Impact of realistic properties of the point spread function on classification tasks to reveal a possible distribution shift

Patrick Müller
Hochschule Düsseldorf -
University of Applied Sciences

Alexander Braun
Hochschule Düsseldorf -
University of Applied Sciences

Margret Keuper
University of Siegen &
Max Planck Institute for Informatics

Abstract

Image classification is a long-standing task in computer vision with deep neural networks (DNN) producing excellent results on various challenges. However, they are required not only to perform highly accurate on benchmarks such as ImageNet, but also to robustly handle images in adverse conditions, such as modified lighting, sharpness, weather conditions and image compression. Various benchmarks aimed to measure robustness show that neural networks perform differently well under distribution shifts. While datasets such as ImageNet-C model for example common corruptions such as blur and adverse weather conditions, we argue that the properties of the optical system and the potentially resulting complex lens blur are insufficiently well studied in the literature. This study evaluates the impact of realistic optical corruptions on the ImageNet classification. The proposed complex corruption kernels are direction and wavelength dependent and include chromatic aberration, which are all to be expected in realistic scenarios such as autonomous driving applications. Our experiments on twelve different DNN models show significant differences of more than 5% in the top1 classification accuracy, when compared to the model performances on matched ImageNet-C blur kernels.

1 Introduction

Neural networks for computer vision tasks are used in a wide range of applications and are expected to be reliable in safety-critical situations such as in autonomous driving [1, 2, 3]. Therefore, it is crucial that neural networks are able to generalize from the training distribution to unknown images and slight domain shifts. In order to test such behavior, various benchmarks have been proposed that introduce targeted corruptions [4, 5, 6, 7] or mimic common corruptions and adverse weather conditions [8, 9]. The simulation of effects such as lighting conditions, noise or image sharpness have been addressed for examples by Hendrycks et al. [8]. In [9], Kar et al. aim for more realistic modeling by proposing additional 3D corruptions such as change of field of view, camera roll or far focus.

While these papers cover a wide range of corruptions, they introduce necessary simplifications for each of them. As a result, they do not consider complex corruptions as they can result from the optical recording system. In this paper, we argue that the modeling of such realistic optical blur kernels that include chromatic aberrations and astigmatism should be considered, since they are very likely to eventually occur in practice.

Workshop on Distribution Shifts, 36th Conference on Neural Information Processing Systems (NeurIPS 2022).

Thus, we examine the effects of two complex kernel types on ImageNet and evaluate on the resulting corruptions the behavior of existing, robust and non-robust classification models, using ImageNet-C [8] as a baseline.

Practically, the image of an object plane produced by a lens can be described using linear systems theory: If an idealized point light source is propagated through a camera's objective lens, the point function is spread to some degree on the image side. This observed intensity function is therefore called the intensity point *spread* function (PSF). In the best case, the PSF is diffraction-limited and therefore small and rotationally-symmetric. As a rule, however, the optical system is not perfectly balanced and aberrations arise. These aberrations directly affect the observed PSF. Thus, non-rotationally symmetric shapes that differ in wavelength can often occur and show the effects of e.g. astigmatism and chromatic aberration. PSF examples are given in the appendix A. In general, the PSF varies over the image field. The superposition of all these point responses weighted with the scene's intensity yields the observed image.[10, 11] By means of simulation, such PSFs applied to an image can mimic the capture with the virtual lens model. For simplicity, we assume isoplanarity such that a single PSF represents a virtual lens, but with a focus on directional and wavelength dependence in contrast to approximate rotational symmetric blur kernels such as assumed in ImageNet-C[8].

In other words, here, the lens effect is mimiced by convolution of an image with a single PSF kernel representing the whole lens. This assumes independence of the incident angle of light. Further, this post-process filter is applied to an already projected scene excluding distortion and we do not model vignetting or effects like ghosts or scattering. Further, convolution assumes a large depth of field resulting in a PSF, which does not depend on depth. Also, camera sensor simulation is not covered in this study, excluding e.g. noise and color filter array effects. We focus on the shape of realistic PSFs compared to disk-like blur kernels.

2 Analysis of the impact of realistic PSF properties to classification

This article investigates the impact of directional and wavelength dependent PSFs at different severities. The optically motivated kernels will be called optical kernels here in contrast to the corruption kernels from ImageNet-C [8].

We take the defocus blur kernels from ImageNet-C and compare to them in size approximate matched optical kernels. Subsequently the image datasets (degraded with defocus blur from ImageNet-C and the proposed optical kernels) are classified by different pretrained neural networks. We compare the results for all severities for classification and image degradation. All operations -are done in python and pytorch. Google Colab GPUs are used to perform the inference for the image datasets and to perform the different image dataset comparisons.

2.1 Metrics

To compare the data, different metrics are used following the dataflow from a set of PSFs to images (IMG) and classification (CLS). First, PSFs are generated, see Figure 1(left). From this, the modulation transfer function (MTF) can be generated, which is an established optical quality metric.[12, 13] The two-dimensional function is given in pairs of orthogonal slices in a particular direction, e.g. in direction of the coordinate axes or in diagonal direction. There exist different functionals derived from the MTF curve such as MTF50, which denotes the frequency value where the MTF falls to $50\,\%$, and area under the curve (AUC), which is the integral of the MTF.[14] The metrics are meaningful, if curves of similar shape are compared to ensure for example for AUC, that the intensity is concentrated in similar frequency regions. These curves are used to match the corruption kernels and optical kernels.

