https://doi.org/10.2352/EI.2026.38.16.AVM-118

Same Scene, Different Pipeline: ISP Impact on Automotive Detection at Range

Tejus Vijayakumar[1,4,5], Tim Brophy[2,3,4,5], Brian Deegan[3,4], Ciarán Eising[2,4,5], and Patrick Denny[1,3,4,5]

[1]Dept. of Computer Science & Information Systems (CSIS), University of Limerick, Ireland
[2]Dept. of Electronic & Computer Engineering, University of Limerick, Ireland
[3]Connaught Automotive Research (CAR) Group, Dept. of Electrical & Electronic Engineering, University of Galway, Ireland
[4]Lero, the Research Ireland Centre for Software, Ireland
[5]D2iCE (Data Driven Computer Engineering) Research Centre, University of Limerick, Ireland

Abstract

Image signal processors in automotive cameras are typically tuned for human visual perception, yet these same cameras increasingly serve as the primary input to safety-critical object detection systems. In this study, we evaluate the sensitivity of object detection to ISP parameter variation under nighttime conditions. We process raw Bayer data from the nighttime subset of the G-MIND dataset through 21 ISP configurations spanning gain, gamma correction, saturation, bilateral noise filtering, and edge enhancement, and additionally test raw Bayer input with and without gamma correction. For each configuration, we fine-tune four detector architectures representing three design families (single-stage CNN, two-stage CNN, and transformer-based), yielding 92 models evaluated using mAP_{50-95} per class across five distance bins from 0 to 75 metres. Gamma and gain have negligible effects when models are retrained. Saturation is the most critical parameter: YOLOv8m loses 26.2 mAP points across the saturation range while Faster R-CNN loses only 2.2. Raw Bayer input performs on par with the default ISP for single-stage detectors while eliminating all ISP processing cost, suggesting that a full human-tuned ISP is not optimal for nighttime machine perception. No ISP variant reverses detection degradation with distance. These findings demonstrate that ISP sensitivity is architecture-dependent, that a full human-tuned ISP is not optimal for nighttime machine perception, and that there is scope to develop leaner, perception-aware ISP pipelines tailored to the downstream detector.

Introduction

Camera-based perception is fundamental to advanced driver assistance systems and automated driving. As illustrated in Figure 1, the imaging pipeline from scene to detection involves multiple stages, each of which can influence downstream perception performance. Among the many challenges, nighttime operation remains one of the most critical. Nighttime conditions account for a disproportionate share of pedestrian and cyclist fatalities despite significantly lower traffic volumes. In the United States, 78.8% of pedestrian deaths caused by motor vehicles in 2017 occurred during dark, dawn, or dusk conditions, with just 21.2% occurring in daylight. [3]. The ability to detect vulnerable road users at a sufficient range to allow safe braking is a core safety requirement, and camera systems are expected to contribute meaningfully to this task. Yet nighttime represents one of the most challenging conditions for camera-based perception.

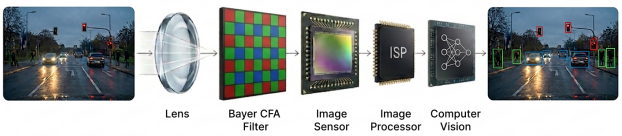


Figure 1. *Illustration of an imaging pipeline from scene to object detection.*

At night, the camera sensors become photon-limited, and the signal-to-noise ratio degrades sharply. This degradation is most severe for small, distant objects that occupy only a few pixels on the sensor. The image that ultimately reaches the object detection model is not simply a function of the scene and the optics, but also of how the raw sensor data is processed through the camera's ISP. As shown in Figure 2, the ISP transforms raw Bayer mosaic data into a usable RGB image through a chain of processing blocks including demosaicing, black level correction, gain control, gamma correction, noise filtering, edge enhancement, and colour processing [26] [1]. Each of these stages alters the pixel-level representation that a detector must interpret, and under photon-limited conditions, the choices made within the ISP may significantly affect the discrimination between object and background under severely degraded signal-to-noise ratios.

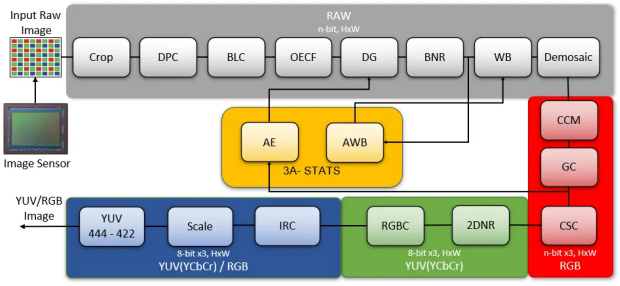


Figure 2. *Typical modules in an ISP pipeline (source: [1]).*

Despite this, ISP pipelines are typically tuned subjectively by imaging experts for optimal human visual perception [21]. Object detection models trained on post-ISP images therefore learn features that are implicitly coupled with subjectively chosen ISP settings. This raises a practical question: are the ISP configurations that produce visually pleasing nighttime images also the ones that maximise detection performance, particularly at longer ranges where the task is most safety-critical?

In recent years, there have been several studies examining the relationship between ISP configuration and object detection. Buckler et al. [2] were among the first to investigate how ISP blocks affect CNN-based vision tasks, finding that demosaicing and gamma compression are the most critical stages. Yahiaoui et al. [31] conducted the first automotive-focused ISP tuning study and reported a 14% improvement in pedestrian detection accuracy through sharpening parameter optimisation. Mosleh et al. [24] and Robidoux et al. [28] demonstrated hardware-in-the-loop ISP optimisation for object detection, achieving 30% and 33% improvements in mAP, respectively, by jointly optimising ISP parameters with detection objectives. Molloy et al. [21] presented the most comprehensive characterisation to date, evaluating 14 object detection models across eight ISP blocks on a custom raw dataset and proposing ISP variation as a data augmentation strategy. Hansen et al. [7] showed that ISP processing can improve classification accuracy by up to 12% on MobileNet architectures, with tone mapping identified as a particularly important stage.

