

Sensitivity analysis of AI-based algorithms for autonomous driving on optical wavefront aberrations induced by the windshield

Dominik Werner Wolf

Volkswagen Group

dominik.werner.wolf
@volkswagen.de

Markus Ulrich

Karlsruhe Institute of Technology

markus.ulrich@kit.de

Nikhil Kapoor

CARIAD SE

nikhil.kapoor
@cariad.technology

Abstract

Autonomous driving perception techniques are typically based on supervised machine learning models that are trained on real-world street data. A typical training process involves capturing images with a single car model and windshield configuration. However, deploying these trained models on different car types can lead to a domain shift, which can potentially hurt the neural networks performance and violate working ADAS requirements. To address this issue, this paper investigates the domain shift problem further by evaluating the sensitivity of two perception models to different windshield configurations. This is done by evaluating the dependencies between neural network benchmark metrics and optical merit functions by applying a Fourier optics based threat model. Our results show that there is a performance gap introduced by windshields and existing optical metrics used for posing requirements might not be sufficient.

1. Introduction

The aspiration to launch level-4-ready autonomous vehicles within this decade drives new challenges in the automotive world. In order to increase the perception performance w.r.t. the frontal far field, the car will be equipped with high spatial resolution cameras. Advanced Driver Assistance Systems (ADAS) cameras with telephoto lenses and high-resolution sensors provide a high pixel resolution per field angle, wherefore they are more sensitive to optical aberrations within the optical path. Since car windshields are typically curved they will act as an additional lens in the optical path. Unfortunately, the curvature and thickness characteristics of the windshield are not sufficiently controllable on small domains [15]. This indicates inherent optical aberrations in the optical path of ADAS cameras and impacts the recoverable information content of camera-based ADAS systems.

From the physics point of view, the imaging process of an optical system is entirely determined by the convolution

of the raw image with the Point Spread Function (PSF) because of the superposition principle, which arises from the linear nature of the Helmholtz equation (1). The PSF of an aberrated optical system can be parameterized by the wavefront error in terms of Zernike coefficients [12]. This paper presents a methodology to define an optical threat model based on Fourier optics to reflect the perturbations induced by windshields. This difficulty becomes important if the training dataset is taken by a camera mounted on the vehicle roof but the network inference is performed on images from a camera behind the windshield. Hence, the optical aberrations can induce a significant dataset domain shift and might affect the model performance. This paper focuses on two primary research questions. First of all, how sensitive is a neural network to optical perturbations and are those sensitivities reflected sufficiently by optical merit functions, like the refractive power of the windshield for example. In order to tackle this question, we utilize a common metric in explainable AI, namely the Shapley values [29], which quantify the contribution or impact of a particular feature on the merit function of interest. The analysis of different windshield configurations will lead to a Shapley distribution for every merit function and Zernike coefficient. In order to synchronize the development efforts regarding the optical quality of windshields in the light of neural network performance, we are aiming for an optical merit function that reflects a congruent Shapley distribution as the distribution imposed by the AI benchmark metric. Secondly, we are investigating the correlation between neural network and optical system benchmark metrics. From a quality assurance perspective, we would like to determine a bijective function between neural network and optical Key Performance Indicators (KPIs). This would allow us to derive optical system requirements for the level-4 functionality. We are addressing this issue by generating different threat model attacks on the neural network architecture by Monte-Carlo sampling from uniformly distributed Zernike coefficients of second order.

The intertwining between optical characteristics and the

neural network predictive power has manifested itself as a new scientific branch denoted as deep optics [5]. The essential idea is to trim the PSF during training by minimizing the loss function. This results in the most optically informative PSF [31], which might differ from the PSF with the highest imaging fidelity. For example, if the task consists of performing a depth estimation of objects by a single 2D image, it might be beneficial to code the PSF with an artificial defocusing blur [5]. The blurring will then affect objects differently depending on their depth position. Hence, optical aberrations can be utilized as a feature for improving the information decoding capabilities of neural networks. As a downside, this methodology requires task-specific end-to-end optimization, which is not compatible with multi-task architectures typically used in the autonomous driving industry [1, 19, 28]. Commonly used multi-task architectures consist of a pre-trained backbone model, which is trained on a joint dataset and is based on unified learning across multiple tasks in the encoder step. This is sequentially followed by different adaption models or simply heads that are trained on downstream, task-specific datasets, e.g., classification, segmentation and detection [6, 34]. This hybrid architecture increases the run-time performance by the joint encoder utilization and enhances the generalization capability by incorporating data heterogeneity, which ultimately strengthens the model's robustness in inference [25]. As a downside, the jointly used backbone model might induce a lack of information capacity, which would result in lower task-specific KPIs [28]. If we would like to make use of the deep optics approach for multi-task networks we would need to train the heads individually. As a result, the most optically informative PSF would be task-dependent, which can not be satisfied by a single optical element. Even in the case of a single-task network, the deep optics approach is economically unfeasible in the context of car windshields because of the manufacturing process, which focuses on industrial macro parameters instead of physical micro parameters that drive optical aberrations.

The results of this paper indicate that optical aberrations of the windshield can significantly deteriorate the model's performance by a domain shift and evidence on the insufficiency of existing optical working requirements for ADAS systems was found.

2. Scope and research motivation

For the homologation of autonomous driving vehicles, it will be necessary to perform a holistic analysis of the entire functional chain. The sensitivity analysis of image-based deep neural networks on optical aberrations induced by the windshield is only one aspect of the entire challenge. Other impact factors might also be critical, like weather conditions, out-of distribution events or lighting conditions. Those effects are not discussed in detail in this paper, which

does not imply a judgment on the relative severity. The main motivation of focusing on wavefront aberrations of the windshield in this paper is based on the question: "What makes a windshield smart and level-4 ready?"

This question can only be answered if a most informative optical metric is found, which allows for deriving component requirements for safeguarding level 4 functionalities.

