https://doi.org/10.2352/EI.2024.36.9.IQSP-257

ISP Tuning for Improved Image Quality in Machine Vision

Diarmaid Geever[1], Tim Brophy[1], Dara Molloy[2], Martin Glavin[1], Edward Jones[1], and Brian Deegan[1]

[1]School of Engineering, University of Galway, University Road, Galway, Ireland.
[2]Valeo, Tuam, Co. Galway H54 Y276, Ireland

Abstract

This paper investigates the relationship between image quality and computer vision performance. Two image quality metrics, as defined in the IEEE P2020 draft Standard for Image quality in automotive systems, are used to determine the impact of image quality on object detection. The IQ metrics used are (i) Modulation Transfer function (MTF), the most commonly utilized metric for measuring the sharpness of a camera; and (ii) Modulation and Contrast Transfer Accuracy (CTA), a newly defined, state-of-the-art metric for measuring image contrast. The results show that the MTF and CTA of an optical system are impacted by ISP tuning. Some correlation is shown to exist between MTF and object detection (OD) performance. A trend of improved AP5095 as MTF50 increases is observed in some models. Scenes with similar CTA scores can have widely varying object detection performance. For this reason, CTA is shown to be limited in its ability to predict object detection performance. Gaussian noise and edge enhancement produce similar CTA scores but different AP5095 scores. The results suggest MTF is a better predictor of ML performance than CTA.

Introduction

Machine vision has become a critical technology in ADAS and autonomous driving systems. Cameras provide these systems with information about the environment around the vehicle. It stands to reason that the images captured by these cameras must be of a sufficient quality at which useful information can be extracted from them. The IEEE P2020 working group is cuurently creating a set of standardised image quality metrics for use in automotive applications. In this study, two metrics, which are proposed by the P2020 group [1], are evaluated in terms of their ability to predict object detection performance. These are Modulation Transfer Function (MTF) and Contrast Transfer Accuracy (CTA).

A key task in machine vision is Detection Recognition Identification (DRI). This refers to tasks such as object detection (OD) and Optical Character Recognition (OCR). Commonly, these tasks are performed using deep learning networks. In this study, we investigate how such a network is affected by the quality of the input images and how an Image Signal Processor (ISP) can be tuned to improve relevant image quality metrics.

The study uses raw Bayer images, an open source software ISP [2], sharpness analysis software [3], contrast analysis software [4] and various detection models. Benchmarking different ISP configurations with detection models reveals relationships between ISP settings and key performance indicators. The investigation explores whether detection algorithms perform similarly with varying image quality and how these results can indicate the optimum ISP configuration.

Related Works

The P2020 working group was established to address the considerable ambiguity in the measurement of image quality of automotive imaging systems [5]. A draft standard was developed with the goal of applying relevant metrics and key performance indicators (KPIs) to automotive image quality, thus enabling customers and suppliers to efficiently define, measure, and communicate image quality of their imaging systems[1].

Several KPIs have been proposed since the formation of the P2020 working group. In 2018, Geese *et al*. [6] proposed the contrast detection probability (CDP). In more recent literature this metric is called CTA. Geese *et al*. put forward that a good requirement for the imaging chain is to guarantee the reproduction of a particular level of contrast to a specified probability, e.g., an 80% chance that a system can reproduce a contrast of 0.6. CTA is independent of the system under test. In 2019, Artmann *et al*. [7] showed how CTA can be used to describe the performance of the imaging chain by using test targets with varying light intensities. In 2023, Klein *et al*. [8] evaluated the ability of CTA to predict the performance of a computer vision algorithm. This study correlates OCR and CTA using licence plates as the target. CTA is benchmarked against the Contrast Signal-to-Noise Ratio (CSNR) and Frequency of Correct Resolution (FCR). In this study, the CTA is shown to have correlation with the performance of the OCR algorithm across the range of contrast and luminance. CTA is shown to have a greater correlation with OCR than CSNR; however, at low levels of contrast, CTA reaches a point of saturation, whereas CSNR is more sensitive. In Klein *et al*. [8] the impact of ISP on CTA is not evaluated.

Molloy *et al*. [9] evaluated the impact of ISP tuning on object detection by running a grid search of ISP blocks, varying the parameters of each block. Among others, the impact of Bi-lateral Noise Filtering (BNF) and Edge Enhancement (EEH) were evaluated. An open source software-based ISP, fast-openISP [2] was used. It was found that the more the ISP configuration differs from the default configuration, the greater the degradation in object detection performance.

Evaluating Impact of ISP on Object Detection

Dataset

Figure 1: A sample image from 8.9 MP camera used in the data acquisition

This study uses an original data set, which will be publicly available in the near future [10]. A sequence of raw images capturing a typical driving scene, from the perspective of an infrastructure node, i.e. the camera was mounted on a stand four metres vertically from the ground, at the side of the road, pointing towards the road surface, a similar aspect to that of a CCTV camera. An example of the captured images is shown in Figure 1. The camera used was a BLACKFLY S BFS-U3-89S6 8.9 MP camera manufactured by Teledyne FLIR.