A related question is whether image quality metrics can predict detection performance, which would enable system designers to reason about ISP configurations without exhaustive model retraining. The IEEE P2020 working group [10] [9] is developing standardised image quality metrics for automotive applications, but a clear metric that reliably correlates with machine vision performance has yet to emerge. Geever et al. [4] investigated one of the P2020 metrics, Contrast Transfer Accuracy (CTA), along with Modulation Transfer Function (MTF) for the object detection task and found that, although MTF shows some correlation, the relationship is not clear. CTA was found to be particularly limited as a predictor, as scenes with similar CTA scores produced widely varying detection performance. More recently, Geever et al. [5] studied Shannon Information Capacity (SIC) [14] as an alternative predictor of machine vision performance. Their results showed that SIC exhibits a substantially stronger correlation with detection accuracy than MTF50, with a Spearman coefficient of 0.98, outperforming MTF by 35% for the same model. This suggests that information-theoretic metrics that jointly account for spatial frequency response and noise may be more appropriate for characterising camera systems intended for machine perception. These findings are relevant to our study because they indicate that traditional image quality reasoning, which centers on sharpness and contrast as perceived by humans, may not translate directly to machine perception, particularly under challenging conditions such as nighttime imaging.

However, the existing literature on ISP and detection has mostly been conducted under daytime or well-lit conditions. The interaction between ISP tuning and detection performance under photon-limited nighttime conditions remains unexplored.

In this paper, we present a systematic evaluation of the sensitivity of the ISP parameters for the detection of automotive objects at night. We process raw Bayer sensor data from the nighttime subset of the G-MIND dataset [22] through 23 ISP configurations spanning six parameter groups: gain, gamma correction, saturation, bilateral noise filtering, edge enhancement, as well as raw Bayer input with and without gamma correction. For each of the 23 ISP variants, we independently fine-tune four object detection architectures representing three architectural families: YOLOv8m and YOLO26m as single-stage CNN detectors, Faster R-CNN with ResNet-50 FPN as a two-stage CNN detector, and RT-DETR-L as a transformer-based detector. This results in 92 trained models in total. We evaluate all models using mAP_{50-95} per object class (person, bicycle, car) across five distance bins from 0 to 75 metres, computed using known camera geometry. Our aim is twofold: to identify which ISP parameters and configurations most significantly affect detection at range under nighttime conditions, and to determine whether this sensitivity is architecture-dependent.

Experimental Setup

Dataset

This study uses the nighttime subset of the G-MIND dataset (Galway Multimodal Infrastructure Node Dataset) [22]. The data used for this study were collected using a FLIR BlackFly-S 8.9 megapixel camera mounted on a fixed infrastructure stand at a height of four metres from the ground with a 20-degree downward tilt, at the University of Galway, Ireland. The camera captured controlled scenarios involving pedestrians, cyclists, and cars traversing a car park in night conditions. Critically, the raw 8-bit Bayer sensor data was preserved during acquisition, allowing ISP processing to be applied retrospectively under full experimental control. Sample images from the dataset are shown in Figure 3.

Figure 3. *Sample images from the G-MIND Dataset*

Three object classes are evaluated: person, bicycle, and car. The distance of the object from the camera was calculated using the known camera geometry, specifically the mounting height, tilt angle, and calibrated intrinsic parameters, which allow the ground-plane distance to be estimated for each annotated bounding box. Based on this, the objects were grouped into five distance bins: 0 to 15 m, 15 to 30 m, 30 to 45 m, 45 to 60 m, and 60 to 75 m. The number of ground truth instances varies across bins, with the majority of annotations falling in the 0 to 30 m range and substantially fewer instances at longer distances. The dataset split and class distribution are summarised in Table 1.

Table 1. Dataset split and class distribution.

Class	Training (3,456 frames)	Test (864 frames)
Person	7,910	1,041
Bicycle	1,295	181
Car	45,183	12,234
Total	**54,388**	**13,456**

ISP Pipeline and Variants

All raw Bayer frames were processed using fast-openISP [26], an open-source software ISP that implements a standard set of processing blocks. The default configuration provided with fast-openISP is intended as a general-purpose starting point and is not tuned for any specific camera or scene condition. We instead adopt the default ISP configuration from Molloy et al. [21], which was subjectively tuned by imaging experts for optimal human visual perception on the same G-MIND nighttime data used in this study. This configuration serves as our baseline.

We generated 23 distinct ISP configurations by varying individual processing blocks while holding the remainder at their default values. The parameter ranges for each block were informed by the experimental designs of Molloy et al. [21] and Geever et al. [4]. To select specific parameter steps, we generated sample images across a sweep of values for each module and identified the points at which a visually distinct change from the default output was observable. This was done to ensure that each configuration step produces a meaningful shift in image appearance, allowing us to capture the overall trend in detection performance as each ISP module is progressively varied from its default setting.

The configurations are organised into five groups, with the specific parameter values for each listed in Tables 2 through 6. Gamma correction, defined in Equation 1, applies a non-linear transfer function between raw sensor values and output pixel intensity, and is the standard tool for adjusting the tonal response of images under low-light conditions. Digital gain uses the same formulation (Equation 1 with $\gamma = 0.45$), applying a linear multiplier to amplify pixel intensity prior to further processing; higher values brighten the image but also amplify noise. Colour saturation, defined in Equation 2, controls the vividness of colours in the output image, with a value of 0 producing a desaturated grayscale result. Bilateral noise filtering (BNF) [29], used in this study, suppresses sensor noise by weighting nearby pixels based on both spatial distance and intensity similarity, smoothing flat regions while preserving edges. The degree of filtering is controlled by intensity sigma, spatial sigma, and kernel size. The Edge enhancement (EEH) algorithm used in this study is unsharp mask [25], which sharpens the image by extracting high-frequency detail via Gaussian subtraction and selectively amplifying edges using adaptive thresholding, where a flat threshold suppresses noise amplification and an edge threshold controls the gain applied to strong edges.

$$V_{out}(c) = A \cdot V_{in}(c)^{\gamma}, \quad c \in \{R,G,B\} \tag{1}$$

where A is a gain multiplier applied to the input, γ is the gamma parameter, and c denotes the colour channel.

$$(Cb', Cr') = k \cdot ((Cb, Cr) - 128) + 128 \tag{2}$$

where k is the saturation ratio, and Cb, Cr and Cb', Cr' are the input and output chroma channels respectively.