The application of each PSF kernel to ImageNet yields a modified image dataset for each severity, which is compared to the unmodified ImageNet validation dataset. Every image of a modified dataset is compared to the unmodified original. These comparisons are evaluated with mean squared error (MSE) and the structural similarity index (SSIM).[15] To obtain a manageable overview, the means of these metrics are taken and are denoted as mMSE and mSSIM respectively. This asseses each dataset and severity with a single number.

The image datasets are then fed to the different classifiers. The classifiers are assessed with the top1-Accuracy, which takes the mean of all falsely classified images subtracted from 1 to have a similarity measure. These values are then compared for the different models and for each severity.

2.2 Blur models

The corruption kernels represent the reproduced "defocus blur" corruption kernels from ImageNet-C for five severities. The generation of these kernels is a two step process: First, a disk shape is created, which is then blurred with Gaussian blur.[8] This yields approximate rotationally symmetric blur kernels, which do not depend on the color channel and model constant defocus.

To compare such kernels with more realistic optical kernels, we parameterize the color-dependent central PSF of a Cooke-Triplet lens model [16]. The below Eq. 1 models the optical system as a black box, where Eq. 1 propagates the wave containing the effect of the lens from the exit pupil to the imager position at focus z_i. At this distance an intensity PSF h can be observed, representing imaging with diffraction and aberration [11]:

$$h(u, v, \lambda) = |\mathcal{F}\big\{(P(x, y) \cdot e^{-j\frac{2\pi}{\lambda z_i} W_\lambda(x,y,\lambda)}\big\}|^2, \tag{1}$$

The complex argument of the Fourier Transform $\mathcal{F}$ contains the information about the lens consisting of the pupil shape P and its phase with the wavefront aberration W_λ.[11] The model assumes scalar diffraction theory and therefore no polarization. Since we use a l_1-normed discrete PSF to preserve energy in the final image, any scalar weights are suppressed. In this article, we further restrict the model to represent the lens with a single PSF and no dependence on the object distance and angle and no magnification to concentrate on the kernel's shape. We observe the PSF h in a region $u, v \in [0, 24)$ pixels.

To obtain the specific PSF, the description of the wavefront aberration W_λ representing the lens as available from [16] in terms of Zernike Polynomials Z_n^m [10] is parameterized:

$$W_\lambda(x, y, \lambda) = \lambda \cdot \sum_{n,m} A_n^m(\lambda) \cdot Z_n^m \tag{2}$$

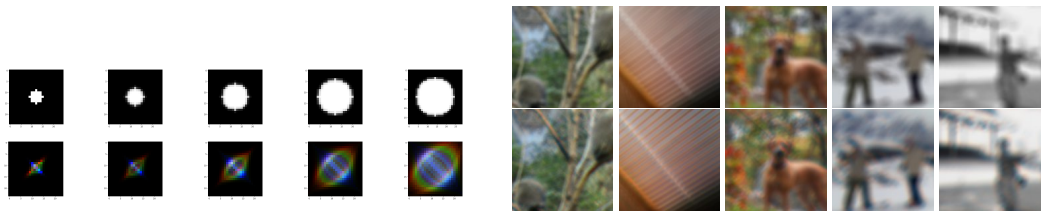
Each coefficient A_n^m in multiples of the wavelengths $\lambda_i \in \{0.4861\,\mu\text{m}, 0.5876\,\mu\text{m} \text{ and } 0.6563\,\mu\text{m}\}$ represents the contribution of a particular type of aberration. This allows to parameterize different aspects such as the amount of coma, astigmatism or defocus. To show directional dependence, oblique astigmatism A_2^{-2} is added to the original wavefront aberration W_λ, which may also lead to chromatic aberration as the wavefront aberration W_λ depends on wavelength. The experimentally determined parameterization $A_2^{-2} \in \{1, 1.4, 2, 3, 3.8\}\lambda_i$ leads to optical kernels comparable to the corruption kernels both in terms of PSF metrics (MTF50, AUC, MTF curves) as well as IMG metrics (SSIM, MSE). Note that the central PSF contains little aberration and the resulting shape is mainly controlled by the parameterization.

We also include two additional studies in appendix C for "straight" astigmatism and a closer matched version for oblique astigmatism with comparable trend.

3 Experiments on ImageNet

We select 12 pretrained neural networks for classification with common architectures [17] from Tab. 1. All models are pretrained on ImageNet and publicly available from Pytorch Vision's model zoo. We also include three models with ResNet50 architecture from the RobustBench's model zoo [7] ranked under the top-5 networks at the ImageNet leaderboard for corruptions.[7] These neural networks are robust on ImageNet-C [8] and ImageNet-3DC [9] corruptions respectively. Hendrycks2020AugMix is trained on ImageNet together with AugMix, which randomly mixes different augmentations based on AutoAugment [18], such as posterizing, rotation and translation, while ImageNet-C corruptions are excluded.[19] Hendrycks2020Many augments ImageNet with DeepAugment, which stochastically perturbates images by zeroing, convolving, changing activation functions and others in the augmentation network's forward pass and weights. This creates visually diverse distortions, which preserve semantics.[20] Geirhos2020SIN+IN is trained on Stylized-ImageNet (SIN) and ImageNet (IN).[21]

The exact listing of the network architectures can be found in the appendix D in Tab. 1.



(a) (b) (c) (d) (e)

Figure 1: Left: PSFs representing the different "severities" 1-5 from left to right: First row visualizes the corruption kernels, the second row astigmatic optical kernels showing chromatic aberration. Right: Visual examples from the modified datasets with increasing severities: 2 (a), 3 (b-c) and 4 (d-e). The first row visualizes the effects of corruption kernels and the second of optical kernels. Read example: Chromatic aberration visible at the reddish and greenish branches (a) and at the window blind (b), while the simple kernels do not create color fringes.