The final dataset contains 324 raw images taken from the overall data collection. The provided images have undergone annotation at full resolution using CVAT (Computer Vision Annotation Tool) [11] subsequent to processing from an ISP with default parameters. The total number of car, person, and bicycle objects present in the dataset is 10285, as seen in Table 1.

Class	Instances
Pedestrian	2273
Bicycle	783
Car	7229

Table 1: Data set Annotations break down

Object Detection Model Selection

This study investigates five object detection models, described in Table 2. The goal of this selection was to achieve a of the object detection field. All algorithms were pretrained on the COCO dataset for 300 epochs. The detection models were sourced from Ultralytics GitHub [12] and the PyTorch library [13], with default parameters. The models utilised in this study are Faster RCNN with a ResNet50 FPN backbone, FCOS with a ResNet50 FPN backbone, RetinaNet with a ResNet50 FPN backbone, YOLOv5x and YOLOv5m.

Model	Backbone	Stages	Params(M)	Year
FCOS	ResNet50 FPN	1	32.3	2019
Faster RCNN	ResNet50 FPN	2	41.8	2016
RetinaNet	ResNet50 FPN	1	34	2017
YOLOv5x	CSP	1	86.7	2020
YOLOv5m	CSP	1	21.1	2020

Table 2: Object Detection models

In 2014 Girshick et al. [14] created Regions with CNN features (RCNN). RCNN was the first object detection architecture to utilize CNNs. RCNN development was continued in Fast RCNN [15] and Faster RCNN [16], with each bringing iterative improvements compared to previous versions. Faster RCNN uses a Region Proposal Network (RPN), a CNN-based method for obtaining bounding box coordinates. Faster RCNN is a two-stage detector, the two stages being localisation and classification.

The YOLO architecture was created by Redmond et al. in 2016 [17]. YOLO was the first prominent architecture to use a single-stage design. A single CNN performs both localisation and classification, so the whole input image is used to predict bounding boxes and class labels in a single pass. Moving to a single stage with YOLO increased inference speed while delivering slightly lower performance than two-stage algorithms such as Faster RCNN [17]. Several versions of YOLO have been released since the initial version, with increasing performance and speed [18, 19, 20, 12, 21, 22], and for this study, we utilise YOLOv5 [12]. There are five different versions of YOLOv5 with different numbers of parameters, from 1.9 M to 86.7 M, made for different computing capabilities. In this study, two versions are used to investigate the effect of varying the image quality against the same architecture with different number of parameters.

Lin et al. [23], in 2018, investigated why single-stage detectors could not achieve the same level of performance as two-stage detectors and identified foreground-background class imbalance during training as the main reason. To mitigate the loss in performance, Focal Loss was created and utilised in their single-stage algorithm, RetinaNet to achieve state-of-the-art COCO AP performance. RetinaNet uses a Feature Pyramid Network (FPN) [24] with a ResNet architecture. This FPN and ResNet backbone combination is also utilised in this study with a Faster RCNN model and the Fully Convolutional One-Stage Object Detection (FCOS) [25] model. FPNs allow object recognition at multiple scales. FPNs make use of a top-down pathway as well as lateral connections to achieve multiscale high-resolution feature extraction that, when paired with ResNet, achieves state-of-the-art results.

Software ISP

In this study, an open source software ISP, called fast openISP [26, 2], is utilized. This ISP implementation contains a standard set of ISP blocks. With this open source software ISP, the full ISP implementation is available to view, making this study repeatable to test future object detection algorithms or datasets. Completing this analysis with an open-source software ISP instead of a hardware ISP allows for more control over the exact ISP parameters.

The default configuration for this ISP, provided with openISP [26, 2], was subjectively tuned for optimal human perception by an ISP expert for Canon 600D and Lumix S1 individually. Three ISP techniques were used in this experiment. Two of these, Edge Enhancement and Bilateral Noise filtering, are included in fast-openISP. A third block; Gaussian noise, was also used. This is typically not used in ISP systems optimised for human perception, but is used here as a technique to reduce contrast in the images. A grid search of parameters for the chosen ISP blocks was selected and is shown in Tables 3 and 4, ranging from smaller increments directly surrounding the default ISP configuration to

larger increments at the extremes of each parameter. The parameters are taken from Molloy *et al*. [9]. The blocks chosen from this ISP are EEH and BNF, the functionality and architecture of these blocks is described below. These two blocks were chosen as they are both commonly used ISP blocks and each can have a significant impact on the sharpness of an image.

Bilateral Noise Filtering

There are many sources of noise when capturing digital images, such as photon shot noise, dark noise, quantisation noise, and pixel response non-uniformity [27]. Noise filtering is carried out in the ISP to combat this problem [28, 29, 30]. Bilateral Noise Filtering (BNF) [28] is utilised in this study to investigate the impact of noise filtering on object detection performance. BNF removes noise by applying a Gaussian blur kernel, but only in areas with no edges. There are three parameters associated with BNF: the Gaussian filter kernel size, sigma value (corresponding to the kernel's standard deviation) and intensity sigma value corresponding to the amplitude of edges that are to be preserved. A set of BNF values has been determined, ranging from no noise filtering to significant noise filtering, that results in severe image blurring. The impact of this range of values is evaluated in terms of sharpness and contrast. The particular values are listed in Table 3.