In addition to these ISP-processed variants, two configurations bypass the ISP pipeline entirely. The first, referred to as Bayer, feeds the raw Bayer mosaic directly to the detector with no processing applied. The second, referred to as Bayer + GC, applies only gamma correction (Equation 3) uniformly across all pixel values on the raw Bayer mosaic before passing it to the detector. Since all four detector architectures were pre-trained on three-channel RGB data, the single-channel Bayer mosaic was replicated across three channels to match the expected input format. These two conditions were included to test whether full ISP processing is necessary for nighttime detection, or whether detectors can learn effective representations directly from minimally processed sensor data.

$$V_{out}(i,j) = A \cdot V_{in}(i,j)^{\gamma} \tag{3}$$

where A is a gain multiplier applied to the input, γ is the gamma parameter, and (i,j) is the pixel location on the single-channel Bayer image.

For each of the 23 ISP configurations, a separate training dataset was generated from the same set of raw source frames. This means every model was trained and evaluated on images that depict identical scenes but differ only in how the raw data was processed. This controlled approach isolates the effect of ISP configuration from all other variables.

Table 2. Gamma correction configurations. Values below 1 brighten the image overall, especially dark/mid regions; values above 1 darken the overall image.

Step	Gamma
−2	0.1
−1	0.25
Default	0.45
+1	1
+2	1.5

Table 3. Digital gain configurations. Values represent a fixed-point multiplier where 256 corresponds to 1.0× (no change). A value of 128 halves pixel intensity; 512 doubles it, amplifying the linear sensor signal before gamma correction.

Step	Digital Gain
−2	8
−1	32
Default	256
+1	512
+2	1024

Table 4. Colour saturation configurations. Values represent a fixed-point multiplier where 256 corresponds to 1.0× (no change), 0 produces an unsaturated image; 512 doubles the saturation.

Step	Colour Saturation
−1	0
Default	256
+1	512
+2	1024
+3	2048

Detector Architectures

Four object detection architectures were selected to represent three distinct architectural families commonly deployed in

Table 5. Bilateral Noise Filtering (BNF) configurations. Higher values produce stronger blurring.

Step	Intensity Sigma	Spatial Sigma	Kernel Size
−1	0.35	0.3	5
Default	0.8	0.8	5
+1	6	6	7
+2	16	16	13
+3	72	72	25

Table 6. Edge Enhancement (EEH) configurations. Higher edge gain produces more aggressive sharpening.

Step	Edge Gain	Flat Threshold	Delta Threshold	Kernel Size
Default	384	12	64	5
+1	768	8	64	7
+2	1408	6	128	13
+3	2048	2	128	21

automotive perception systems. YOLOv8m [11] is a single-stage CNN detector with an anchor-free, decoupled detection head. It is widely used as an established baseline for object detection in automotive research. YOLO26m [13] is a more recent single-stage CNN detector designed specifically for edge and low-power deployment, incorporating attention mechanisms, and operating without non-maximum suppression. Including both YOLO variants allows us to examine whether architectural updates within the same detector family alter ISP sensitivity.

Faster R-CNN [27] with a ResNet-50 backbone and Feature Pyramid Network (FPN) [16] is a two-stage CNN detector that generates region proposals before classification. The FPN enables multi-scale feature extraction, which is particularly relevant for detecting objects at farther distances. Two-stage detectors are generally considered more accurate than single-stage alternatives, though at a higher computational cost.

RT-DETR-L [32] is a transformer-based, end-to-end set prediction detector that uses global self-attention rather than convolutional feature extraction. Including a transformer architecture allows us to assess whether the fundamentally different feature encoding mechanism of transformers leads to different ISP sensitivity characteristics compared to CNN-based detectors.

All four architectures were pre-trained on the COCO dataset [17] and subsequently fine-tuned on each of the 23 ISP variant datasets independently. This resulted in 92 trained models in total (23 ISP variants × 4 architectures).

All models were fine-tuned using their respective default training frameworks: Ultralytics [12] for YOLOv8m, YOLO26m, and RT-DETR-L, and TorchVision [20] for Faster R-CNN. YOLOv8m and YOLO26m were trained for 75 epochs with a batch size of 64 and a learning rate of 0.001. Faster R-CNN was trained for 30 epochs with a batch size of 16 and a learning rate of 0.005. RT-DETR-L was trained for 100 epochs with a batch size of 32 and a learning rate of 0.0001. The number of training epochs for each architecture was chosen to ensure convergence given the dataset size, and the best-performing checkpoint on the test set was selected for final evaluation. A summary of the detector architectures is provided in Table 7.

Table 7. Object detection architectures evaluated in this study.

Model	Backbone	Stages	Params (M)	Year
YOLOv8m	CSP	1	25.9	2023
YOLO26m	CSP	1	20.4	2026
Faster R-CNN	ResNet50 FPN	2	41.8	2016
RT-DETR-L	HGNetV2	1	32.0	2023

Evaluation Metrics

Detection performance is measured using mean average precision at IoU thresholds from 0.50 to 0.95 in increments of 0.05 (mAP_{50-95}), following the standard COCO evaluation protocol [17]. This metric was chosen over the simpler mAP_{50} because mAP_{50-95} penalises poor localisation accuracy in addition to missed detections, providing a more complete picture of detection quality. Performance is reported per object class (person, bicycle, car) and per distance bin (0 to 15 m, 15 to 30 m, 30 to 45 m, 45 to 60 m, 60 to 75 m), enabling analysis of how ISP sensitivity interacts with both object type and range.

To support reproducibility, the code used to generate ISP variants, train models, and evaluate results is publicly available.[1]

Computational Setup

All ISP processing was performed on an Intel Xeon Gold 6330 CPU, with each ISP variant taking ∼3.5 hours to process the subset of G-MIND dataset used in this work. Model training was conducted on an NVIDIA A100 40GB GPU. Training times per ISP variant were ∼50 minutes for YOLOv8m, ∼70 minutes for YOLO26m, ∼2.5 hours for RT-DETR-L, and ∼3 hours for Faster R-CNN. The total computational cost of this study was ∼80 CPU-hours for ISP processing and ∼173 GPU-hours for model training.