3. Optical merit functions

The foundation of Fourier optics is based on the Helmholtz equation [12]. An electromagnetic field wave $\rho(\vec{x})$ has to satisfy the wave equation, which results in the time-independent Helmholtz equation:

$$(\Delta + k^2)\rho(\vec{x}) = 0 \quad , \quad \text{with: } k := \frac{2\pi f}{c} = \frac{2\pi}{\lambda} . \quad (1)$$

A unit amplitude spherical wave satisfies the Helmholtz equation (1) and is commonly known as the free-space Green's function [12]. Generally, Green's functions are the physical version of the impulse response function in control systems engineering and are applicable to linear differential operators. If the system can be characterized by a Green's function then the system output is given by the convolution of the driving term or input signal with the Green's function. This theoretical mechanism is the causal reason for the validity of the superposition principle in optics. Therefore, the imaging process of an optical system is determined by:

$$\begin{aligned} \rho(\vec{x}_o) &= \iint_{\mathbb{R}^2} |h(\vec{x}_o - \vec{x}_s)|^2 \cdot \rho(\vec{x}_s) \, dx_s^2 \\ \Leftrightarrow \mathcal{F}[\rho](\vec{k}) &= \mathcal{F}[|h|^2](\vec{k}) \cdot \mathcal{F}[\rho](\vec{k}) . \end{aligned} \quad (2)$$

The Green's function of an optical system $|h(\vec{x}_o)|^2$ is commonly denoted as the PSF. It describes the image of an infinitesimal light pulse given by a Dirac delta distribution. Since we are considering an imaging system under incoherent light incidence, only the squared magnitude of the electrical field or the intensity of the light pulse matters. Essentially, with incoherent light there are no interference effects. If the Fresnel approximation is valid [12], then the PSF $|h(\vec{x}_o)|^2$ of an optical system is determined by:

$$\begin{aligned} h(\vec{x}_o) &\approx \iint_{\mathbb{R}^2} P(\lambda d_z \vec{k}_a) \cdot e^{-2\pi i \vec{x}_o \cdot \vec{k}_a} \, dk_a^2 \\ \Leftrightarrow h(\vec{x}_o) &= \mathcal{F}\left[P(\lambda d_z \vec{k}_a)\right] \quad , \quad \text{with: } \vec{k}_a := \frac{\vec{x}_a}{\lambda \cdot d_z} . \end{aligned} \quad (3)$$

Here, $P(\vec{x}_a)$ denotes the aperture function of the optical system. In the case of an ADAS camera, the aperture stop of the objective lens is considered. Furthermore, d_z quantifies the distance between the observation plane at z_o and the position of the aperture stop at z_a . If there are no inherent optical aberrations, then the system is diffraction limited and the incoherent impulse response function (PSF)

of a one-dimensional rectangular aperture is given by the squared sinc function [12]. Unfortunately, optical systems in the automotive industry are not diffraction limited especially if the windshield is included. Therefore, the concept of the aperture function $P(\vec{x}_a)$ has to be extended to the generalized aperture function¹ $\mathcal{P}(\vec{x}_a)$, given by:

$$\mathcal{P}(\vec{x}_a) := P(\vec{x}_a) \cdot \exp \left[\frac{2\pi i}{\lambda} \cdot W(\vec{x}_a) \right]. \quad (4)$$

Here, $W(\vec{x}_a)$ denotes the wavefront aberration map on the aperture surface. Physically, the wavefront aberration map is given by the optical path difference between the expected wavefront and the observed wavefront. In the case of a windshield, the expected wavefront is given by a plane wave. Even in the case of non-diffraction limited systems, the superposition principle is applicable since the differential operator remains linear, wherefore aberrated optical systems are characterized by:

$$\begin{aligned} h(\vec{x}_o) &\stackrel{(3)}{=} \mathcal{F} \left[\mathcal{P} \left(\lambda d_z \vec{k}_a \right) \right] \\ \Rightarrow \mathcal{F} [\rho] (\vec{k}) &\stackrel{(2)}{=} \mathcal{F} [|h|^2] (\vec{k}) \cdot \mathcal{F} [\rho] (\vec{k}). \end{aligned} \quad (5)$$

So far, the influence of optical aberrations on the imaging process has been discussed in detail and the governing physical equations were presented. For quality assurance purposes in the light of reliable autonomous driving, this physical process has to be mapped to measurable physical quantities on which quality requirements can be imposed. The following subsections elaborate on different optical metrics, which serve this objective.

3.1. Refractive power

Historically, refractive power measurements have been utilized as the primary quality criterion for windshields [4]. The refractive power D_{x_i} quantifies the curvature of the wavefront aberration map $W(\vec{x}_a)$ along the axis of interest x_i if a plane wave is expected, as it is the case for windshields [33]. Consequently, D_{x_i} is given by:

$$D_{x_i}(\vec{x}_a) = \frac{\partial^2}{\partial x_i^2} W(\vec{x}_a). \quad (6)$$

Current optical requirements in terms of the refractive power are typically expressed as the maximum absolute value over both transversal axes.

3.2. Modulation Transfer Function (MTF)

Generally, the Green's function entirely determines the behaviour of a Linear and Time-Invariant (LTI) system [18]. As a consequence, it is insightful to further analyse the PSF.

¹Generally, we adopt the Feynman slash notation if optical field quantities are assumed to be non-diffraction limited.

The MTF is defined as the real part of the Fourier transform of the PSF normalized to one at $\vec{k} = \vec{0}$, as stated by:

$$\text{MPFF}(\vec{k}) = \frac{\left| \mathcal{F} [|h|^2] (\vec{k}) \right|}{\left| \mathcal{F} [|h|^2] (\vec{k} = 0) \right|}. \quad (7)$$

Hence, if the real-valued intensity PSF is considered as a light distribution in the observer plane then the MTF corresponds to the characteristic function in statistics [3], which determines the moments of the light distribution, e.g., gray values centroid, intensity variance etc. Tier-1 ADAS suppliers are recently defining functional requirements in terms of the MTF at a spatial frequency of half-Nyquist.