Edge Enhancement

Edge enhancement (EEH) improves the perceived sharpness of the image by first identifying and then increasing the contrast directly on sharp edges. The EEH algorithm utilized in this study is unsharp mask [31]. Unsharp mask identifies high-frequency details, or edges, in the image by Gaussian blurring the image and then subtracting the blurred image, containing the low-frequency details from the original, leaving an edge map of high-frequency details. High-frequency details below a defined threshold, the flat threshold, are removed from the edge map to reduce image noise amplification. Once filtered by the flat threshold, the remaining edge map is multiplied by a gain factor. The edge map is clipped by the delta threshold and added back to the original image. The key parameters associated with unsharp mask are Gaussian filter kernel size; edge gain (which is the multiplication factor of the edge map); flat threshold (which defines a minimum threshold for the edge map, below which the edges are not amplified), preventing weaker edges and noise from being amplified; and the delta threshold (which is the minimum and maximum value by which an edge can be altered). Due to the large number of parameters associated with this block, several parameter sets have been used, varying the edge enhancement from no enhancement to a level associated with significant over-sharpening. The particular values are shown in Table 4.

Object Detection Metrics

AP50 is a commonly used metric for object detection performance. AP50 is given as the area under the precision and recall curve. This curve is generated by varying the confidence threshold for a set of predictions. At a lower confidence threshold, there will be fewer missed objects and fewer false negatives, but more false positives, leading to higher recall and lower precision. The inverse is true for higher confidence thresholds; with fewer false positives, it leads to a higher precision with a lower recall. The different precision and recall values are plotted and interpolated, and the area under the curve is AP50. The definition of a true positive in AP50 is an intersection over union (IoU) between the model's predicted bounding box and ground truth bounding box being greater than 50%. The primary metric in the COCO dataset, AP5095, varies the IoU threshold in increments of 5 % from 50 % to 95 % and then averages the result. This adds a positive bias for bounding boxes with a higher localisation accuracy. AP5095 is the primary metric used in this study as it is a good measure of overall object detection performance and encompasses true positives, false positives, false negatives, algorithm confidence, and the IoU threshold.

The score value reported by the object detection models for each prediction represents how much confidence the detector has in the prediction being correct. In this study, the average score of all predictions, with a score $>1\%$, is evaluated to determine whether the performance degradation that occurs due to ISP variations results in reduced confidence for those predictions.

Image quality Metrics

A key element of this study is to investigate the correlation between image quality and object detection performance. The P2020 working group seeks to compile a standard set of metrics suitable for use in evaluating image quality for automotive applications. In this study, two metrics proposed by the working group are used.

The first metric is the MTF. This is a well-established metric that is used to evaluate the ability of an optical system to capture sharpness. MTF is given as a function of spatial frequency. Four single-value metrics are commonly extracted from this function; these are MTF10; MTF50; Nyquist/2, and Nyquist/4. MTF10 is called the limiting resolution of the camera. It is measured by reading horizontally from the 0.1 cy/px mark on the Y-axis across the chart to the function; the spatial frequency value at this point is the MTF10 value. It is generally accepted that the eye is insensitive to detail at spatial frequencies where MTF is 10 % or less [32]. The limiting resolution describes the maximum spatial frequency detail that a camera can capture and faithfully reproduce detail. MTF50, like MTF10 is measured in terms of spatial frequency and is sometimes considered a better metric than MTF10 for some applications as it deals with the spatial frequency range with which the human eye is most concerned. Nyquist/2 and Nyquist/4 are measures of the SFR at half the Nyquist limit and a quarter the Nyquist limit, respectively. The Nyquist limit for any digital camera is 0.5 Line pairs/pixel. These metrics tend to be used less than the MTF10 and 50 values, but are still useful to measure Spatial Frequency Response (SFR) at a point MTF50 is used primarily in this study to evaluate the sharpness of the images, as shown in Figure 4

The second metric used is CTA. This is a measure of contrast proposed in [6]. CTA evaluates the ability of an optical system to capture different levels of contrast. Figure 2 shows a selection of CTA plots and, on the left, the graphs from which they are calculated. An array of pixels is extracted from an image with the pixels having a relatively uniform luminance level. The values are plotted as a probability histogram that should have the shape of a Gaussian curve. This process is repeated for a second patch, resulting in two bell-shaped curves that describe the luminance

levels. The green boxes in the upper left image of Figure 2 show how pixels are extracted from an image. The difference between the two curves is calculated as a probability distribution function of contrast. Bounds are taken from the centre of the function, typically 0.1, and the area under the curve and within the bounds is the CTA score for that level of contrast. The score ranges from 0 to 1. The top right plot in Figure 2 is a typical CTA heat map, showing scores for many different levels of contrast and luminance. The luminance is the average luminance of the two patches. The CTA score is used as a single-value metric to describe the plots. Figure 3 shows how a CTA score can be achieved for a particular level of contrast and luminance. A contrast value of 0.5 and an average Luminance Value of $Log_{10}4.4cd/m^2$ were used to extract the CTA score from each plot. These values produce a single-value metric, which shows sufficient variation to plot against AP5095. It is shown in Figure 3 how CTA score can be extracted from a heatmap.