3.1 PSF comparisons

Here, the complex optical kernels shall first be compared to the optical kernels visually in Fig 1 for all severities. Visually, the severities are comparable, even though there is a dependence on the color channel, as well as a break in rotational symmetry. A criterion to quantify the similarity are also the MTF50 values in appendix B, Fig. 8. The values are similar and indicate an averaged comparability between the kernels. The MTF curves for the kernels can also be found in the appendix B. Note that for the optical kernels six different curves exist, which occur because of the direction dependence and wavelength dependence, whereas the complexity of the corruption kernels can be reduced to a single MTF curve. Astigmatic behavior can be read, for example, in severity 1 in appendix B, Fig. 6 in the red and blue channels, where either the $45°$ or the $-45°$ curves proceed low. In addition, the orientations are just reversed, resulting in chromatic aberrations.

3.2 Image comparisons

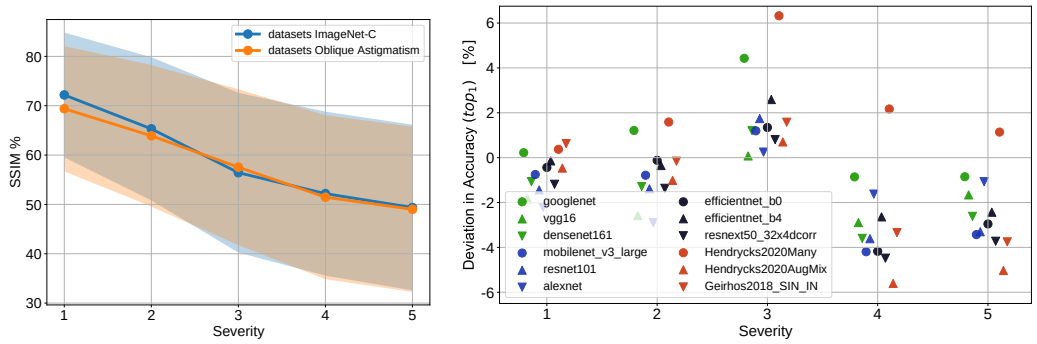
The different filter kernels are then used to convolve the images from ImageNet and yield ten datasets of 50k images each. Both, images from corruption kernels and optical kernels are saved with the same slight JPEG compression as in ImageNet-C to avoid a corruption bias between lossless and lossy image data compression and to keep the processing of 500k images manageable. The usual transformations center crop and downscale to RGB images of size 224×224 are done as preprocessing step to the bigger original ImageNet images to ensure that no additional smoothing by the scaling operation is applied to the filtered images. The image samples in Fig. 1 visualize differences in the corruption datasets and optical datasets. The blur in the images increases significantly with severities. Looking at optical datasets 2-4 color fringes at edges are readily visible and directional dependent blur occurs.

However, the mean values mMSE and mSSIM listed in Fig. 2 are very close to each other, which indicate similar and comparable image quality.

3.3 Classification on ImageNet

The images from the corruption datasets and optical datasets are then processed with the different neural networks for classification from appendix D, Tab. 1. Again, the processing is set such that no additional transformation besides image normalization is done, because the input size matches already the target size of 224×224. Fig. 20 visualizes the absolute accuracies for all studies, which shows a clear trend. While the neural networks of the Pytorch Vision model zoo were trained only on ImageNet, the networks of the RobustBench model zoo marked with red labels are additionally trained with different augmentations. As a result, these networks stand out as the most robust.

Further, to directly compare the effect of the optical kernels to the corruption kernels Fig. 2b shows the differences of Accuracies for the selected 12 neural networks and all severities. Looking at the mean of the deviation in accuracy μ_{diff}, this bias can possibly be due to the small variation between the kernel types: The image results show slight variations in mSSIM for the different



(a) (b)

Figure 2: (a) Difference in mSSIM for all severities. (b) Difference in Accuracy between modified and unmodified ImageNet for corruption datasets (blue) and optical datasets (orange). The average difference in mSSIM is below $1.1\,\%$ for severities 3-5. Classification results (b) with a $+$ denote higher Accuracy for optical datasets with oblique astigmatism.

severities. For instance, severity 3 shows for the optical datasets a higher mSSIM ($1.11\,\%$) and for severity 4, the mSSIM for optical datasets is lower than for corruption datasets ($-0.720\,\%$). However, significant differences between the networks can be found within these conditions. For example, Hendrycks2020Many (red dot) [20, 22], as well as in most cases GoogLeNet (green dot) [23], can handle the blur of optical kernels better than that of corruption kernels (higher Accuracy). Hendrycks2020AugMix (red triangle) [19, 22] does mostly the opposite. The total deviation in accuracy between these two neural networks is above 7% for severity 4. Comparing the ImageNet only trained networks, indicated by blue and green colors, GoogLeNet can handle the blurred optical datasets more than 4% better than corruption datasets for severity 3 compared to VGG16 (green triangle). Note that the differences in mSSIM, mMSE, MTF50 and AUC are for severities 1-2 greater, but the Accuracy deviation is smaller for these less blurred images.

4 Conclusion

This article compares the ImageNet validation dataset in several modified versions degraded with corruption kernels from ImageNet-C and optical kernels, which are directional and wavelength dependent. Although the blur kernels are of comparable size and produce similar image datasets, some classification results differ significantly. We conclude that realistic optical corruptions should be considered in our standard benchmarks towards model robustness.

References

[1] Dario Amodei et al. *Concrete Problems in AI Safety*. July 25, 2016. arXiv: `1606.06565[cs]`.