3.3. Strehl Ratio (SR)

Instead of specifying only a single spatial frequency requirement for the MTF it might be advisable to consider the entire spectrum. In order to do so, the area under the MTF curve can be evaluated. The spectral integral of the MTF in relation to the diffraction limited MTF area is defined as the Strehl ratio [12] and is given by:

$$\text{SR}_{x_i} := \frac{\int_{\mathbb{R}} \text{MPFF}(k_{x_i}) dk_{x_i}}{\int_{\mathbb{R}} \text{MTF}(k_{x_i}) dk_{x_i}} \bigg|_{\vec{x}_o = \vec{0}} \frac{|h|^2}{|h|^2}. \quad (8)$$

An equivalent definition of the Strehl ratio is given by the quotient of the aberrated PSF over the diffraction-limited PSF, evaluated at the optical axis ($\vec{x}_o = \vec{0}$).

3.4. Optical Informative Gain (OIG)

Unfortunately, there is still a drawback in the definition of the Strehl ratio because it does not incorporate the knowledge about the shape of the PSF, which entirely characterizes the optical system. Therefore, an optical merit function would be desirable that shows a dependency on higher-order moments of the PSF as well. One possible metric that considers this constraint is introduced in this paper as the Optical Informative Gain (OIG):

$$\text{OIG}_{x_i} := \frac{\int_{\mathbb{R}} |\text{MPFF}|^2 dk_{x_i}}{\int_{\mathbb{R}} |\text{MTF}|^2 dk_{x_i}} \frac{\int_{\mathbb{R}} |h|^4 dx_{o_i}}{\int_{\mathbb{R}} |h|^4 dx_{o_i}}. \quad (9)$$

Equation (9) takes advantage of the Plancherel theorem [8]. If the OIG is evaluated by measurement data then the domain of the MTF is restricted by the Nyquist frequency. Hence, the OIG incorporates the resolution limitation given by the image sensor and relates to the amount of photonic energy, which can be spatially discriminated in relation to the diffraction-limited case.

4. Neural network merit functions

Previous studies on the effect of dataset shifts [26] and noise corruptions [14] on image classification underpin the

importance of optical robustness analyses for autonomous driving algorithms. The impact of dataset shifts induced by optical aberrations of the windshield on traffic sign classification has already been quantified as an accuracy drop of up to ten percent [7]. In contrast, our paper focuses on the performance and the network calibration reliability for semantic segmentation. Due to the pixel-wise class prediction, it can be hypothesized that the sensitivities for optical aberrations are amplified in relation to macro-level predictions in image classification.

4.1. Intersection over Union (IoU)

The governing benchmark metric for semantic segmentation is given by the Jaccard similarity coefficient or also commonly known as the Intersection over Union (IoU) [22]. In this paper we will make use of multi-class segmentation datasets, wherefore the mean of the IoU is computed over all classes (N_c), denoted as mIoU.

4.2. Expected calibration Error (ECE)

Standard neural networks typically yield non-calibrated predictions, which can be transformed into calibrated confidence scores using post-hoc calibration methods [30]. Nevertheless, modern neural networks tend to yield systematically overconfident predictions [13]. A metric that assesses the calibration quality of neural networks is given by the Expected calibration Error (ECE) [24]. For non-binary datasets, the metric is generalized as the mean over all classes (mECE).

4.3. Shapley values

Deep convolutional neural networks are inherently highly non-linear, wherefore it is generally difficult to assess the global sensitivity of the model predictions on single input features. One way to tackle this problem is by considering the outcome of the model with and without a particular feature. If all input feature subsets \mathcal{S} are considered regarding the marginal contribution of feature i to the sub-coalition performance then the correlations between different features are inherently incorporated. Averaging the weighted marginal contribution of feature i over all possible input feature coalitions of different cardinality results in a sensitivity metric, which fulfills all fairness properties in game theory namely the efficiency-, symmetry-, linearity- and the null player condition [27]. This sensitivity metric was initially introduced by Shapley [29] in the field of economics and has been widely adopted in the explainable Artificial Intelligence (AI) world since an approximative evaluation method was found by Lundberg & Lee [20]. In general, the Shapley value φ for feature i and objective function $\hat{\Xi}$ is determined under a particular feature set \mathcal{M}_f . Hence, the Shapley value is a local explanation method [23], which describes the feature effect by quanti-

fying the direction and magnitude of the local gradient in the feature space. As a consequence, if the entire feature space is sampled equidistantly the Shapley values will generate a distribution. The shape of the Shapley distribution for feature i in contrast to feature j might indicate differences in the feature importance for the neural network inference. The Shapley values:

$$\varphi_i(\hat{\Xi}) := \sum_{\mathcal{S} \subseteq \mathcal{M}_f \setminus \{i\}} \binom{|\mathcal{M}_f| - 1}{|\mathcal{S}|}^{-1} \frac{[\hat{\Xi}(\mathcal{S} \cup \{i\}) - \hat{\Xi}(\mathcal{S})]}{|\mathcal{M}_f|}, \quad (10)$$

are determined by weighting the individual coalition merit with the inverse of the binomial coefficient, which quantifies the number of sub-coalitions with cardinality $|\mathcal{S}|$.

5. Experimental setup

In order to examine the sensitivities and dependencies of semantic segmentation predictions on optical wavefront aberrations induced by the windshield, a proper testing environment has to be established. First of all, we will elaborate on the physical imaging model used in this paper, which utilizes Fourier optical principles to translate the wavefront aberrations induced by the windshield into image degradations. Secondly, the network architectures are introduced for conducting the evaluation experiments.

5.1. Optical threat model

The optical threat model, which simulates the optical aberrations of the windshield, is based on Fourier optics [12]. Inspired by the work of Chang et al. [5], we extend the proposed optical threat model to 4k ADAS cameras with telephoto objective lenses. Generally, the optical threat model assumes that the wavefront aberration map $W(\vec{x}_a)$ is known in advance either by measurement or optical simulation. The wavefront aberrations are parameterized by a set of Zernike coefficients $\{\omega_n\}$, which decompose the wavefront aberration map W on the unit circle:

$$W(\rho, \phi) = \sum_{n=0}^{\infty} \omega_n Z_n(\rho, \phi) \quad , \quad \omega_n := \langle W, Z_n \rangle \quad . \quad (11)$$

In this paper, the aperture stop of the objective lens is circular, wherefore the orthonormal² Zernike polynomials Z_n are selected as a basis³ obeying the orthogonality relation $\langle Z_n, Z_m \rangle = \delta_{nm}$. Eq. (11) is parameterized by the normalized radius ρ and the polar angle ϕ of the circular aperture.