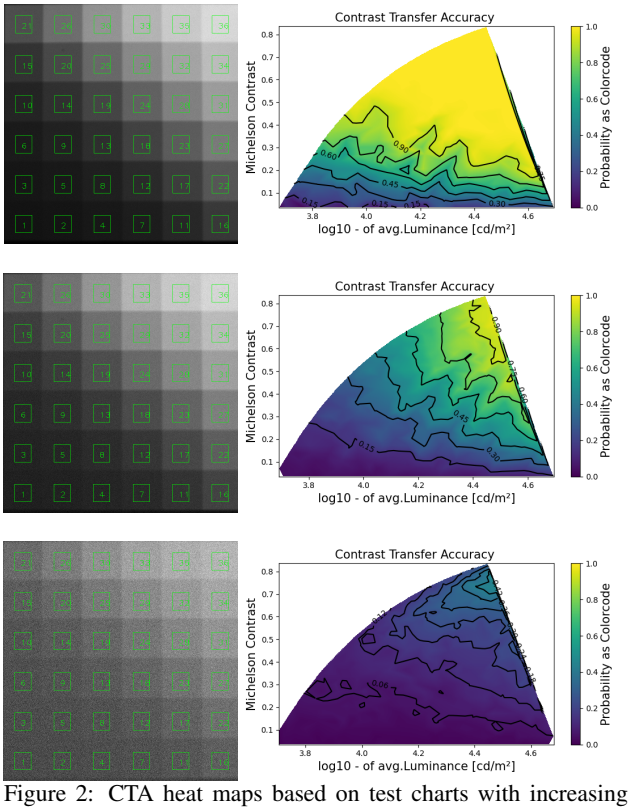

Figure 2: CTA heat maps based on test charts with increasing levels of noise

Image Quality Analysis

The analysis pipeline in this study is adapted from that described by Molloy *et al.* [9], except in this implementation a step to measure image quality is also included. In this study, image quality is measured using specific test charts for sharpness and contrast. MTF is calculated using the slanted edge method, taken from the Imatest SFR target. For contrast, an Image Engineering 36-step slide backlit with a Vega module is used. The backlit slide is shown at the top-left of Figure 2. The raw dataset is processed with fast OpenISP. In Klein *et al.* [8] the impact of ISP on CTA

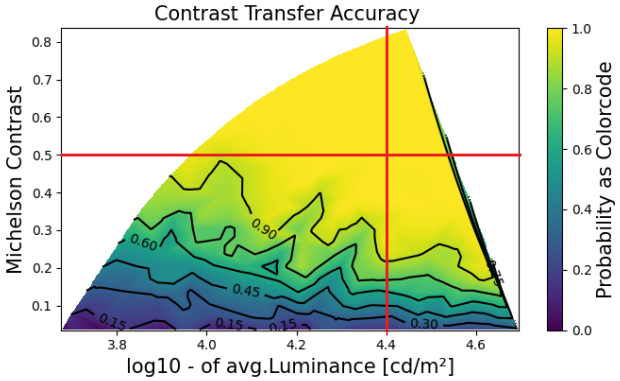

Figure 3: A CTA Score can be extracted for a particular level of luminance and contrast.

was not evaluated. The luminance values recorded are taken as average values of the entire frame. In this paper, the image quality is varied using a selection of ISP techniques, and the impact of these techniques is systematically evaluated to determine the robustness of CTA. Figure 4 shows how varying two ISP parameters; BNF and EEH impact MTF50. 0 on the X-axis is the default ISP setting, tuned for human vision. Each whole number on the positive and negative aspects of the x-axis indicates a step away from the default setting. The exact ISP steps used are listed in Tables 3 and 4. The values range range from smaller increments directly surrounding the default ISP configuration to larger increments at the extremes of each parameter. The Default ISP settings are applied at position 0. It is evident from Figure 4, that when a level of EEH is applied to the test images which is above the default level, the sensor shows improved sharpness performance in terms of MTF50. When BNF is applied, sharpness decreases, as indicated by the green branch of the chart.

position	intensity sigma	spatial sigma	BNF kernel size
-9	72	72	25
-8	36	36	21
-7	16	16	13
-6	8	8	9
-5	6	6	7
-4	4	4	5
-3	1.2	1	5
-2	0.5	0.4	5
-1	0.35	0.3	5
0	0.8	0.8	5