[2] Philip Koopman and Michael Wagner. "Challenges in Autonomous Vehicle Testing and Validation". In: *SAE International Journal of Transportation Safety* 4.1 (Apr. 5, 2016), pp. 15–24.

[3] Philip Koopman, Aaron Kane, and Jen Black. "Credible Autonomy Safety Argumentation". In: *Proceedings of the Safety-Critical Systems Symposium 2019* (), p. 27.

[4] Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy. *Explaining and Harnessing Adversarial Examples*. Mar. 20, 2015. arXiv: `1412.6572[cs,stat]`.

[5] Nicholas Carlini and David Wagner. "Adversarial Examples Are Not Easily Detected: Bypassing Ten Detection Methods". In: *Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security*. CCS '17: 2017 ACM SIGSAC Conference on Computer and Communications Security. Dallas Texas USA: ACM, Nov. 3, 2017, pp. 3–14.

[6] Nicolas Papernot et al. "The Limitations of Deep Learning in Adversarial Settings". In: *2016 IEEE European Symposium on Security and Privacy (EuroS&P)*. 2016 IEEE European Symposium on Security and Privacy (EuroS&P). Mar. 2016, pp. 372–387.

[7] Francesco Croce et al. "RobustBench: a standardized adversarial robustness benchmark". In: *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*. 2021.

[8] Dan Hendrycks and Thomas Dietterich. "Benchmarking Neural Network Robustness to Common Corruptions and Perturbations". In: *arXiv:1903.12261 [cs, stat]* (Mar. 28, 2019). arXiv: `1903.12261`.

[9] Oguzhan Fatih Kar et al. "3D Common Corruptions and Data Augmentation". In: (), p. 12.

[10] Max Born and Emil Wolf. *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*. 7th expanded ed. Cambridge ; New York: Cambridge University Press, 1999. 952 pp.

[11] Joseph W. Goodman. *Introduction to Fourier optics*. Fourth edition. New York: W.H. Freeman, Macmillan Learning, 2017. 546 pp.

[12] Glenn D. Boreman. *Modulation Transfer Function in Optical and Electro-Optical Systems*. SPIE, July 1, 2001.

[13] *ISO12233:2017, Photography — Electronic still picture imaging — Resolution and spatial frequency responses*. Standard. Volume: 2017. Geneva, CH: International Organization for Standardization, 2017.

[14] HH Nasse. "How to read mtf curves, Part II". In: *Carl Zeiss, Camera Lens Division* (2009).

[15] Z. Wang et al. "Image Quality Assessment: From Error Visibility to Structural Similarity". In: *IEEE Transactions on Image Processing* 13.4 (Apr. 2004), pp. 600–612.

[16] Patrick Müller and Alexander Braun. "Simulating optical properties to access novel metrological parameter ranges and the impact of different model approximations". In: *2022 IEEE International Workshop on Metrology for Automotive (MetroAutomotive)*. 2022 IEEE International Workshop on Metrology for Automotive (MetroAutomotive). July 2022, pp. 133–138.

[17] Alfredo Canziani, Adam Paszke, and Eugenio Culurciello. *An Analysis of Deep Neural Network Models for Practical Applications*. Apr. 14, 2017. arXiv: `1605.07678[cs]`.

[18] Ekin D. Cubuk et al. "AutoAugment: Learning Augmentation Strategies From Data". In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, June 2019, pp. 113–123.

[19] Dan Hendrycks et al. *AugMix: A Simple Data Processing Method to Improve Robustness and Uncertainty*. Feb. 17, 2020. arXiv: `1912.02781[cs,stat]`.

[20] Dan Hendrycks et al. "The Many Faces of Robustness: A Critical Analysis of Out-of-Distribution Generalization". In: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada: IEEE, Oct. 2021, pp. 8320–8329.

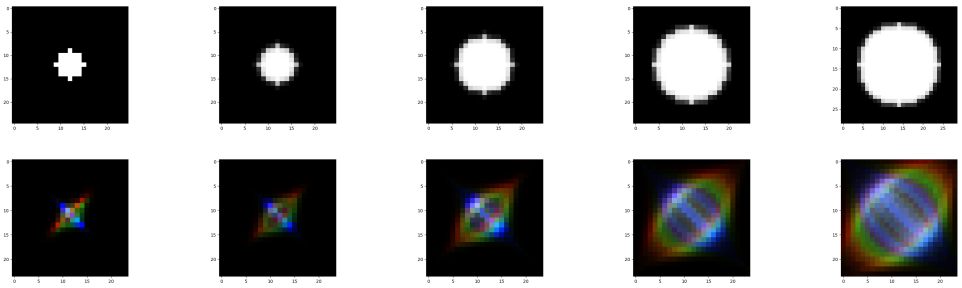
[21] Robert Geirhos et al. *ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness*. Jan. 14, 2019. arXiv: `1811.12231[cs,q-bio,stat]`.

[22] Francesco Croce et al. *RobustBench: a standardized adversarial robustness benchmark*. Oct. 31, 2021. arXiv: `2010.09670[cs,stat]`.

[23] Christian Szegedy et al. *Going Deeper with Convolutions*. arXiv:1409.4842. type: article. arXiv, Sept. 16, 2014. arXiv: `1409.4842[cs]`.