In general, Zernike polynomials of zeroth- and first-order only induce a phase modulation, which does not impact the measured intensity distribution on the image sensor [33]. As a result, the incoherent MTF is not affected by

²The measured Zernike coefficients might need to be renormalized according to ISO24157 [9].

³If a square aperture stop is considered, then the Legendre polynomials would be utilized as a basis for the decomposition of W .

the Zernike coefficients ω_0 to ω_2 . This is physically sound because the zeroth order term describes a longitudinal offset of the wavefront, which does not influence the image. Secondly, the first-order terms physically describe a deflection of the light beam, wherefore the image is displaced but not structurally perturbed since the wavefront curvature is not affected. As a consequence, the studies of this paper are restricted to second-order Zernike coefficients. Higher-order terms are neglected because the magnitude of the coefficients decays with increasing order, which reflects the convergence of the series expansion in Equation (11). Future studies might also investigate terms of the truncation order, e.g., coma and trefoil.

With the knowledge of the wavefront aberration map of the windshield and the aperture stop of the camera under consideration, the generalized aperture function \mathcal{P} can be constructed by applying Equation (4). Based on \mathcal{P} the incoherent, non-diffraction limited PSF $|\mathcal{H}|^2$ is computed based on Equation (5), which entirely characterizes the optical system. The perturbed image ρ is finally given by convolving the clean image ρ with the perturbed PSF $|\mathcal{H}|^2$. From the measured wavefront aberration map and the deduced PSF, the entire ensemble of optical merit functions introduced in Section 3 can be derived.

Figure 1 demonstrates the effect of the optical threat model. The Zernike coefficients for the parameterization of the wavefront aberration map were determined by a Shack-Hartmann wavefront measurement of a test sample windshield. The black square target within the image has been utilized for a slanted edge analysis according to ISO12233 [10]. The MTF of the perturbed image is normalized by the MTF of the undistorted image to retrieve the net effect of the induced optical aberrations. The resulting MTF curves for the horizontal and vertical direction are compared to the MTF parameterized by the optical threat model in Figure 2. In addition, the refractive power triggered by the curvature modulation of the wavefront can be evaluated by Equation (6), which has been benchmarked by a reference refractive power measurement using the Moiré pattern technique [32]. In conclusion, the measurement results for the physical test sample are sufficiently reflected by the optical threat model, which underpins the validity of the implemented Fourier optics approach.

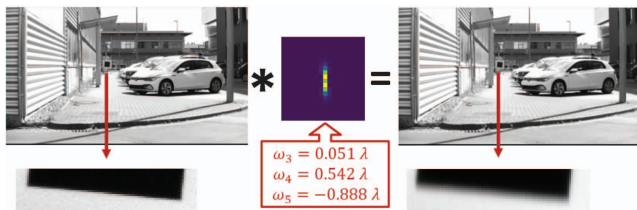


Figure 1. Toy example demonstrating the effect of the optical threat model applied to a real-world scene. The slanted edge targets are shown enlarged on the bottom.

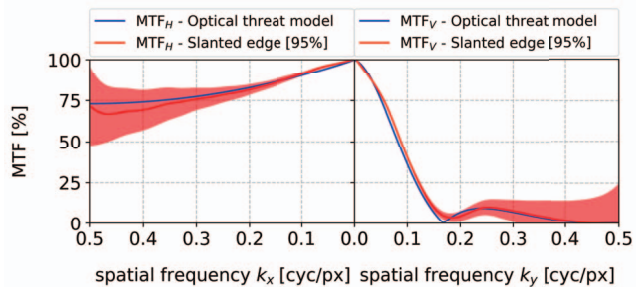


Figure 2. Validation of the optical threat model based on MTF measurements with the slanted edge method. The confidence bands are reflecting the Poisson noise of the image sensor.

5.2. Architecture of the evaluation networks

In this paper, we make use of a publically available deep convolutional neural network trained on the KITTI dataset [11] from TensorFlow Hub and we study a state-of-the-art multi-task network from CARIAD.

5.2.1 High-Resolution Network (HRNet)

The High-Resolution Network (HRNet) architecture has been invented by Microsoft [16] and the TensorFlow Hub model was adapted and trained by Google [2, 21]. The selection of the HRNet architecture as an evaluation model is based on the fact that future ADAS functionalities will most likely rely on 4k high-resolution cameras. In general, a model architecture can be tuned in three dimensions: depth (e.g., more layers), wideness (e.g., more channels) or fineness (e.g., higher resolution images). The standard sequential encoder-decoder architecture in deep convolutional neural networks lacks on information capacity in the condensed low-resolution feature map. Hence, the standard encoder-decoder architecture is typically extended for highly spatially sensitive applications like autonomous driving. The HRNet tackles this challenge by switching the information propagation from serial to parallel [16]. In detail, convolutions are performed in parallel on multiple resolutions to improve the information capacity of the model architecture. Therefore, the high-resolution representation of the input information is maintained throughout the whole process. Repeated fusion steps between parallel streams of different resolutions ensure an information flow across the levels.

5.2.2 Multi-Task Learning (MTL) model

The in-house developed Multi-Task Learning (MTL) model consists of a large shared encoder with several feature extraction layers followed by five decoder heads, each for a specific task, referred to as task head. These task heads are mainly of two types: segmentation heads and object detection heads. In detail, the parallelized decoders address the following tasks:

- **Semantic segmentation head:** Provides a pixel-wise classification across the image for several classes. The head’s performance is quantified by the mIoU.
- **Blockage detection head:** Provides a binary segmentation mask that detects if a certain region of the image is blocked or not. The evaluation metric is given by the IoU.
- **Traffic Light Detection and Classification:** At first, 2D bounding boxes for traffic lights in the image are predicted. Subsequently, the pixels within a single 2D bounding box are segmented to either belong to the class ”traffic light bulb” or ”housing”. Finally, the pixels belonging to the class ”traffic light bulb” are used to classify the signal color of the corresponding traffic light. For quantifying the performance of this multi-step classification task, a head-specific combined metric is evaluated, which relies, i.a., on the average accuracy and the area under the precision-recall curve.
- **Traffic Sign Detection and Classification:** Predicts 2D bounding boxes for traffic signs within the image. Afterwards, the subimages are used to classify the corresponding traffic sign type. Similar to the traffic light classification head a combined metric is assessed for the performance of the multi-step traffic sign classification task.
- **Vehicle Detection:** Provides a categorized 3D bounding box across two types of vehicles: large vehicles (e.g., trucks, buses, etc.) and passenger cars. The head’s performance is evaluated by the average precision metric.