Results

Overall mAP by ISP Variant

Figure 4 presents the overall ΔmAP_{50-95} from the Default ISP for each detector.

Gamma correction, despite being the standard tool for improving the appearance of nighttime images [8], has remarkably little effect on detection. Across all four architectures, the gamma variants produce mAP scores within 1 to 2 points of the default ISP. Gain variants behave similarly within the 32 to 1024 range. The exception is Gain 8, which clips most pixel values to zero, yet retrained models still perform within 3 to 6 mAP points of the default. The images in Gain 8 are approximately five times smaller in storage, confirming the extent of information discarded.

The most striking result concerns saturation. All four architectures exhibit a strong negative correlation between saturation level and mAP (Spearman $\rho = -1.0$, $p < 0.001$), but the magnitude differs dramatically. YOLOv8m drops from 0.457 at Saturation 0 to 0.194 at Saturation 2048, a loss of 26.2 points. YOLO26m falls similarly from 0.449 to 0.250, but Faster R-CNN loses only 2.2 points (0.491 to 0.470), a 12x difference in sensitivity. RT-DETR-L shows intermediate sensitivity, dropping 13.7 points. Saturation 0 (grayscale) outperforms the default ISP for

[1] `https://github.com/tejus-vignesh/GMIND-ISP-Impact-Analysis`

both YOLO models, suggesting that colour information may not be beneficial for nighttime detection.

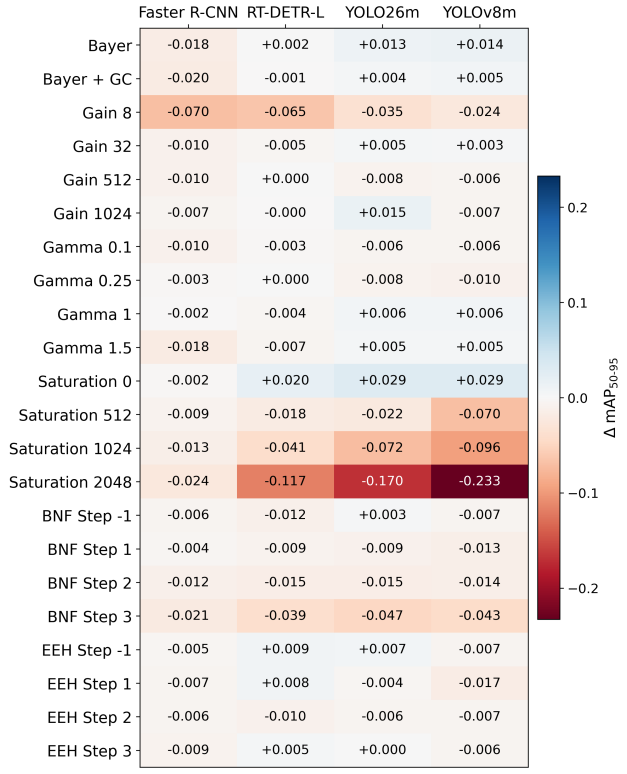


Figure 4. *ΔmAP_{50-95} from Default ISP for all 4 detector architectures.*

The coefficient of variation of mAP across all 23 ISP configurations quantifies this: Faster R-CNN exhibits 2.87%, RT-DETR-L 6.13%, YOLO26m 9.71%, and YOLOv8m 12.81%. This indicates that the degree to which ISP tuning affects detection performance varies considerably across architectures.

Raw Bayer input is competitive with the default ISP across all architectures. For YOLOv8m, Bayer achieves 0.442 compared to 0.427 for the default, slightly outperforming the fully processed pipeline. YOLO26m shows a similar pattern. Figure 5 presents the per-module latency breakdown of the full ISP pipeline alongside the ΔmAP_{50-95} for all four architectures when the ISP is partially (Bayer + GC) or fully bypassed (Bayer). For the YOLO models, bypassing the entire ISP eliminates all pipeline latency while maintaining or marginally improving detection accuracy, suggesting that the full ISP pipeline offers limited perceptual benefit for these single-stage detectors under nighttime conditions. However, Faster R-CNN exhibits a drop of 1.8-2.0 mAP points when the ISP is bypassed, indicating that the value of ISP processing is architecture-dependent.

Distance-Binned Per-Class Analysis

Figures 6 and 7 present AP_{50-95} by class and distance bin for YOLOv8m and Faster R-CNN models. Car detection is the most robust across all architectures, peaking near 0.95 at 15 to 30 m. For YOLOv8m, car AP drops to approximately 0.4-0.5 by 45 to 60m, whereas Faster R-CNN maintains AP above 0.6 at the same range. Person detection degrades steadily with distance across both architectures, though Faster R-CNN retains noticeably higher AP at longer ranges. Bicycle detection is the most challenging class, with AP falling sharply at 15 to 30m across most ISP configurations.

No ISP variant reverses the degradation with distance. The spread between best and worst ISP variants at each bin is consistently much smaller than the drop between adjacent bins. For the YOLO models, the high-saturation variants sit at the bottom of the spread across all bins and classes, while for Faster R-CNN, the saturation variants remain tightly clustered with the rest.

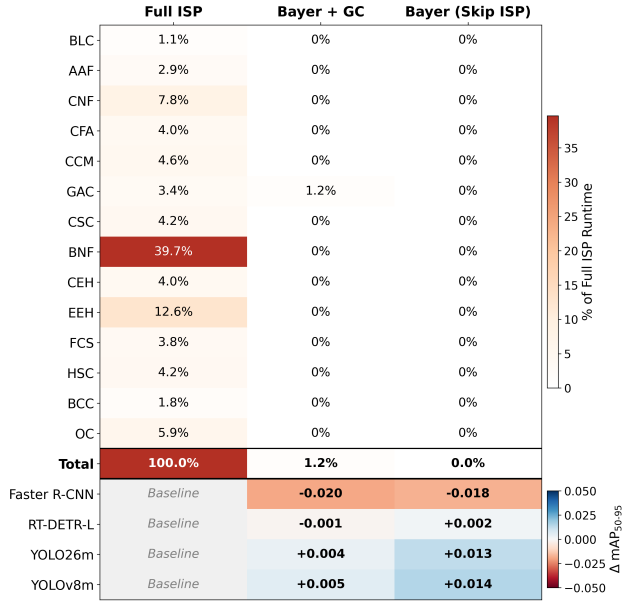
Faster R-CNN maintains noticeably higher AP at longer ranges, particularly for the person class, consistent with the two-stage architecture's region proposal mechanism and FPN backbone.