Table 3: ISP Parameters for each step of BNF

In the CTA test plan, three parameters are used to vary the contrast in the images. EEH and BNF are used with the same steps as listed in Tables 3 and 4. In addition to these, Gaussian noise was added to the images, as a method of reducing contrast. Figure 2 shows three variants of the 36-step chart, with varying levels of Gaussian noise applied. Shown are also the resulting CTA plots. As more noise is applied, the CTA decreases. The yellow areas, which indicate a high CTA score become darker as the patches become less distinguishable from each other. The impact of EEH, BNF and noise are shown in Figure 5. The level of each block increases in steps outwards from the default ISP

position	edge gain	flat Threshold	delta Threshold	kernel size	sigma
0	384	4	64	5	3
1	0	16	32	5	3
2	256	16	32	5	3
3	384	12	64	5	3
4	512	10	64	7	3
5	768	8	64	7	3
6	1024	8	64	9	3
7	1408	6	128	13	3
8	1792	4	128	17	3
9	2048	2	128	21	3

Table 4: ISP Parameters for each step of EEH

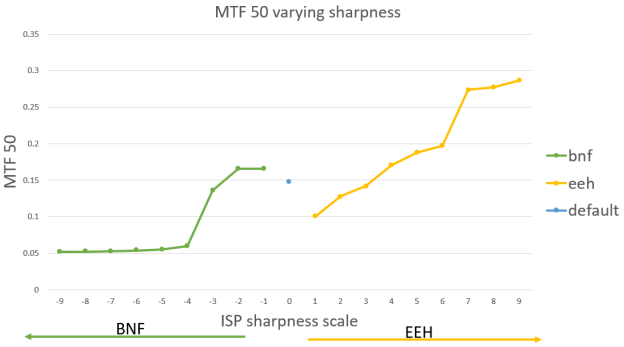

Figure 4: MTF50 increasing across the 'sharpness' scale

configuration which is located at 0. Note EEH and noise have similar trends while applying BNF, results in much higher values for CTA.

Object Detection Performance

The same ISP settings are applied to the object detection dataset as for the sensor characterisation images. Object detection performance is described with the AP5095 metric. To gather these performance statistics, first, a set of ISP parameters is taken from the grid search, and the openISP configuration is updated. The raw dataset is then processed via openISP with the updated ISP configuration to produce post-ISP RGB images as seen in Figure 1. Each object detection model infers over each image and returns a set of predictions that are converted to the COCO prediction format. Each prediction set contains a single model's predictions of all images in the dataset for one ISP configuration. These predictions are compared against our ground truth using the official COCO API. The average confidence score of each prediction in the prediction data is also calculated for further analysis.

Results and Discussion

AP5095 is used to review the overall object detection performance due to variations in sharpness, as shown in Figure 6. The different models have a relatively similar performance, all between approximately 35% AP5095 to 45%. The best-performing algorithms among those tested are YOLOv5x, FCOS with the ResNet50 FPN backbone, and Faster RCNN with the ResNet50 FPN backbone. This chart shows that across 4 of the 5 models, there is an increase in performance as MTF50 improves. The default ISP setting correlates to an MTF50 score of 0.15 as shown in Figure 4. The trend among the 4 similar plots is that AP5095

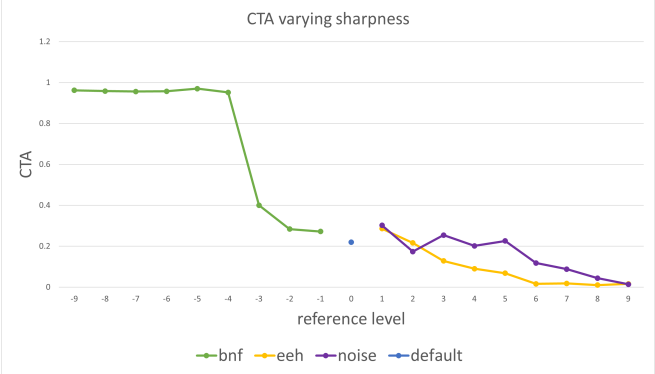

Figure 5: Variation in CTA due to different image processing blocks

improves with EEH and worsens with BNF. A notable exception to this trend is YOLOv5x. The model has only a small change in performance, approximately 1% across the entire sharpness scale. This might indicate that the model is more robust to changes in the ISP. This model shows a degradation in performance as the ISP steps away from the default configuration. Of the 5 models used, YOLOv5x has the most parameters. YOLOv5x has 86.7 m parameters compared to Faster RCNN which has the second most parameters with 41.8. It may be the case that YOLOv5x is more susceptible to degradations in certain features, due to the additional layers.

Figure 7 shows the performance of each model as a function of CTA Score. As seen in the previous set of results, images processed with EEH have a higher AP5095, while images processed with BNF tend to have worse performance. A point of interest in these results is that higher CTA scores do not correlate with better object detection. Intuitively, one might expect that an image with a high level of contrast might have better object detection than one without. This result can be somewhat attributed to the specific ISP parameters which were applied and the way in which CTA is calculated. When EEH is applied to an image, the edges are highlighted due to increased contrast at areas of high spatial frequency. EEH also adds noise. Some features in the image are interpreted as edges and the pixel intensity values are altered. In the 36-step chart, this noise reduces the contrast between patches, and hence the CTA score is lower. This metric does not take into account the part of the image which is improved, i.e. the edges. The Edges, however, are a critical feature for object detection, and so the object detection improves while the CTA worsens. This logic can also be applied to the BNF arm of the graph, in that BNF smoothes each patch, resulting in a more precise luminance distribution, but the edges suffer. Images with higher CTA scores have a slightly lower AP5095 score.