For a consolidated evaluation, we first determine the task-specific metrics (i.e., average precision for object detection and mIoU for semantic segmentation). However, to convey a holistic model performance, the head-specific loss functions are integrated using weighted averaging after normalization culminating into an overall combined multi-task loss, ranging from 0 (worst performance) to 1 (perfect performance). This aggregated score reflects the model’s collective efficiency across all tasks. In order to prevent a single task from being dominant in the learning process, the individual, task-specific head losses can be integrated by an uncertainty-based weighting scheme to obtain a more robust combined metric [17].

5.3. Evaluation datasets

Typically, datasets for autonomous driving are taken with cameras mounted behind the windshield. As a consequence, the images are inherently perturbed by optical aberrations, which leads to an unknown dataset domain shift that makes it impossible to quantitatively assess the impact of different windshield configurations without prior knowledge about the inherent aberrations. Hence, it is eminently beneficial that the HRNet was trained on the KITTI dataset,

where the camera had been mounted on the car roof [11]. The evaluation images from the KITTI dataset are characterized by a resolution of 370×1224 px.

The MTL model is trained on a joint dataset, i.e., each head is trained on a task-specific dataset with corresponding labels. The multi-task dataset from CARIAD features images of the dimensions 1024×2048 px.

6. Evaluation results

The results obtained from employing the optical threat model on two distinct neural network architectures are summarized in this section. In general, the results for the HRNet and the MTL model are primarily coherent, e.g., the dependency of the model performance on the optical merit functions introduced in Section 3 and the network performance sensitivity on different Zernike coefficients.

6.1. Sensitivity analysis

The Shapley studies envisioned by Figure 6 on different optical merit functions and neural network benchmark measures indicate a non-linear mapping of the sensitivities between the AI world and the optical world. For comparability reasons, the Shapley values have been normalized to the effect of ω_4 , which physically represents defocus. The behavior of the mIoU and the mECE with increasing perturbation magnitude is physically sound and predicted but the symmetry is remarkable. The mirror symmetry w.r.t. the abscissa originates from the observation that the mECE is dominated by the accuracy degradation, as illustrated by Figure 3. In addition, the refractive power shows no sensitivity regarding ω_3 as expected, which reflects the fact that the merit function has been explicitly restricted to the x - and y -axis. In general, it can be concluded that ω_4 aberrations have the biggest impact on the performance of the studied merit functions. Furthermore, the Shapley distributions regarding the MTF, the Strehl ratio and the OIG are very similar in terms of their codomains but they reveal slightly different probability allocations, which indicates different statistical moments.

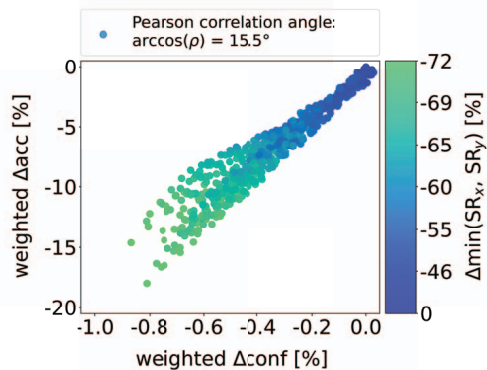


Figure 3. Pearson correlation between the weighted confidences and the weighted accuracies (bin cardinality weighting scheme).

6.2. Correlation analysis

The dependencies between optical merit functions and neural network benchmark measures are directly contrasted in Figure 7 for the HRNet. It is clearly noticeable that from the refractive power and the MTF it is not possible to infer the performance of the neural network unambiguously. On the contrary, the Strehl ratio and the OIG indicate a functional relationship to the mIoU as well as the mECE within the uncertainty intervals. The uncertainty bars are given by the standard deviation of the mean of the mIoU and the mECE regarding the test image batch of size 40. In addition, the MTL model shows similar performance trends regarding the Fourier optical metrics as the HRNet if the envelope functions and the subdomain fluctuations in Figure 4 are considered. Here, each data point represents the effect of a windshield, parameterized by a set of Zernike coefficients in the range of $\omega_n \in [-\lambda, \lambda]$, on a test image batch of size 20. The envelope function is obtained by binning the data and assigning the minimum value within a bin to the envelope function. This procedure can be performed for all head-specific KPIs leading to the colored stack plot in Figure 4 after applying the uncertainty-weighting. In order to indicate the local performance spread, additional boxplots

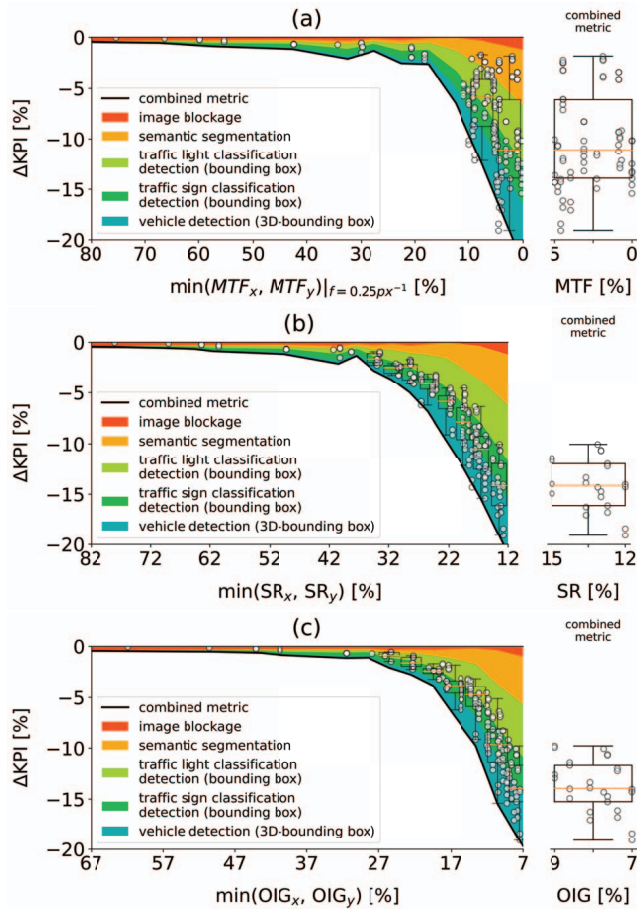


Figure 4. Multi-task performance versus (a) the MTF at half-Nyquist, (b) the Strehl ratio and versus (c) the OIG.