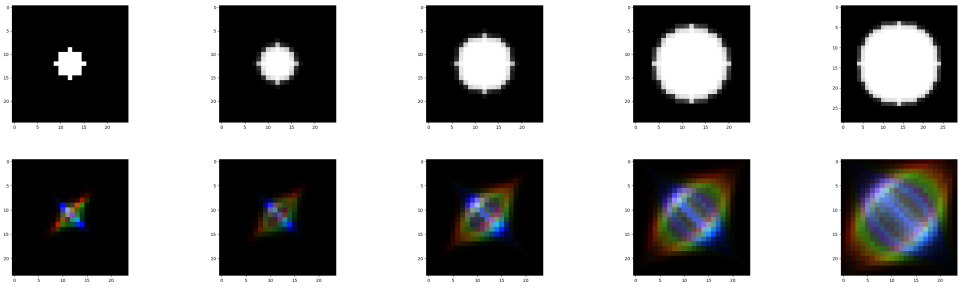
A PSFs

The different used kernels are visualized in Figs. 3, 4 and 5.



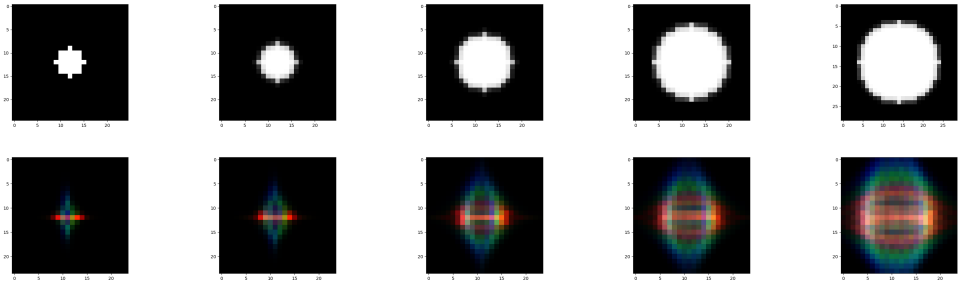
(a) (b) (c) (d) (e)

Figure 3: Average PSFs representing the different "severities" 1-5: (a-e) corruption kernels, (f-j) optical kernels with oblique astigmatism and chromatic aberrations.



(a) (b) (c) (d) (e)

Figure 4: Average PSFs representing the different "severities" 1-5: (a-e) corruption kernels, (f-j) optical kernels with oblique astigmatism and chromatic aberrations. These kernels have higher match quality and minimize mSSIM on the datasets, although visually hard to distinct from Fig. 3

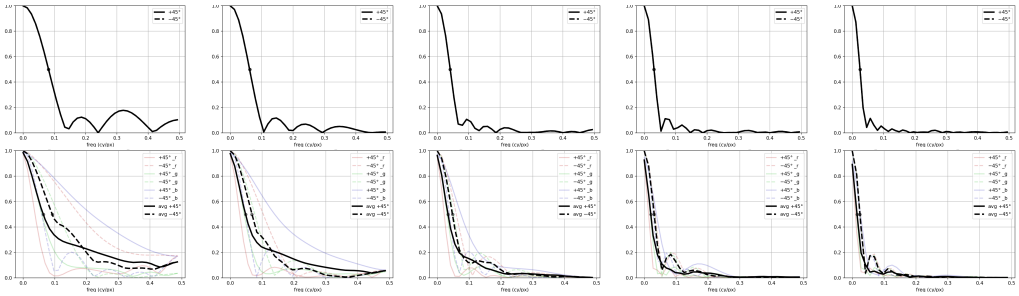


(a) (b) (c) (d) (e)

Figure 5: Average PSFs representing the different "severities" 1-5: (a-e) corruption kernels, (f-j) optical kernels with straight astigmatism and chromatic aberrations.

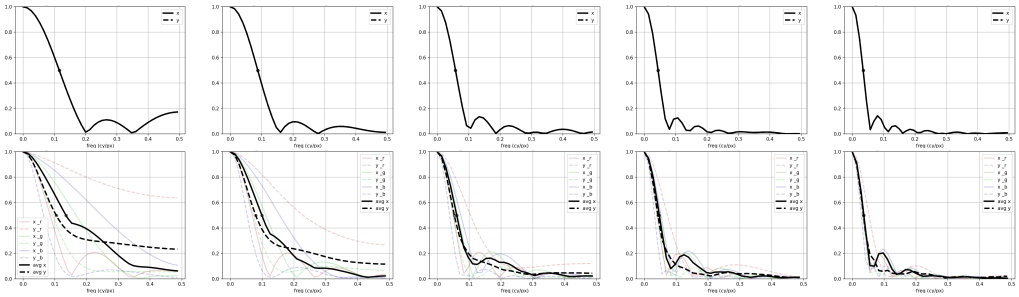
B MTF curves

Here, the direction in coordinate axes is denoted with (x, y) and the diagonal direction is denoted with $\pm 45°$. Note that sometimes curves for the optical kernels in different directions proceed differently. This astigmatic behaviour is color channel dependent and therefore creates chromatic aberrations.



(a) (b) (c) (d) (e)

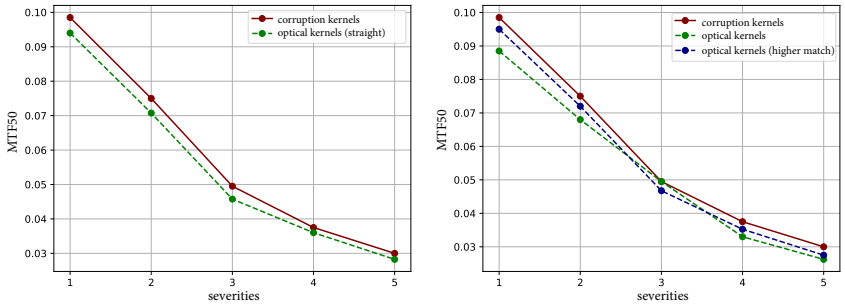
Figure 6: MTF slices from PSFs for severities (1-5, a-e): First row corruption kernels, second row optical kernels with oblique astigmatism from Fig. 3