Qualitative Examples

Figure 8 shows the same scene processed with the default ISP and Saturation 2048, with YOLOv8m detections overlaid. The high-saturation image produces exaggerated colours, and the detector misses a substantial number of objects. Figure 9 compares the default ISP with Gain 8; despite the visibly dark image, the retrained YOLO26m detector identifies many objects that remain visible as bright regions against the dark background.

Discussion

As mentioned in the previous section, YOLOv8m loses 26.2 mAP points across the saturation range while Faster R-CNN loses only 2.2 points under the same conditions. Both YOLO architectures exhibit nearly identical sensitivity despite substantial internal design differences (decoupled head vs. attention mechanisms, NMS vs. NMS-free), suggesting the vulnerability is rooted in the single-stage paradigm itself rather than specific components



BLC: Black Level Compensation; AAF: Anti-Aliasing Filter; CNF: Chroma Noise Filter; CFA: Demosaicing; CCM: Colour Correction Matrix; GAC: Gamma Correction; CSC: Colour Space Conversion; BNF: Bilateral Noise Filter; CEH: Contrast Enhancement (CLAHE); EEH: Edge Enhancement; FCS: False Colour Suppression; HSC: Hue/Saturation Control; BCC: Brightness/Contrast Control; OC: Output Conversion

Figure 5. *ISP module latency breakdown as a percentage of total pipeline runtime, with ΔmAP_{50-95} from Default ISP for all 4 detector architectures.*

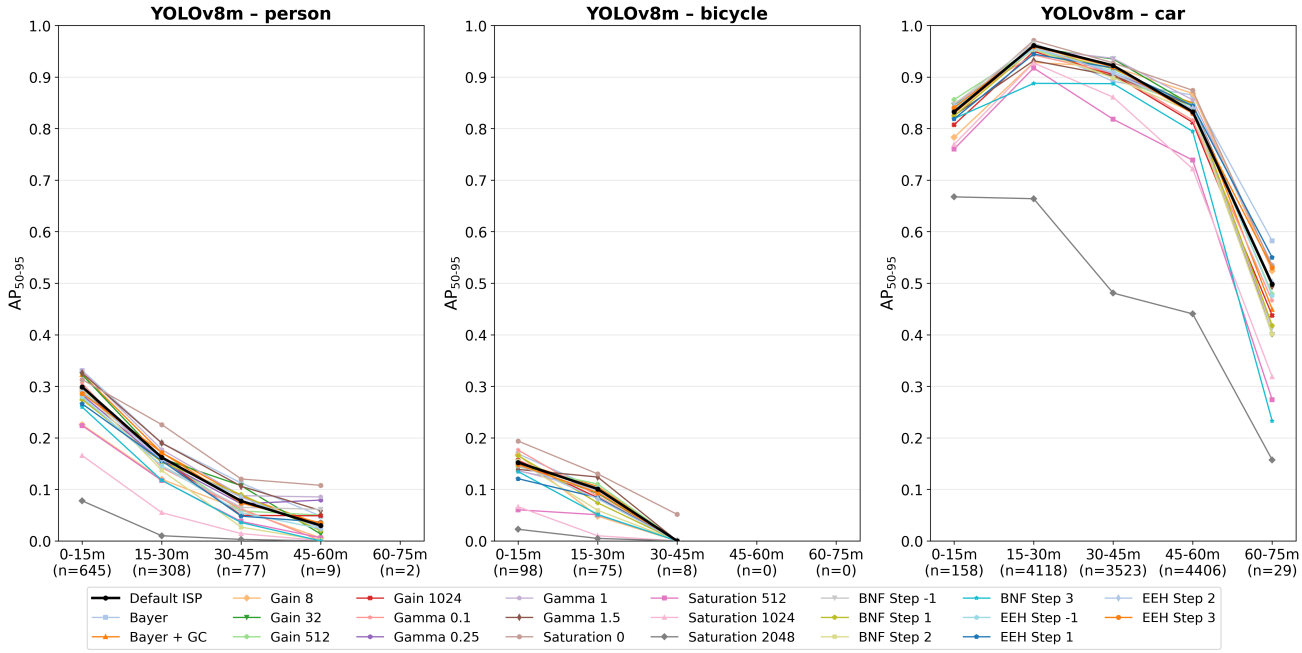


Figure 6. *Per-class AP_{50-95} of YOLOv8m as a function of object distance for person, bicycle, and car classes, with n indicating the ground truth instance count in each case.*