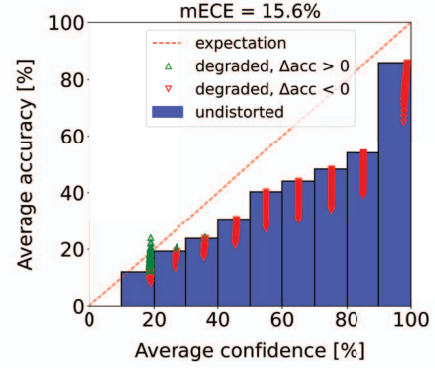


Figure 5. Calibration curve for the HRNet from Google.

are provided for the most relevant bins. The bin comprising the most severe optical aberrations is highlighted on the right-hand side in Figure 4. It can be concluded, that the statistical mass allocations within each bin are significantly more clustered in the case of the Strehl ratio and the OIG as compared to the MTF at Nyquist half frequency. Hence, Figure 4 shows evidence for the superiority of the Strehl ratio and the OIG as a quality metric in contrast to the MTF at Nyquist half frequency. The results clearly indicate that the information density of the PSF is beneficial for defining an optical ADAS working requirement. For quality assurance purposes it would be required to ensure that the quality metric is bijective, which is the subject of future studies. Overall, the MTF at Nyquist half frequency seems to be insufficient as a safety quality criterion for windshields.

6.3. Calibration robustness

The effect of optical aberrations on the reliability curve of the HRNet is visualized in Figure 5. In the diffraction limited case ($\omega_n = 0 \forall n \in \mathbb{N}_0$), the HRNet shows an mECE of 15.6%. If optical aberrations are considered, then the average accuracy and the average confidence decrease with increasing perturbation magnitude but the reduction is non-coherent, as demonstrated by Figure 3. Consequently, the neural network becomes more and more overconfident, which is underpinned by the increasing mECE in Figure 6. In the case of low prediction confidences and low prediction accuracies, the binned network accuracy seems to slightly increase if aberrated test data is used. This behavior is counterintuitive and represents an artifact of the visualization method. Essentially, optical aberrations drive the probability flow of the confidence distribution towards lower values shifting the predictions into low-confidence bins. Since the domain is limited and quantized, the bin composition varies affecting the binned accuracy. Physically, low-confidence predictions in semantic segmentation mostly correspond to class area borders. If the image is perturbed, then the contours are getting blurred as illustrated by Figure 1, which results in lower prediction confidences for these pixels.

7. Conclusion

The results presented in this paper reflect an initial assessment of the functional relationship between neural network benchmark metrics and optical aberrations parameterized by Zernike coefficients. From our experiments, we report evidence on the superiority of the Strehl ratio and the OIG as an optical quality indicator for image-based neural network performance in contrast to present functionality requirements in terms of the refractive power and the MTF at half-Nyquist frequency. In addition, the studies demonstrate that a pure defocus is influencing the performance of a semantic segmentation algorithm more than astigmatic aberrations. Furthermore, the investigations on

the HRNet from Google and the studies on the MTL model from CARIAD show similar sensitivities and functional dependencies on optical aberrations, which leads to the hypothesis, that the results presented in this paper could be network architecture independent.

Finally, it has to be emphasized that the optical threat model applied in this paper was tuned for telephoto objective lenses. As a consequence, the scalar product in the exponent of Equation (3) is assumed to be given by the product of the vector magnitudes. If wide-angle cameras are considered, then the optical threat model has to be adjusted by including the dependency on the field angle ψ as: $\vec{x}_o \cdot \vec{k}_a = \|\vec{x}_o\|_2 \cdot \|\vec{k}_a\|_2 \cdot \cos \psi$.

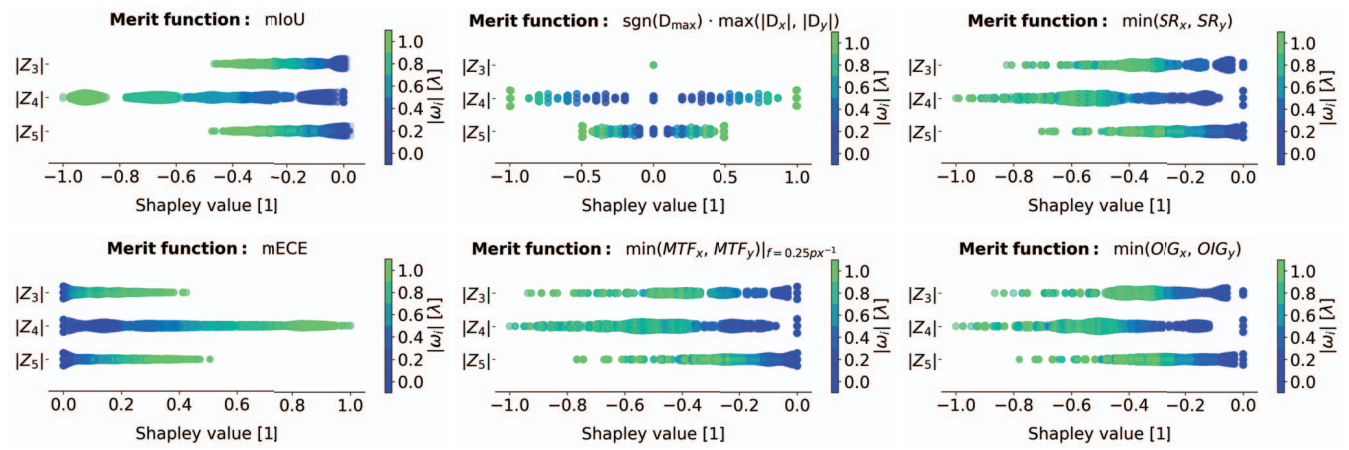


Figure 6. The sensitivities of different convolutional neural network benchmark metrics as well as the sensitivities of several optical KPIs on wavefront aberrations, parameterized by Zernike coefficients, are quantified and visualized in terms of Shapley values. The impact of an induced defocus (Z_4) surpasses the effect of oblique- (Z_3) and vertical astigmatism (Z_5) for all merit functions studied in this paper.