(a) (b) (c) (d) (e)

Figure 7: MTF slices from PSFs for severities (1-5, a-e): First row corruption kernels, second row optical kernels with straight astigmatism (0°,90°) from Fig. 5

B.1 Aggregated MTF metrics - differences for MTF50



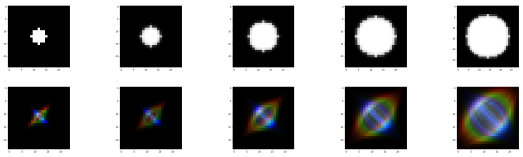
(a) (b)

Figure 8: MTF50 values for all severities and the different studies: (a) visualizes corruption kernels vs. optical kernels with straight astigmatism from Fig. 5 and (b) displays the optical kernels from Fig. 3 and 4. Each point is acquired by averaging color, orientation and direction to get a single MTF from the above MTF curves from Figs. 7 and 6. From this resulting MTF the MTF50 value is taken. We also show the MTF50 values from curves with minimum mSSIM on the datasets.

C Additional studies for different optical kernels

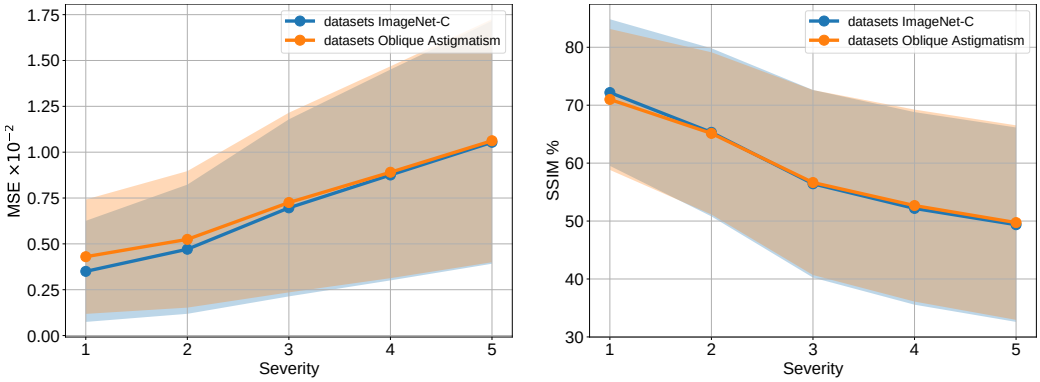
C.1 Oblique astigmatism (higher match quality)

We provide here a version with closer match in mSSIM, mMSE, MTF50 and AUC for all severities with the parameterization $A_2^{-2} \in \{0.9, 1.3, 2.1, 2.8, 3.6\}\lambda_i$. The mSSIM for severities 3-5 is below $0.5\,\%$, but although significantly smaller differences occur compared to Fig. 14 still the DNNs can handle the blur types differently well, which indicates a distribution shift. Lower severities tend to be more robust to changes in mSSIM and the improved match quality has little effect.



(a) (b) (c) (d) (e)

Figure 9: Average PSFs representing the different "severities" 1-5. First row: corruption kernels with higher match quality, second row: optical kernels with chromatic aberrations.



(a) (b)

Figure 10: Average MSE (a) and SSIM (b) between corrupted image database and unmodified original.

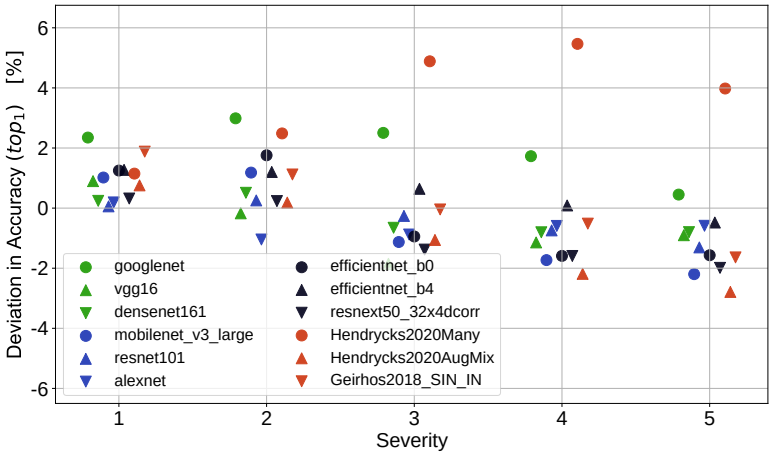
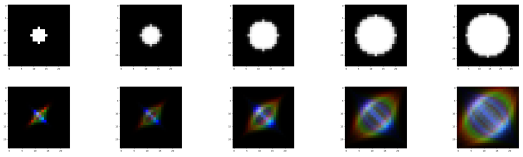


Figure 11: Deviation in Accuracy for all neural networks and severities. "+" means that optical datasets have higher Accuracy than corruption datasets.