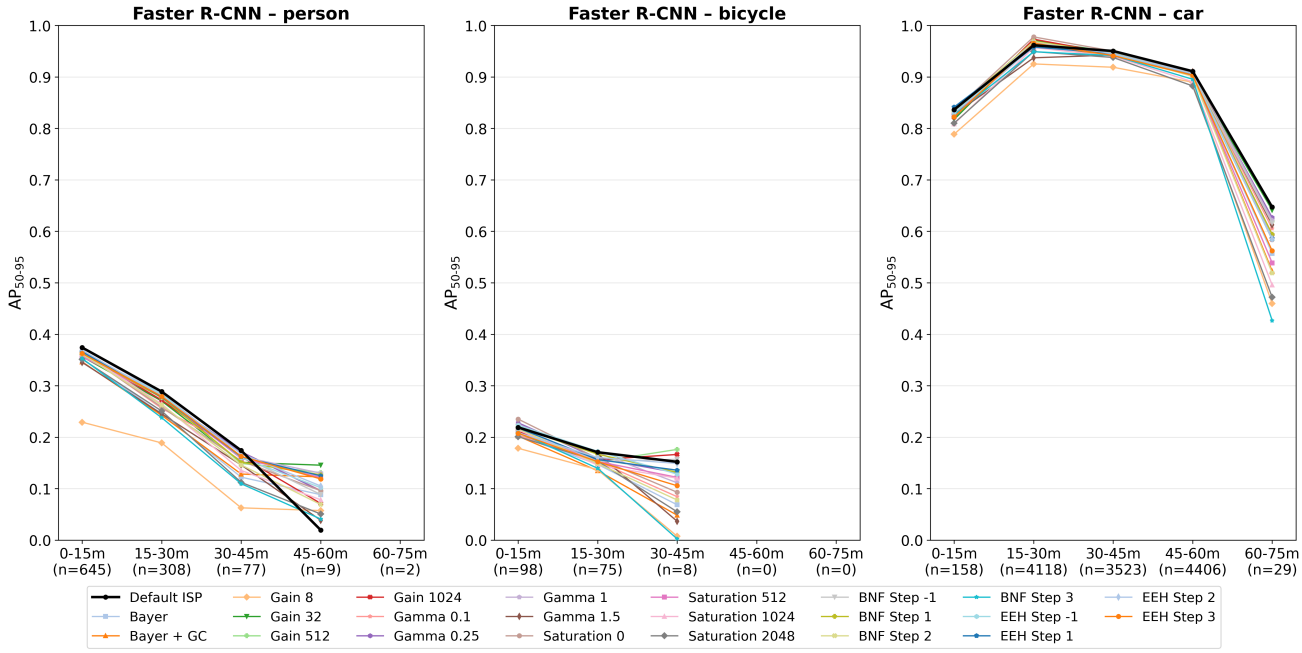


Figure 7. *Per-class AP_{50-95} of Faster R-CNN (ResNet-50 FPN) as a function of object distance for person, bicycle, and car classes, with n indicating the ground truth instance count in each case.*

within it. We hypothesise that single-stage detectors, which process the full feature map in a single pass, encode colour relationships in their convolutional filters that become unreliable when saturation is boosted under nighttime conditions where colour fidelity is already poor. Faster R-CNN's two-stage design, which first proposes candidate regions and then classifies them using pooled FPN features, may be inherently less dependent on absolute colour values and more reliant on structural and spatial features. This is consistent with Molloy et al. [21], who found that ResNet50-FPN backbones exhibit different ISP sensitivity profiles, and with Liu et al. [18], who reported that colour-change corruptions are particularly threatening and that single-stage and two-stage detectors exhibit fundamentally different robustness characteristics.

The finding that grayscale outperforms the default ISP for YOLO aligns with the AdaptiveISP system [30], which uses reinforcement learning to optimise ISP configurations per image and was found to frequently desaturate colour in low-light scenarios.

Figure 8. *Detection results from YOLOv8m on the same scene processed with Saturation 2048 (top) and the default ISP (bottom). Green dashed: matched ground truth; orange dashed: missed ground truth; red solid: predictions.*

Figure 9. *Detection results from YOLO26m on the same scene processed with Gain 8 (top) and the default ISP (bottom). Green dashed: matched ground truth; orange dashed: missed ground truth; red solid: predictions.*

Our empirical result provides independent corroboration. Practically, this is relevant to the development of cameras for Asian markets, where ISP pipelines are often tuned with boosted saturation to match regional aesthetic preferences. Automotive systems pairing such cameras with YOLO-based detectors may inadvertently degrade nighttime perception performance.

The strong performance of raw Bayer input relative to the default ISP warrants further discussion. As shown in Figure 5, bypassing the ISP entirely eliminates 100% of the pipeline processing cost while yielding a ΔmAP of +0.014 for YOLOv8m and +0.013 for YOLO26m. Even applying only gamma correction, which accounts for just 1.2% of the full ISP runtime, produces comparable results. This suggests that the remaining 98.8% of ISP computation, including bilateral noise filtering (39.7% of runtime) and edge enhancement (12.6%), provides no measurable benefit for single-stage nighttime detection. For Faster R-CNN the picture is slightly different: bypassing the ISP reduces mAP by 1.8 to 2.0 points, a small but consistent penalty, indicating that the two-stage architecture does extract some value from ISP processing that the YOLO models do not. RT-DETR-L falls in between, with negligible change in either direction. These results extend the findings of Ljungbergh et al. [19] and Li et al. [15] to nighttime conditions, where the ISP's noise filtering and contrast enhancement capabilities would presumably matter most.

Rather than concluding that the ISP should be removed entirely, these findings suggest that a full ISP tuned for human perception is not the optimal configuration for machine vision. There is an opportunity to develop leaner pipelines that retain only the processing stages that truly enhance detection performance, or to explore lightweight neural ISP [23, 6] that are optimised for the downstream perception task rather than for visual appearance. Such modules also improve model generalisability to unseen data from different sensors, since they learn sensor-agnostic representations rather than relying on sensor-specific data [23]. Notably, Guo et al. [6] demonstrated that their Dark-ISP module not only outperforms a conventional ISP but also surpasses raw Bayer input, suggesting that the optimal approach lies not in removing ISP processing but in replacing human-tuned stages with perception-aware alternatives.

Limitations

The G-MIND dataset was collected from a fixed infrastructure node rather than a vehicle-mounted camera, and results may not generalise directly to onboard advanced driver-assistance systems (ADAS). The single-location car park with artificial lighting does not represent all nighttime scenarios and has a lack of scene variability relative to other automotive datasets. Our 23 configurations are univariate perturbations; multivariate ISP block interactions remain unexplored [21]. Ground truth counts at long range are considerably low, particularly for person and bicycle beyond 45 m. All four architectures were pre-trained on the COCO dataset, which consists of ISP-processed RGB images. This introduces a potential bias: models may retain a preference for ISP-processed representations despite fine-tuning, which could disadvantage the raw Bayer and extreme ISP configurations where the input distribution differs most from COCO. Training from scratch would be needed to fully disentangle this bias, which was not feasible in this study. Finally, hand-selected parameter values may miss improvements achievable through gradient-based co-

optimisation of ISP and detector parameters.