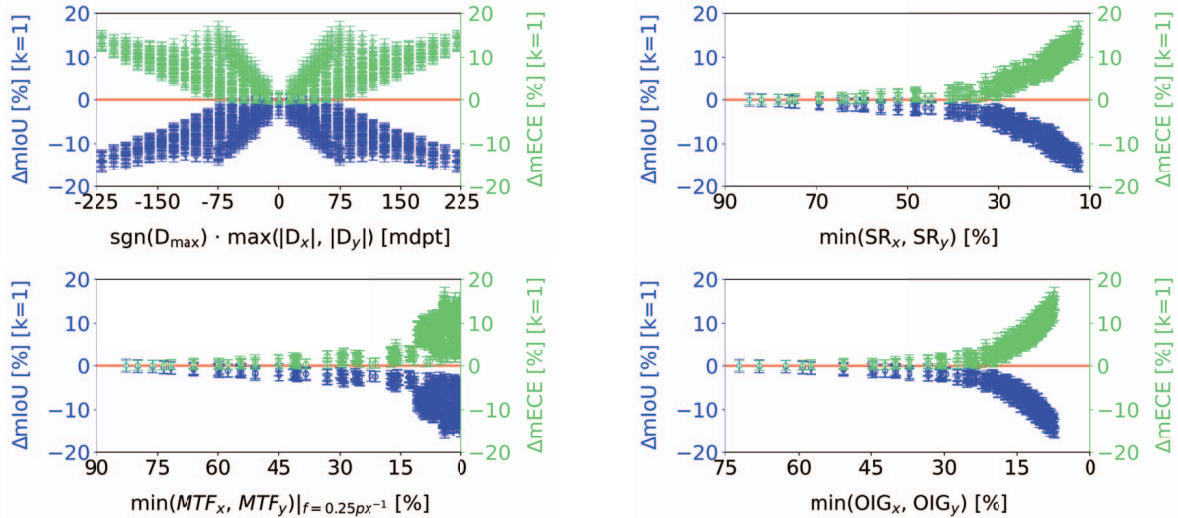


Figure 7. The dependency of the mIoU and the mECE on different optical merit functions is shown. The results are almost symmetrical around the baseline if the trend of the mIoU is considered in relation to the mECE, which is scientifically justified by Figure 3. In summary, the refractive power and the MTF at half-Nyquist frequency do not demonstrate a functional relationship w.r.t. the mIoU and the mECE. In contrast, the Strehl ratio and the OIG indicate a functional relationship, which might even fulfill the required bijectivity criterion.

References

- [1] Waseem Abbas, Muhammad Fakhir Khan, Murtaza Taj, and Arif Mahmood. Statistically correlated multi-task learning for autonomous driving, 2021. *Neural Computing and Applications*, Volume 33, Issue 19. URL: <https://doi.org/10.1007/s00521-021-05941-8>. 2
- [2] Andrea Agostinelli, Michal Pándy, Jasper Uijlings, Thomas Mensink, and Vittorio Ferrari. How stable are transferability metrics evaluations?, 2022. URL: <https://arxiv.org/abs/2204.01403>. 5
- [3] Mark Bakker, Kees Maas, and Jos R. Von Asmuth. Calibration of transient groundwater models using time series analysis and moment matching, 2008. *Water Resources Research*, Vol.44. URL: <https://doi.org/10.1029/2007WR006239>. 3
- [4] Manfred-Andreas Beeck and Bernhard Schittek. Optical properties of automotive glazing—design and feasibility limitations, 2003. *Volkswagen AG, Glass Processing Days 2023*, p. 502-504. 3
- [5] J. Chang and G. Wetzstein. Deep optics for monocular depth estimation and 3d object detection, 2019. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 10193-10202). URL: https://openaccess.thecvf.com/content_ICCV_2019/papers/Chang_Deep_Optics_for_Monocular_Depth_Estimation_and_3D_Object_Detection_ICCV_2019_paper.pdf. 2, 4
- [6] Qing Chang, Junran Peng, Lingxie Xie, Jiajun Sun, Haoran Yin, Qi Tian, and Zhaoxiang Zhang. Data: Domain-aware and task-aware self-supervised learning, 2022. *Computer Vision and Pattern Recognition (CVPR)*. URL: <https://arxiv.org/abs/2203.09041>. 2
- [7] Krebs Christian, Müller Patrick, and Braun Alexander. Impact of windshield optical aberrations on visual range camera based classification tasks performed by cnns, 2021. In *Proc. IS&T London Imaging Meeting 2021: Imaging for Deep Learning*, pp 83 - 87. URL: <https://doi.org/10.2352/issn.2694-118X.2021.LIM-83>. 4
- [8] A. Deitmar and S. Echterhoff. *Principles of harmonic analysis*, 2008. Springer New York. ISBN: 9780387854687. 3
- [9] International Organization for Standardization. *Ophthalmic optics and instruments - reporting aberrations of the human eye*, 2008. ISO Standard No. 24157:2008. URL: <https://www.iso.org/obp/ui/#iso:std:iso:24157:ed-1:v1:en>. 4
- [10] International Organization for Standardization. *Photography - electronic still picture imaging - resolution and spatial frequency responses*, 2017. ISO Standard No. 12233:2017. URL: <https://www.iso.org/obp/ui/#iso:std:iso:12233:ed-3:v1:en>. 5
- [11] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The KITTI dataset, 2013. *International Journal of Robotics Research (IJRR)*. 5, 6
- [12] Joseph W. Goodman. *Introduction to fourier optics*. McGraw-Hill, San Francisco, 1968. Stanford University, ISBN: 0-07-023776-X. URL: <https://d-nb.info/1143249038>. 1, 2, 3, 4
- [13] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On calibration of modern neural networks, 2017. URL: <https://arxiv.org/abs/1706.04599>. 4
- [14] Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations, 2019. *International Conference on Learning Representations (ICLR)*. URL: <https://arxiv.org/abs/1903.12261>. 3
- [15] M. J. Irland. Windshield optics, 1969. *Appl. Opt.* 8, 1787-1790. URL: <https://opg.optica.org/ao/abstract.cfm?URI=ao-8-9-1787>. 1
- [16] Wang Jingdong, Sun Ke, Cheng Tianheng, Jiang Borui, Deng Chaorui, Zhao Yang, Liu Dong, Mu Yadong, Tan Mingkui, Wang Xinggang, Liu Wenyu, and Xiao Bin. Deep high-resolution representation learning for visual recognition, 2020. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. URL: <https://arxiv.org/abs/1908.07919>. 5
- [17] Alex Kendall, Yarin Gal, and Roberto Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics, 2018. *Conference on Computer Vision and Pattern Recognition (CVPR)*. URL: <https://arxiv.org/abs/1705.07115>. 6
- [18] Michael C. Khoo. *Physiological control systems : Analysis, simulation, and estimation*, 2018. IEEE Press series in biomedical engineering, ISBN: 978-1-119-05879-3. 3
- [19] Der-Hau Lee and Jinn-Liang Liu. End-to-end multi-task deep learning and model based control algorithm for autonomous driving, 2021. URL: <https://arxiv.org/abs/2112.08967>. 2