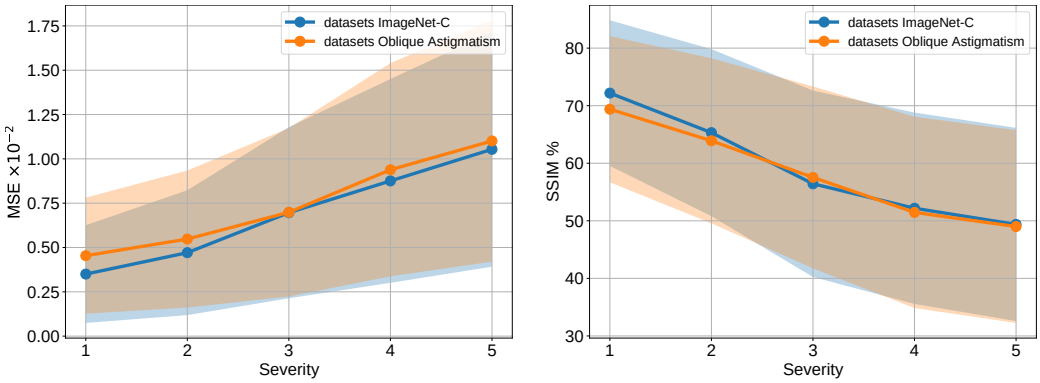
C.2 Oblique astigmatism

We include here the study on oblique astigmatic optical kernels as used in the main article. The size of the kernels in Fig. 3 look very similar to Fig. 4, but create a slightly larger mSSIM $\leq 1.1\,\%$ for severities 3-5.



(a) (b) (c) (d) (e)

Figure 12: Average PSFs representing the different "severities" 1-5. First row: corruption kernels, second row: optical kernels with chromatic aberrations.



(a) (b)

Figure 13: Average MSE (a) and SSIM (b) between corrupted image database and unmodified original.

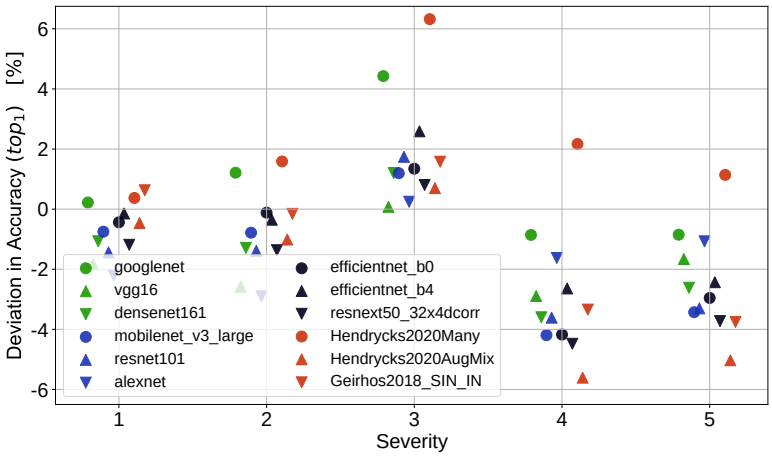
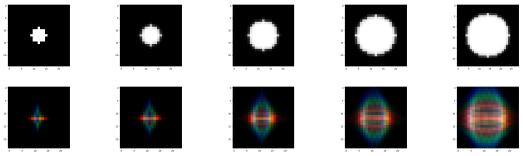


Figure 14: Deviation in Accuracy for all neural networks and severities. "+" means that optical datasets have higher Accuracy than corruption datasets.

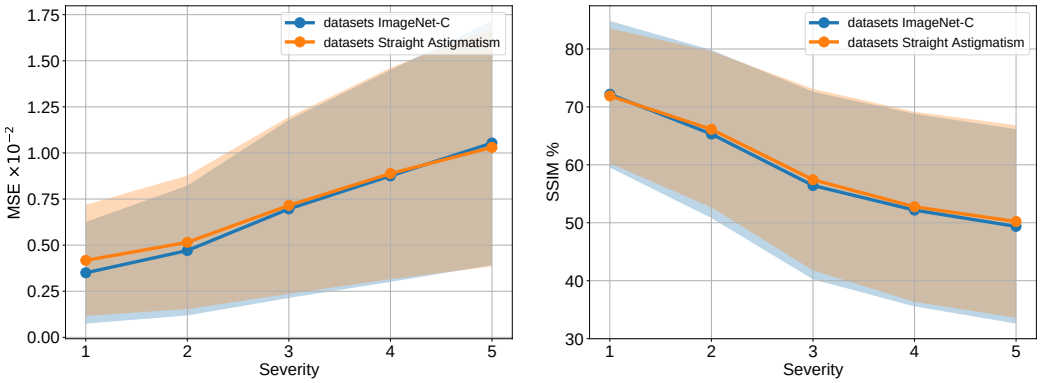
C.3 Straight astigmatism

This study shows the effect of the parameterization of straight astigmatism $A_2^{+2} \in \{1.0, 1.4, 2.2, 2.8, 3.6\}\lambda_i$, which leads to comparable results as in Fig. 11.



(a) (b) (c) (d) (e)

Figure 15: Average PSFs representing the different "severities" 1-5. First row: corruption kernels, second row: optical kernels with chromatic aberrations.



(a) (b)

Figure 16: Average MSE (a) and SSIM (b) between corrupted image database and unmodified original.

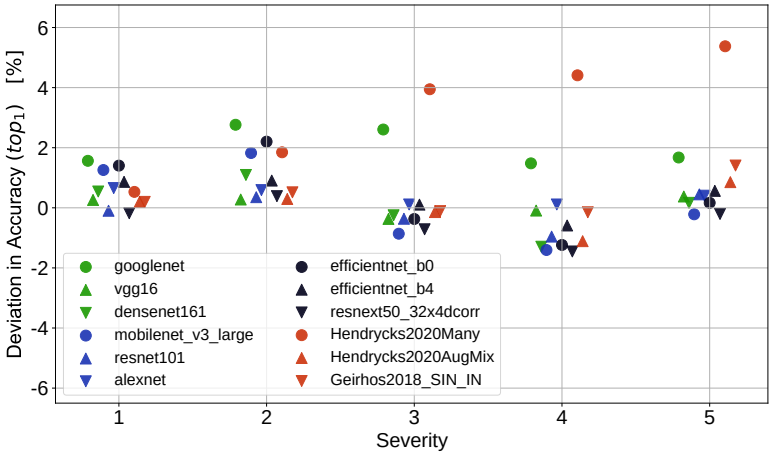


Figure 17: Deviation in Accuracy for all neural networks and severities. "+" means that optical datasets have higher Accuracy than corruption datasets. Difference in mSSIM $\leq 1\,\%$.