The results of the images with Gaussian noise applied show both a decrease in the CTA score and the OD performance. Both the planes and the edges in the image degrade.

Conclusion and Future Work

In this study, the performance of a range of object detection models due to varying image quality have been characterised. These results show how a selection of CNN-based object detection models are impacted by variations in image quality due to ISP tuning. The results suggest that altering image sharpness using

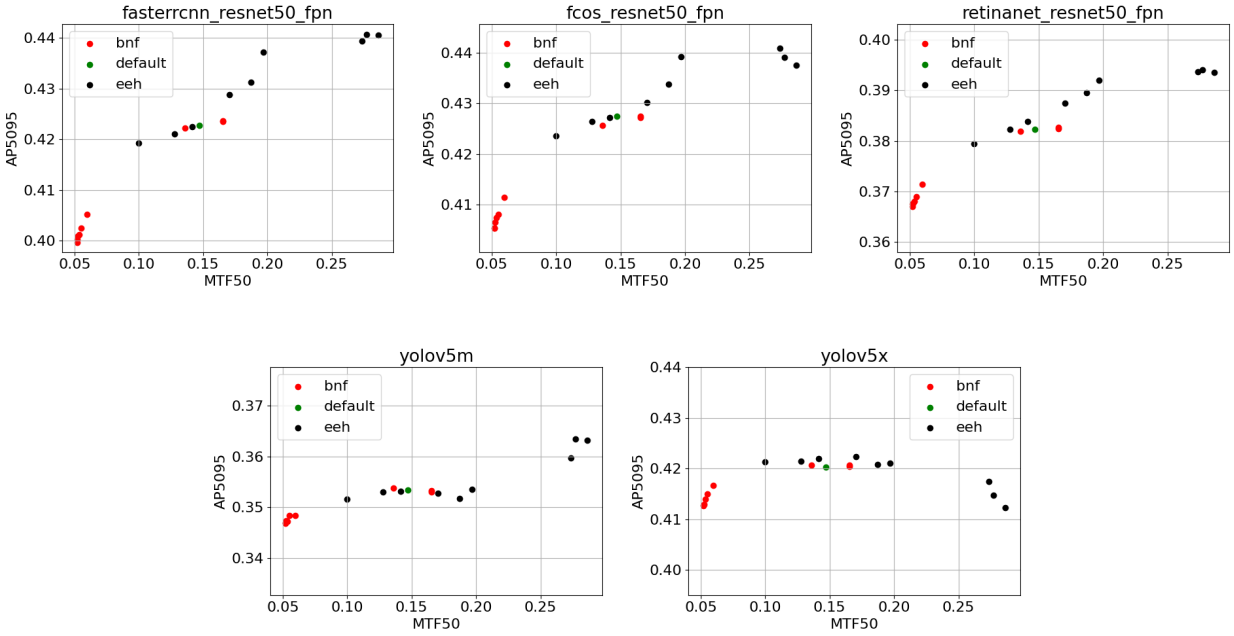

Figure 6: Object Detection Performance vs MTF50

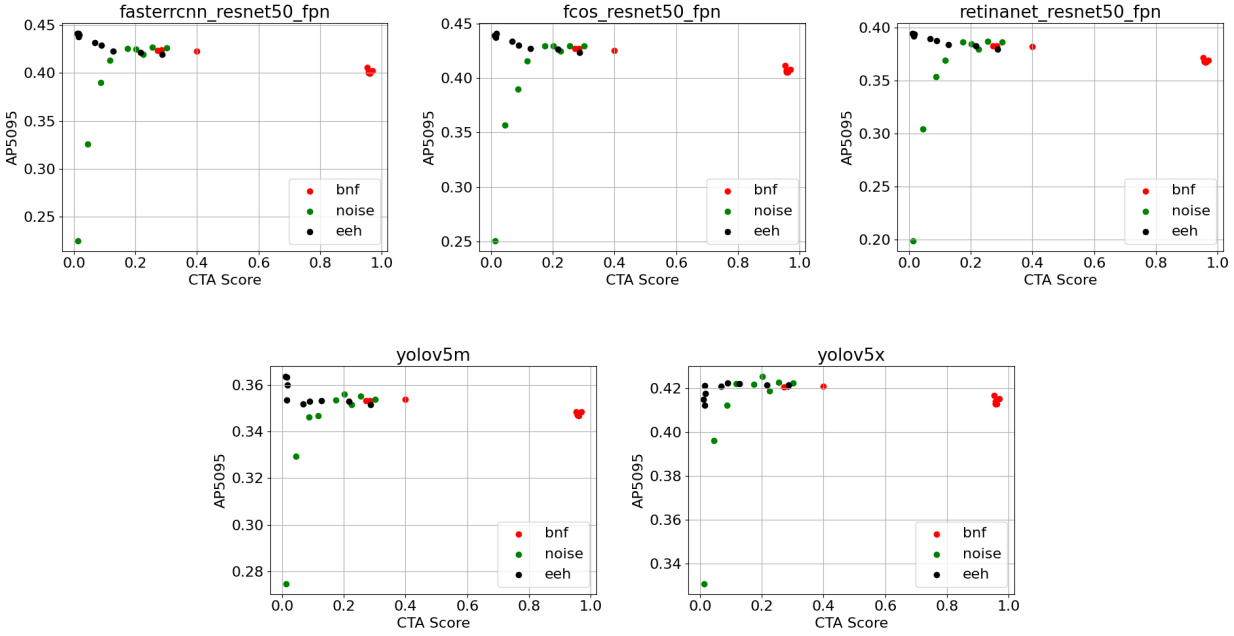

Figure 7: Object Detection Performance vs CTA Score

ISP tuning can improve object detection performance, and there is a correlation between MTF and object detection performance.

CTA is also identified as a gap item in the P2020 white paper, as part of research on the tonal response of a camera sensor. CDP/CTA is a relatively new metric that was proposed in 2018 in Geese *et al.*[7]. CTA was developed with the aim of predicting sensor performance for safety-critical use cases. In this project, the P2020 CDP Evaluator was used to calculate CTA. The images were processed to degrade their contrast. It was found that noisy images make CTA worse, while denoising can improve CTA. It was also shown that a caveat of these results is that CTA does not show the full picture. This was shown when EEH was applied to the images. EEH adds a certain level of noise to an image, but also, as per its description, improves contrast at the edges. As was shown in the sharpness results, applying EEH improved object detection performance.

Additional ISP blocks should be investigated, such as anti-aliasing, additional denoising algorithms, and HDR tone map-

ping. Future work should also include a multivariate analysis, in which multiple ISP blocks are altered, to characterise the impact of the object detection performance due to coupled ISP blocks.

Other data sets could be substituted in the toolchain, with different characteristics, e.g. from a vehicle perspective or with different types of vehicles such as trucks. Other sensor characteristics should be investigated in a similar way. CTA is a metric that has been generated as part of the P2020 research. There will likely be more metrics. The robustness of these metrics will need to be tested. In the P2020 white paper [5] a gap analysis is performed to evaluate the deficiencies in the existing KPIs when applied to the autonomous driving use case. It should be investigated how independent MTF is of the data on which it is calculated. A potential caveat of this