- [20] Scott Lundberg and Su-In Lee. A unified approach to interpreting model predictions, 2017. URL: <https://arxiv.org/abs/1705.07874>. 4
- [21] Thomas Mensink, Jasper Uijlings, Alina Kuznetsova, Michael Gygli, and Vittorio Ferrari. Factors of influence for transfer learning across diverse appearance domains and task types, 2021. IEEE Transactions on Pattern Analysis and Machine Intelligence. URL: <https://arxiv.org/abs/2103.13318>. 5
- [22] Shervin Minaee, Yuri Boykov, Fatih Porikli, Antonio Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos. Image segmentation using deep learning: A survey, 2020. URL: <https://arxiv.org/abs/2001.05566>. 4
- [23] Christoph Molnar, Gunnar König, Julia Herbinger, Timo Freiesleben, Susanne Dandl, Christian A. Scholbeck, Giuseppe Casalicchio, Moritz Grosse-Wentrup, and Bernd Bischl. General pitfalls of model-agnostic interpretation methods for machine learning models, 2020. URL: <https://arxiv.org/abs/2007.04131>. 4
- [24] Hauskrecht M. Naeini MP, Cooper GF. Obtaining well calibrated probabilities using bayesian binning, 2015. Proc Conf AAAI Artif Intell. URL: https://people.cs.pitt.edu/~milos/research/AAAI_Calibration.pdf. 4
- [25] Conference on Computer Vision, Marvin Pattern Recognition (CVPR). Klingner, Andreas Bär, and Tim Fingscheidt. Improved noise and attack robustness for semantic segmentation by using multi-task training with self-supervised depth estimation, 2020. 2
- [26] Yaniv Ovadia, Emily Fertig, Jie Ren, Zachary Nado, D Sculley, Sebastian Nowozin, Joshua Dillon, Balaji Lakshminarayanan, and Jasper Snoek. Can you trust your model’s uncertainty? evaluating predictive uncertainty under dataset shift, 2019. Advances in Neural Information Processing Systems (NeurIPS). URL: <https://papers.nips.cc/paper/2019/hash/8558cb408c1d76621371888657d2eb1d-Abstract.html>. 3
- [27] Benedek Rozemberczki, Lauren Watson, Péter Bayer, Hao-Tsung Yang, Olivér Kiss, Sebastian Nilsson, and Rik Sarkar. The shapley value in machine learning, 2022. URL: <https://arxiv.org/abs/2202.05594>. 4
- [28] Sebastian Ruder. An overview of multi-task learning in deep neural networks, 2017. URL: <https://arxiv.org/abs/1706.05098>. 2
- [29] Lloyd S. Shapley. A value for n-person games, 1952. RAND Corporation, Santa Monica, CA. URL: <https://www.rand.org/pubs/papers/P295.html>. 1, 4
- [30] Christian Tomani, Daniel Cremers, and Florian Buetner. Parameterized temperature scaling for boosting the expressive power in post-hoc uncertainty calibration, 2022. European Conference on Computer Vision (ECCV) 2022. URL: https://www.ecva.net/papers/eccv_2022/papers_ECCV/papers/136730554.pdf. 4
- [31] Klinghoffer Tzofi, Somasundaram Siddharth, Tiwary Kushagra, and Raskar Ramesh. Physics vs. learned priors: Rethinking camera and algorithm design for task-specific imaging, 2022. International Conference on Computational Photography (ICCP). URL: <https://arxiv.org/pdf/2204.09871.pdf>. 2
- [32] Dominik Werner Wolf, Markus Ulrich, and Alexander Braun. Novel developments of refractive power measurement techniques in the automotive world, 2023. Submitted to Metrologia. 5
- [33] Dominik Werner Wolf, Markus Ulrich, and Alexander Braun. Windscreen optical quality for ai algorithms: Refractive power and mtf not sufficient, 2023. Accepted at the IEEE ITSC-2023. URL: <https://arxiv.org/abs/2305.14513v1>. 3, 4
- [34] Lu Yuan, Dongdong Chen, Yi-Ling Chen, Noel Codella, Xiyang Dai, Jianfeng Gao, Houdong Hu, Xuedong Huang, Boxin Li, Chunyuan Li, Ce Liu, Mengchen Liu, Zicheng Liu, Yumao Lu, Yu Shi, Lijuan Wang, Jianfeng Wang, Bin Xiao, Zhen Xiao, Jianwei Yang, Michael Zeng, Luowei Zhou, and Pengchuan Zhang. Florence: A new foundation model for computer vision, 2021. URL: <https://arxiv.org/abs/2111.11432>. 2