Honeine, "The OmniScape dataset," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2020, pp. 1603–1608.

[116] B. Berenguel-Baeta, J. Bermudez-Cameo, and J. J. Guerrero, "OmniSCV: An omnidirectional synthetic image generator for computer vision," *Sensors*, vol. 20, no. 7, p. 2066, Apr. 2020.

[117] J. W. Dickson Ang, C. K. Seow, K. Subramanian, and S. Pranata, "Big data scenarios simulator for deep learning algorithm evaluation for autonomous vehicle," in *Proc. IEEE Global Commun. Conf.*, Dec. 2020, pp. 1–6.

[118] D. Scaramuzza, A. Martinelli, and R. Siegwart, "A toolbox for easily calibrating omnidirectional cameras," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2006, pp. 5695–5701.

[119] E. U. Samani, F. Tao, H. R. Dasari, S. Ding, and A. G. Banerjee, "F2BEV: Bird's eye view generation from surround-view fisheye camera images for automated driving," 2023, *arXiv:2303.03651*.

[120] H. Caesar et al., "nuScenes: A multimodal dataset for autonomous driving," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11621–11631.

[121] P. Sun et al., "Scalability in perception for autonomous driving: Waymo open dataset," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2443–2451.

[122] A. Amini et al., "VISTA 2.0: An open, data-driven simulator for multimodal sensing and policy learning for autonomous vehicles," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2022, pp. 2419–2426.

[123] A. Amini et al., "Learning robust control policies for end-to-end autonomous driving from data-driven simulation," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 1143–1150, Apr. 2020.

[124] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, "Domain randomization for transferring deep neural networks from simulation to the real world," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2017, pp. 23–30.

[125] A. Bewley et al., "Learning to drive from simulation without real world labels," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2019, pp. 4818–4824.

[126] M. Bojarski et al., "End to end learning for self-driving cars," 2016, *arXiv:1604.07316*.

[127] S. R. Richter, V. Vineet, S. Roth, and V. Koltun, "Playing for data: Ground truth from computer games," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 102–118.

Daniel Jakab received the M.E. degree in electronic and computer engineering from the University of Limerick, Ireland, in 2022, where he is currently pursuing the Ph.D. degree with the Department of Electronic and Computer Engineering. From 2019 to 2020, he was an Intern with Black Box Research and Development, for nine months. His Ph.D. study is under the sponsorship of Lero, the Science Foundation Ireland-funded project, "Treating fisheye cameras natively using CNNs to improve production tolerances of cameras."

Brian Michael Deegan received the bachelor's degree in computer engineering and the M.Sc. degree in biomedical engineering from the University of Limerick, Ireland, in 2004 and 2005, respectively, and the Ph.D. degree in biomedical engineering from the University of Galway, Ireland, in 2011, with a focus on the relationship between blood pressure and cerebral blood flow in humans. From 2011 to 2022, he was with Valeo Vision Systems as a Vision Research Engineer, with a focus on image quality. In 2022, he joined the Department of Electrical and Electronic Engineering, University of Galway, as a Lecturer and a Researcher. His research interests include high dynamic range imaging, LED flicker, top view harmonization algorithms, and the relationship between image quality and machine vision.

Enda Ward received the B.E. degree in electronic engineering from the University of Galway, in 1999, and the M.Eng.Sc. degree in research in electronic engineering, with a focus on biomedical electronics, in 2002. He was a Lecturer of electronics and computing systems with Atlantic Technological University, Ireland, for a number of years. He is currently responsible for defining the camera product roadmap for surround and automated driving applications with Valeo Vision Systems. He was with key technology experts across the supply chain and within OEMs to define optimal system architectures. Later, he moved to industry, working in the biomedical space, and has spent the last 16 years in automotive ADAS design. He holds several patents in the area of automotive vision.

Sushil Sharma received the B.Tech. degree in electronics and instrumentation from ITM University, India, in 2014, and the M.Sc. degree in automatic control and robotics, specializing in smart aerospace and autonomous systems, from Université Paris-Saclay, France, and Poznań University of Technology, Poland, in 2017. He is currently pursuing the Ph.D. degree with the Department of Electronic and Computer Engineering, University of Limerick, Ireland. His research interests include "Automotive Parking Systems" involve a broad range of areas, such as multicamera perception, trajectory prediction, and machine-based algorithms.

Pepijn van de Ven received the M.Sc. degree in electronic engineering from Eindhoven University of Technology, The Netherlands, in 2000, and the Ph.D. degree in artificial intelligence for autonomous underwater vehicles from the University of Limerick, Ireland, in 2005. Currently, he is a Professor of Artificial Intelligence with the Department of Electronic and Computer Engineering, University of Limerick.

Eoin Martino Grua was born in Cork, Ireland, in 1993. He received the B.S. degree in liberal arts and sciences from Amsterdam University College, Amsterdam, The Netherlands, in 2015, the M.S. degree in computer science from Swansea University, Swansea, U.K., in 2016, and the Ph.D. degree in computer science from Vrije Universiteit Amsterdam, Amsterdam, in 2021. In 2021, he was a Research Assistant with the University of Limerick, Limerick, Ireland, where he is currently a Post-Doctoral Researcher with the Department of Electronic and Computer Engineering. His research interests include artificial intelligence, software engineering and architecture, and sustainability.

Anthony Scanlan received the B.Sc. degree in experimental physics from the National University of Ireland, Galway, Galway, Ireland, in 1998, and the M.Eng. and Ph.D. degrees in electronic engineering from the University of Limerick, Limerick, Ireland, in 2001 and 2005, respectively. He is currently a Senior Research Fellow with the Department of Electronic and Computer Engineering, University of Limerick. He has been a principal investigator for several research projects in the areas of signal processing and data converter design. His current research interests include artificial intelligence, computer vision, and industrial and environmental applications.

Jonathan Horgan is a Manager of computer vision and deep learning architecture and a Senior Expert with Valeo Vision Systems. He has worked in the field of computer vision for over 16 years with a focus over the last ten years on automotive computer vision for advanced driver assistance systems (ADAS), automated parking, and automated driving. He is currently working on next-generation advanced computer vision and deep learning with the ultimate goal of achieving fully autonomous driving and parking. He has 25 publications in peer-reviewed journals and conferences and over 100 patents published in the field of automotive computer vision.
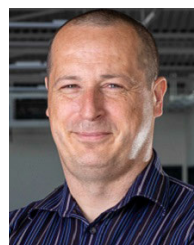
Ciarán Eising (Senior Member, IEEE) received the B.E. degree in electronic and computer engineering and the Ph.D. degree from the National University of Ireland, Galway, in 2003 and 2010, respectively. From 2009 to 2020, he was a Computer Vision Architect Senior Expert with Valeo Vision Systems. In 2020, he joined the University of Limerick as an Associate Professor of AI.