study which has not been examined is how do ISP techniques manifest in different images. In theory, the MTF calculated on the test target is an MTF of the optical system; however, this is based only on one edge of one image. A toolkit exists for extracting MTF from natural scenes [33]. It should be investigated how MTF extracted in this way compares to that based on the test target. The same can also be said for the CTA charts. The key question is: how much are image quality metrics biased by the images they are captured from? Some of the other areas mentioned in the white paper are LED flicker, low light performance, distinguishability of traffic-relevant colours, lens distortion. These areas and others require research and investigation to be carried out in the coming years, as the P2020 standard is developed and updated.

References

[1] IEEE P2020/D3, June 2022 , 1 (2022).

[2] QiuJueqin, fast-openisp, `https://github.com/QiuJueqin/fast-openISP`, 2021.

[3] P. Burns, sfrmat4, `http://burnsdigitalimaging.com/software/sfrmat/`, 2022.

[4] Continental, CDP Evaluator, 2020.

[5] IEEE P2020 Automotive Imaging White Paper, Technical report, 2018.

[6] M. Geese, U. Seger, and A. Paolillo, Electronic Imaging **2018**, 148 (2018).

[7] U. Artmann, M. Geese, and M. Gäde, Electronic Imaging **31**, 1 (2019).

[8] V. Klein et al., Electronic Imaging **35**, 1 (2023).

[9] D. Molloy et al., Journal of Imaging **9**, 260 (2023).

[10] D. Molloy, ISP Object Detection Benchmark, https://doi.org/10.5281/zenodo.7802651.

[11] B. Sekachev et al., OpenCV/CVAT: v1.1.0, https://doi.org/10.5281/zenodo.4009388, 2020.

[12] G. Jocher et al., Ultralytics/YOLOv5: v6.1 - TensorRT, TensorFlow Edge TPU and OpenVINO Export and Inference, https://doi.org/10.5281/zenodo.6222936, 2022.

[13] TorchVision: PyTorch's Computer Vision library, https://github.com/pytorch/vision, 2016.

[14] R. Girshick, J. Donahue, T. Darrell, and J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, 2014, arXiv:1311.2524 [cs].

[15] R. Girshick, Fast R-CNN, 2015, arXiv:1504.08083 [cs].

[16] S. Ren, K. He, R. Girshick, and J. Sun, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, 2016, arXiv:1506.01497 [cs].

[17] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, You Only Look Once: Unified, Real-Time Object Detection, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 779–788, Las Vegas, NV, USA, 2016, IEEE.

[18] J. Redmon and A. Farhadi, YOLO9000: Better, Faster, Stronger, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6517–6525, Honolulu, HI, 2017, IEEE.

[19] J. Redmon and A. Farhadi, Yolov3: An incremental improvement, 2018, cite arxiv:1804.02767.