Conclusion and Future Work

We presented a systematic evaluation of ISP sensitivity for nighttime object detection, fine-tuning 92 models across 23 ISP configurations and four detector architectures on the G-MIND dataset. Saturation is one of the most critical ISP parameters, but its effect is strongly architecture-dependent: YOLOv8m loses 26.2 mAP points across the saturation range while Faster R-CNN loses 2.2, a difference reflected in a coefficient of variation of 2.87% for Faster R-CNN compared to 12.81% for YOLOv8m across all 23 ISP configurations. Gamma and gain have negligible effects when models are retrained. Raw Bayer input matches or outperforms the default ISP for single-stage detectors while eliminating all ISP processing cost, suggesting that a full ISP tuned for human perception is not the optimal configuration for nighttime machine perception. No ISP variant reverses detection degradation with distance, indicating that improvements in long-range nighttime detection will likely require approaches beyond the ISP, such as higher-resolution sensors or sensor fusion with complementary modalities.

Several directions for future work emerge from these findings. The architecture-dependent saturation sensitivity warrants deeper investigation to understand how much each architecture relies on inter-channel colour information. Our ISP configurations are univariate perturbations of individual blocks; exploring multivariate combinations where multiple blocks are altered simultaneously would capture inter-block interactions not visible in this study. Given the strong raw Bayer performance, developing lightweight neural ISP modules that are optimised for detection rather than visual appearance, and evaluating their generalisability across different sensors, is a promising direction.

Acknowledgments

This publication has emanated from research conducted with the financial support of Taighde Éireann – Research Ireland under Project Agreement SoWP2-TP0028 and Grant number 13/RC/2094_2. The authors wish to acknowledge the support of Qualcomm and Lero, the Research Ireland Centre for Software at the University of Limerick, for this work, which is part of the QIQ4CV project.

References

[1] Taimur Bilal, Bakhtawar Amjad, Talha Iqbal, Bilal Zafar, Khurram Usman, Sohaib Imran Bhatti, Abdul Rahman Qureshi, Muqudas Rafiq, Maria Nadeem, Junaid Amjad, et al. Infinite-ISP: An open source hardware image signal processor platform for all imaging needs. *Electronic Imaging*, 37:1–6, 2025.

[2] Mark Buckler, Suren Jayasuriya, and Adrian Sampson. Reconfiguring the imaging pipeline for computer vision. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 975–984, 2017.

[3] Nicholas N Ferenchak and Masoud Ghodrat Abadi. Nighttime pedestrian fatalities: A comprehensive examination of infrastructure, user, vehicle, and situational factors. *Journal of Safety Research*, 79:14–25, 2021.

[4] Diarmaid Geever, Tim Brophy, Dara Molloy, Martin Glavin, Edward Jones, and Brian Deegan. ISP tuning for improved image quality in machine vision. *Electronic Imaging*, 36:1–8, 2024.

[5] Diarmaid Geever, Tim Brophy, Dara Molloy, Enda Ward, Roshan George, Norman Koren, Martin Glavin, Edward Jones, and Brian Deegan. Information capacity as a predictor of perception performance. *IEEE Open Journal of Vehicular Technology*, 2026.

[6] Jiasheng Guo, Xin Gao, Yuxiang Yan, Guanghao Li, and Jian Pu. Dark-ISP: Enhancing RAW Image Processing for Low-Light Object Detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9583–9593, 2025.

[7] Patrick Hansen, Alexey Vilkin, Yury Krustalev, James Imber, Dumidu Talagala, David Hanwell, Matthew Mattina, and Paul N Whatmough. ISP4ML: The role of image signal processing in efficient deep learning vision systems. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 2438–2445. IEEE, 2021.

[8] Kuo-Feng Hung and Kang-Ping Lin. Bio-inspired dark adaptive nighttime object detection. *Biomimetics*, 9(3):158, 2024.

[9] IEEE. IEEE Standard for Automotive System Image Quality. IEEE 2020-2024. `https://standards.ieee.org/ieee/2020/6765/`, 2025. Published 2025-03-21.

[10] IEEE P2020 Working Group. IEEE P2020 Automotive Imaging White Paper. `https://ieeexplore.ieee.org/document/8439102`, August 2018. pp. 1–26.

[11] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics YOLOv8, 2023.

[12] Glenn Jocher, Ayush Chaurasia, Alex Stoken, Jirka Borovec, Yonghye Kwon, Jiacong Fang, Kalen Michael, Diego Montes, Jebastin Nadar, Piotr Skalski, et al. ultralytics/YOLOv5: v6. 1-tensorrt, tensorflow edge tpu and openvino export and inference. *Zenodo*, 2022.

[13] Glenn Jocher and Jing Qiu. Ultralytics YOLO26, 2026.

[14] Norman Koren. Measuring camera information capacity with slanted edges. `https://www.imatest.com/wp-content/uploads/2022/11/Information_capacity_slanted_edge.pdf`, 2023. Accessed: 2026-03-31.

[15] Zhihao Li, Ming Lu, Xu Zhang, Xin Feng, M Salman Asif, and Zhan Ma. Efficient visual computing with camera raw snapshots. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(7):4684–4701, 2024.

[16] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2117–2125, 2017.

[17] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft COCO: Common Objects in Context. In *European Conference on Computer Vision*, pages 740–755. Springer, 2014.

[18] Jiawei Liu, Zhijie Wang, Lei Ma, Chunrong Fang, Tongtong Bai, Xufan Zhang, Jia Liu, and Zhenyu Chen. Benchmarking object detection robustness against real-world corruptions. *International Journal of Computer Vision*, 132(10):4398–4416, 2024.

[19] William Ljungbergh, Joakim Johnander, Christoffer Petersson, and Michael Felsberg. Raw or Cooked? Object Detection on Raw Images. In *Scandinavian Conference on Image Analysis*, pages 374–385. Springer, 2023.

[20] TorchVision maintainers and contributors. Torchvision: Pytorch's computer vision library. `https://github.com/pytorch/vision`, 2016.

[21] Dara Molloy, Brian Deegan, Darragh Mullins, Enda Ward, Jonathan Horgan, Ciaran Eising, Patrick Denny, Edward Jones, and Martin Glavin. Impact of ISP tuning on object detection. *Journal of Imag-*

ing, 9(12):260, 2023.

[22] Dara Molloy, Roshan George, Tim Brophy, Brian Deegan, Darragh Mullins, Enda Ward, Jonathan Horgan, Ciaran Eising, Patrick Denny, Edward Jones, et al. G-MIND: Galway Multimodal Infrastructure Node Dataset for Intelligent Transportation Systems. *IEEE Open Journal of Vehicular Technology*, 2025.