D Additional tables

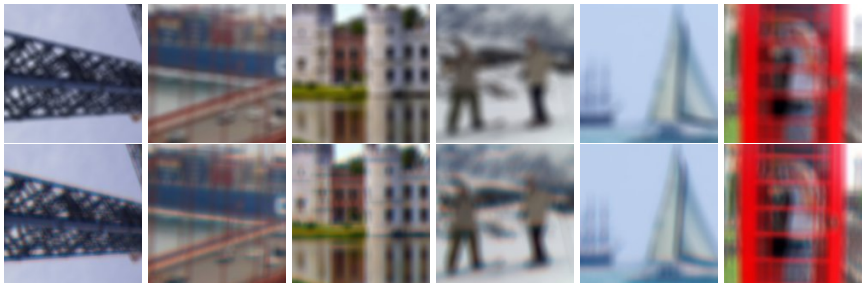
Table 1: Selection of pretrained models from pytorch vision and RobustBench model zoos with corresponding number of trainable parameters. The selected neural networks from RobustBench use a Resnet50.

Name	Number of Parameters $[\times 10^6]$	Name	Number of Parameters $[\times 10^6]$
VGG16	138.4	ResNeXt50	25.0
AlexNet	61.1	EfficientNet_B4	19.3
ResNet101	44.5	GoogLeNet	6.6
DenseNet161	28.7	MobileNet_v3	5.5
Resnet50	25.6	EfficientNet_B0	5.3

E Additional image examples

(a) (b) (c) (d) (e) (f)

Figure 18: All images are zoomed in. First row corruption datasets, second row optical datasets. Severity 3 (a-c): Chromatic aberration visible at the airplanes wheels, the dog's face, the window blind and the boat's mast. Note also the diagonal astigmatism. Severity 5: (d-e) visible chromatic aberration at the fish scales as orange stripes, the eagle, and as color fringe at the baby's hair.

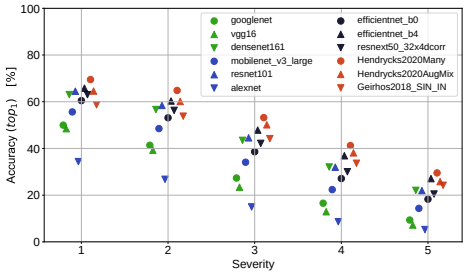


(a) (b) (c) (d) (e) (f)

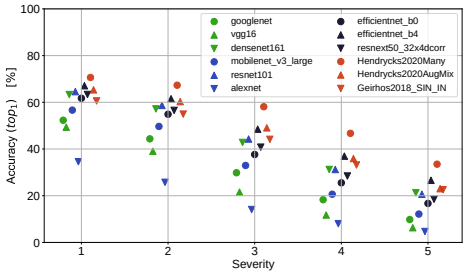
Figure 19: First row corruption datasets, second row optical datasets (straight). Severity 3 (a-c): Chromatic aberration visible at the airplanes wheels, the dog's face, the window blind and the boat's mast. Note also the diagonal astigmatism. Severity 5: (d-e) visible chromatic aberration at the fish scales as orange stripes, the eagle, and as color fringe at the baby's hair.

F Absolute classification Accuracies on the different modified dataset versions

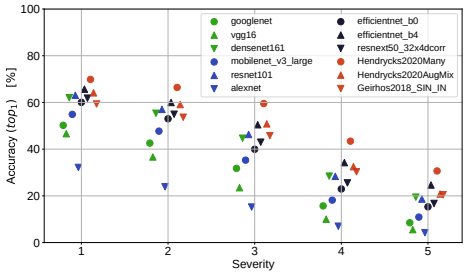
Below Fig. 20 shows the classification accuracies for all studies from appendix C. Note that while AlexNet (blue triangle downward) has always the worst accuracy, Hendrycks2020Many (red dot) achieves in all cases the best results. Additionally, the pretrained models from RobustBench's model zoo (red markers) outperfom most of the PyTorch Vision models.



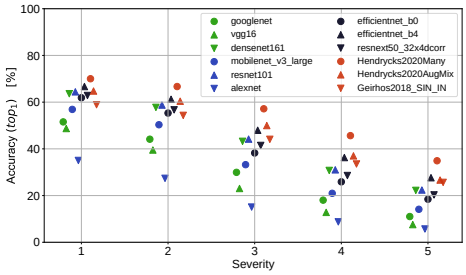
(a) Accuracy for corruption kernels from upper row in Fig. 3 for severities 1-5.



(b) Accuracy for optical kernels (oblique astigmatism) from Fig. 3 for severities 1-5. Difference in mSSIM $\leq 1.1\,\%$ for severities 3-5.



(c) Accuracy in percentage points: optical datasets with optical kernels from Fig. 4 *with higher match quality* for severities 1-5. Difference in mSSIM $\leq 0.5\,\%$ for severities 3-5.



(d) Accuracy in percentage points for optical datasets with straight astigmatism from Fig. 5 for severities 1-5. Difference in mSSIM $\leq 1\,\%$.

Figure 20: Classification results: Accuracies for the different studies from appendix C.