[20] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, YOLOv4: Optimal Speed and Accuracy of Object Detection, 2020, arXiv:2004.10934 [cs, eess].

[21] 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, 2022.

[22] G. Jocher, A. Chaurasia, and J. Qiu, Ultralytics yolov8, https://github.com/ultralytics/ultralytics, 2023.

[23] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, Focal Loss for Dense Object Detection, 2018, arXiv:1708.02002 [cs].

[24] T.-Y. Lin et al., Feature Pyramid Networks for Object Detection, 2017, arXiv:1612.03144 [cs].

[25] Z. Tian, C. Shen, H. Chen, and T. He, FCOS: Fully Convolutional One-Stage Object Detection, in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9626–9635, Seoul, Korea (South), 2019, IEEE.

[26] cruxopen, OpenISP: Image signal processor, https://github.com/cruxopen/openISP, 2019.

[27] E. M. V. Association, Emva standard 1288, https://doi.org/10.5281/zenodo.3951558, 2005.

[28] C. Tomasi and R. Manduchi, Bilateral filtering for gray and color images, in *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, pages 839–846, 1998.

[29] R. Chan, C.-W. Ho, and M. Nikolova, IEEE Transactions on Image Processing **14**, 1479 (2005).

[30] S. Chang, B. Yu, and M. Vetterli, IEEE Transactions on Image Processing **9**, 1532 (2000).

[31] A. Polesel, G. Ramponi, and V. Mathews, IEEE Transactions on Image Processing **9**, 505 (2000).

[32] D. Stump, Mtf, resolution, contrast, and nyquist theory, in *Digital Cinematography*, pages 111–126, Routledge, 2014.

[33] O. van Zwanenberg, S. Triantaphillidou, R. B. Jenkin, and A. Psarrou, Electronic Imaging **34** (2021).

Author Biography

Diarmaid Geever *received a B.E. in Electronic and Computer Engineering from the University of Galway (2023). He is currently pursuing a PhD from the University of Galway, with the Connaught Automotive Research (CAR) Group. His areas of research are machine vision and image quality for autonomous driving applications.*

Tim Brophy *received the B.E. in Computer and Electronic Engineering from the University of Galway, in 2018. He is currently pursuing a Ph.D. degree at the University of Galway. Tim is currently working as a member of the Connaught Automotive Research (CAR) group under the supervision of Prof. Edward Jones and Prof. Martin Glavin. His research interests include computer vision and sensor availability within an autonomous vehicle context*

Dara Molloy *received the B.E. (Hons.) degree from the University of Galway, in 2018. He is currently pursuing a Ph.D. degree at the University of Galway. Dara is currently working as a member of the Connaught*

Automotive Research (CAR) group under the supervision of Prof. Martin Glavin and Prof. Edward Jones. His research interests include computer vision and sensor availability within an autonomous vehicle context.

Martin Glavin *received the B.E. degree in electronic engineering and the Ph.D. degree in the area of algorithms and architectures for high-speed data communications systems from the University of Galway, Ireland, in 1997 and 2004, respectively, and the Higher Diploma in Third Level Education in 2007. He was a Lecturer (fixed term contract) from September 1999 to December 2003 and became a permanent member of the academic staff in January 2004. He is currently the Joint Director of the Connaught Automotive Research (CAR) Group, University of Galway. He is also a Funded Investigator in Lero, the Irish Software Research Centre. He currently has a number of Ph.D. students and Post-Doctoral Researchers working in collaboration with industry in the areas of signal processing and embedded systems for automotive and agricultural applications.*

Edward Jones *received the B.E. and Ph.D. degrees in electronic engineering from the University of Galway, Ireland. He is currently a Professor of Electrical & Electronic Engineering in the School of Engineering at the University of Galway. From 2009 to 2010, he was a Visiting Researcher at the Department of Electrical Engineering, Columbia University, New York, NY, USA, and has also been appointed a Visiting Fellow at the School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, Australia. He also has a number of years of industrial experience in senior positions, in both start-ups and multinational companies, including Toucan Technology Ltd., PMC-Sierra Inc., Innovada Ltd., and Duolog Technologies Ltd. He has also been a member of international standardization groups ANSI T1E1.4 and ETSI TM6. His current research interests are in DSP algorithm development and embedded implementation for applications in connected and autonomous vehicles, biomedical engineering, and speech and audio processing. He is a Chartered Engineer and a Fellow of the Institution of Engineers of Ireland.*

Brian Deegan *received a Bachelor's Degree in Computer Engineering from the University of Limerick in 2004, an MSc In Biomedical Engineering from the University of Limerick in 2005 and a Ph.D. in Biomedical Engineering from the University of Galway in 2011. The focus of his research was the relationship between blood pressure and cerebral blood flow in humans. From 2011 to 2022 Brian worked in Valeo Vision Systems as a Vision Research Engineer focusing on Image Quality. Brian's research focus is on high dynamic range imaging, LED flicker, Topview harmonization algorithms, and the relationship between image quality and machine vision. In 2022 Brian joined the Department of Electrical & Electronic Engineering at the University of Galway as a Lecturer and Researcher.*