[23] Igor Morawski, Yu-An Chen, Yu-Sheng Lin, Shusil Dangi, Kai He, and Winston H Hsu. GenISP: Neural ISP for Low-Light Machine Cognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 630–639, 2022.

[24] Ali Mosleh, Avinash Sharma, Emmanuel Onzon, Fahim Mannan, Nicolas Robidoux, and Felix Heide. Hardware-in-the-loop end-to-end optimization of camera image processing pipelines. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7529–7538, 2020.

[25] Andrea Polesel, Giovanni Ramponi, and V John Mathews. Image enhancement via adaptive unsharp masking. *IEEE Transactions on Image Processing*, 9(3):505–510, 2000.

[26] Jueqin Qiu. fast-openISP: A faster re-implementation of openISP. `https://github.com/QiuJueqin/fast-openISP`, 2021. Accessed: 2026-03-29.

[27] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, 28, 2015.

[28] Nicolas Robidoux, Luis E Garcia Capel, Dong-eun Seo, Avinash Sharma, Federico Ariza, and Felix Heide. End-to-end high dynamic range camera pipeline optimization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6297–6307, 2021.

[29] Carlo Tomasi and Roberto Manduchi. Bilateral filtering for gray and color images. In *Sixth international conference on computer vision (IEEE Cat. No. 98CH36271)*, pages 839–846. IEEE, 1998.

[30] Yujin Wang, Tianyi Xu, Fan Zhang, Tianfan Xue, and Jinwei Gu. AdaptiveISP: Learning an adaptive image signal processor for object detection. *Advances in Neural Information Processing Systems*, 37:112598–112623, 2024.

[31] Lucie Yahiaoui, Jonathan Horgan, Brian Deegan, Senthil Yogamani, Ciarán Hughes, and Patrick Denny. Overview and empirical analysis of ISP parameter tuning for visual perception in autonomous driving. *Journal of Imaging*, 5(10):78, 2019.

[32] Yian Zhao, Wenyu Lv, Shangliang Xu, Jinman Wei, Guanzhong Wang, Qingqing Dang, Yi Liu, and Jie Chen. DETRs beat yolos on real-time object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16965–16974, 2024.

Author Biography

Tejus Vijayakumar is a Ph.D. candidate at the University of Limerick, Ireland, co-funded by Qualcomm and Lero SFI. He received his B.Tech. in Mechanical Engineering from Kalasalingam University, India in 2017 and his M.Sc. (Hons.) in Artificial Intelligence and Machine Learning from the University of Limerick in 2023. From 2018 to 2022, he worked as a Machine Learning Engineer at Tech Mahindra, developing and deploying machine learning solutions for State Bank of India. His research interests include Image Signal Processing, computer vision, and image sensors within an automated vehicle context.

Tim Brophy received his B.Eng. (Hons.) in 2018 and Ph.D. from the University of Galway. From 2018 to 2025, he was part of the Connaught Automotive Research (CAR) Group, focusing on sensing and perception for automated vehicles. In 2025, he joined the University of Limerick as a Postdoctoral Researcher. His research interests include sensor availability, computer vision, and artificial intelligence, with a particular emphasis on their application in automated vehicle systems.

Brian Deegan received his Bachelor's degree in Computer Engineering and M.Sc. in Biomedical Engineering from the University of Limerick in 2004 and 2005, and his Ph.D. in Biomedical Engineering from the University of Galway in 2011. He worked at Valeo Vision Systems from 2011 to 2022 as a Vision Research Engineer focused on image quality. In 2022, he joined the University of Galway as a Lecturer. His research interests span cerebral blood flow, HDR imaging, LED flicker, and machine-vision image quality.

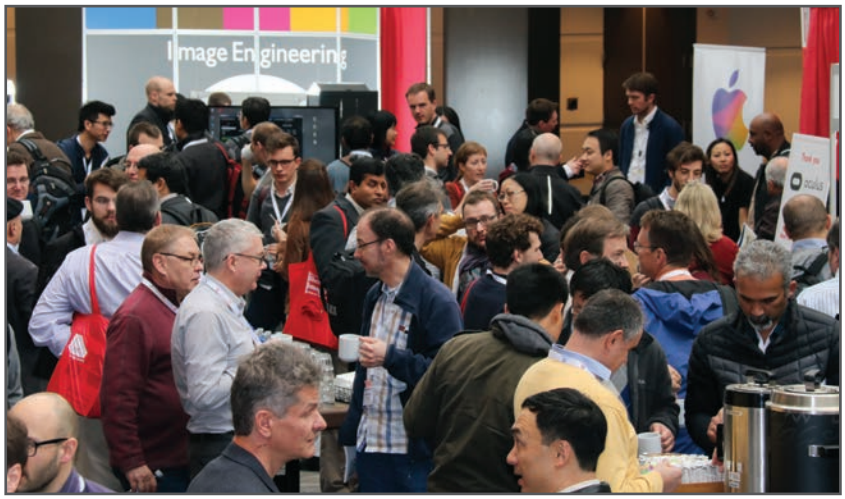
Ciarán Eising received his electronic and computer engineering degree and Ph.D. from the University of Galway in 2003 and 2010. From 2009 to 2020, he served as Computer Vision Team Lead, Architect, and Senior Expert with Valeo Vision Systems. He became an Adjunct Lecturer at the University of Galway in 2016. In 2020, he joined the University of Limerick, where he is now a Professor in the Department of Electronic and Computer Engineering.

Patrick Denny holds degrees in Physics and Mathematics from NUI Maynooth and the University of Galway, where he also earned his Ph.D. in Physics in 2000. His career spans RF Engineering at AVM GmbH, Supercomputing at Compaq-HP, and two decades leading RF and imaging innovation at Valeo. He became Adjunct Professor at the University of Galway, later becoming Associate Professor of AI and Imaging at the University of Limerick. He co-founded several key automotive imaging standards and conference groups.

JOIN US AT THE NEXT EI!

electronic IMAGING

Imaging across applications . . . Where industry and academia meet!



• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

ADVANCING IMAGING SCIENCE 75 1947 IS&T 2022 imaging.org